Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 1–16.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 5–21.
Original Russian Text Copyright © 2000 by Vergeles.

GRAVITATION, ASTROPHYSICS

Canonical Quantization of Two-Dimensional Gravity

S. N. Vergeles
Landau Institute of Theoretical Physics, Russian Academy of Sciences,
Chernogolovka, Moscow oblast, 142432 Russia;
e-mail: vergeles@itp.ac.ru

Received March 23, 1999

Abstract—A canonical quantization of two-dimensional gravity minimally coupled to real scalar and spinor Majorana fields is presented. The physical state space of the theory is completely described and calculations are also made of the average values of the metric tensor relative to states close to the ground state. © *2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The quantum theory of gravity in four-dimensional space–time encounters fundamental difficulties which have not yet been surmounted. These difficulties can be arbitrarily divided into conceptual and computational. The main conceptual problem is that the Hamiltonian is a linear combination of first-class constraints. This fact makes the role of time in gravity unclear. The main computational problem is the nonrenormalizability of gravity theory. These difficulties are closely intertwined. For instance, depending on the computational procedure, the constraint algebra may or may not contain an anomalous contribution (central charge). The presence or absence of an anomaly in the first-class constraint algebra has a decisive influence on the quantization procedure and the ensuing physical picture.

These fundamental problems can be successfully resolved using relatively simple models of generally covariant theories in two-dimensional space–time. These models particularly include two-dimensional gravity models, both pure and interacting with matter, and also two-dimensional string models (see, for example [1–6] and the literature cited therein).

In the present paper we present a canonical quantization of two-dimensional gravity minimally coupled to real scalar and spinor Majorana fields. All the constructions and calculations are given before the final result is obtained. The physical states of the theory are fully described. The complete state space has similar properties to the multidimensional Fock space in which boson and fermion operators are acting. The calculations begin with the average values of the metric tensor relative to states close to the ground state.

However, the present study should merely be perceived as one of the first steps along the way to the anomaly-free quantization (if this is at all possible) of some of the theories which, in more traditional quantizations, are anomalous. Only further studies, including a systematic study of the discrepancies in the predictions (which are encountered in different approaches to quantization) for observables in the highest orders, will serve as a criterion for the correctness of the selected quantization route.

The progress achieved in the construction of a two-dimensional quantum theory of gravity is associated with two ideas. These ideas will be formulated below after the necessary notation has been introduced.

We shall postulate that space–time is topologically equivalent to a two-dimensional cylinder. The time coordinate t varies between minus infinity and plus infinity while the spatial coordinate σ varies between 0 and 2π. All these functions are periodic with respect to the coordinate σ. The set of coordinates (t, σ) is denoted as $\{x^\mu\}$. The metric tensor in space–time is denoted by $g_{\mu\nu}$ so that the square of the interval is written as

$$ds^2 = g_{00}dt^2 + g_{11}d\sigma^2 + 2g_{01}dtd\sigma. \qquad (1.1)$$

Most of the formulas and notation in the Introduction are taken from [2]. The metric tensor is then parametrized as follows:

$$g_{\mu\nu} = e^{2\rho}\begin{pmatrix} u^2 - v^2 & v \\ v & -1 \end{pmatrix}, \qquad (1.2)$$

$$g \equiv \det g_{\mu\nu} = -u^2e^{4\rho}.$$

Let $i, j = 0, 1$ and $\eta_{ij} = \mathrm{diag}(1, -1)$. We introduce the orthonormalized basis $\{e_i^\mu\}$ so that

$$g_{\mu\nu}e_i^\mu e_j^\nu = \eta_{ij}. \qquad (1.3)$$

To be specific we take

$$e_0^\mu = \frac{1}{u}e^{-\rho}\begin{pmatrix} 1 \\ v \end{pmatrix}, \quad e_1^\mu = \begin{pmatrix} 0 \\ e^{-\rho} \end{pmatrix}. \qquad (1.4)$$

The dyad $\{e^i_\mu\}$ is uniquely determined by the equations $e^i_\mu e^\nu_i = \delta^\nu_\mu \longleftrightarrow e^i_\mu e^\mu_j = \delta^i_j$. Taking into account (1.4), we have

$$e^0_\mu = \begin{pmatrix} ue^\rho \\ 0 \end{pmatrix}, \quad e^1_\mu = e^\rho \begin{pmatrix} -v \\ 1 \end{pmatrix}. \tag{1.5}$$

We analyze the action

$$\begin{gathered} S = \int dt \int_0^{2\pi} d\sigma \sqrt{-g}\left\{\frac{1}{4\pi G}(\eta R - 2\lambda)\right. \\ \left. + \frac{1}{2}g^{\mu\nu}\partial_\mu f \partial_\nu f + \frac{i}{2}e^\mu_j \bar{\psi}\gamma^j \mathcal{D}_\mu \psi\right\}. \end{gathered} \tag{1.6}$$

Here G is the gravitational constant, λ is the cosmological constant, R is the scalar curvature of space–time, η and f are the real scalar fields, ψ is the two-component spinor Majorana field, and $\{\gamma^i\}$ are two-dimensional Dirac matrices. We then assume that

$$\begin{gathered} \gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \\ \gamma^5 \equiv \gamma^0\gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{gathered} \tag{1.7}$$

The Majorana nature of the spinor field implies that $\psi = \gamma^0 \bar{\psi}^t$ (the superscript t indicates transposition). In our case, we have

$$\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix}, \quad \phi = \phi^\dagger, \quad \chi = \chi^\dagger. \tag{1.8}$$

The covariant differentiation operation of the spinor field is determined in accordance with the formula

$$\begin{gathered} \mathcal{D}_\mu \psi = \left(\frac{\partial}{\partial x^\mu} + \frac{1}{2}\omega_{ij\mu}\sigma^{ij}\right)\psi, \\ \sigma^{ij} = \frac{1}{4}[\gamma^i, \gamma^j]. \end{gathered} \tag{1.9}$$

The form of connectedness $\omega_{ij\mu}$ is obtained uniquely from

$$d\omega^i + \omega^i_j \wedge \omega^j = 0,$$

where $\omega^i \equiv e^i_\mu dx^\mu$. Hence, we find

$$\begin{gathered} \omega_{01} = \left[u' + u\rho' - \frac{v}{u}(\dot{\rho} + \rho' v + v')\right]dt \\ + \frac{1}{u}(\dot{\rho} + \rho' v + v')d\sigma. \end{gathered} \tag{1.10}$$

Here and subsequently the dot and prime denote the partial derivatives $\partial/\partial t$ and $\partial/\partial\sigma$, respectively. Using the Cartan structure equation, we can easily establish the relationship

$$\sqrt{-g}R dt \wedge d\sigma = 2d\omega^{01}. \tag{1.11}$$

Since the fields ϕ and χ in (1.8) are real and belong to Grassmann algebra, we have

$$\phi(x)\phi(x) = 0, \quad \chi(x)\chi(x) = 0. \tag{1.12}$$

Using (1.12), we can make the substitution $\mathcal{D}_\mu\psi \longrightarrow (\partial/\partial x^\mu)\psi$ in (1.6). Thus, the fermion component of the action is proportional to the expression

$$e^\rho\{\phi\dot{\phi} + (u + v)\phi\phi' + \chi\dot{\chi} - (u - v)\chi\chi'\}.$$

In this last expression the factor e^ρ may be eliminated by substituting

$$\phi \longrightarrow e^{-\rho/2}\phi, \quad \chi \longrightarrow e^{-\rho/2}\chi.$$

On account of (1.12), no additional derivatives of the field ρ appear in the action when this substitution is made. Consequently, the Lagrangian of this system has the form

$$\begin{gathered} \mathcal{L} = \int d\sigma \left\{\frac{1}{2\pi G}\left[u^{-1}\dot{\eta}(\dot{\rho} + v\rho' + v')\right.\right. \\ \left. + \eta'\left(\frac{v}{u}(\dot{\rho} + v\rho' + v') - u' - u\rho'\right) - \lambda u e^{2\rho}\right] \\ + \frac{1}{2u}[\dot{f}^2 + 2v\dot{f}f' - (u^2 - v^2)f'^2] \\ \left. + \frac{i}{2}[\phi\dot{\phi} + (u + v)\phi\phi' + \chi\dot{\chi} - (u - v)\chi\chi']\right\}. \end{gathered} \tag{1.13}$$

We denote by π_η, π_ρ, and π the fields canonically conjugate to the fields π, ρ, and f, respectively. The fields u and v in (1.13) are Lagrangian factors. We obtain the Hamiltonian of the system (1.13) by a standard procedure:

$$\begin{gathered} \mathcal{H} = \int d\sigma(u\mathcal{E} + v\mathcal{P}), \\ \mathcal{E} = 2\pi G\pi_\eta\pi_\rho + \frac{1}{2\pi G}[-(\eta'' - \eta'\rho') + \lambda e^{2\rho}] \\ + \frac{1}{2}(\pi^2 + f'^2) + \frac{i}{2}(-\phi\phi' + \chi\chi'), \\ \mathcal{P} = -\left(\pi_\eta\eta' + \pi_\rho\rho' + \pi f' + \frac{i}{2}\phi\phi' + \frac{i}{2}\chi\chi'\right). \end{gathered} \tag{1.14}$$

We make the following canonical transposition of variables:

$$\lambda r^0 = -\frac{1}{2}\sqrt{\frac{\lambda}{4\pi G}}e^{-\rho}(2\eta'\cosh\Sigma - 4\pi G\pi_\rho\sinh\Sigma),$$

$$\begin{gathered}\lambda r^1 = -\frac{1}{2}\sqrt{\frac{\lambda}{4\pi G}}e^{-\rho}(4\pi G\pi_\rho\cosh\Sigma - 2\eta'\sinh\Sigma),\\ \pi_0 - r^{1'} = \sqrt{\frac{\lambda}{\pi G}}e^{\rho}\sinh\Sigma,\\ \pi_1 + r^{0'} = -\sqrt{\frac{\lambda}{\pi G}}e^{\rho}\cosh\Sigma.\end{gathered} \tag{1.15}$$

Here we have

$$\Sigma(\sigma) = 2\pi G\int_0^\sigma d\tilde{\sigma}\pi_\eta(\tilde{\sigma}).$$

The variables describing the matter remain unchanged. In the new variables the Hamiltonian (1.14) has the form

$$\begin{gathered}\mathcal{E} = \frac{1}{2}\left[-\left(\pi_0^2 + (r^{0'})^2\right) + \left(\pi_1^2 + (r^{1'})^2\right)\right]\\ + \frac{1}{2}[\pi^2 + f'^2 + i(-\phi\phi' + \chi\chi')],\end{gathered} \tag{1.16}$$

$$\mathcal{P} = -(\pi_0 r^{0'} + \pi_1 r^{1'}) - \left(\pi f' + \frac{i}{2}\phi\phi' + \frac{i}{2}\chi\chi'\right).$$

So far the analysis has been classical. In order to quantize the system, we must first begin by defining the commutation relations for the canonically conjugate variables. In our case, we have

$$\begin{gathered}[r^0(\sigma), \pi_0(\sigma')] = [r^1(\sigma), \pi_1(\sigma')]\\ = [f(\sigma), \pi(\sigma')] = i\delta(\sigma - \sigma').\end{gathered} \tag{1.17}$$

For the fermion degrees of freedom we have the anticommutation relations

$$\{\phi(\sigma), \phi(\sigma')\} = \{\chi(\sigma), \chi(\sigma')\} = \delta(\sigma - \sigma'). \tag{1.18}$$

All the other commutators or anticommutators of the fundamental fields r^a, π_a, f, π, and ψ are zero. It is easy to check that the Heisenberg equations $i\dot{\mathbb{O}} = [\mathbb{O}, \mathcal{H}]$ obtained using the commutation relations (1.17) and (1.18) are the same as the Lagrange equations, where $\mathbb{O}$ is any operator.

Since the fields u and v in (1.14) are Lagrangian factors, the quantities (1.16) are constraints In terms of classical analysis, they are first-class constraints. However, it is well-known that as a result of quantization, anomalies or a central charge may appear in this system: the algebra of simultaneous commutators of $\mathcal{E}$ and $\mathcal{P}$ contains a central charge. The existence of this central charge in the constraint algebra radically complicates the quantization problem. In particular, the system (1.16)–(1.18) may be nonself-consistent.

Recently an anomaly-free approach to the quantization of the system (1.16)–(1.18) has been proposed in various studies [1–4,6]. In this approach no central charge is present in the quantum algebra of the quantities (1.16). This implies that all the operators (1.16) may be treated as first-class constraints in the Dirac sense. This new approach is used in the present study.

The idea of a new approach to quantization arose when studying a model describing pure gravity. This model is obtained from the model (1.16) by deleting the second terms on the right-hand sides of the system (1.16). It was shown in [1–4, 6] that in pure gravity theory the central charge is zero if the scalar product is positive definite in the entire scalar state space. The reason for this phenomenon is that in the new approach the operator ordering procedure in the quantities $\mathcal{E}$ and $\mathcal{P}$ differs radically from the ordering in traditional quantization.

We now formulate the assumptions used as the basis to develop the new quantization method.

1) The entire state space H_C in which the fundamental operator fields r^a, π_a, f, π, and ψ act has a positive definite scalar product. No indefinite metric is present in the H_C space.

In order to formulate the next assumption, we denote the set of operators (1.16) by L and the set of all the material fields f, ψ by ψ. From the operators L we eliminate the degrees of freedom describing the material fields and we denote the set of operators thus obtained by $L^{(0)}$. Hence the operators $L^{(0)}$ only determine the dynamics of the gravitational degrees of freedom and

$$[L^{(0)}, \psi] = 0. \tag{1.19}$$

2) In the theory (1.16) there exists a unitary transformation U such that

$$UL^{(0)}U^\dagger = L. \tag{1.20}$$

As a result of (1.19) and (1.20) the fields

$$\Psi = U\psi U^\dagger \tag{1.21}$$

commute with all the operators L.

We shall clarify the important role of this last assumption in the quantization of this system. We postulate that in the theory (1.16) there is a state $|0\rangle$ which annuls all the operators $L^{(0)}$ and all the annihilation operators of the fields ψ. According to the reasoning put forward above this is possible. Then the state $U|0\rangle$ is annulled by all the operators L and all the annihilation operators of the fields Ψ. The physical space of the states annulling all the operators L is constructed from the ground state $U|0\rangle$ using the creation operators of the fields Ψ. Consequently the problem of quantizing the system (1.16)–(1.18) is solved completely.

In the second assumption, the properties of the unitary transformation U of interest to us are only described in broad outline. Subsequently this unitary transformation is constructed explicitly for the model of two-dimensional gravity studied in the present article. The equivalent of formula (1.20) then has a more complex form. Nevertheless, fairly good progress can be made in the calculations by using the constructed unitary transformation.

2. QUANTIZATION OF PURE GRAVITY

The problem of quantizing two-dimensional pure gravity was studied in [1–4]. In [4, 6] the author described an anomaly-free quantization of a two-dimensional string whose constraint system is the same as the constraint system of two-dimensional pure gravity in the representation (1.16). This gives us the possibility of using the methods developed in [4, 6].

Let us assume that $a = 0, 1$ and $\eta_{ab} = \mathrm{diag}(-1, 1)$. In the gauge $u = 1$, $v = 0$ the Heisenberg equations for the fields r^a, $\pi^a = \eta^{ab}\pi_b$ have the form

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \sigma^2}\right) r^a = 0, \quad \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \sigma^2}\right) \pi^a = 0. \tag{2.1}$$

Consequently, the fields r^a and π^a contain both positive- and negative-frequency modes:

$$\begin{aligned} r^a(\sigma) &= \frac{x^a}{\sqrt{4\pi}} + \frac{i}{\sqrt{4\pi}} \sum_{n \neq 0} \frac{1}{n} (\alpha_n^a e^{in\sigma} + \bar{\alpha}_n^a e^{-in\sigma}), \\ \pi^a(\sigma) &= \frac{p^a}{\sqrt{\pi}} + \frac{i}{\sqrt{4\pi}} \sum_{n \neq 0} (\alpha_n^a e^{in\sigma} + \bar{\alpha}_n^a e^{-in\sigma}). \end{aligned} \tag{2.2}$$

We assume that $\alpha_0^a \equiv \bar{\alpha}_0^a \equiv p^a$. From the conditions that the fields (2.2) are real it follows that

$$x^{a\dagger} = x^a, \quad \alpha_n^{a\dagger} = \alpha_{-n}^a, \quad \bar{\alpha}_n^{a\dagger} = \bar{\alpha}_{-n}^a. \tag{2.3}$$

In order to satisfy the commutation relations (1.17), we impose the constraint that the nonzero commutators of the new variables should have the form

$$\begin{gathered} [\alpha_m^a, \alpha_n^b] = [\bar{\alpha}_m^a, \bar{\alpha}_n^b] = m\eta^{ab}\delta_{m+n}, \\ [x^a, p^b] = i\eta^{ab}. \end{gathered} \tag{2.4}$$

The set of operators (1.16) is equivalent to the two series of operators

$$\begin{aligned} L_n &= \frac{1}{2}\int_0^{2\pi} d\sigma e^{-in\sigma}(\mathcal{E} + \mathcal{P}), \\ \bar{L}_n &= \frac{1}{2}\int_0^{2\pi} d\sigma e^{in\sigma}(\mathcal{E} - \mathcal{P}), \quad n = 0, \pm 1, \ldots \end{aligned} \tag{2.5}$$

Using (2.2) we find

$$\begin{aligned} L_n &= \frac{1}{2} : \sum_m \alpha_{n-m}^a \alpha_{am} :, \\ \bar{L}_n &= \frac{1}{2} : \sum_m \bar{\alpha}_{n-m}^a \bar{\alpha}_{am} :. \end{aligned} \tag{2.6}$$

The ordering of the operators in (2.6) is determined in accordance with the general quantization conditions and plays a decisive role. The aim of quantization is to search for that space of physical states on which all the operators (2.6) go to zero and in which there is a mathematically correct and positive definite scalar product.

In the present article we adopt two approaches to the quantization of this system.

The first approach is well-known. It was formulated by Dirac and is described by the following scheme. Let us assume that $\{\chi_n\}$ is a complete set of first-class constraints. The physical states then satisfy the conditions

$$\chi_n | \rangle_{PD} = 0. \tag{2.7}$$

The conditions for consistency of the theory then follow from (2.7)

$$[\chi_m, \chi_n] = c_{mn}^l \chi_l. \tag{2.8}$$

In (2.8) the coefficients c_{mn}^l may be operator quantities and should be positioned to the left of the constraints χ_l.

The following difficulty is encountered when quantizing (2.2)–(2.8) (for further details see [6]). It follows from the conditions (2.7) that all the physical states do not depend on certain initial dynamic variables. For this reason the following problems arise:

a) Determining the scalar product on the physical state space;

b) Calculating the matrix elements relative to the physical states.

This is because not all the initial dynamic variables are operators in the physical state space. Hence the matrix elements of these variables are not determined in physical space. Although the observable quantities do not depend on these dynamic variables, nevertheless, serious difficulties may arise when the matrix elements of the observable quantities are calculated in physical space.

We shall subsequently call the quantization (2.7)–(2.8) the first method of quantization.

In [6] a different method of quantization was applied to the system (2.4), (2.6). The idea of this method involves somewhat weakening the Dirac conditions (2.7) by replacing them with the conditions

$$\langle P | \chi_m | P \rangle = 0. \tag{2.9}$$

Here P numbers the physical states. The quantization conditions (2.9) are similar to the Gupta–Bleuler conditions in electrodynamics when the equality $\partial_\mu A^\mu = 0$

is only satisfied in the sense of the mean value, and also to the quantization conditions in ordinary strong theory when the Virasoro algebra generators also only satisfy the conditions $L_n = 0$ in the sense of the mean value. In this case, averaging is performed relative to the physical states.

The fundamental difference between the quantization method proposed here and the Gupta–Bleuler quantization and the generally accepted string quantization is that in our approach the entire state space has a positive definite scalar product. We subsequently show that this fact can be used to make an anomaly-free quantization of a two-dimensional string.

The conditions for consistency of the theory, which replace the Dirac conditions (2.8), now have the form

$$\langle P|[\chi_m, \chi_n]|P\rangle = 0. \qquad (2.10)$$

The physical meaning of the conditions (2.10) is as follows. Let us assume that the Hamiltonian of the system has the form used in the generally covariant theories:

$$\mathcal{H}_T = \sum v_m \chi_m.$$

We assume that at time t the conditions (2.9) are satisfied. At an infinitely close time $t + \delta t$ the constraint χ_n is given by

$$\chi_n(t + \delta t) = \chi_n(t) + i\delta t \sum_m v_m[\chi_m, \chi_n](t).$$

Thus the self-consistency conditions (2.10) yield the equality (2.9) at any time. The method of quantization (2.9), (2.10) is subsequently called the second method of quantization. We initially apply the first method of quantization to the model (2.6). We introduce the notation

$$\begin{gathered} x_\pm = x^0 \pm x^1, \quad \alpha_m^{(\pm)} = \alpha_m^0 \pm \alpha_m^1, \\ \bar{\alpha}_m^{(\pm)} = \bar{\alpha}_m^0 \pm \bar{\alpha}_m^1. \end{gathered} \qquad (2.11)$$

The nonzero commutation relations of the new variables are obtained using (2.4):

$$\begin{gathered} [\alpha_m^{(+)}, \alpha_n^{(-)}] = -2m\delta_{m+n}, \quad [\bar{\alpha}_m^{(+)}, \bar{\alpha}_n^{(-)}] = -2m\delta_{m+n}, \\ [x_+, \alpha_n^{(-)}] = [x_-, \alpha_n^{(+)}] = -2i\delta_n. \end{gathered} \qquad (2.12)$$

We write the operators (2.6) in the new variables:

$$L_n = -\frac{1}{2}\sum_m \alpha_{n-m}^{(+)}\alpha_m^{(-)} = -\frac{1}{2}\sum_m \alpha_{n-m}^{(-)}\alpha_m^{(+)}. \qquad (2.13)$$

Where possible subsequently, for the operators with a bar we do not write those relations which are exactly the same as those for the operators without a bar. By definition, in (2.13) the ordering operation implies that either the elements $\alpha^{(+)}$ are positioned to the left of all the operators $\alpha^{(-)}$ or the converse is true. Both these orders are equivalent (see [4, 6]).

We perform the canonical transformation

$$\alpha \longrightarrow U_M \alpha U_M^\dagger, \quad \bar{\alpha} \longrightarrow U_M \bar{\alpha} U_M^\dagger, \quad x \longrightarrow U_M x U_M^\dagger,$$

$$U_M = \exp\left(\frac{iM^2 x_+}{2p_+}\right).$$

In this canonical transformation only the variables $\alpha_0^{(-)}$ and x_- change, as given by the formulas

$$\begin{gathered} \tilde{\alpha}_0^{(-)} = U_M \alpha_0^{(-)} U_M^\dagger = \alpha_0^{(-)} - \frac{M^2}{p_+}, \\ \tilde{x}_- = U_M x_- U_M^\dagger = x_- + M^2\frac{x_+}{p_+}. \end{gathered} \qquad (2.14)$$

For uniformity, in this section we introduce the notation $\tilde{\alpha}_m^{(-)} = \alpha_m^{(-)}$ for $m \neq 0$.

Subsequently, instead of the operators (2.13) we use the operators

$$L_n = -\frac{1}{2}\sum_m \alpha_{n-m}^{(+)}\tilde{\alpha}_m^{(-)}. \qquad (2.15)$$

The reason for this substitution will become clear in the following sections.

We define the vector state space in which the dynamic variables of the system act, as linear operators. We represent the entire state space H_{CD} as the tensor product of the gauge state space H_G and the physical state space H_{PD}:

$$H_{CD} = H_G \otimes H_{PD}. \qquad (2.16)$$

The space H_G is generated by its vacuum vector $|0; G\rangle$ which is determined by the following properties:

$$\begin{gathered} \alpha_{-m}^0|0; G\rangle = 0, \quad \alpha_m^1|0; G\rangle = 0, \quad m > 0, \\ \langle 0; G|0; G\rangle = 1. \end{gathered} \qquad (2.17)$$

The basis of the space H_G consists of vectors of the type

$$\alpha_m^0 \ldots \bar{\alpha}_n^0 \ldots \alpha_{-l}^1 \ldots \bar{\alpha}_{-r}^1|0; G\rangle, \quad m, n, l, r > 0. \qquad (2.18)$$

Therefore H_G is a Fock space with a positive-definite scalar product.

The basis in the physical Dirac state space H_{PD} consists of two series of states $|k\rangle_D$, $k = (k^0, k')|$ possessing the following properties:

$$\tilde{\alpha}_m^{(-)}|k\rangle_D = \tilde{\bar{\alpha}}_m^{(-)}|k\rangle_D = 0, \quad m = 0, \pm 1, \ldots \qquad (2.19)$$

The relationships (2.12) with $m = 0$ are rewritten in the form

$$(p_a^2 + M^2)|k\rangle_D = 0. \qquad (2.20)$$

From this it can be seen that vectors of the basis set $|k\rangle_D$ are split into two series of vectors $|k\pm\rangle_D$, each parametrized by a single continuous real parameter. For instance,

$$p^1|k\pm\rangle_D = \pm k|k\pm\rangle_D,$$
$$p^0|k\pm\rangle_D = \pm\sqrt{k^2+M^2}|k\pm\rangle_D, \qquad (2.21)$$
$$-\infty < k < +\infty.$$

Since the operators p^a are Hermitian, the scalar products

$$\langle k\pm|k'\pm\rangle_D = \delta(k-k'), \quad \langle k-|k'+\rangle_D = 0 \qquad (2.22)$$

are self-consistent.

Quantization conditions similar to (2.19) were used in [1, 2, 4, 6] and much earlier by Dirac in electrodynamics (see [8]).

Note. We stress that the variables α_n^a, $\bar{\alpha}_n^a$, $n \neq 0$, being linear operators in the space H_G, are not generally operators in the space H_{PD}. In fact, as a result of the action of the operators $\alpha_n^{(+)}$ and $\bar{\alpha}_n^{(+)}$ on the vectors $|k-\rangle_D$ we obtain vectors which do not belong to the physical space H_{PD}.

As a result of (2.13) and (2.19) the following equalities hold

$$L_n|p\rangle_D = \bar{L}_n|p\rangle_D = 0, \quad |p\rangle_D \in H_{PD}. \qquad (2.23)$$

Hence the model (2.6) is quantized using the first method.

We shall now quantize this model using the second method. We postulate that the entire state space H_C in which the initial variables act, is expressed as the tensor product

$$H_C = H_G \otimes H_P. \qquad (2.24)$$

Here the space H_G is defined in accordance with (2.17), (2.18). The space H_P has a basis with the properties (2.20)–(2.22). If the vector is $|p\rangle \in H_P$, it satisfies the conditions (2.19) with $m = 0$.

We draw attention to the fact that the operators $\alpha_n^{(+)}$ and $\bar{\alpha}_n^{(+)}$ with $n \neq 0$ (or combinations of these) act in the space H_G but their action is not determined on any one vector from the space H_P. This distinguishes H_P from H_{PD} (see (2.19)). Thus, the entire state space (2.24) is the tensor product of the spaces in which the corresponding operators act. Quite clearly, the space (2.24) has a positive definite scalar product.

For the following calculations we need to determine the ordering of the operators. The ordering (2.15) is equivalent to the ordering

$$L_0 = \frac{1}{2}(p_a^2 + M^2) - \sum_{m>0}(\alpha_m^0\alpha_{-m}^0 - \alpha_{-m}^1\alpha_m^1), \qquad (2.25)$$

which we shall use subsequently.

In our opinion, in this particular model the most convenient physical states satisfying the conditions (2.9) are states which are coherent in terms of gauge degrees of freedom. In H_G space we analyze the coherent state

$$|z, \bar{z}; G\rangle = \prod_{m>0}\exp\Biggl\{-\frac{1}{2m}(|z_{-m}^0|^2 + |z_m^1|^2 + |\bar{z}_{-m}^0|^2 + |\bar{z}_m^2|^2) \qquad (2.26)$$
$$+ \frac{1}{m}(z_{-m}^0\alpha_m^0 + z_m^1\alpha_{-m}^1 + \bar{z}_{-m}^0\bar{\alpha}_m^0 + \bar{z}_m^1\bar{\alpha}_{-m}^1)\Biggr\}|0; G\rangle.$$

Here z_m^a and $\bar{z}_m^a$ are complex numbers. We subsequently assume that $z_0^a = \bar{z}_0^a$ and $z_m^{a*} = z_{-m}^a$. The asterisk indicates complex conjugation. By definition we have

$$z_m^{(\pm)} = z_m^0 \pm z_m^1, \quad \bar{z}_m^{(\pm)} = \bar{z}_m^{(0)} \pm \bar{z}_m^1.$$

We also introduce the notation

$$\tilde{z}_m^{(-)} = z_m^{(-)}, \quad \tilde{\bar{z}}_m^{(-)} = \bar{z}_m^{(-)}, \quad m \neq 0,$$
$$\tilde{z}_0^{(-)} = z_0^{(-)} - \frac{M^2}{z_0^{(+)}}. \qquad (2.27)$$

Throughout this article it is assumed that $z_0^0 > 0$.

We denote the basis vectors in H_P space as $|z_0\pm\rangle_P$, $z_0^a \equiv k^a$. We have

$$p^a|z_0\pm\rangle_P = \pm z_0^a|z_0\pm\rangle_P, \quad \tilde{z}_0^{(-)} = 0. \qquad (2.28)$$

The last equality in (2.28) is a consequence of the relationship (2.19) with $m = 0$.

Let us assume that the sets of complex numbers $\{z, \bar{z}\}$ satisfy the equations (2.28) with the upper sign and

$$L_n(z) \equiv -\frac{1}{2}\sum_m z_{n-m}^{(+)}\tilde{z}_m^{(-)} = 0, \qquad (2.29)$$
$$L_n(\bar{z}) = 0, \quad n = 0, \pm 1, \ldots$$

Using the notation

$$z^{(+)}(\sigma) \equiv \frac{1}{\sqrt{2\pi}}\sum_n e^{in\sigma}z_n^{(+)},$$

$$\bar{z}^{(+)}(\sigma) \equiv \frac{1}{\sqrt{2\pi}}\sum_n e^{-in\sigma}\bar{z}_n^{(+)}$$

these equations can be rewritten in the more convenient form:

$$z^{(+)}(\sigma)\tilde{z}^{(-)}(\sigma) = 0, \quad \bar{z}^{(+)}(\sigma)\tilde{\bar{z}}^{(-)}(\sigma) = 0. \qquad (2.29')$$

The functions $z^{(+)}(\sigma)$, $\tilde{z}^{(-)}(\sigma)$, $\bar{z}^{(+)}(\sigma)$ and $\tilde{\bar{z}}^{(-)}(\sigma)$ are real and periodic and their zero harmonics satisfy (2.27) and (2.28).

The states

$$|z, \bar{z}\pm\rangle_P \equiv |\pm z, \pm\bar{z};\, G\rangle \otimes |z_0\pm\rangle_P \tag{2.30}$$

are called basis physical states if conditions (2.28) and (2.29') are satisfied.

According to these definitions, the basis physical states have the following properties:

$$\begin{aligned} \alpha^0_{-m}|z, \bar{z}\pm\rangle_P &= \pm z^0_{-m}|z, \bar{z}\pm\rangle_P, \\ \alpha^1_m|z, \bar{z}\pm\rangle_P &= \pm z^1_m|z, \bar{z}\pm\rangle_P, \quad m \ge 0. \end{aligned} \tag{2.31}$$

It rapidly follows from formulas (2.25), (2.29), and (2.31) that in our case the conditions (2.9) are satisfied:

$$\begin{aligned} \langle z, \bar{z}\pm|L_n|z, \bar{z}\pm\rangle_P &= 0, \\ \langle z, \bar{z}\pm|\bar{L}_n|z, \bar{z}\pm\rangle_P &= 0. \end{aligned} \tag{2.32}$$

We check whether the self-consistency conditions (2.10) are satisfied. For this it is sufficient to confirm that

$$\langle z, \bar{z}\pm|(L_nL_{-n} - L_{-n}L_n)|z, \bar{z}\pm\rangle_P = 0. \tag{2.33}$$

Simple calculations show that

$$\begin{aligned} L_nL_{-n} = \frac{1}{2}\sum_{m=1}^{n} m(n-m) + n(\tilde{\alpha}^1_0)^2 + 2n\sum_{m=1}^{n}\tilde{\alpha}^1_{-m}\tilde{\alpha}^1_m \\ + \sum_{m=n+1}^{\infty}(n+m)\tilde{\alpha}^1_{-m}\tilde{\alpha}^1_m \\ + \sum_{m=n+1}^{\infty}(m-n)\tilde{\alpha}^0_m\tilde{\alpha}^0_{-m} + :L_nL_{-n}:. \end{aligned} \tag{2.34}$$

Similarly

$$\begin{aligned} L_{-n}L_n = \frac{1}{2}\sum_{m=1}^{n} m(n-m) + n(\tilde{\alpha}^0_0)^2 + 2n\sum_{m=1}^{n}\tilde{\alpha}^0_m\tilde{\alpha}^0_{-m} \\ + \sum_{m=n+1}^{\infty}(n+m)\tilde{\alpha}^0_m\tilde{\alpha}^0_{-m} \\ + \sum_{m=n+1}^{\infty}(m-n)\tilde{\alpha}^1_{-m}\tilde{\alpha}^1_m + :L_{-n}L_n:. \end{aligned} \tag{2.35}$$

Here the operators $\tilde{\alpha}^a$ are expressed in terms of the operators $\alpha^{(+)}$, $\tilde{\alpha}^{(-)}$ in the same way that the operators α^a are expressed in terms of the operators $\alpha^{(+)}$, $\alpha^{(-)}$.

Since $:L_nL_{-n}:\equiv:L_{-n}L_n:$, from the last two equalities we have

$$L_nl_{-n} - L_{-n}L_n = 2nL_0. \tag{2.36}$$

The ordering on the right-hand side of equality (2.36) is defined in accordance with (2.25). It can be seen from (2.36) that the equations (2.33) are satisfied, i.e., the self-consistency conditions (2.10) are satisfied.

Note that generally

$$\langle z, \bar{z}\pm|L_n|z, \bar{z}\pm\rangle_P \ne 0. \tag{2.37}$$

We can see that the second method also yields a self-consistent quantum theory of the model (2.6).

We shall briefly discussion the superposition principle in the second method of quantization.

We assume that the states $|z, \bar{z}\pm\rangle_P$ and $|z', \bar{z}'\pm\rangle_P$ are physical. Is the state

$$|z, \bar{z}\pm\rangle_P + |z', \bar{z}'\pm\rangle_P \tag{2.38}$$

physical?

In our view, the superposition principle need not necessarily be extended to nonphysical gauge degrees of freedom. Thus, if in more complex theories using the second method of quantization, the superposition principle in H_G space is bounded, in our opinion this does not invalidate the method. In physical state space the superposition principle is fully obeyed.

3. INCLUSION OF MATTER

It can be seen from the commutation relations (1.17) and the anti-commutation relations (1.18) that boson and fermion fields have the following expansions in terms of modes [cf. (2.2)–(2.4)]:

$$\begin{aligned} f(\sigma) &= \frac{x}{\sqrt{4\pi}} + \frac{i}{\sqrt{4\pi}}\sum_{n\ne0}\frac{1}{n}(\alpha_n e^{in\sigma} + \bar{\alpha}_n e^{-in\sigma}), \\ \pi(\sigma) &= \frac{p}{\sqrt{\pi}} + \frac{1}{\sqrt{4\pi}}\sum_{n\ne0}(\alpha_n e^{in\sigma} + \bar{\alpha}_n e^{-in\sigma}), \end{aligned} \tag{3.1}$$

$$\begin{aligned} \phi(\sigma) &= \frac{1}{\sqrt{2\pi}}\sum_n \beta_n e^{in\sigma}, \\ \chi(\sigma) &= \frac{1}{\sqrt{2\pi}}\sum_n \bar{\beta}_n e^{-in\sigma}. \end{aligned} \tag{3.2}$$

We further assume that $\alpha_0 \equiv \bar{\alpha}_0 \equiv p$. Since the fields (3.1) and (3.2) are real, we have

$$\begin{aligned} x^\dagger = x, \quad \alpha^\dagger_n = \alpha_{-n}, \quad \bar{\alpha}^\dagger_n = \bar{\alpha}_{-n}, \\ \beta^\dagger_n = \beta_{-n}, \quad \bar{\beta}^\dagger_n = \bar{\beta}_{-n}. \end{aligned} \tag{3.3}$$

The commutation relationships (1.17) and (1.18) are equivalent to

$$[\alpha_m, \alpha_n] = [\bar{\alpha}_m, \bar{\alpha}_n] = m\delta_{m+n}, \quad [x, p] = i, \tag{3.4}$$

$$\{\beta_m, \beta_n\} = \{\bar{\beta}_m, \bar{\beta}_n\} = \delta_{m+n}. \tag{3.5}$$

We only write the nonzero commutation relationships. We shall subsequently denote the operators (2.6) by $L_n^{(0)}$ and $\bar{L}_n^{(0)}$, respectively. Taking into account the contribution of the material degrees of freedom, the Fourier components (2.5) have the form

$$\begin{aligned} L_n &= L_n^{(0)} \\ &+ \frac{1}{2}\sum_m\left[\alpha_{n-m}\alpha_m + \left(m-\frac{n}{2}\right)\beta_{n-m}\beta_m\right]. \end{aligned} \tag{3.6}$$

It is convenient to begin constructing the unitary transformation indicated at the end of the Introduction, which solves the quantization problem, by defining the creation and annihilation operators of the field (1.21). In other words, our first task is to construct boson and fermion creation and annihilation operators of matter which commute with all the operators (3.6). We can see that the problem with slightly weaker conditions has a solution. This is sufficient for our purposes.

We shall analyze the "gravitationally dressed" operators of the material fields:

$$A_m = \sum_n \mathcal{M}_{m,n}\alpha_n, \quad \bar{A}_m = \sum_n \bar{\mathcal{M}}_{m,n}\bar{\alpha}_n, \tag{3.7}$$

$$B_m = \sum_n {}^F\mathcal{M}_{m,n}\beta_n, \quad \bar{B}_m = \sum_n {}^F\bar{\mathcal{M}}_{m,n}\bar{\beta}_n. \tag{3.8}$$

The infinite-dimensional matrices $\mathcal{M}_{m,n}$, $\bar{\mathcal{M}}_{m,n}$, ${}^F\bar{\mathcal{M}}_{m,n}$, and ${}^F\mathcal{M}_{m,n}$ in (3.7) and (3.8) are determined in the Appendix. The elements of these matrices depend on the operators (x_+/p_+, $\alpha_m^{(+)}/p_+$, $\bar{\alpha}_m^{(+)}/p_+$) whose relative commutators are zero. Thus, all the matrix elements in (3.7) and (3.8) mutually commute

It is easy to check by means of direct calculations that the nonzero commutators of the operators (3.7) and (3.8) have the following form:

$$[A_m, A_n] = [\bar{A}_m, \bar{A}_n] = m\delta_{m+n}, \tag{3.9a}$$

$$\{B_m, B_n\} = \{\bar{B}_m, \bar{B}_n\} = \delta_{m+n}. \tag{3.9b}$$

Using (3.4), (3.7), and (A.14) we find

$$\begin{aligned} [A_m, A_n] &= \sum_l l\mathcal{M}_{m,l}\mathcal{M}_{n,-l} \\ &= -n\sum_l \mathcal{M}_{m,l}\mathcal{M}^{-1}_{l,-n} = m\delta_{m+n}. \end{aligned} \tag{3.10}$$

The equalities (3.9a) are thereby established. As a result of (3.5), (3.8), and (A.14') we have

$$\{B_m, B_n\} = \sum_l {}^F\mathcal{M}_{m,l}\,{}^F\mathcal{M}_{n,-l} = \sum_l {}^F\mathcal{M}_{m,l}\,{}^F\mathcal{M}^{-1}_{l,-n}. \tag{3.11}$$

From this it follows that the commutation relations (3.9b) are valid.

The operators (3.7) and (3.8) introduced here differ only negligibly from the DDF operators used in the string theory (see [9, 10]).

It is easy to see from these definitions that:

$$\begin{aligned} &[\alpha_m^{(+)}, A_n] = [\alpha_m^{(+)}, \bar{A}_n] \\ &= [\alpha_m^{(+)}, B_n] = [\alpha_m^{(+)}, \bar{B}_n] = 0. \end{aligned} \tag{3.12}$$

The relations (3.12) remain valid if instead of $\alpha_m^{(+)}$ we substitute $\bar{\alpha}_m^{(+)}$ or x_+.

Now, instead of $\alpha_m^{(-)}$ we must introduce the variables $\tilde{\alpha}_m^{(-)}$ into the theory, which conserve the previous form of the commutation relations with the variables $\alpha_m^{(+)}$ and have zero commutators with the new variables (3.7), (3.8).

From the definition (3.7) we find

$$[\alpha_m^{(-)}, A_n] = \sum_l [\alpha_m^{(-)}, \mathcal{M}_{n,l}]\alpha_l.$$

We use (A.16) and also reverse the inequalities (3.7). As a result, we obtain

$$[\alpha_m^{(-)}, A_n] = -\frac{2n}{p_+}\sum_p\left(\sum_l \mathcal{M}_{n,l-m}\mathcal{M}^{-1}_{l,p}\right)A_p, \quad m \neq 0.$$

Here the sum in brackets is calculated using formulas (A.3), (A.10), and (A.12). Thus, we find

$$\begin{aligned} [\alpha_m^{(-)}, A_n] &= -\frac{2n}{p_+}\sum_p \mathcal{M}^{-1}_{m,p-n}A_p, \quad m \neq 0, \\ [\alpha_0^{(-)}, A_n] &= -\frac{n}{p_+}A_n. \end{aligned} \tag{3.13}$$

Here the equality (3.11) was taken into account. The following relations also hold

$$\begin{aligned} [\alpha_0^{(-)}, \bar{A}_n] &= -\frac{n}{p_+}\bar{A}_n, \\ [\alpha_m^{(-)}, \bar{A}_n] &= 0, \quad m \neq 0. \end{aligned} \tag{3.14}$$

Similarly, using formulas (3.8), (A.4), (A.11), (A.12), and (A.14) we obtain

$$
\begin{aligned}
[\alpha_m^{(-)}, B_n] &= -\frac{1}{p_+}\sum_p (n+p)\mathcal{M}^{-1}_{m,\,p-n}B_p, \\
[\alpha_m^{(-)}, \bar{B}_n] &= 0, \quad m \neq 0, \\
[\alpha_0^{(-)}, B_n] &= -\frac{n}{p_+}B_n, \\
[\alpha_0^{(-)}, \bar{B}_n] &= -\frac{n}{p_+}\bar{B}_n.
\end{aligned}
\tag{3.15}
$$

It follows directly from formulas (3.13)–(3.15) and also (3.9) and (3.12) that the variables

$$
\begin{aligned}
\tilde{\alpha}_m^{(-)} &= \alpha_m^{(-)} - \frac{1}{p_+}\sum_{p,\,q}\mathcal{M}^{-1}_{m,\,p+q}\left(A_pA_q - \frac{p-q}{2}B_pB_q\right), \\
& m \neq 0, \\
\tilde{\alpha}_0^{(-)} \equiv \tilde{\bar{\alpha}}_0^{(-)} &= \alpha_0^{(-)} - \frac{1}{2p_+} \\
&\times \sum_p \{A_pA_{-p} + \bar{A}_p\bar{A}_{-p} + p(B_{-p}B_p + \bar{B}_{-p}\bar{B}_p) + 2M^2\}
\end{aligned}
\tag{3.16}
$$

commute with all the operators A_n, $\bar{A}_n$, B_n, and $\bar{B}_n$:

$$
\begin{aligned}
[\tilde{\alpha}_m^{(-)}, A_n] &= [\tilde{\alpha}_m^{(-)}, \bar{A}_n] \\
= [\tilde{\alpha}_m^{(-)}, B_n] &= [\tilde{\alpha}_m^{(-)}, \bar{B}_n] = 0.
\end{aligned}
\tag{3.17}
$$

The number M^2 in braces in (3.16) can also be taken as the result of normal ordering [cf. (2.14)].

If all quantities without a bar in formulas (3.13)–(3.17) are replaced by the same quantities with a bar and at the same time all quantities with a bar are replaced by the same quantities without a bar, these formulas remain valid.

We shall now determine the normal ordering of the creation and annihilation operators (3.7) and (3.8). These operators are by definition assumed to be normally ordered if all the creation operators $A_{-|n|}$, $\bar{A}_{-|n|}$, $B_{-|n|}$, $\bar{B}_{-|n|}$ are positioned to the left of all the annihilation operators $A_{|n|}$, $\bar{A}_{|n|}$, $B_{|n|}$, and $\bar{B}_{|n|}$.

The right-hand sides of the equalities (3.16) contain quadratic forms of the operators (3.7) and (3.8). These quadratic forms are represented as sums which are not normally ordered. However the right-hand sides of the equalities (3.16) can in fact be considered to be normally ordered, since the following equalities are satisfied

$$
\begin{aligned}
&\sum_{p,\,q}\mathcal{M}^{-1}_{m,\,p+q}\left(A_pA_q - \frac{p-q}{2}B_pB_q\right) \\
&=: \sum_{p,\,q}\mathcal{M}^{-1}_{m,\,p+q}\left(A_pA_q - \frac{p-q}{2}B_pB_q\right) :.
\end{aligned}
\tag{3.18}
$$

In order to prove the equalities (3.18) it is sufficient to establish that

$$
\begin{aligned}
\sum_p (A_pA_{-p} + pB_{-p}B_p) &= \sum_p : (A_pA_{-p} + pB_{-p}B_p) \\
&:=: \sum_p (A_pA_{-p} + pB_{-p}B_p) :.
\end{aligned}
\tag{3.19}
$$

The first equality in (3.19) follows directly from the definition of ordering and the commutation relations (3.9). In order to prove the second equality in (3.19), it should be borne in mind that formally

$$
\sum_{n=1}^{\infty} n = \zeta(-1),
$$

where $\zeta(s)$ is a Riemann zeta function:

$$
\zeta(s) = \sum_{n=1}^{\infty} n^{-s}. \tag{3.20}
$$

The zeta function has a unique analytic continuation to the point $s = -1$ where $\zeta(-1) = -1/12$. This regularization of the divergent sum $\sum_{n=1}^{\infty} n$ is now generally accepted. Thus, we can assume that

$$
\left(\sum_{n=1}^{\infty} n\right) - \left(\sum_{n=1}^{\infty} n\right) = \zeta(-1) - \zeta(-1) = 0.
$$

Here the first divergent sum appears as a result of the ordering of the boson operators and the second appears as a result of the ordering of the fermion operators. From this follows the second equality in (3.19).

Thus, the right-hand sides of the equalities (3.16) can equally well be taken to be normally ordered relative to the operators (3.7) and (3.8) or disordered and written in the form of sums contained in (3.16). For some calculations the disordered variant of the right-hand sides of (3.16) is more convenient.

We shall now prove the following commutation relation;

$$
[\tilde{\alpha}_m^{(-)}, \tilde{\alpha}_n^{(-)}] = [\tilde{\bar{\alpha}}_m^{(-)}, \tilde{\bar{\alpha}}_n^{(-)}] = [\tilde{\alpha}_m^{(-)}, \tilde{\bar{\alpha}}_n^{(-)}] = 0. \tag{3.21}
$$

Let $m \neq 0$ and $n \neq 0$. We take the variables $\tilde{\alpha}_m^{(-)}$ in the form (3.16). We then have

$$[\tilde{\alpha}_m^{(-)}, \tilde{\alpha}_n^{(-)}] = [\tilde{\alpha}_m^{(-)}, \alpha_n^{(-)}]$$
$$-\frac{1}{p_+}\sum_{p,q}[\tilde{\alpha}_m^{(-)}, \mathcal{M}^{-1}_{n,\,p+q}]\left(A_pA_q - \frac{p-q}{2}B_pB_q\right). \tag{3.22}$$

Here we use the definition (3.16) for $\tilde{\alpha}_n^{(-)}$ and the commutation relations (3.17). On the right-hand side of (3.22) we replace $\tilde{\alpha}_m^{(-)}$ by its value as given by (3.16)

$$[\tilde{\alpha}_m^{(-)}, \tilde{\alpha}_n^{(-)}]$$
$$= \frac{1}{p_+}\sum_{p,q}\{[\tilde{\alpha}_n^{(-)}, \mathcal{M}^{-1}_{m,\,p+q}] - [\tilde{\alpha}_m^{(-)}, \mathcal{M}^{-1}_{n,\,p+q}]\}$$
$$\times\left(A_pA_q - \frac{p-q}{2}B_pB_q\right)$$
$$+\frac{1}{p_+}\sum_{p,q}\mathcal{M}^{-1}_{m,\,p+q}\left\{[\alpha_n, A_pA_q] - \frac{p-q}{2}[\alpha_n^{(-)}, B_pB_q]\right\}.$$

Using formulas (A.17), (3.13), and (3.15), after redefining the notation of the indices in some sums we transform this last expression to give

$$\frac{1}{p_+^2}\left\{\sum_{p,q}\left[2m\mathcal{M}_{-(p+q),\,-(m+n)}\left(A_pA_q - \frac{p-q}{2}B_pB_q\right)\right.\right.$$
$$-2\sum_r r\mathcal{M}^{-1}_{m,\,q+r}\mathcal{M}^{-1}_{n,\,p-r}A_pA_q \tag{3.23}$$
$$\left.\left.+\frac{1}{2}\sum_r (r+p)(r-q)\mathcal{M}^{-1}_{m,\,q+r}\mathcal{M}^{-1}_{n,\,p-r}B_pB_q\right]\right\}$$
$$-\frac{1}{p_+^2}\{m \longleftrightarrow n\}.$$

Here the components of the sums over r which are antisymmetric in terms of the indices m and n are obtained using the relations (A.18). As a result, all the terms in (3.23) are mutually reduced. Thus, we have proven that the commutator (3.22) is zero. The relations

$$[\tilde{\bar{\alpha}}_m^{(-)}, \tilde{\bar{\alpha}}_n^{(-)}] = 0, \quad m \neq 0, \quad n \neq 0,$$

are proven by exactly repeating these procedures.

Similarly, it is established that

$$[\bar{\alpha}_0^{(-)}, \tilde{\alpha}_m^{(-)}] = [\tilde{\alpha}_0^{(-)}, \tilde{\bar{\alpha}}_m^{(-)}] = 0.$$

The equalities

$$[\tilde{\alpha}_m^{(-)}, \tilde{\bar{\alpha}}_n^{(-)}] = 0, \quad m \neq 0, \quad n \neq 0$$

follow trivially from the fundamental commutation relations (3.4), (3.5) and the definitions (3.16). The validity of the commutation relations (3.21) is thus completely proven. In addition, from the definitions (3.16) and the commutation relations (3.12) we have

$$\begin{gathered}[\alpha_m^{(+)}, \tilde{\alpha}_n^{(+)}] = [\bar{\alpha}_m^{(+)}, \tilde{\bar{\alpha}}_n^{(-)}] = -2m\delta_{m+n},\\ [x_+, \tilde{\alpha}_n^{(-)}] = [\tilde{x}_-, \tilde{\alpha}_n^{(+)}] = -2i\delta_n.\end{gathered} \tag{3.24}$$

The explicit form of the new variable $\tilde{x}_-$ is not given here since this variable is not used below.

The initial variables $\{x_\pm, \alpha_m^{(\pm)}, \bar{\alpha}_m^{(\pm)}, \alpha_m, \bar{\alpha}_m, \beta_m, \bar{\beta}_m\}$, (or more accurately their linear combinations) are canonical. It follows from the commutation relations (3.9), (3.12), (3.17), and (3.21) that the set of variables

$$\{x_+, \tilde{x}_-, \alpha_m^{(+)}, \tilde{\alpha}_m^{(-)}, \bar{\alpha}_m^{(+)}, \tilde{\bar{\alpha}}_m^{(-)}, A_m, \bar{A}_m, B_m, \bar{B}_m\} \tag{3.25}$$

is also canonical.

We shall now determine the unitary transformation appearing in (1.21). We define the unitary operator U using the following equalities:

$$\begin{gathered}Ux_+ = x_+U, \quad Ux_- = \tilde{x}_-U,\\ U\alpha_m^{(+)} = \alpha_m^{(+)}U, \quad U\alpha_m^{(-)} = \tilde{\alpha}_m^{(-)}U,\\ U\alpha_m = A_mU, \quad U\beta_m = B_mU,\\ U\bar{\alpha}_m^{(+)} = \bar{\alpha}_m^{(+)}U, \ldots\end{gathered} \tag{3.26}$$

and so on for the other operators with a bar. We know that equalities of the type (3.26) uniquely determine the linear operator U and this operator is unitary [11].

We must express the operators (3.6) in terms of the new variables. To do this we represent the operators $L_n^{(0)}$ from (3.6) in the form (2.13) and express the operators $\alpha_m^{(-)}$, α_m, and β_m in terms of the operators x_+, $\alpha_m^{(+)}$, $\tilde{\alpha}_m^{(-)}$, A_m, and B_m using formulas (3.7), (3.8), and (3.16). As a result of simple calculations using the sum rules (A.20) and (A.21) we arrive at the following formulas:

$$L_n = -\frac{1}{2}\sum_m \alpha_{n-m}^{(+)}\tilde{\alpha}_m^{(-)} + \frac{1}{4p_+}\alpha_n^{(+)}\mathcal{N},$$
$$\bar{L}_n = -\frac{1}{2}\sum_m \bar{\alpha}_{n-m}^{(+)}\tilde{\bar{\alpha}}_m^{(-)} - \frac{1}{4p_+}\bar{\alpha}_n^{(+)}\mathcal{N}, \tag{3.27}$$
$$\mathcal{N} = \sum_l \{A_{-l}A_l - \bar{A}_{-l}\bar{A}_l + l(B_{-l}B_l - \bar{B}_{-l}\bar{B}_l)\}.$$

Repeating the reasoning put forward to prove (3.19), we arrive at the equality

$$\mathcal{N} =: \mathcal{N} :. \tag{3.28}$$

The validity of the equalities (3.27) can be confirmed by a simpler method. This involves calculating the commutators of the operators (3.7), (3.8), (3.16) with the operators L_n in the representations (3.26) and (3.27). The results of these calculations agree.

Using (3.26) and (3.27), we find that

$$L_n = U\left\{L_n^{(0)} + \frac{1}{2p_+}\alpha_n^{(+)}\mathcal{N}_0\right\}U^\dagger,$$

$$\bar{L}_n = U\left\{\bar{L}_n^{(0)} - \frac{1}{2p_+}\bar{\alpha}_n^{(+)}\mathcal{N}_0\right\}U^\dagger, \tag{3.29}$$

$$\mathcal{N}_0 = \sum_{l>0}[\alpha_{-l}\alpha_l - \bar{\alpha}_{-l}\bar{\alpha}_l + l(\beta_{-l}\beta_l - \bar{\beta}_{-l}\bar{\beta}_l)].$$

The relations (3.29) are an exact variant of (1.20). Although the formulas (3.29) are slightly more complex than (1.20), the new method of quantization based on the assumptions put forward at the end of the Introduction can be developed further.

We note that the operators (3.7), (3.8) used here differ substantially from the gravitationally dressed operators in [2]. For the operators (3.7), (3.8) we have

$$[A_m, L_n] = \frac{m}{2p_+}\alpha_n^{(+)}A_m \tag{3.30}$$

and so on. We stress that it is impossible to construct a set of operators which can be expressed linearly in terms of the operators of the material fields and also commute with all the operators L_n and $\bar{L}_n$. (A similar situation is encountered in closed string theory when the transverse degrees of freedom are described by *DDF* operators.) A different approach to the model being studied was applied in [2]: the authors introduced new operators L'_n and $\bar{L}'_n$ which differ from the operators L_n and $\bar{L}_n$ by values proportional to the Planck constant. The operators L'_n and $\bar{L}'_n$ have the same algebra as L_n and $\bar{L}_n$. At the same time there is a set of gravitationally dressed operators $\{C_n, \bar{C}_n\}$ which describe the degrees of freedom of the material fields and are linearly coupled to them, and also commute with all operators L'_n and $\bar{L}'_n$. This factor simplifies the formal quantization procedure. However, significant difficulties are encountered when we attempt to express the initial dynamic variables in terms of those operators used for quantization.

In the present paper, unlike [2], we give explicit formulas which can be used to express the initial variables in terms of the new variables (see (3.16)). This makes it possible to calculate the matrix elements of the metric tensor (1.2) (see Section 5).

4. PHYSICAL STATE SPACE

The formulas obtained in Sections 2 and 3 can be used to quantize the model being studied.

As a result of the existence of a unitary transformation having the properties (3.26) and (3.29), we can confirm that the state space of the system (3.6) is isomorphic with the state space of the two interacting systems: pure gravity and free fields (3.1), (3.2). Bearing in mind this unitary transformation, we can construct the physical state space directly in the theory with interaction.

We first apply the first quantization method. For this we define two families of states using the formulas [cf. (2.19)]

$$\tilde{\alpha}_m^{(-)}|k\pm\rangle_D = \tilde{\bar{\alpha}}_m^{(-)}|k\pm\rangle_D = 0, \qquad m = 0, \pm1, \ldots, \tag{4.1}$$

$$A_n|k\pm\rangle_D = \bar{A}_n|k\pm\rangle_D = B_n|k\pm\rangle_D = \bar{B}_n|k\pm\rangle_D = 0, \qquad n > 0. \tag{4.2}$$

We also impose the constraint

$$A_0|k\pm\rangle_D = \bar{A}_0|k\pm\rangle_D = 0,$$

which is not necessary and is merely used to simplify the formulas. In addition, we assume that relations (2.21) and (2.22) are satisfied.

The reason why the quantity $M^2 > 0$ was introduced in formulas (2.14) and (3.16) now becomes clear: as a result of this constraint and condition (4.1) with $m = 0$ the operator p_+ has no zero eigenvalues in the physical space H_{PD}. Thus, the unitary transformation (3.26) is defined correctly and the operators A_n can act on the states $|k\pm\rangle_D$.

All the physical states are linear combinations of basis states having the form

$$|k\pm; n_i, \bar{n}_i, m_i, \bar{m}_i\rangle_D = (A_{-n_1}\ldots\bar{A}_{-\bar{n}_1}\ldots B_{-m_1}\ldots\bar{B}_{-\bar{m}_1}\ldots)|k\pm\rangle_D \in H_{PD}^{(\pm)}, \qquad n_i, \bar{n}_i, m_i, \bar{m}_i > 0, \tag{4.3}$$

$$\left(\sum_i n_i + \sum_i m_i - \sum_i \bar{n}_i - \sum_i \bar{m}_i\right) = 0. \tag{4.4}$$

The entire physical state space is expressed as the direct sum

$$H_{PD} = H_{PD}^{(+)} \otimes H_{PD}^{(-)}. \tag{4.5}$$

As a result of relations (4.1), (4.4), and (3.9), we have

$$\mathcal{N}|\ \rangle_P = 0, \quad |\ \rangle_P \in H_{PD}. \tag{4.6}$$

Now using (3.27), (4.1), (4.2), and (4.6) we obtain

$$L_n|\ \rangle_P = \bar{L}_n|\ \rangle_P = 0. \tag{4.7}$$

It follows from the commutation relations (3.9) and equalities (4.2), (2.22) that the scalar product in the $H_{PD}^{(+)}$ and $H_{PD}^{(-)}$ states is positive definite and these states are mutually orthogonal.

This implies that an anomaly-free quantization of the system (3.6) has been performed using the first method.

We now apply the second method of quantization. By definition, the state space is generated by two series of states $|z, \bar{z}\pm\rangle_P$ having the following properties. The states $|z, \bar{z}\pm\rangle_P$ satisfy equations (2.28), (4.2) and also [cf. (2.31)] the equalities

$$\begin{gathered}\frac{1}{2}(\alpha_{-m}^{(+)} + \tilde{\alpha}_{-m}^{(-)})|z, \bar{z}\pm\rangle_P = \pm z_{-m}^0|z, \bar{z}\pm\rangle_P,\\ \frac{1}{2}(\alpha_m^{(+)} - \tilde{\alpha}_m^{(-)})|z, \bar{z}\pm\rangle_P = \pm z_m^1|z, \bar{z}\pm\rangle_P,\\ m > 0,\end{gathered} \tag{4.8}$$

and similarly for the quantities with a bar.

It follows from (2.28) that

$$\tilde{\alpha}_0^{(-)}|z, \bar{z}\pm\rangle_P = 0. \tag{4.9}$$

The basis states in the physical space are denoted by $|z, \bar{z}\pm;\ n_i, \bar{n}_i, m_i, \bar{m}_i\rangle_P$. They are constructed using the operators $A_{-m}, \ldots, m > 0$ in accordance with (4.3) and (4.4). The physical state space H_P is expanded as a direct sum of the orthogonal subspaces $H_P^{(+)}$ and $H_P^{(-)}$. The scalar product in the H_P space is positive definite.

As a result of the commutation relations (3.12) and (3.17), equations (4.8) and (4.9) also hold for physical states if these physical states are "pure" in relation to the gauge degrees of freedom. Here we take "purity" to mean that all the physical states have the same set of parameters $\{z_m^a, \bar{z}_m^a\}$.

We shall not demonstrate in detail that in this case the quantization conditions (2.9) and (2.10) are satisfied. This follows directly from all the reasoning put forward above. Here we merely stress the following important fact: averaging in (2.9) and (2.10) can only be performed in the space of the gauge degrees of freedom H_G [see (2.17), (2.18)]; averaging in (2.9) and (2.10) need not be performed over variables described by the operators A_m, $\bar{A}_m$, B_m, $\bar{B}_m$.

5. CALCULATION OF AVERAGES

In the model studied the most interesting quantity is the average value of the metric tensor (1.2) relative to the physical states. For this, in accordance with (1.2), we need to calculate the average of the expression $\exp(2\rho)$ since the parameters u and v are Lagrange numerical factors.

In order to begin our calculations we postulate that the formulas (1.15) valid in classical theory also hold in quantum theory. This assumption allows us to express the unknown quantity in terms of the quantum fields π^a, r^a (2.2) as follows:

$$\frac{\lambda}{\pi G}e^{2\rho} = \pi_a\pi^a - r'_a r^{a\prime} + 2(\pi^0 r^{1\prime} - \pi^1 r^{0\prime}).$$

Taking into account (2.2) this inequality is rewritten in the convenient form for us:

$$e^{2\rho(\sigma)} = -\frac{G}{\lambda}\left(\sum_m \alpha_m^{(+)} e^{im\sigma}\right)\left(\sum_n \bar{\alpha}_n^{(-)} e^{-im\sigma}\right). \tag{5.1}$$

In (5.1) the variables $\bar{\alpha}_n^{(-)}$ should be expressed in terms of the new variables $\bar{\alpha}_n^{(+)}$, $\tilde{\bar{\alpha}}_n^{(-)}$ and $\bar{A}_m$, $\bar{B}_m$, A_m, and B_m in accordance with (3.16). Then we can calculate the average values of the expression (5.1) using the results of the previous section.

On the right-hand side of (5.1) all the variables in the first set of brackets commute with all the variables in the second set of brackets. Bearing this in mind and also the results of Section 4, we can confirm that the following formula is valid to calculate the averages relative to the basis vectors

$$\langle e^{2\rho}\rangle = -\frac{G}{\lambda}\left\langle \sum_m \alpha_m^{(+)} e^{im\sigma}\right\rangle\left\langle \sum_n \bar{\alpha}_n^{(-)} e^{-im\sigma}\right\rangle. \tag{5.2}$$

We use the second method of quantization to calculate the averages in expression (5.2). We first calculate the averages relative to the ground state $|z, \bar{z}\pm\rangle_P$. We express the variables $\alpha_m^{(+)}$ in the form

$$\begin{gathered}\alpha_m^{(+)} = a_m^0 + a_m^1, \quad a_m^0 = \frac{1}{2}(\alpha_m^{(+)} + \tilde{\alpha}_m^{(-)}),\\ a_m^1 = \frac{1}{2}(\alpha_m^{(+)} + \tilde{\alpha}_m^{(-)})\end{gathered} \tag{5.3}$$

and use formulas (4.8) and (2.28). Thus, we obtain

$$\left\langle z, \bar{z}+\left|\sum_m \alpha_m^{(+)} e^{im\sigma}\right|z, \bar{z}+\right\rangle_P = \sum_m z_m^{(+)} e^{im\sigma}. \tag{5.4}$$

Similarly we find

$$\left\langle z, \bar{z}+\left|\sum_n \tilde{\bar{\alpha}}_n^{(-)} e^{-in\sigma}\right|z, \bar{z}+\right\rangle_P = \sum_n \tilde{\bar{z}}_n^{(-)} e^{-in\sigma}. \tag{5.5}$$

It follows from (3.16), (3.18), and (4.2) that

$$\left\langle z, \bar{z}+\left|(\bar{\alpha}_n^{(-)} - \tilde{\bar{\alpha}}_n^{(-)})\right|z, \bar{z}+\right\rangle_P = 0. \tag{5.6}$$

Combining formulas (5.2), (5.4), (5.5), (5.6), and (2.29'), we obtain

$$\left\langle z, \bar{z}+\left|e^{2\rho(\sigma)}\right|z, \bar{z}+\right\rangle_P = -\frac{2\pi G}{\lambda} z^{(+)}(\sigma)\tilde{\bar{z}}^{(-)}(\sigma). \qquad (5.7)$$

In order to find more complex matrix elements, we need to make additional calculations.

As a result of (A.2), we have

$$u^l = z^l \exp\left(il\frac{x_+}{2p_+}\right)\exp(\xi_+ + \xi_-),$$

$$\xi_+ = \frac{l}{p_+}\sum_{m>0}\frac{1}{m}(-z^{-m}a_m^0 + z^m a_{-m}^1), \qquad (5.8)$$

$$\xi_- = \frac{l}{p_+}\sum_{m>0}\frac{1}{m}(z^m a_{-m}^0 - z^{-m}a_m^1).$$

Here a_m^0 and a_m^1 are defined in accordance with (5.3). As a result of the relations (4.8), when calculating the matrix elements we need to order the operators in (5.8) so that the functions of ξ_+ are to the left of the functions of ξ_-. This is easily achieved since the commutator

$$\frac{1}{2}[\xi_+, \xi_-] = \left(\frac{l}{p_+}\right)^2 \sum_{m>0}\frac{1}{m} \qquad (5.9)$$

is a c-number which commutes with ξ_+ and ξ_-. Then using the well-known Baker–Hausdorff formula we obtain

$$u^l = \exp\left[-\left(\frac{l}{p_+}\right)^2\sum_{m>0}\frac{1}{m}\right] z^l \exp\left(il\frac{x_+}{2p_+}\right)e^{\xi_+}e^{\xi_-}. \qquad (5.10)$$

Now the averages of expression (5.10) are calculated using relations (4.8):

$$\left\langle z, \bar{z}+\left|u^l\right|z, \bar{z}+\right\rangle_P \sim \exp\left[-\left(\frac{l}{p_+}\right)^2\sum_{m>0}\frac{1}{m}\right] \times \exp\left(-\frac{l}{p_+}\sum_{n\neq 0}\frac{z_n^{(+)}}{n}z^{-n}\right). \qquad (5.11)$$

Since the argument of the first exponential function on the right-hand side of (5.11) is an infinitely large negative number, the average of (5.11) is zero. This implies that in calculations of the matrix elements below the average sign the elements of the matrix $\mathcal{M}_{n,l}^{-1}$ are in fact only nonzero for $l = 0$. This fact which applies equally to the elements of the matrix $\overline{\mathcal{M}}_{n,l}^{-1}$ substantially simplifies the calculations. Using (A.9) and (A.10) we obtain

$$\mathcal{M}_{n,l}^{-1} = \frac{\bar{\alpha}_n^{(+)}}{p_+}. \qquad (5.12)$$

Bearing this observation in mind and also formula (A.12) we note that the matrix elements should be calculated using the following relations

$$\bar{\alpha}_m^{(-)} = \tilde{\bar{\alpha}}_m^{(-)} + \frac{\bar{\alpha}_m^{(+)}}{p_+^2} : \sum_p(\bar{A}_{-p}\bar{A}_p + p\bar{B}_{-p}\bar{B}_p):, \quad m \neq 0, \qquad (5.13)$$

and the second formula (3.16) (for $m = 0$). Using (5.2), (5.4), and (5.13), we can calculate the diagonal matrix element of the metric tensor relative to the basis state $|p\rangle = |z, \bar{z}+; n_i, \bar{n}_i, m_i, \bar{m}_i\rangle$ in the form (4.3), (4.4):

$$\left\langle p\left|e^{2\rho(\sigma)}\right|p\right\rangle = -\frac{2\pi G}{\lambda}z^{(+)}(\sigma) \times\left\{\tilde{\bar{z}}^{(-)}(\sigma) + \frac{\bar{z}^{(+)}(\sigma)}{(z_0^{(+)})^2}\left(\sum_i n_i + \sum_i m_i\right)\right\}. \qquad (5.14)$$

The functions $z^{(+)}(\sigma)$, $\bar{z}^{(+)}(\sigma)$, and $\tilde{\bar{z}}^{(-)}(\sigma)$ satisfy (2.29'). To derive formula (5.14) we used equality (4.4) and assumed that the basis vector is normalized.

Further calculations of averages together with their study and interpretation are outside the scope of the present study.

Calculations of the matrix elements from the metric tensor using the first method encounter serious difficulties. This is because the variables $\alpha_m^{(+)}$ are not operators in the physical state space H_{PD}. Thus averages having the form

$$\left\langle\sum_m \alpha_m^{(+)}e^{im\sigma}\right\rangle_{PD}, \quad \left\langle\sum_m \bar{\alpha}_m^{(+)}e^{-im\sigma}\right\rangle_{PD} \qquad (5.15)$$

are not determined. As a result of the quantization conditions (4.1) and (4.2) we could assume that the average (5.2) is zero if the matrix element is calculated relative to the generating states $|k\pm\rangle_D$. However, this does not rescue the general situation since averages of the type (5.15) (see (5.12) and (5.14)) must be calculated to calculate the matrix elements relative to the excited states.

6. CONCLUSIONS

In the present paper we have applied two methods of quantization to the theory of two-dimensional gravity. The first method using the Dirac approach allows us to achieve complete quantization. This implies

(a) constructing a physical state space with a positive definite scalar product;

(b) explicitly expressing physically meaningful quantities in terms of those operators used to construct the physical state space.

However the averages of the metric tensor cannot be calculated using the first method for fundamental reasons. This statement holds to a lesser extent when the physical state space is constructed using the operators A_n, B_n, …, determined in Section 3.

In addition to the problems (a) and (b) the second method of quantization can also solve the following problem:

(c) calculating the averages of the metric tensor relative to the physical states.

On the basis of these results we can conclude that the second method of quantization should be used subsequently to study other models.

To conclude we make the following observation.

The model (3.6) can easily be quantized in a "light cone" gauge. In the terms used in the present study using this gauge implies

imposing second-class constraints:

$$\alpha_m^{(+)} = 0, \quad \tilde{\alpha}_m^{(-)} = 0, \quad \bar{\alpha}_m^{(+)} = 0, \quad \tilde{\bar{\alpha}}_m^{(-)} = 0, \quad m \neq 0. \tag{6.1}$$

Then, in accordance with (5.2) and (5.6), the average of the metric tensor relative to the ground state is zero. On the other hand, in the second method of quantization, in accordance with (5.7) and (2.29'), the average of the metric tensor relative to the ground state is generally nonzero. From this it can be seen that imposing special calibration conditions which simplify the solution of the quantization problem should be avoided if possible. In our case, imposing the constraints (6.1) significantly degrades the results.

This work was supported by the VNSh program (grant No. 96-1596821).

APPENDIX

Let us assume that τ is some "time-like" parameter and $z = e^{i\tau}$. We introduce the following operator functions:

$$\begin{aligned} q(\tau) \equiv q(z) &= \frac{x_+}{2p_+} + \frac{1}{i}\ln z + \frac{i}{p_+}\sum_{n\neq 0}\frac{1}{n}\alpha_n^{(+)} z^{-n}, \\ \bar{q}(\tau) \equiv \bar{q}(z) &= \frac{x_+}{2p_+} + \frac{1}{i}\ln z + \frac{i}{p_+}\sum_{n\neq 0}\frac{1}{n}\bar{\alpha}_n^{(+)} z^{-n}. \end{aligned} \tag{A.1}$$

By definition we have

$$\begin{gathered} u(z) \equiv e^{iq(z)} = z\exp\left(i\frac{x_+}{2p_+}\right)\exp\left(-\sum_{n\neq 0}\frac{\alpha_n^{(+)}}{np_+}z^{-n}\right), \\ \bar{u}(z) \equiv e^{i\bar{q}(z)}. \end{gathered} \tag{A.2}$$

Let the contour C in the plane of the complex variable z go counterclockwise once around the point $z = 0$. We define four infinite-dimensional matrices according to the formulas

$$\mathcal{M}_{m,n} = \frac{1}{2\pi i}\oint_C \frac{dz}{z} z^{-n} u^m, \tag{A.3}$$

$${}^F\mathcal{M}_{m,n} = \frac{1}{2\pi i}\oint_C \frac{dz}{z} z^{-n} u^m \sqrt{\dot{q}}. \tag{A.4}$$

The definitions of the matrixes $\bar{\mathcal{M}}_{m,n}$ and ${}^F\bar{\mathcal{M}}_{m,n}$ are obtained from (A.3) and (A.4) by substituting $q \longrightarrow \bar{q}$ and $u \longrightarrow \bar{u}$. Here and subsequently we have

$$\dot{q} = \frac{d}{d\tau}q = iz\frac{d}{dz}q = 1 + \frac{1}{p_+}\sum_{n\neq 0}\alpha_n^{(+)} z^{-n}. \tag{A.5}$$

We draw attention to the fact that all the quantities (A.1)–(A.5) should be considered to be formal series in the elements of free associative commutative involutive algebra $\mathcal{A}^{(+)}$ with the generatrices $\{x_+/p_+, \alpha_m^{(+)}/p_+, \bar{\alpha}_m^{(+)}/p_+\}$. This assumption holds until any averages are calculated relative to physical states. The coefficients at the monomials relative to the generatrices of the algebra $\mathcal{A}^{(+)}$ in the expansions of (A.1)–(A.5) are finite polynomials in z and z^{-1}. Hence, the integrals in (A.3) and (A.4) are determined correctly. The matrix elements of the matrices (A.3) and (A.4) thus belong to the algebra $\mathcal{A}^{(+)}$.

From (A.2) we have

$$\operatorname*{var}_C \ln u(z) = \operatorname*{var}_C \ln z - \operatorname*{var}_C\left(\sum_{n\neq 0}\frac{1}{n}\frac{\alpha_n^{(+)}}{p_+}z^{-n}\right).$$

Here $\operatorname{var}_C F(z)$ implies the change in the function $F(z)$ (generally nonunique around the contour C) for a single counterclockwise circuit around the contour C. For this definition of the contour the second term on the right-hand side of the last equality makes no contribution and we have

$$\operatorname*{var}_C \ln u(z) = \operatorname*{var}_C \ln z = 2\pi i. \tag{A.6}$$

From (A.2) and (A.5) we find

$$\frac{du(z)}{dz} = \frac{u(z)}{z}\sum_{n=-\infty}^{+\infty}\frac{\alpha_n^{(+)}}{p_+}z^{-n} \neq 0. \tag{A.7}$$

This last inequality is a consequence of the fact that in $\mathcal{A}^{(+)}$ algebra there are no relations apart from those derived from its commutativeness.

The transform of the contour C in $\mathcal{A}^{(+)}$ algebra for the mapping (A.2) is denoted by C^*.

The mapping (A.2) can be inverted. This is a consequence of inequality (A.7). This inversion implies that there is an analytic function $z(u)$ of the variable u which inverts the equation (A.2) into an identity. The function $z(u)$ is a formal series of the variables $\{\alpha^{(+)}/p_+\}$. Equation (A.2) may be inverted by an iterative method in powers of the variables $\alpha^{(+)}/p_+$. Let us assume that

$$z(u) = z^{(0)}(u) + z^{(1)}(u) + \ldots,$$

where

$$z^{(i)}(u) = \sum_{n=-\infty}^{+\infty} z_n^{(i)} u^{-n},$$

and $z_n^{(i)}$ are homogeneous functions of the variables $\alpha^{(+)}/p_+$ to the power i. Using equation (A.2) we obtain

$$\begin{gathered}
z^{(0)}(u) = ue^{-i\delta}, \quad \delta = \frac{x_+}{2p_+}, \\
z^{(1)}(u) = ue^{-i\delta}\sum_{n \neq 0}\frac{\alpha_n^{(+)}}{np_+}(e^{-i\delta}u)^{-n}, \\
z^{(2)}(u) = ue^{-i\delta}\left\{\frac{1}{2}\left[\sum_{n \neq 0}\frac{\alpha_n^{(+)}}{np_+}(e^{-i\delta}u)^{-n}\right]^2\right. \\
\left.-\left[\sum_{n \neq 0}\frac{\alpha_n^{(+)}}{p_+}(e^{-i\delta}u)^{-n}\right]\left[\sum_{m \neq 0}\frac{\alpha_m^{(+)}}{mp_+}(e^{-i\delta}u)^{-m}\right]\right\}
\end{gathered} \tag{A.8}$$

and so on. Equation (A.2) uniquely determines each successive term $z^{(i)}(u)$ in terms of the previous ones.

It also follows from (A.6) that for a single counterclockwise circuit of the variable u around the circuit C^*, the variable z goes once counterclockwise around the circuit C. From this it follows that a closed integral over the variable z along the circuit C may be converted into a closed integral over the variable u along the circuit C^* and conversely. In accordance with (A.7) and (A.5) we have

$$\frac{du}{u} = \frac{dz}{z}\left\{1 + \sum_{n \neq 0}\frac{\alpha_n^{(+)}}{p_+}z^{-n}\right\} = \frac{dz}{z}\dot{q}. \tag{A.9}$$

The matrices which are the inverse of (A.3) and (A.4) are represented most simply in the form

$$\mathcal{M}_{n,l}^{-1} = \frac{1}{2\pi i}\oint_{C^*}\frac{du}{u}u^{-l}z^n, \tag{A.10}$$

$${}^F\mathcal{M}_{n,l}^{-1} = \frac{1}{2\pi i}\oint_{C^*}\frac{du}{u}u^{-l}z^n(\dot{q})^{-1/2}. \tag{A.11}$$

In order to prove the equalities $\mathcal{M}\mathcal{M}^{-1} = \mathcal{M}^{-1}\mathcal{M} = 1$, ${}^F\mathcal{M}{}^F\mathcal{M}^{-1} = {}^F\mathcal{M}^{-1F}\mathcal{M} = 1$ we use the identities

$$\begin{gathered}
\left(z\frac{d}{dz}\right)^s f(z, u(z)) = \frac{1}{2\pi i}\oint_C\frac{dz_1}{z_1}f(z_1, u(z_1))\sum_{n=-\infty}^{+\infty}n^s\left(\frac{z}{z_1}\right)^n, \\
\left(u\frac{d}{du}\right)^s f(u, z(u)) \\
= \frac{1}{2\pi i}\oint_{C^*}\frac{du_1}{u_1}f(u_1, z(u_1))\sum_{n=-\infty}^{+\infty}n^s\left(\frac{u}{u_1}\right)^n, \\
s = 0, 1, \ldots
\end{gathered} \tag{A.12}$$

In (A.12) the function $f(z, u)$ is expanded as a Laurent series in terms of its arguments and integration is performed in the counterclockwise direction. The identities (A.12) are based on the obvious equalities

$$\oint_C\frac{dz}{z}z^n = \oint_{C^*}\frac{du}{u}u^n = 2\pi i\delta_n.$$

As a result of (A.6) and (A.12) for $s = 0$ we have

$$\begin{gathered}
\sum_{l=-\infty}^{+\infty}\mathcal{M}_{m,l}\mathcal{M}_{l,n}^{-1} = \frac{1}{2\pi i}\oint_{C^*}\frac{du_1}{u_1}u_1^{-n} \\
\times\frac{1}{2\pi i}\oint_C\frac{dz}{z}u^m\sum_{l=-\infty}^{+\infty}\left(\frac{z_1}{z}\right)^l = \frac{1}{2\pi i}\oint_{C^*}\frac{du_1}{u_1}u_1^{m-n} = \delta_{m-n}.
\end{gathered} \tag{A.13}$$

In (A.13) it is implied that $u = u(z)$ and $z_1 = z(u_1)$. Thus, the equality $\mathcal{M}\mathcal{M}^{-1} = 1$ is established. The relations $\mathcal{M}^{-1}\mathcal{M} = 1$, ${}^F\mathcal{M}{}^F\mathcal{M}^{-1} = 1$, and ${}^F\mathcal{M}^{-1F}\mathcal{M} = 1$ are established similarly.

Formulas (A.3) and (A.10) directly yield the relation

$$l\mathcal{M}_{n,l}^{-1} = n\mathcal{M}_{-l,-n}, \tag{A.14}$$

which is valid for all l and n. From (A.4), (A.9), and (A.11) we obtain

$${}^F\mathcal{M}_{n,l}^{-1} = {}^F\mathcal{M}_{-l,-n}. \tag{A.14'}$$

The nonzero commutators $\alpha_m^{(-)}$ and $\bar{\alpha}_m^{(-)}$ with $u(z)$ and $\bar{u}(z)$ are obtained using (A.2) and (2.12):

$$\begin{gathered}
[\alpha_m^{(-)}, u(z)] = -\frac{2}{p_+}z^m u(z), \quad m \neq 0, \\
[\alpha_0^{(-)}, u(z)] = -\frac{1}{p_+}z^m u(z)
\end{gathered} \tag{A.15}$$

and similarly for barred quantities.

From the definitions (A.3) and (A.4) allowing for (A.15) it follows that

$$[\alpha_m^{(-)}, \mathcal{M}_{n,l}] = -\frac{2n}{p_+}\mathcal{M}_{n,l-m},$$

$$[\alpha_m^{(-)}, {}^F\mathcal{M}_{n,l}] = -\frac{1}{2\pi i p_+}$$

$$\times \oint\limits_C \frac{dz}{z} z^{-l+m} u^n \left(2n(\dot{q})^{1/2} + m(\dot{q})^{-1/2}\right), \qquad \text{(A.16)}$$

$$m \neq 0,$$

$$[\alpha_0^{(-)}, \mathcal{M}_{n,l}] = -\frac{n}{p_+}\mathcal{M}_{n,l},$$

$$[\alpha_0^{(-)}, {}^F\mathcal{M}_{n,l}] = -\frac{n}{p_+}{}^F\mathcal{M}_{n,l}.$$

The commutation relations (A.16) conserve their form if barred quantities are substituted everywhere in (A.16). The relations (A.16) exhaust all the nonzero commutators between $\alpha^{(-)}$, $\bar{\alpha}^{(-)}$, $\mathcal{M}$, $\overline{\mathcal{M}}$, ${}^F\mathcal{M}$, and ${}^F\overline{\mathcal{M}}$.

From (A.14) and (A.16) we then obtain

$$[\alpha_m^{(-)}, \mathcal{M}_{n,l}^{-1}] = \frac{2n}{p_+}\mathcal{M}_{-l,-(m+n)}, \quad m \neq 0,$$

$$[\alpha_0^{(-)}, \mathcal{M}_{n,l}^{-1}] = \frac{n}{p_+}\mathcal{M}_{-l,-n} = \frac{l}{p_+}\mathcal{M}_{n,l}^{-1}. \qquad \text{(A.17)}$$

In Section 3 we use the following formulas:

$$\sum_q \mathcal{M}_{m,l+q}^{-1}\mathcal{M}_{n,p-q}^{-1} = \mathcal{M}_{(m+n),(l+p)}^{-1}, \qquad \text{(A.18a)}$$

$$\sum_q q(\mathcal{M}_{m,l+q}^{-1}\mathcal{M}_{n,p-q}^{-1} - \mathcal{M}_{n,l+q}^{-1}\mathcal{M}_{m,p-q}^{-1})$$
$$= (m-n)\mathcal{M}_{-(l+p),-(m+n)}, \qquad \text{(A.18b)}$$

$$\sum_q q^2(\mathcal{M}_{m,l+q}^{-1}\mathcal{M}_{n,p-q}^{-1} - \mathcal{M}_{n,l+q}^{-1}\mathcal{M}_{m,p-q}^{-1})$$
$$= (m-n)(p-l)\mathcal{M}_{-(l+p),-(m+n)}. \qquad \text{(A.18c)}$$

Using formulas (A.10) and (A.12) for $s = 1$ we can obtain the following relation:

$$\sum_{q=-\infty}^{+\infty} q\mathcal{M}_{m,l+q}^{-1}\mathcal{M}_{n,p-q}^{-1} = \frac{1}{2\pi i}\oint\limits_{C^*} du\, u^{-p} z^n \frac{d}{du}(u^{-l}z^m). \qquad \text{(A.19)}$$

Since we are only interested in the antisymmetric component of the relation (A.19) in terms of the indices m and n, on the right-hand side we can remove u^{-l} from beneath the differentiation sign. Also bearing in mind that $(dz/du)du = dz$ we obtain (A.18b). The other formulas [(A.18a) and (A.18c)] are proven similarly.

The following formulas also hold:

$$\sum_{m=-\infty}^{+\infty} \mathcal{M}_{m,s}^{-1}\mathcal{M}_{n-m,r}^{-1} = \frac{1}{p_+}\sum_{m=-\infty}^{+\infty} \alpha_{n-m}^{(+)}\mathcal{M}_{m,r+s}^{-1}, \qquad \text{(A.20)}$$

$$\sum_l l({}^F\mathcal{M}_{n-l,p}^{-1}\,{}^F\mathcal{M}_{l,q}^{-1} - {}^F\mathcal{M}_{n-l,q}^{-1}\,{}^F\mathcal{M}_{l,p}^{-1})$$
$$= \frac{(q-p)}{p_+}\sum_l \alpha_l^{(+)}\mathcal{M}_{(n-l),(p+q)}^{-1}, \qquad \text{(A.21)}$$

which we derived using formulas (A.9), (A.10), and (A.12).

REFERENCES

1. R. Jackiw, E-print Archive, gr-qc/9612052.
2. E. Benedict, R. Jackiw, and H.-J. Lee, Phys. Rev. D **54**, 6213 (1996).
3. D. Cangemi, R. Jackiw, and B. Zwiebach, Ann. Phys. **245**, 408 (1996); D. Cangemi and R. Jackiw, Phys. Lett. B **337**, 271 (1994); D. Cangemi and R. Jackiw, Phys. Rev. D **50**, 3913 (1994); D. Amati, S. Elitzur, and E. Rabinovici, Nucl Phys. B **418**, 45 (1994); D. Louis-Martinez, J. Gegenberg, and G. Kunstatter, Phys. Lett. B **321**, 193 (1994); E. Benedict, Phys. Lett. B **340**, 43 (1994); T. Strobl, Phys. Rev. D **50**, 7346 (1994).
4. S. N. Vergeles, Zh. Éksp. Teor. Fiz. **113**, 1566 (1998).
5. O. Andreev, Phys. Rev. D **57**, 3725 (1998).
6. S. N. Vergeles, E-Print Archive, hep-th/9906024.
7. P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva Univ., New York, 1964).
8. P. A. M. Dirac, *Lectures on Quantum Field Theory* (Yeshiva Univ., New York, 1967; Mir, Moscow, 1971).
9. E. Del Guidice, P. Di Vecchia, and S. Fubini, Ann. Phys. **70**, 378 (1972).
10. M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge Univ. Press, Cambridge, 1987; Mir, Moscow, 1990).
11. P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1958; Nauka, Moscow, 1970).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 102–108.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 115–121.
Original Russian Text Copyright © 2000 by Eritsyan.

FLUIDS

Diffraction Reflection of Light in a Cholesteric Liquid Crystal in the Presence of Wave Irreversibility and Bragg Formula for Media with Nonidentical Forward and Return Wavelengths

O. S. Eritsyan
Yerevan State University, Yerevan, 375049 Armenia;
e-mail: physdep@moon.yerphi.am

Received October 22, 1998

Abstract—An analysis is made of the diffraction reflection of light in a cholesteric liquid crystal in the presence of magnetooptic activity which leads to wave irreversibility and in particular to nonidentical forward and return wavelengths. It is shown that in this particular case, the Bragg formula containing a single wavelength which is the same for the forward and return waves should be written to include two wavelengths. Relations are put forward which generalize the Bragg formula for media with nonidentical forward and return wavelengths and examples of using these relations are considered. A boundary-value problem is solved for a layer of cholesteric liquid crystal. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The propagation of light in cholesteric liquid crystals (CLCs) is known to have diffraction properties [1, 2]. In media whose spatial structure exhibits right–left asymmetry such as gyrotropic media and CLCs, which by definition should also be classed as gyrotropic [3], wave irreversibility occurs in the presence of magnetooptic activity [4]: the wave vector surface has no center of symmetry with the result that the quantities characterizing the properties of the medium (phase velocity, moduli of the angles of rotation of the polarization plane and circular dichroism, and so on) differ for mutually opposite directions of propagation.

If diffraction reflection is analyzed in a CLC as light propagates along the axis of the medium in the presence of a magnetic field, the field induces magnetooptic activity in the medium which leads to the wave irreversibility noted above. In order to avoid effects involving distortion of the structure which occur under the influence of the field, it is advisable to select CLCs with Franck moduli K_{22}, K_{33} [2, p. 244], for which no distortion occurs until the field reaches a certain critical value, and then the Bragg formula

$$2d\sin\varphi = n\lambda$$

becomes meaningless since there is not one wavelength but two different ones (as a result of the wave irreversibility) for the forward and return directions of propagation.

In Section 2 we shall analyze diffraction reflection in a CLC under conditions of wave irreversibility and we shall calculate the wavelengths of the forward and return waves at frequencies coinciding with the boundaries of the frequency range of diffraction reflection. Direct calculations show that the wavelengths of the forward and return waves differ and the form of the relationship which expresses the condition for amplification of waves reflected at periodic inhomogeneities in the medium differs from the Bragg formula. Accordingly, in the Laue equation [5] the moduli of the wave vectors of the forward and return waves are different and the diagram which expresses this equation geometrically [5] is asymmetric relative to the plane perpendicular to the vector of the reciprocal grating and bisecting it.

It is found that wave irreversibility in the sense of no center of symmetry at the wave vector surface is a stringent constraint for changing the form of the Bragg formula. Specifically, this change also occurs in a naturally gyrotropic medium in the presence of periodic inhomogeneity. The different wavelengths of the forward and return waves in this medium is attributed to the different polarization of these waves. Periodically inhomogeneous naturally gyrotropic media are investigated in Section 3.

In Section 4 we analyze diffraction reflection for $k_i \neq k_s$ ($\mathbf{k}_i$ and $\mathbf{k}_s$ are the wave vectors of the forward (incident) and return (scattered) waves) for propagation inclined toward the planes of the layers neglecting interaction between the forward and return waves in the dispersion equation. We also briefly consider the case where an external magnetic field is present in a naturally gyrotropic medium.

In Section 5 we present results for the propagation of light across a CLC layer possessing wave irreversibility.

1063-7761/00/9001-0102$20.00 © 2000 MAIK "Nauka/Interperiodica"

2. DIFFRACTION REFLECTION IN A MAGNETOACTIVE CHOLESTERIC LIQUID CRYSTAL AND THE BRAGG FORMULA

2.1. Dispersion Equation

We consider the propagation of light at frequency ω along the axis of a CLC (z axis) in the presence of an external magnetic field directed along this axis. Using the method of circular components [1] or using a method of converting in the wave equation to the field components relative to the x' and y' axes which rotate together with the structure [6] (the x' axis is everywhere oriented along the director, the y' axis is perpendicular to the x' axis, and the three axes x', y', and z form a right-handed system), in either case we arrive at the following equation:

$$\begin{aligned}\mathcal{K}_g^4 - \left[\frac{\omega^2}{c^2}(\varepsilon_1+\varepsilon_2)+2q^2\right]\mathcal{K}_g^2 - 4q\frac{\omega^2}{c^2}g\mathcal{K}_g \\ + \left(\frac{\omega^2}{c^2}\varepsilon_1 - q^2\right)\left(\frac{\omega^2}{c^2}\varepsilon_2 - q^2\right) - \frac{\omega^4}{c^4}g^2 = 0.\end{aligned} \tag{1}$$

Here $2\pi/\mathcal{K}_g$ is the spatial period of the field in the local system of the medium, ε_1 and ε_2 are the principal values of the permittivity tensor of the CLC along the x' and y' axes, respectively, g is the z-component of the gyration vector directed along the z axis, which is responsible for the magnetooptic activity, $q = 2\pi/\sigma$, and σ is the pitch of the helix. The projections $k_{mz}$of the wave vectors of the circular components of the field are related to the roots of equation (1) by

$$k_{mz}^{\pm} = \mathcal{K}_{mg} \pm q, \tag{2}$$

as in the case of no magnetooptic activity analyzed in [1, 2]: the difference is that the roots of equation (1) for $g \neq 0$ differ from the roots of this equation for $g = 0$.

2.2. Determination of the Boundaries of the Diffraction Reflection Region and the Values of $\mathcal{K}_{mg}$ at these Boundaries

We shall now determine the boundaries ω_{1g} and ω_{2g} of the diffraction reflection region (the subscript g indicates the presence of magnetooptic activity). We shall assume that ε_1, ε_2, and g are real. Outside the region of diffraction reflection the values of $\mathcal{K}_{mg}$ are real, while inside this region they have an imaginary part, i.e., they are complex. The same applies to $k_{mz}^{\pm}$ since the value of q in (2) is real. Since the coefficients of equation (1) are real, its complex roots should be complex-conjugate. Thus, at the boundaries of the diffraction reflection region at which the imaginary parts of the complex roots of equation (1) vanish on leaving this region, these roots should be multiples. (For $g = 0$ the boundaries of the region of diffraction reflection are usually determined from the constraint that equation (1) has zero roots [1, 2] and these are automatically multiples.)

We denote the multiple roots of equation (1) for $g = 0$ by $\mathcal{K}_m$. For $g = 0$ we have $\mathcal{K}_m = 0$ at the boundaries ω_1 and ω_2 of the diffraction reflection region. In view of the smallness of g, the values of $\mathcal{K}_{mg}$ will differ little from their values for $g = 0$, i.e., from zero. (The magnetooptic rotation of the plane of polarization in fields of $\sim 10^4$ G in nonmagnetic dielectrics in the optical frequency range is ~10 deg/cm. At the wavelength (in vacuum) $\lambda \simeq 5 \times 10^{-5}$ cm and the permittivity $\varepsilon \sim 5$, we obtain $g \sim 10^{-5}$.) Neglecting the fourth power of $\mathcal{K}_g$ in (1), we arrive at the following expression for $\mathcal{K}_g$:

$$\mathcal{K}_g = \left(-2q\frac{\omega^2}{c^2}g \pm \eta\right)\left[\frac{\omega^2}{c^2}(\varepsilon_1+\varepsilon_2)+2q^2\right]^{-1}, \tag{3}$$

where

$$\begin{aligned}\eta = \Bigg\{\left[\frac{\omega^2}{c^2}(\varepsilon_1+\varepsilon_2)+2q^2\right]\left(\frac{\omega^2}{c^2}\varepsilon_1 - q^2\right)\left(\frac{\omega^2}{c^2}\varepsilon_2 - q^2\right) \\ + \frac{\omega^4}{c^4}g^2\left[2q^2 - \frac{\omega^2}{c^2}(\varepsilon_1+\varepsilon_2)\right]\Bigg\}^{1/2}.\end{aligned} \tag{4}$$

The roots will be multiples if $\eta = 0$. From the equation $\eta = 0$ we determine the boundaries ω_{1g} and ω_{2g} of the frequency region of diffraction reflection. To within terms containing g to the second power, we have

$$\omega_{1g} = \frac{qc}{\sqrt{\varepsilon_1}}\left[1 + \frac{g^2}{2\varepsilon_1(3\varepsilon_1+\varepsilon_2)}\right], \tag{5}$$

$$\omega_{2g} = \frac{qc}{\sqrt{\varepsilon_2}}\left[1 + \frac{g^2}{2\varepsilon_2(3\varepsilon_2+\varepsilon_1)}\right]. \tag{6}$$

Substituting into (3) $\eta = 0$ and the values of the frequencies (5) and (6), we obtain the values $\mathcal{K}_{1,2g}(\omega_{1g})$ and $\mathcal{K}_{1,2g}(\omega_{2g})$ of the multiple roots, confining ourselves to quantities linear in g:

$$\mathcal{K}_{1,2g}(\omega_{1g}) = -\frac{2qg}{3\varepsilon_1+\varepsilon_2} \tag{7}$$

at the frequency $\omega = \omega_{1g}$ and

$$\mathcal{K}_{1,2g}(\omega_{2g}) = -\frac{2qg}{3\varepsilon_2+\varepsilon_1} \tag{8}$$

at $\omega = \omega_{2g}$. The relative error in the determination of the multiple roots associated with the neglect of $\mathcal{K}_g^4$ in (1) is of the order of g^2/ε^2 ($\varepsilon \sim \varepsilon_1 \sim \varepsilon_2$).

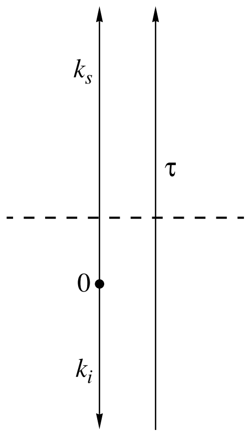


Fig. 1. Diagram describing the Laue equation for a magnetoactive CLC. As a result of the wave irreversibility, the moduli of the wave vectors of the forward and return waves (k_i and k_s, respectively) differ.

2.3. Laue Diagram and Bragg Formula

Using (2), (7), and (8) we obtain for the z-components of the wave vectors of waves with diffraction polarization

$$k_{1z}^{\pm} = k_{2z}^{\pm} = -2qg(3\varepsilon_1 + \varepsilon_2)^{-1} \pm q \quad (9)$$

at frequency ω_{1g} and

$$k_{1z}^{\pm} = k_{2z}^{\pm} = -2qg(3\varepsilon_2 + \varepsilon_1)^{-1} \pm q \quad (10)$$

at frequency ω_{2g}.

To be specific, we shall analyze one of these frequencies, say $\omega = \omega_{1g}$. From (9) we have for the z-component of the wave vector of the forward wave (propagating in the direction of the z axis)

$$k_{zi} = -2qg(3\varepsilon_1 + \varepsilon_2) + q. \quad (11)$$

For the return wave ($k_{zs} < 0$) we have

$$k_{zs} = -2qg(3\varepsilon_1 + \varepsilon_2) - q. \quad (12)$$

In accordance with (11) and (12) we have $|k_{zi}| \neq |k_{zs}|$. Thus, the wavelengths of the forward and return waves, λ_i and λ_s also differ:

$$\lambda_{i,s} = \frac{2\pi}{k_{i,s}} = \frac{2\pi}{|k_{zi,s}|} = \frac{\sigma}{1 \pm 2g(3\varepsilon_1 + \varepsilon_2)^{-1}}. \quad (13)$$

Figure 1 gives a diagram illustrating the Laue equation

$$\mathbf{k}_s - \mathbf{k}_i = \boldsymbol{\tau}, \quad (14)$$

where $\tau = 2\pi n/d$, $d = \sigma/2$ is the period of inhomogeneity of the medium.

A characteristic feature of this diagram is that the moduli of the vectors $\mathbf{k}_i$ and $\mathbf{k}_s$ differ so that there is no symmetry relative to the plane perpendicular to the vector $\boldsymbol{\tau}$ which bisects the latter (this plane is indicated by the dashed line in the figure).

Projecting the Laue equation onto the direction $\boldsymbol{\tau}$, we obtain

$$\frac{2\pi}{\lambda_s} + \frac{2\pi}{\lambda_i} = \frac{2\pi}{d}, \quad (15)$$

which merely yields the Bragg formula in its usual form

$$2d\sin\varphi = n\lambda \quad (16)$$

(in this case $n = 1$, $\varphi = \pi/2$) when $\lambda_i = \lambda_s$.

In the following section we consider an example of another medium for which the Bragg formula in its usual form also cannot be applied because of a difference between λ_i and λ_s.

3. PERIODICALLY INHOMOGENEOUS ISOTROPIC NATURALLY GYROTROPIC MEDIUM

3.1. Material Equations

We shall analyze the propagation of an electromagnetic wave of frequency ω in a medium described by the material equations

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \frac{1}{2}(\Delta\varepsilon)e^{-i\Omega t}[e^{i\tau z} + e^{-i\tau z}]\mathbf{E} + \gamma\,\mathrm{rot}\,\mathbf{E}, \qquad \mathbf{B} = \mathbf{H}, \quad (17)$$

where Ω is the frequency of the wave modulating the permittivity of the medium, $\Delta\varepsilon$ is the percent modulation, $\tau = 2\pi/d$, and d is the period of the inhomogeneity of the medium. The equations (17) describe a naturally gyrotropic isotropic medium [3, 5, 7, 8] with spatially modulated permittivity. This modulation may be created, for example, by a plane ultrasonic wave. We shall assume that in the wave equation for the electromagnetic wave we can neglect the time dependence of the parameters of the medium but allow for their time dependence in the final results. Whereas in the absence of spatial dispersion ($\gamma = 0$) this procedure can be adopted for $\Omega/\omega \ll 1$ [9], for $\gamma \neq 0$ we also need to impose the constraint

$$\left|\frac{\Omega}{\omega}\Delta\varepsilon\right| \ll \left|\frac{\omega^2}{c^2}\gamma\right|, \quad (18)$$

in order to correctly conserve γ in the wave equation, neglecting the time derivatives given above. Assuming that the length of the ultrasonic wave is of the order of the wavelength of light (which is required for diffraction reflection, i.e., $\Omega/v \sim \omega/c$, where v is the velocity of the mechanical wave), we write condition (18) in the form

$$\left|\frac{v}{c\lambda}\Delta\varepsilon\right| \ll \left|\frac{\omega^2}{c^2}\gamma\right|. \tag{19}$$

The usual values of the angle ϑ_0 of rotation of the plane of polarization per unit path length of the ray are a few degrees per centimeter. Assuming $\vartheta_0 \simeq 5$ deg/cm, we obtain $|\omega^2\gamma/c^2| \sim 10^{-1}$. Consequently relation (19) is reduced to the inequality

$$\left|\frac{\omega}{c}\Delta\varepsilon\right| \ll 0.1,$$

which is easily satisfied.

3.2. Diffraction Reflection

We shall express the field of a monochromatic wave propagating in a medium along the z axis in the form

$$\mathbf{E}(z,t) = \left[\mathbf{E}_0\exp(ik_{0z}z) + \sum_m \mathbf{E}_m\exp(ik_{mz}z)\right]\exp(-i\omega t). \tag{20}$$

Using equations (17) we obtain from the wave equation

$$k_{mz} = k_{0z} + m\tau, \quad m = \pm1, \pm2, \dots . \tag{21}$$

In a two-wave approximation which takes into account the spatial components $\mathbf{E}_0\exp[i(k_{0z}z - \omega t)]$ and $\mathbf{E}_{-1}\exp[i(k_{-1z}z - \omega t)]$ we obtain the following equation for k_{0z}:

$$\left[\frac{\omega^2}{c^2}\varepsilon_0 - k_{0z}^2 + \frac{\omega^2}{c^2}\gamma k_{0z}\right] \times\left[\frac{\omega^2}{c^2}\varepsilon_0 - (k_{0z}-\tau)^2 + \frac{\omega^2}{c^2}\gamma(k_{0z}-\tau)\right] - \frac{\omega^4}{c^4}\frac{(\Delta\varepsilon)^2}{4} = 0. \tag{22}$$

In order to determine k_{0z} and k_{-1z} at the boundaries of the diffraction reflection region, we express k_{0z} in the form

$$k_{0z} = \frac{\tau}{2} + x. \tag{23}$$

For k_{-1z} we then have

$$k_{-1z} = -\frac{\tau}{2} + x. \tag{24}$$

Substituting (23) into (22) we obtain the equation for x:

$$x^4 - 2\frac{\omega^2}{c^2}\gamma x^3 - \left[2\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right) - \frac{\omega^4}{c^4}\gamma^2\right]x^2 + 2\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right)\frac{\omega^2}{c^2}\gamma x + \left(\frac{\omega^2}{c^2}\varepsilon_0 - \frac{\tau^2}{4}\right)^2 - \frac{\omega^4}{c^4}\frac{(\Delta\varepsilon)^2}{4} - \frac{\tau^2}{4}\frac{\omega^4}{c^4}\gamma^2 = 0. \tag{25}$$

In an isotropic homogeneous medium the spatial dispersion leads to a change in the moduli of the wave vectors by a value of the order of $\omega^2\gamma/c^2$ [5]. Assuming that x in (25) is a value of this order and neglecting x to the fourth and third powers, we obtain from (25)

$$x = \left[\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right)\frac{\omega^2}{c^2}\gamma \pm \sqrt{u}\right] \times\left[2\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right) + \frac{\omega^4}{c^4}\gamma^2\right]^{-1}, \tag{26}$$

$$u = \left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right)\frac{\omega^4}{c^4}\gamma^2 \times\left[2\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right) + \frac{\omega^4}{c^4}\gamma^2\right] \times\left[\left(\frac{\omega^2}{c^2}\varepsilon_0 - \frac{\tau^2}{4}\right)^2 - \frac{\omega^4}{c^4}\frac{(\Delta\varepsilon)^2}{4} - \frac{\tau^2}{4}\frac{\omega^4}{c^4}\gamma^2\right]. \tag{27}$$

Assuming a relative error of the order

$$\frac{\omega^4}{c^4}\gamma^2\left(\frac{\omega^2}{c^2}\varepsilon_0 + \frac{\tau^2}{4}\right)^{-1},$$

we obtain from (27) the following expression for the multiple roots:

$$x_{1,2} = \frac{1}{2}\frac{\omega^2}{c^2}\gamma. \tag{28}$$

Thus, in accordance with (23), (24), and (28), we have

$$k_{0z} = \frac{\tau}{2} + \frac{1}{2}\frac{\omega^2}{c^2}\gamma, \quad k_{-1z} = -\frac{\tau}{2} + \frac{1}{2}\frac{\omega^2}{c^2}\gamma. \tag{29}$$

Formulas (29) are satisfied when the polarization of the waves is given by

$$E_{0x} - iE_{0y} = 0, \quad E_{-1x} - iE_{-1y} = 0 \tag{30}$$

(the forward wave is right circularly polarized and the return wave is left circularly polarized). When the polarizations of both waves are the reverse, γ in (29) should be replaced by $-\gamma$.

According to (29), the wavelengths of the forward and return waves differ so that the Bragg formula should contain two wavelengths, as in the case of a CLC possessing wave irreversibility.

The frequency boundaries of the diffraction reflection region are determined from the equation $u = 0$ in

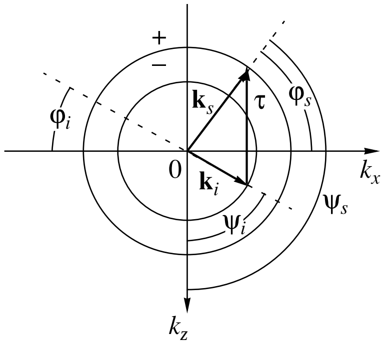


Fig. 2. Diagram describing the Laue equation for an isotropic naturally gyrotropic periodic inhomogeneous medium. The circles give the lines of intersection of the wave vector surface with the plane k_x and k_z.

(27). Assuming a relative error of the order of $\tau^2\gamma^2/\Delta\varepsilon$, for these boundaries we have

$$\frac{\omega_{1,2}}{c} = \frac{\tau}{2}\left(\varepsilon_0 \pm \frac{\Delta\varepsilon}{2}\right)^{-1/2}. \qquad (31)$$

The following relation was assumed to derive (31)

$$\left|\frac{\omega}{c}\gamma\right| \ll \sqrt{2|\Delta\varepsilon|}. \qquad (32)$$

4. OBLIQUE PROPAGATION OF LIGHT. GENERALIZATION OF THE BRAGG FORMULA

In Sections 2 and 3 we considered the propagation of light perpendicular to the layers and we allowed for the interaction of the forward and return waves. For a CLC this can be seen from the fact that we used an exact dispersion equation which allows for all the waves (we selected those solutions corresponding to diffraction polarization). For a medium described by equations (17), we obtained a dispersion equation to determine k_{0z} and k_{-1z} allowing for both waves with the z-components of the wave vectors k_{0z} and k_{-1z}. We shall now consider oblique propagation with respect to the layers neglecting interaction of the waves.

4.1. Condition for which Wave Interaction at Periodic Inhomogeneities in a Medium Can Justifiably Be Neglected

According to (29), the difference between the wavelengths of the forward and return waves is caused by the presence of spatial dispersion which makes the contributions $\sim\omega^2\gamma/2c^2$ to the moduli of the wave vectors of these waves. In order to conserve the effect of different wavelengths we must retain the quantities $\omega^2\gamma/2c^2$ in k_{0z} and k_{-1z}. We now determine the conditions under which the contribution of the spatial dispersion can be correctly retained while neglecting wave interaction at periodic inhomogeneities. Substituting into (29) the value of $\tau/2$ from (31) we obtain

$$k_{0z} = \frac{\omega}{c}\sqrt{\varepsilon_0} + \frac{1}{2}\frac{\omega^2}{c^2}\gamma \pm \frac{\omega}{c}\frac{\Delta\varepsilon}{4\sqrt{\varepsilon_0}} \qquad (33)$$

(the two signs correspond to the two frequency boundaries of the diffraction reflection region). Hence, at the boundaries of the diffraction reflection region wave interaction produces a correction to the wave vector moduli given by $(\omega/c)\Delta\varepsilon/4\sqrt{\varepsilon_0}$ and the spatial dispersion changes these moduli by $\omega^2\gamma/2c^2$. Thus, it is justifiable to neglect wave interaction when determining the wavelengths if

$$\left|\frac{\omega}{c}\gamma\right| \gg \left|\frac{\Delta\varepsilon}{2\sqrt{\varepsilon_0}}\right|. \qquad (34)$$

For light having the wavelength in vacuum $\lambda \simeq 6 \times 10^{-5}$ cm relations (32) and (34) for $\omega^2\gamma/c^2 \sim 0.1$ give

$$\left|\frac{\Delta\varepsilon}{\varepsilon_0}\right| \ll 10^{-6} \ll \sqrt{2|\Delta\varepsilon|}. \qquad (35)$$

These relations are satisfied, for example, for

$$\varepsilon_0 \sim 5, \quad \Delta\varepsilon \simeq 5 \times 10^{-7}.$$

Figure 2 shows a cross section over the plane (passing through the z axis) of the wave vector surface for medium (17) neglecting periodic inhomogeneity. The two spheres correspond to right and left circularly polarized waves: the moduli of the wave vectors for these waves (i.e., the radii of the spheres) are given by:

$$k^- = \frac{\omega}{c}\sqrt{\varepsilon_0} + \frac{1}{2}\frac{\omega^2}{c^2}\gamma, \quad k^+ = \frac{\omega}{c}\sqrt{\varepsilon_0} - \frac{1}{2}\frac{\omega^2}{c^2}\gamma \qquad (36)$$

(the superscripts "–" and "+" correspond to left and right circular polarizations).

4.2. Laue Diagram

We shall now assume that a periodic inhomogeneity is created in the medium and relation (34) is satisfied. Then, assuming a relative error of the order $(\Delta\varepsilon/\varepsilon_0)(\omega\gamma/c)^{-1}$ we can use the expressions (36) for k^+ and k^- and construct a diagram which geometrically describes the Laue equation, assuming that the moduli of the wave vectors are equal to the radii of the spheres.

Figure 2 shows a situation which satisfies the Laue equation. Since no inhomogeneities occur in the direction of the x axis, the tangential components of the wave vectors are the same:

$$k_s\cos\varphi_s = k_i\cos\varphi_i. \qquad (37)$$

This relation is also obtained by projecting the Laue equation onto the plane perpendicular to $\boldsymbol{\tau}$. Since $k_s \neq k_i$, we have

$$\varphi_s \neq \varphi_i. \tag{38}$$

Projecting the Laue equation onto the $\boldsymbol{\tau}$ direction, we obtain

$$k_s \sin\varphi_s + k_i \sin\varphi_i = \frac{2\pi}{d}n. \tag{39}$$

This relation is the phase condition for amplification of waves scattered at periodic inhomogeneities in a medium. Since $\varphi_i \neq \varphi_s$, formula (39) does not reduce to the usual form of the Bragg formula.

4.3. Case Where a Magnetic Field is Present

In the presence of a magnetic field directed along the z axis, the term $i[\mathbf{gE}]$ responsible for the magnetooptic activity formed as a result of the presence of this field is added to the right-hand side of the first of the material equations (17). Assuming a relative error of the order g^2 in the moduli of the wave vectors, for the moduli k^+ and k^- we have [4, 10]

$$k^{\pm} = \frac{\omega}{c}\sqrt{\varepsilon_0}\left[1 \mp \left(\frac{g}{\varepsilon_0}\cos\psi + \frac{(\omega/c)\gamma}{\sqrt{\varepsilon_0}}\right)\right], \tag{40}$$

where ψ is the angle between the direction of propagation of the wave and the z axis along which the external field is applied.

The Bragg formula has the form ($\sin\varphi_i = \cos\psi_i$, $\cos\varphi_i = \sin\psi_i$, $\sin\varphi_s = -\cos\psi_s$, $\cos\varphi_s = \sin\psi_s$, see Fig. 3)

$$\begin{gathered}\frac{\omega}{c}\sqrt{\varepsilon_0}\left[1 \mp \frac{1}{2}\left(\frac{g}{\varepsilon_0}\sin\varphi_i + \frac{\omega}{c}\gamma\frac{1}{\sqrt{\varepsilon_0}}\right)\right]\sin\varphi_i \\ + \frac{\omega}{c}\sqrt{\varepsilon_0}\left[1 \mp \frac{1}{2}\left(-\frac{g}{\varepsilon_0}\sin\varphi_s + \frac{\omega}{c}\gamma\frac{1}{\sqrt{\varepsilon_0}}\right)\right]\sin\varphi_s = \frac{2\pi}{d}n.\end{gathered} \tag{41}$$

To this formula we need to add the following (which replaces the relationship $\varphi_s = \varphi_i$ satisfied for $\lambda_s = \lambda_i$):

$$\begin{gathered}\left[1 \mp \frac{1}{2}\left(\frac{g}{\varepsilon_0}\sin\varphi_i + \frac{\omega}{c}\gamma\frac{1}{\sqrt{\varepsilon_0}}\right)\right]\cos\varphi_i \\ = \left[1 \mp \frac{1}{2}\left(-\frac{g}{\varepsilon_0}\sin\varphi_s + \frac{\omega}{c}\gamma\frac{1}{\sqrt{\varepsilon_0}}\right)\right]\cos\varphi_s.\end{gathered} \tag{42}$$

The four variants of the signs on the left- and right-hand sides of expressions (41) and (42) correspond to the four situations: right and left polarizations of the scattered wave for right and left polarizations of the incident wave.

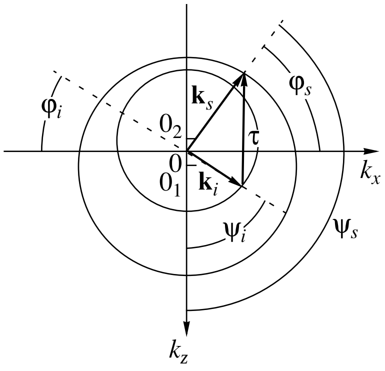


Fig. 3. As Fig. 2 for the case of an external magnetic field which produces wave irreversibility.

A diagram expressing the Laue equation is shown in Fig. 3. Note that the diagrams plotted in Figs. 2 and 3 for light incident normal to the layers have the same form as the diagram in Fig. 1 for a CLC.

4.4. Generalization of the Bragg Formula

The Bragg formula for a periodically inhomogeneous medium in which the wavelengths of the forward and return waves differ has the form

$$\frac{2\pi}{\lambda(\varphi_i)}\sin\varphi_i + \frac{2\pi}{\lambda(\varphi_s)}\sin\varphi_s = \frac{2\pi}{d}n, \tag{43}$$

to which we need to add the following formula because of the presence of the two angles φ_i and φ_s which generally differ

$$\frac{2\pi}{\lambda(\varphi_i)}\cos\varphi_i = \frac{2\pi}{\lambda(\varphi_s)}\cos\varphi_s \tag{44}$$

(instead of $\varphi_i = \varphi_s$).

The dependence of λ on φ is given by the dispersion equation.

5. PROPAGATION OF LIGHT ACROSS A PLANE-PARALLEL LAYER OF CHOLESTERIC LIQUID CRYSTAL WITH WAVE IRREVERSIBILITY

We analyze the normal propagation of plane-polarized light across a plane-parallel CLC layer possessing wave irreversibility. We shall study two variants: propagation in two mutually opposite directions. Figure 4b gives the results of calculating the difference between the transmission coefficients for different wavelengths outside the region of diffraction reflection and Fig. 4a gives the results for this region. The difference $\Delta T = T_1 - T_2$ is plotted on the ordinate where T_1 is the transmission coefficient when light is incident on the layer in the direction of the z axis which lies in the direction of the gyration vector $\mathbf{g}$ perpendicular to the layer boundaries; T_2 is the transmission coefficient for the opposite direction of propagation of the incident light. The components of the permittivity tensor are $\varepsilon_1 = 2.290$, $\varepsilon_2 = 2.143$; $g = 10^{-4}$, the helix pitch is 0.42 μm; the layer

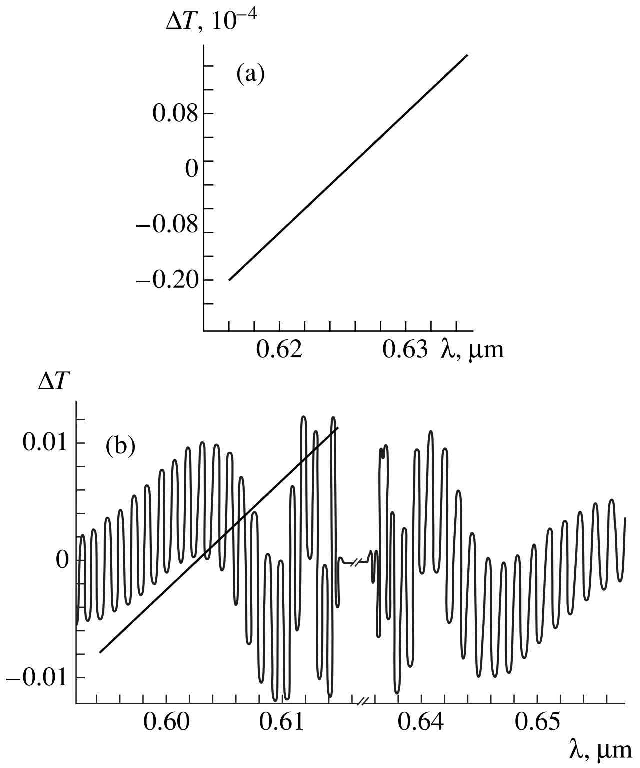


Fig. 4. Difference between the transmission coefficients for different wavelengths inside (a) and outside (b) the region of diffraction reflection.

thickness is 200 μm, and the boundaries of the diffraction reflection region 615 and 630 μm. It can be seen from the figures that the wave irreversibility (different wavelengths of the forward and return waves) resulted in different coefficients T_1 and T_2 for the two mutually opposite directions of propagation of plane-polarized light across the layer.

ACKNOWLEDGMENTS

The author would like to thank A. A. Gevorgyan for the computer calculations and discussions of various results.

REFERENCES

1. E. I. Kats, Zh. Éksp. Teor. Fiz. **59**, 1854 (1970).
2. V. A. Belyakov and A. S. Sonin, *Optics of Cholesteric Liquid Crystals* (Nauka, Moscow, 1982).
3. F. I. Feodorov, *Gyrotropy Theory* (Nauka i Tekhnika, Minsk, 1976).
4. O. S. Eritsyan, Usp. Fiz. Nauk **138**, 645 (1982).
5. L. D. Landau and E. M. Lifshits, *Électrodynamics of Continuous Media* (Pergamon, Oxford, 1984; Nauka, Moscow, 1982).
6. C. W. Oseen, Trans. Faraday Soc. **29**, 883 (1933).
7. V. M. Agranovich and V. L. Ginzburg, *Spatial Dispersion in Crystal Optics and Theory of Excitons* (Nauka, Moscow, 1979).
8. B. V. Bokut’, A. N. Serdyukov, F. I. Feodorov, *et al.*, Kristallografiya **18**, 227 (1973).
9. L. A. Ostrovskiĭ and B. N. Stepanov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **14**, 484 (1981).
10. O. S. Eritsyan, *Optics of Gyrotropic Media and Cholesteric Liquid Crystals* (Aĭastan, Yerevan, 1988), p. 68.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 109–120.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 122–135.
Original Russian Text Copyright © 2000 by Bogdanov, Men'shikov.

SOLIDS

Multiple Small-Angle Neutron Scattering for an Arbitrary Value of the Born Parameter

S. G. Bogdanov and A. Z. Men'shikov
Institute of Metal Physics, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, 620219 Russia

Received May 11, 1999

Abstract—Computer calculations are made of the intensity of multiple small-angle neutron scattering using the general Moliére formula [8] over a wide range of variation of the Born parameter, embracing the diffraction and refraction regimes, and a transition region between diffraction and reflection. A comparison is made with approximate formulas obtained earlier by Maleev *et al.* [9, 10] in the limiting cases of the Born parameter $\alpha \ll 1$ and $\alpha \gg 1$ for the diffraction and refraction regimes, respectively. It is shown that over a wide range of values of α the results of the calculations using the approximate and general formulas are the same. The theoretical conclusions were checked experimentally using data from measurements of small-angle neutron scattering for the domain structure of ferromagnets. Measurements were made of the neutron beam broadening for samples of different thickness and these were used to determine the effective domain sizes in pure iron and nickel exposed to thermal treatment and plastic deformation, and also in the Invar alloys $Fe_{65}Ni_{35}$ and Fe_3Pt. An analysis is made of the angular dependence of magnetic small-angle neutron scattering at the asymptote. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Multiple small-angle scattering is one of the main physical processes accompanying the propagation of corpuscular or wave radiation through inhomogeneous media. It is observed in the broadening of the primary beam as a result of multiple refraction or diffraction from various types of inhomogeneities whose effective radius R_{eff} is considerably greater than the wavelength λ of the incident radiation ($R_{eff} \gg \lambda$). This phenomenon was first analyzed in 1926 by Nardroff [1] for the multiple refraction of X-rays. Nardroff derived a formula linking the broadening of the primary beam with the effective size of the particles at which refraction takes place. Subsequently, an effect involving changes in the width of the X-ray scattering curve was analyzed for diffraction processes of various orders [2].

Multiple neutron scattered was first observed by Hughes *et al.* [3] when neutrons passed through an unmagnetized pure iron plate. In this case, the broadening of the primary neutron beam occurred as a result of multiple refraction and diffraction at the domain structure. This effect was then studied in greater detail in [4, 5] where the effective domain size was determined from the results of the broadening of a primary beam of monochromatic neutrons. Recently this method has also been widely used to determine the size of the scattering particles in nonmagnetic powder and porous materials [6, 7]. A method based on an analysis of the line curvature near the zero scattering angle developed in [7] is highly original in this respect.

The problem of multiple particle collisions was considered theoretically by Moliére (see, for example, [8]) who obtained an expression for the intensity of multiple processes written in an integral form, which is valid for an arbitrary value of the Born parameter. However, this theory has not been widely used to analyze the experimental results because of the complexity of analyzing a general expression. In order to simplify this, the authors of [9, 10] used various approximate solutions of this problem which can give expressions for the multiple scattering intensity in an easier form for analysis. In particular, the case $\alpha \ll 1$ was considered in [9] (the Born approximation or diffraction regime) and the case $\alpha \gg 1$ (non-Born approximation or refraction regime) in [10]. In both approximations multiple scattering was considered as random processes described by a Gaussian function. Both solutions yielded the conclusion that the broadening of a primary neutron beam is proportional to the square root of the scattering order: $\propto (L/l)^{1/2}$, where L is the sample thickness and l) is the neutron mean free path. This showed fairly good agreement with the experimental results [4–6]. However, in the refraction regime where the sample thickness exceeded some critical value ($L \gg L_0$) another law was identified, i.e., a linear dependence of the broadening on thickness. However, the experimental confirmation reported in [11] was cast into doubt by these same authors because of the large experimental error associated with the nonidentical domain structure in samples of different thickness. In addition, the dependence of the broadening on the thickness in the transition region from refraction to diffraction where the parameter α is close to one was left undefined. The angular dependence of the multiple scattering intensity for large scattering vectors was also not determined in this region.

The aim of the present paper is to study multiple small-angle neutron scattering over a wide range of particle sizes or the Born parameter α, covering the diffraction and refraction regimes, and also the transition region from diffraction to refraction, by means of numerical calculations using the Moliére formula. On the basis of these calculations and an analysis of simpler analytic expressions obtained in [9, 10] we discuss the results of experimental investigations of multiple magnetic scattering of neutrons at the domain structure in various ferromagnetic materials such as pure iron and nickel, and also in the Invar alloys $Fe_{65}Ni_{35}$ and Fe_3Pt.

2. NUMERICAL CALCULATIONS

In general, the intensity of multiple small-angle scattering for an arbitrary value of the Born parameter α has been calculated using the Moliére formula [8] written in the following form:

$$I(q, L) = \frac{S}{2\pi}\int_0^\infty \Lambda J_0\left(\Lambda\frac{q}{k}\right)\exp\left[-\left(1-\frac{\sigma_\Lambda}{\sigma_0}\right)\frac{L}{l}\right]d\Lambda, \quad (1)$$

where

$$\sigma_\Lambda = \frac{2\pi}{k^2}\int_0^\infty J_0\left(\Lambda\frac{q}{k}\right)\sigma(q)dq, \quad \sigma_0 = \sigma_\Lambda\big|_{\Lambda=0}, \quad (2)$$

S is the area of the beam cross section, $q = 4\pi\sin\theta/\lambda$ is the transferred neutron momentum, θ is the scattering angle, $k = 2\pi/\lambda$ is the wave number, $J_0(x)$ is a Bessel function, and $\sigma(q) \equiv d\sigma/d\Omega$ is the differential cross section for single scattering which is calculated as the square of the scattering amplitude. Following [12], in general we can write this in the form

$$\sigma(q) = k^2R^4$$
$$\times\left\{\left[\int_0^1(\cos(\alpha\sqrt{1-x^2})-1)J_0(qRx)xdx\right]^2\right. \quad (3)$$
$$\left.+\left[\int_0^1\sin(\alpha\sqrt{1-x^2})J_0(qRx)xdx\right]^2\right\}.$$

In accordance with [13], the total cross section for single scattering per particle σ_0 is given by

$$\sigma_0 = 4\pi R^2\left(\frac{1}{2}-\frac{\sin\alpha}{\alpha}-\frac{\cos\alpha-1}{\alpha^2}\right), \quad (4)$$

where α is the Born parameter, defined as

$$\alpha \equiv 2\nu = \frac{U}{E}kR, \quad (5)$$

U is the energy associated with the inhomogeneity, E is the neutron energy, and R is the inhomogeneity radius. In a different formulation this parameter α has the following form [10]:

$$\alpha = 2\lambda R\delta(Nb), \quad (6)$$

where $\delta(Nb)$ is the contrast, $\delta(Nb) = |N_1b_1 - N_0b_0|$. Here N is the nuclear density and b is the amplitude of coherent neutron scattering. The subscripts "0" and "1" refer the matrix and the inhomogeneity, respectively.

Formulas from [9, 10] written in the following form were used as approximate solutions of the multiple neutron scattering problem:

$$I(q, L) = I_0(q, L) + I_1(q, L) + I_2(q, L), \quad (7)$$

where for $\alpha \ll 1$ we have

$$I_0(q, L) = I_0(0, L)\exp(-q^2/q_1^2), \quad (8)$$

$$q_1 = \frac{1}{R}\sqrt{\frac{BL}{l_d}}. \quad (9)$$

In this case, the parameter B was defined as $B \simeq \ln(4L/l_d) + \ln\ln(4L/l_d)$, where $l_d = 1/(n\sigma_d)$ is the mean free path, n is the volume density of the inhomogeneities, and $\sigma_d = \pi\alpha^2R^2/2$ is the integral cross section for single scattering.

We used the following expressions for the terms I_1 and I_2 [5]:

$$I_1(q, L) = \frac{1}{2}SLn\frac{d\sigma_{as}}{d\Omega}, \quad (10a)$$

$$I_2(qL) \simeq \frac{4\nu^2}{\pi}SLn\frac{d\sigma_{as}}{d\Omega}\frac{1}{qR}, \quad (10b)$$

where $d\sigma_{as}/d\Omega = \alpha^2k^2/2q^4$ is the differential cross section for single scattering at the asymptotic limit for spherical particles of radius R having smooth surfaces.

For the case $\nu \gg 1$ (refraction and diffraction in the non-Born approximation) the differential and total cross section for single scattering was calculated in accordance with [13] using the formulas

$$\frac{d\sigma_r}{d\Omega} = k^2R^4\left[\frac{\alpha^2}{(\alpha^2+(qR)^2)^2}+\frac{J_1^2(qR)}{(qR)^2}\right.$$
$$\left.-\frac{2\alpha}{\alpha^2+(qR)^2}\frac{J_1(qR)}{qR}\sin\sqrt{\alpha^2+(qR)^2}\right], \quad (11)$$

$$\sigma_r = 2\pi R^2. \quad (12)$$

In expression (11), the first term describes refraction, the second describes Fraunhofer diffraction, and the third has an interference nature.

Neglecting the third term in (11), the authors of [10] obtained different expressions for the multiple scattering intensity for $L \ll L_0$ and $L \gg L_0$, where L_0 is a certain critical thickness

$$L_0 = l_r \alpha^2 \ln \alpha. \tag{13}$$

Here $l_r = 1/(n\sigma_r)$ is the neutron mean free path in the refraction regime. For $L \ll L_0$ the expressions for the intensity of the multiple processes, and therefore the broadening of the primary neutron beam, were similar to these expressions for the diffraction regime:

$$I_0^r(q, L) = \frac{S}{2\pi}\left(\frac{k}{q_2}\right)^2 \exp\left(-\frac{q^2}{2q_2^2}\right), \tag{14}$$

$$q_2 = \frac{\alpha}{2R}\sqrt{\frac{L}{l_r}\ln\frac{L}{2l_r}}, \tag{15}$$

and the expression for $I_1^r(q, L)$ is the same as formula (10a).[1]

For samples of thickness $L \gg L_0$ the multiple scattering intensity was described by formulas of a different type:

$$I^d(q, L) = \frac{S}{2\pi}\frac{k^2 q_3}{(q^2 + q_3^2)^{3/2}}, \tag{16}$$

$$q_3 = \left(1 + \frac{1}{\pi}\right)\frac{1}{2R}\frac{L}{l_r}. \tag{17}$$

Here the multiple scattering intensity at the asymptotic limit is proportional to q^{-3} and the broadening has a linear dependence on the sample thickness.

Thus, in addition to comparing the results of the calculations using general and approximate formulas, it would be interesting to check the validity of the conclusions that the different types of dependences of the broadening on the thickness and of the intensities on the wave vector exist for the refraction regime when $L < L_0$ and $L > L_0$.

2.1. Dependence of Neutron Beam Broadening on Sample Thickness

Numerical calculations of neutron beam broadening as a result of multiple processes were made for a hypothetical sample comprising a system of spherical particles of the same size, distributed uniformly in vacuum. The density of nuclei inside the particle was assumed to be $N = 8 \times 10^{22}$ cm^{-3} and the amplitude of coherent neutron scattering was assumed to be $b = 1 \times 10^{-12}$ cm. According to the Babinet principle, this system may be represented as a matrix having the appropriate density in which pores of radius R are uniformly distributed. In both cases the nuclear contrast is $\delta(NB) = NB$. The sample thickness was taken to be 2 and 10 cm. The calculations were made for neutron wavelengths close to the experimental conditions $\lambda = 0.5$ and 1 nm. The particle or pore size was varied between 10 nm and 1 mm. The volume fraction of inhomogeneities ρ was taken to be 0.02 and 0.3 which is valid for systems having low and high inhomogeneity concentrations, respectively.

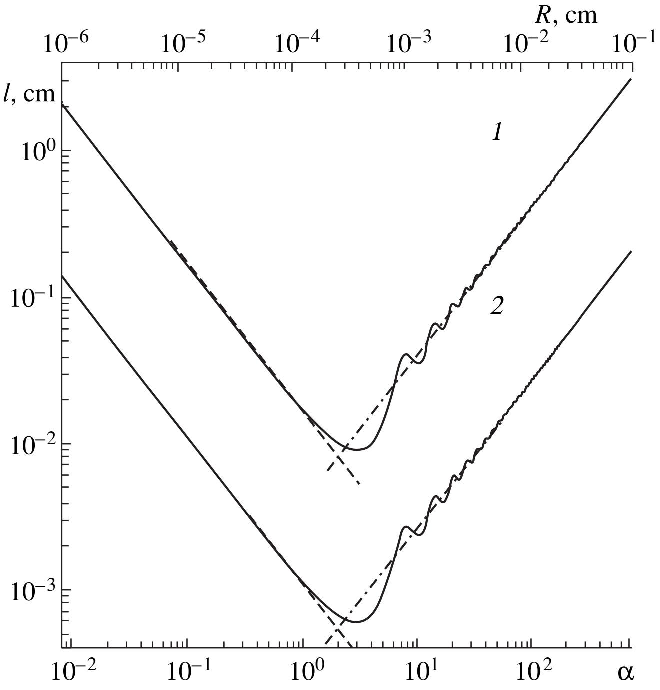


Fig. 1. Dependences of neutron mean free path l on Born parameter α, calculated using various formulas: (18), (4)—solid curves, (19)—dashed curves, (20)—dot-dash curves. The scale for the inhomogeneity radius is shown above. The calculations were made for volume concentrations of inhomogeneities $\rho = 0.02$ (*1*) and 0.3 (*2*).

The neutron mean free path was calculated using the formula

$$l = \frac{1}{n\sigma_0}, \tag{18}$$

where $1/n = 4\pi R^3/(3\rho)$ and σ_0 is described by formula (4). The mean free path was calculated as a function of α or R and is given by the solid curve in Fig. 1 on a logarithmic scale for two values of ρ: 0.02 (*1*) and 0.3 (*2*). The dashed and dot-dash curves also give the mean free paths l_d and l_r calculated using the approximate formulas valid in the diffraction and refraction regimes, respectively:

$$l_d = \frac{2}{3}\frac{1}{\rho\lambda^2 N^2 b^2 R}, \tag{19}$$

$$l_r = \frac{2}{3}\frac{R}{\rho}. \tag{20}$$

It can be seen that the lines l_d and l_r intersect at $\alpha = 2$ ($\nu = 1$) whereas the true curve $l(R)$ has a minimum at a slightly higher value of α and exhibits oscillating

---

[1] An expression for $I_1^r(q, L)$ was obtained by Yu. N. Skryabin.

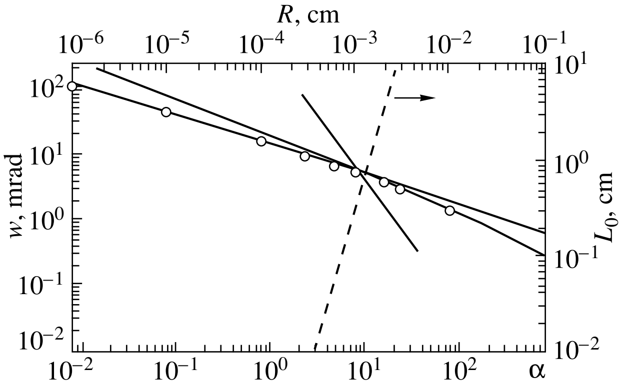


Fig. 2. Dependences of the primary neutron beam broadening on the inhomogeneity size or Born parameter α for a hypothetical sample of thickness $L = 2$ cm in the diffraction (*1*) and refraction (*2* and *3*) regimes for a volume inhomogeneity concentration $\rho = 0.3$. The dashed curve gives the critical thickness L_0 introduced in [10] to determine the range of validity of the dependence $w_3(L)$.

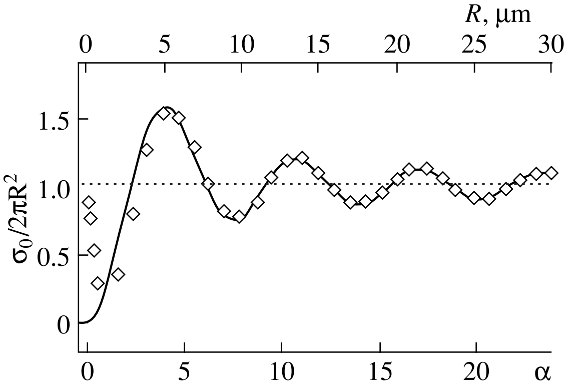


Fig. 3. Integral cross section for single scattering calculated using the general formula (4) (solid curve) and in the refraction limit using formula (2) allowing for (circles) and neglecting (dashed curve) the interference term in formula (11) for the single differential cross section for neutron scattering.

behavior in the range $\alpha \simeq 2$–20. It can be seen that the character of the curves depends weakly on the volume concentration of inhomogeneities ρ.

In order to explain the dependences of the primary neutron beam broadening on the size of the inhomogeneities, we first calculated the angular broadening of the neutron beam using formulas (9), (15), and (17) obtained in the diffraction and refraction regimes

$$w_1 = 0.325\sqrt{\rho}Nb\lambda^2\sqrt{\frac{L}{R}\left(\ln\frac{4L}{l_d} + \ln\ln\frac{4L}{l_d}\right)}, \quad \alpha \ll 1, \tag{21a}$$

$$w_1 = 0.459\sqrt{\rho}Nb\lambda^2\sqrt{\frac{L}{R}\left(\ln\frac{L}{2l_r} + \ln\ln\frac{L}{2l_r}\right)}, \quad \alpha \gg 1, \quad L \ll L_0, \tag{21b}$$

$$w_3 = 0.241\rho\lambda\frac{L}{R^2}, \quad \alpha \gg 1, \quad L \gg L_0, \tag{21c}$$

where w_i is the line width at half height. The results of the calculations for the volume concentration of inhomogeneities $\rho = 0.3$ are plotted in Fig. 2 (solid curves *1*, *2*, and *3*). Similar dependences were obtained for $\rho = 0.02$. The dashed curve gives the dependence of the critical thickness L_0 on the inhomogeneity radius R or the Born parameter α calculated using formula (13). It was found that the range of sizes where the condition $L \gg L_0$ can be satisfied in practice and where the dependence $w_3(R)$ should be used, is very narrow and corresponds to $\alpha \sim 2$–10.

Naturally the reliability of these dependences can be checked by means of numerical calculations, using the general formulas (1)–(3) which are valid for an arbitrary value of the Born parameter. For this purpose we first calculated the differential cross section for single scattering $d\sigma/d\Omega$ for each particle size R and then calculated the Fourier transform of the scattering cross section σ_Λ. We then determined the angular dependence of the multiple scattering cross section $I(q)$ and the full width at half-height of this curve $w(R)$. The calculations were made for the sample thickness $L = 2$ cm, which satisfied the condition $L \gg L_0$ for $R \simeq 5 \times 10^{-4}$ cm. The results of these calculations are given by the circles in Fig. 2. It can be seen that the values of $w(R)$ obtained using the Moliére formula are accurately described by the sections of the curves $w_1(R)$ and $w_2(R)$ calculated using the approximate formulas (21a) and (21b). Moreover, the broadening only depends monotonically on the size and no anomalies are observed in the transition region. Similar calculations made for a sample of thickness $L = 10$ cm and for $\lambda = 1$ nm revealed the same dependence. This suggests that the dependence $w_3(L)$ for $L \gg L_0$ does not in fact exist, although the mathematical solution of the problem in [10], made assuming that the third term in formula (11) makes a small contribution, was performed correctly.

In order to show the influence of the third term in expression (11) on the final results of the calculations of the total neutron scattering cross section, we made numerical calculations of the single scattering cross section in the crossover region. Figure 3 gives various dependences of this cross section in arbitrary units. The first, shown by the solid curve, was calculated using the general formula (4). The second dependence, shown by the circles, was obtained from formula (2) using expression (11) where all terms including the interference term were taken into account. It can be seen that for $\alpha > 2$ these two dependences are very similar. However, if we neglect the third term in formula (11), we obtain the expression $\sigma_0 = 2\pi R^2$, shown by the dashed curve in Fig. 3, which differs substantially from the true scattering cross section given by the solid curve in the range of α values between 2 and 30. Hence, neglecting the interference term in formula (11) leads to apprecia-

Table 1. Beam broadening calculated using formulas (1) and (2) using the single scattering cross section in the refraction limit (11) allowing for (column 3) and neglecting (column 5) the interference term in the cross section (11) and results obtained using formulas (21)

α	R, μm	$w(3)$, mrad	w_2, mrad	$w(2)$, mrad	w_3, mrad
1	2	3	4	5	6
2.4	3	15	13	83	81
4.8	6	10	8.5	24	20
8	10	6.4	6.3	11	7.2
16	20	5.0	4.2	5.0	1.8
24	30	3.5	3.3	3.5	0.8
80	100	1.4	1.6	1.4	0.07

ble inaccuracies in the calculations of the total scattering cross section in this range of α values.

Similar inaccuracies are observed in the calculations of the primary beam broadening. Table 1 gives the primary beam broadening calculated from formula (1) using the differential single scattering cross section (11) allowing for (column 3) and neglecting (column 5) the third term; results of calculations using the formulas (21) are also given for comparison. It can be seen that if the third term is neglected [$w(2)$, column 5], then for $\alpha < 10$ (the condition $L \gg L_0$ is satisfied), the values of the neutron beam broadening are in fact close to those obtained from the formula for $w_3(L)$ (column 6). If this term is taken into account in the calculations, the values of w [$w(3)$, column 3] are closer to the broadening values corresponding to the curve $w_2(R)$ (column 4). At the same time, as α increases, the contribution of the interference term to the cross section (11) decreases and the values given in columns 3 and 5 become comparable and close to the values of $w_2(R)$ calculated using formula (21b).

Consequently, the conclusion reached in [10] that a critical thickness L_0 exists and that the dependences $w_3 \propto L$ and $I(q) \propto q^{-3}$ are obtained for $L \gg L_0$ should be attributed to neglecting the interference term in the single scattering cross section in the refraction regime. However, this neglect of the interference term does not distort the main results derived from the numerical calculations of broadening using the general Moliére formula. This is that the approximate formulas obtained in [9, 10] for broadening in the diffraction regime for any L and in the refraction regime for $L < L_0$ can be jointly used over the entire range of variation of α, including the transition regime from diffraction to refraction. The diffraction approximation can be applied as far as $\alpha \sim 10$–15 while the formulas obtained in the refraction approximation hold for larger α.

2.2. Angular Dependence of Multiple Scattering

We shall now check the conclusions of the approximate theories [9, 10] in relation to the predicted dependences of the multiple scattering intensity on the transferred momentum. It can be seen from expression (7) that the multiple scattering intensity can be represented as the sum of three components I_0, I_1 and I_2. The first of these accurately describes the broadening of the primary neutron beam as a result of multiple processes since the profile of the multiple scattering curve (1) is close to Gaussian. Differences between the curves $I(q)$ calculated using the general formula (1) and formula (8) only occur for large and very small q. We shall first discuss the difference between these curves for large q which is directly relevant to the experimentally measured angular dependences of small-angle neutron scattering. The range of very small transferred momenta is less easily measured.

In the approximate theories [9, 10], the deviation of the dependences $I_0(q, L)$ from the Moliére line profile is taken into account by the terms I_1 and I_2, for which analytic expressions are given by formulas (10a) and (10b) where I_1 and I_2 are proportional to q^{-4} and q^{-5}, respectively. Calculations of the dependence of the multiple scattering cross section on the transferred momentum for particles of different sizes using the Moliére formula showed that at the asymptote the multiple scattering intensity is described by the law $I(q) \propto q^{-4}$. In this case, the contribution from I_2 is not significant. This can be seen clearly from Fig. 4 which gives various dependences $I(q)$ for 100 nm particles.

It also follows from the results of the numerical calculations that the macroscopic cross sections for single and multiple neutron scattering at the asymptote have the same value. From the data plotted in Fig. 4 we can form an opinion on the quantitative determination of the “asymptotic limit.” If the value of the scattering vector q_{as} at which the multiple scattering cross section agrees with its asymptotic value to within 10% is taken as the asymptotic limit, we obtain $q_{as} \simeq 10w$. However, on the experimental curves for multiple neutron scattering typical dependences for the asymptotic limit are observed at angles an order of magnitude smaller. This characteristic of the experimental dependences can naturally be explained by the experimental line profile for multiple scattering at the domain structure of ferromag-

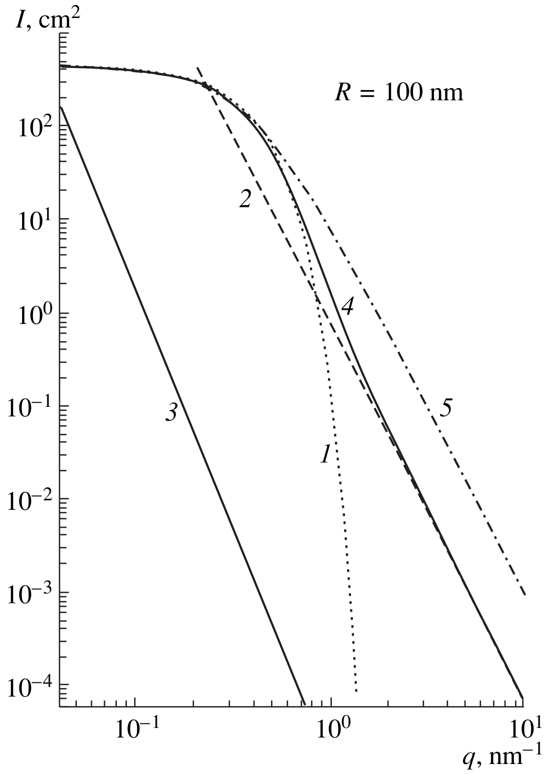


Fig. 4. Differential cross sections for multiple scattering $I_0(q)$ (curve *1*), $I_1(q)$ (*2*), $I_2(q)$ (*3*), and scattering cross section calculated using the general Moliére formula (curve *4*) for 100 nm particles. Curve *5* gives the squared Lorentz function most frequently used to describe the experimental curves for multiple scattering at the domain structure of ferromagnets.

nets. This is usually described by a squared Lorentz function which reaches its asymptotic limit $I \propto q^{-4}$ more rapidly that the Moliére multiple scattering curve.

3. MULTIPLE MAGNETIC SMALL-ANGLE NEUTRON SCATTERING

Multiple small-angle neutron scattering effects at magnetic inhomogeneities can be analyzed theoretically as for nuclear ones. The only difference is that in this case, the energy associated with the inhomogeneity is given by $U = -\boldsymbol{\mu} \cdot \delta\mathbf{B}$ and the Born parameter α in the formulas for the single scattering cross sections (3) and (4) is written in the form

$$\alpha = \frac{2\pi}{\lambda}\frac{\boldsymbol{\mu} \cdot \delta\mathbf{B}}{E}R. \qquad (22)$$

Here E and μ are the neutron energy and magnetic moment, respectively, $\delta\mathbf{B} = \mathbf{B}_1 - \mathbf{B}_0$ is the magnetic contrast, and $\mathbf{B}_0$ and $\mathbf{B}_1$ are the magnetic inductions of the matrix and the inhomogeneity, respectively. For a ferromagnet with a chaotic domain structure the magnetic domains at whose boundaries multiple refraction takes place act as inhomogeneities.

In order to use the conclusions of the theory described above to analyze magnetic multiple neutron scattering at the domain structure in a ferromagnet, we need to make some observations. The first relates to the fact that in general the domains are not spherical. However, all the formulas given above imply that they are. The possibility of applying the theory to aspherical inhomogeneities was discussed in [9, 13] for chaotically oriented particles of arbitrary shape. It was shown in analyses of multiple scattering using this theory that the shape of the inhomogeneities does not play a significant role. Particles of any shape can be considered as spherical with the characteristic dimension R_{eff}, which is of the order of magnitude of the cube root of the inhomogeneity volume: $R_{eff} \sim V^{1/3}$.

The second observation relates to the concept of contrast in multiple scattering at a domain structure. Since in this case the magnetic domains themselves are the inhomogeneities, the contrast is determined by the magnetic induction of the domain which is considered to be a vector. Then for the two neighboring domains in a ferromagnet with 180-degree domain walls we have $\mathbf{B}_1 = -\mathbf{B}_2$ when the inhomogeneity dimension R is the same as the domain thickness. For a ferromagnet with a chaotic domain structure the domain thickness is replaced by some effective radius R_{eff} which is equal to half the wavelength of an averaged alternating-sign step function of amplitude is B, and the corresponding contrast is $|\delta\mathbf{B}| = 2B$.

The third observation concerns the initial conditions of the theory which assumes that multiple neutron scattering takes place at dilute systems, for which the ratio of the volume occupied by the inhomogeneities ΔV to the sample volume V is much less than one ($\Delta V/V \ll 1$). A multidomain ferromagnet is undoubtedly a concentrated system since, as follows from the above reasoning, the analysis of the domain structure as an inhomogeneous system is made using the Ising model where $\rho = 0.5$. This may give rise to some characteristic features mainly associated with the experimental profile of the multiple scattering curve. This may be why the profile of the experimental curve is closer to the squared Lorentz function than the Moliére scattering curve.

4. EXPERIMENT

In order to check the conclusions of this multiple scattering theory experimentally, we used the results of an investigation of small-angle magnetic neutron scattering in ferromagnetic materials such as pure nickel and iron, and also in the alloys $Fe_{65}Ni_{35}$ and Fe_3Pt.

The experiment was carried out using the D6 small-angle neutron scattering device mounted in the horizontal channel of the IVV-2M reactor (Zarechnyĭ) at wavelength $\lambda = 0.478$ nm and angular instrumental line width $w_0 \simeq 8$ min. We used a neutron beam formed in a slit geometry. The beam width was 1 mm and the height varied between 8 and 30 mm. The samples were

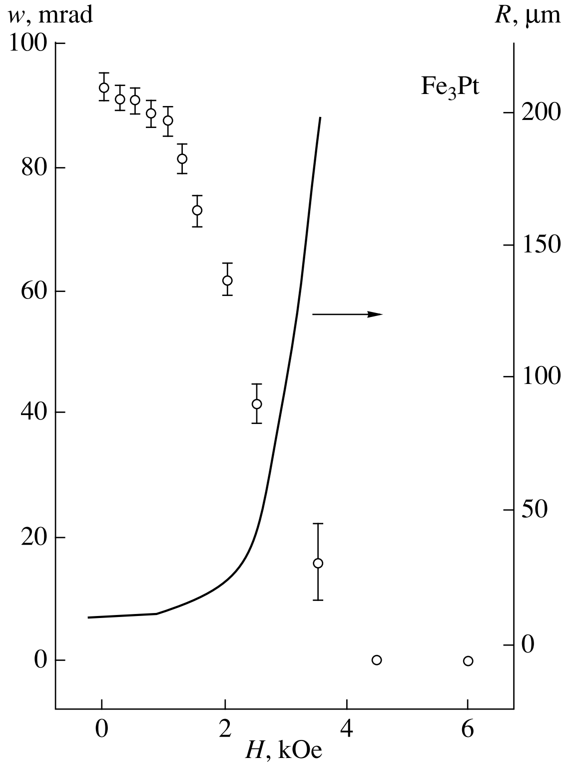


Fig. 5. Field dependences of neutron beam broadening and domain size for the Invar alloy Fe$_3$Pt. The domain size was calculated assuming that the magnetic contrast in the sample remains constant as the external magnetic field increases,

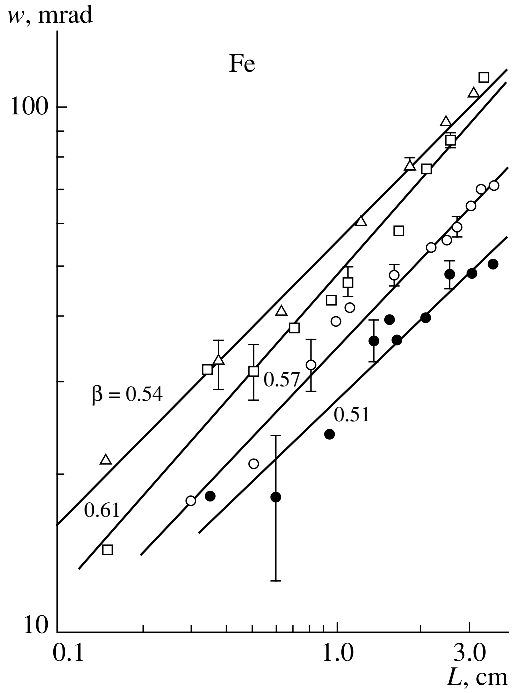


Fig. 6. Experimental dependences of neutron beam broadening on the thickness of pure iron samples after various treatment: (●) annealing at 1000°C for 2 h; (○) 10% deformation; (□) 15% deformation; (△) 25% deformation. The straight lines show the fitting of the experimental points to power dependences.

parallelepipeds or disks of various thickness. In order to vary the domain size, the metal samples were subjected to plastic deformation and heat treatment, and a nickel sample with a fine-grained structure was obtained by thermoplastic deformation. All the measurements were made at room temperature.

The experimentally measured neutron beam broadening effects are only of a magnetic nature. This can be seen from Fig. 5 which gives the field dependence of the neutron beam broadening for Invar alloy Fe$_3$Pt as an example. Here and subsequently the neutron beam broadening is taken as the angular value $w = (w_1^2 - w_0^2)^{1/2}$, where w_1 and w_0 are the total half-height widths of the curve giving the intensity as a function of the scattering angle with and without the sample, respectively. An external magnetic field of several kilo-oersted causes the neutron beam broadening effect to disappear and appreciably reduces the small-angle scattering intensity.

4.1. Dependence of the Neutron Beam Broadening on the Sample Thickness

Typical dependences of the neutron beam broadening as a function of the sample thickness for pure iron are shown in Fig. 6 using a log–log plot. It can be seen that for all the iron samples exposed to various treatment to change the domain structure, the dependence of the broadening on the thickness is described by the power function $w \propto L^\beta$ with the exponent β close to 0.5–0.6. This agrees with the conclusions of the approximate theories put forward above, where the radicand expression in formulas (21a) and (21b) has a logarithmic cofactor in addition to the thickness L. In order to determine how significantly this influences the profile of the dependence $w(L)$, we calculated the broadening directly using the Moliére formula (1) for this hypothetical sample and expressed the results in two variants: $w = f(L^\beta)$ and $w = f(L^\beta[\ln(4L/l) + \ln\ln(4L/l)]^\beta)$. It was found that in the first case the exponent β is 10–15% higher than 0.5. Thus, we can assume that the experimentally determined values of β fall within the limits of the theoretical thickness dependences and the experimental accuracy. The spread of points in Fig. 6 can be attributed to the fact that when the thickness dependences of the neutron beam broadening are measured using a set of samples, the domain structure is not identical. For this reason it is best to study the thickness dependence of the broadening by rotating a plane-parallel sample about the vertical axis in the beam. The results of such an experiment obtained using a plate of ordered Fe$_3$Pt alloy are plotted in Fig. 7. The experimental dependence is accurately described by the

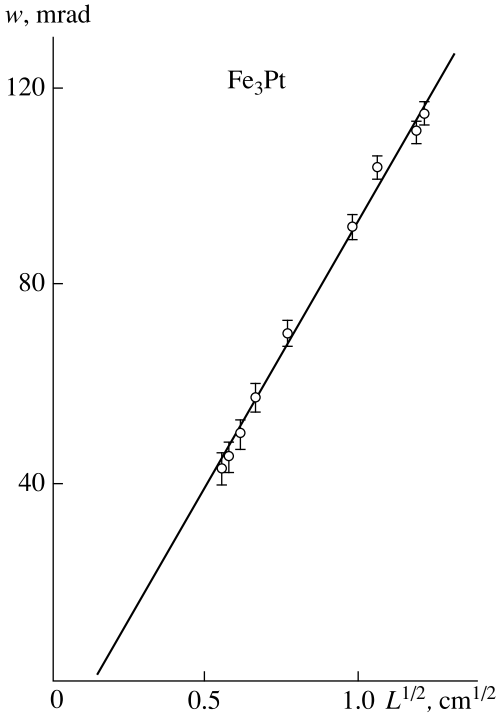


Fig. 7. Thickness dependence of the broadening obtained by rotating a planar sample of Fe_3Pt alloy measuring $3 \times 15 \times 15$ mm^3 about the vertical axis.

power law $w \propto L^{0.61}$ which is similar to that observed for pure iron. In addition, it is easily seen that extrapolating the experimental dependence to zero thicknesses gives a nonzero cutoff on the abscissa. The existence of a cutoff was discussed in [9] and is a consequence of the fact that the argument of the root function in the expression for $w(L)$ has a logarithmic cofactor in addition to the thickness. After calculating the broadening using the general formula (1) and expressing this as a function of the argument $(L\ln(\ldots))^{1/2}$ we obtained a linear dependence which, when extrapolated to $L = 0$, passes through the origin. (Note that the value of $w(0)$ is not determined since the multiple scattering theory is valid for $L \gg l$). This dependence is observed for any value of the parameter α.

On the basis of the agreement between the experimental and calculated values of the thickness dependences, we subsequently made a complex analysis of the experimental data in order to determine the effective sizes of the magnetic domains and the neutron mean free path using formulas similar to (21a) and (21b). In this case, we expressed these in the form:

$$w_d = k_d \frac{\delta B}{E} \sqrt{\frac{L}{R}\left(\ln\frac{4L}{l_d} + \ln\ln\frac{4L}{l_d}\right)}, \quad \alpha \ll 1, \qquad (23a)$$

$$w_r = k_r \frac{\delta B}{E} \sqrt{\frac{L}{R}\left(\ln\frac{L}{2l_r} + \ln\ln\frac{L}{2l_r}\right)}, \quad \alpha \gg 1, \qquad (23b)$$

where $k_d = 6.97 \times 10^{-24}$ erg/G and $k_r = 9.85 \times 10^{-24}$ erg/G.

The procedure involved first defining the magnetic contrast in accordance with the known values of the induction for these materials, and constructing graphs of the broadening as a function of the inhomogeneity size for a single sample thickness over a wide range of sizes similar to the graphs plotted in Fig. 2. Then, the experimental values of the broadening were used to determine the typical domain size and scattering regime. The formulas (23) were then used to fit the theoretical thickness dependences to the experimental ones to determine the final domain size and the neutron mean free path. An example of this fitting is shown in Fig. 8.

A similar treatment of the experimental data was made for all the samples. The results presented in Table 2 indicate that iron samples annealed at 1000°C and deformed by 10% and 15% scatter in the refraction regime and the deformed nickel samples scatter in the diffraction regime. However, the alloys $Fe_{65}Ni_{35}$ and

Table 2. Results of an analysis of experimental data on the thickness dependence of broadening. The second column gives the experimental broadening for samples of thickness $L \simeq 1$ cm, the third column gives the exponent in the thickness dependence of the broadening, the fourth column gives the Born parameter, the fifth column gives the effective domain thickness, and the sixth gives the neutron mean free path. The error in the determination of the exponent is $\Delta\beta = \pm 0.05$

Sample 1	w, mrad 2	β 3	α 4	R, μm 5	l, μm 6	Scattering regime 7
Fe annealed	20	0.51	60	62	82	refraction
Fe 10% deformed	34	0.57	38	39	52	refraction
Fe 15% deformed	36	0.61	22	22	29	refraction
Fe 25% deformed	47	0.54	17	18	23	transition
Ni 40% deformed	38	0.57	1	4	19	diffraction
Ni fine-grained	22	0.45	0.3	1	76	diffraction
$Fe_{65}Ni_{35}$	31	0.52	6	11	15	transition
Fe_3Pt	65	0.52	8	8	12	transition

Fe_3Pt and 25% deformed iron are located in the transition region. It can also be seen from Table 2 that substantial neutron beam broadening is observed for Fe_3Pt. This can be attributed to the fairly small effective size of the magnetic domains $R \simeq 8$ μm and the high magnetic induction $B = 21600$ G. The smallest average domain size $R \simeq 1$ μm with a mean free path $l \simeq 76$ μm was obtained for fine-grained nickel.

By analyzing the neutron beam broadening in an external magnetic field, we can follow the change in the domain size during magnetization of the sample. This is shown in Fig. 5 for Fe_3Pt alloy. It can be seen that initially when grain boundary displacement processes predominate, the size of the domains varies only slightly whereas when a field of around 2 kOe is reached, the domain size increases sharply. This result was obtained assuming that the magnetic contrast remains constant with increasing magnetic field. In fact in the region of the rotation processes the contrast decreases so that the increase in the domain size at the concluding stage was not as steep as that shown in the figure.

4.2. Dependence of Scattering Intensity on Transferred Neutron Momentum

It follows from the previous analysis that the broadening of a primary neutron beam after its propagation through an unmagnetized ferromagnetic plate fairly accurately confirms the conclusions of the theory of multiple neutron scattering at domains as at inhomogeneities of effective radius R_{eff}. The experimentally observed angular dependence of the magnetic scattering intensity is not so obviously related to the multiple processes. This is primarily because the multiple scattering cross section at the asymptote goes over to the single scattering cross section, which also varies as q^{-4} (Porod law). In addition, for high values of q single scattering effects may be observed at small-scale inhomogeneities located within a single domain. Finally, for small transferred momenta below the critical value q_{cr}, inelastic scattering effects take place at spin waves which may distort the angular dependences of the multiple scattering intensity.

In the present study we attempted to isolate these contributions and estimate their absolute values for pure metals (iron and nickel) where no small-scale fluctuations of the magnetization occur inside the domain, and also for Invar alloys $Fe_{65}Ni_{35}$ and Fe_3Pt where such fluctuations may occur as a result of the presence of an inhomogeneous magnetic structure [14] attributed, for example, to antiferromagnetic interaction between Fe–Fe atoms against the background of ferromagnetic interaction between Ni–Ni and Ni–Fe atoms [15].

Before analyzing the experimentally measured angular dependences, we make some observations on the possible influence of vertical divergence on the dependence $I(q)$ since a slit geometry was used in our experiment. Taking into account the specific geometry, we made a theoretical estimate of the influence of vertical divergence on the angular intensity distribution using well-known techniques [16]. This showed that for the geometric parameters of our system the influence of vertical divergence on the function $I(q)$ is negligible. This conclusion also follows from the results of checking these dependences experimentally for various slit heights before the sample and the counter. These experiments showed that in our geometry the resolution of the diffractometer in the vertical plane has no significant influence on the measured angular dependence of the intensity, which distinguished our system from a double crystal spectrometer where these effects are substantial [17].

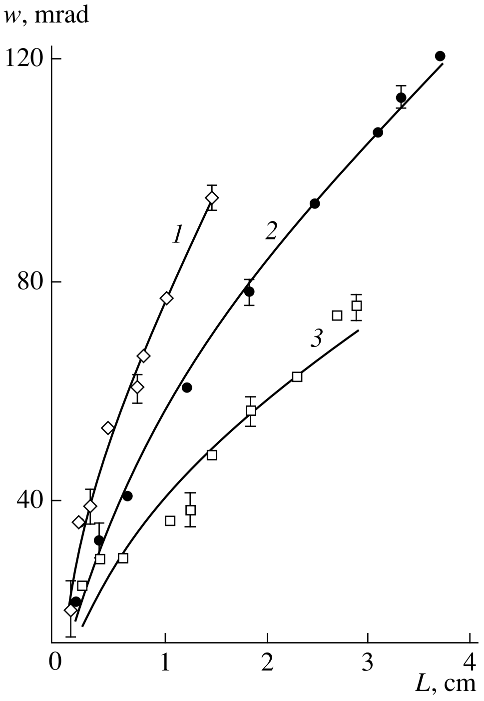


Fig. 8. Dependences of neutron beam broadening on thickness. Circles—measured values, solid curves—fitting using formulas (23): (*1*) Fe_3Pt; (*2*) 25%-deformed iron; (*3*) $Fe_{65}Ni_{35}$.

The results of an experimental investigation of the dependences of the small-angle neutron scattering intensity on q can be seen for samples of pure iron of different thickness in the annealed and deformed states (Fig. 9). In the range of transferred momenta $q < 0.2$ nm^{-1} these dependences are described by the power law $I(q) \propto q^{-4}$, which agrees with the conclusions of the approximate theories [9, 10] and with the numerical calculations using the Moliére formula (1). It can be seen from Fig. 10 that the dependences on q are very sensitive to the external magnetic field as a result of changes in the domains under magnetization. The absolute value of the intensity in small angles is reduced by several factors of 10 when the sample is transferred from the

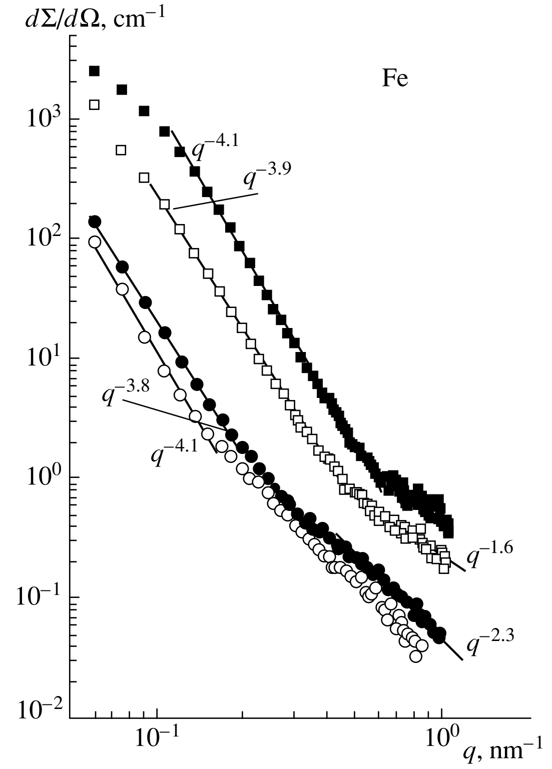


Fig. 9. Experimental dependences of the differential neutron scattering cross section on the transferred momentum at the asymptote for annealed (○, ●) and 25% deformed (□, ■) iron samples of thickness $L = 1.2$ cm (○, □) and 2.5 cm (●, ■).

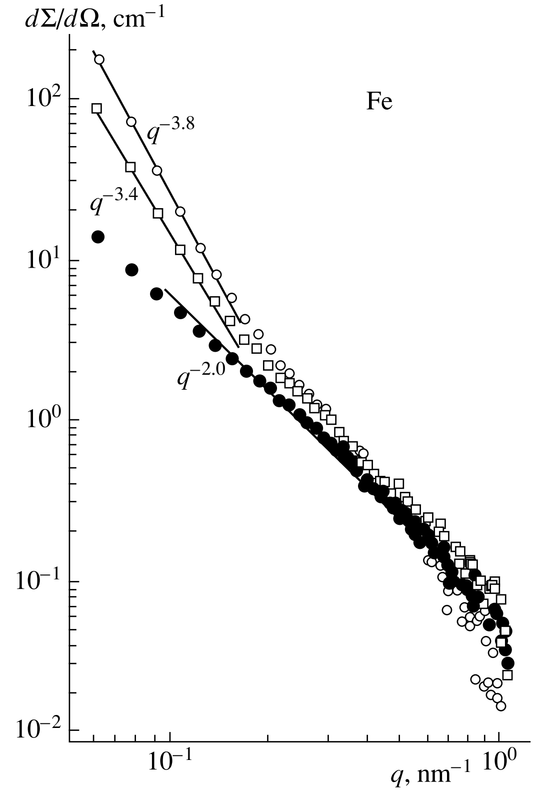


Fig. 10. Angular dependences of the multiple neutron scattering cross section for an annealed iron sample in a zero magnetic field (○) and in fields of 0.5 (□) and 2 kOe (●).

demagnetized to the magnetized state and the dependence $I(q) \propto q^{-n}$ with $n \simeq 4$ becomes distorted in the direction of decreasing n when a many-domain sample is converted to a single-domain one.

Similar dependences are observed for $Fe_{65}Ni_{35}$ alloy, as can be seen from Fig. 11 which gives the profile of the scattering curves as a function of the direction of the magnetic field relative to the neutron scattering vector. However, in the range of small transferred momenta where multiple processes are observed at the domain structure, an increase in the external magnetic field changes the behavior of $I(q)$ whereas for large q this influence is small. Here the dependence $I(q)$ is described by a power law with $n \sim 1.5$–2 and is attributed to single neutron scattering at magnetic fluctuations of the paramagnetic inclusion type in a ferromagnetic matrix. It was shown in [18] that in Invar alloys scattering at these inhomogeneities is described by a Lorentz function and the characteristic size of the inhomogeneities is of the order of a few nanometers. This is significantly smaller than the effective radius of the ferromagnetic domain which, according to Table 2, is tens of micron for these alloys.

Experiments to study the angular dependences of the scattering intensity in a magnetic field in various directions relative to the scattering vector can also be used to estimate the contribution to small-angle diffraction from inelastic scattering of neutrons at spin waves. This scattering exists in the range of angles below the critical value $\theta_c = \hbar^2/(2mD)$, where D is the coefficient of exchange hardness in the quadratic dispersion law for the spin waves $E = Dq^2$, m is the neutron mass, and $\hbar$ is Planck's constant. The procedure for isolating the small-angle magnetic inelastic scattering described in [19] involves investigating the angular dependences for two directions of the external magnetic field on the sample, parallel and perpendicular to the scattering vector. Figure 11 gives these angular dependences for $Fe_{65}Ni_{35}$ alloy in the magnetic field $H = 4$ kOe. It can be seen that for scattering vectors $q < 0.1$ nm^{-1} in the absence of a magnetic field $I_{\mathbf{H}=0}(q)$ reaches 10^5–10^6 arb. units whereas the values of $I_{\mathbf{H}\parallel\mathbf{q}}(q)$ and $I_{\mathbf{H}\perp\mathbf{q}}(q)$ do not exceed 10^4 arb. units. Hence, the inelastic neutron scattering intensity is approximately two orders of magnitude lower than the multiple scattering intensity at the domain structure.

By investigating the angular dependences of the intensity for two directions of the external magnetic field, we can determine the critical angle θ_c of the scattering processes at spin waves. It can be seen from Fig. 12 that the difference scattering intensity near zero angle is negative, then changes sign as the angle increases, passes through a maximum and then, after

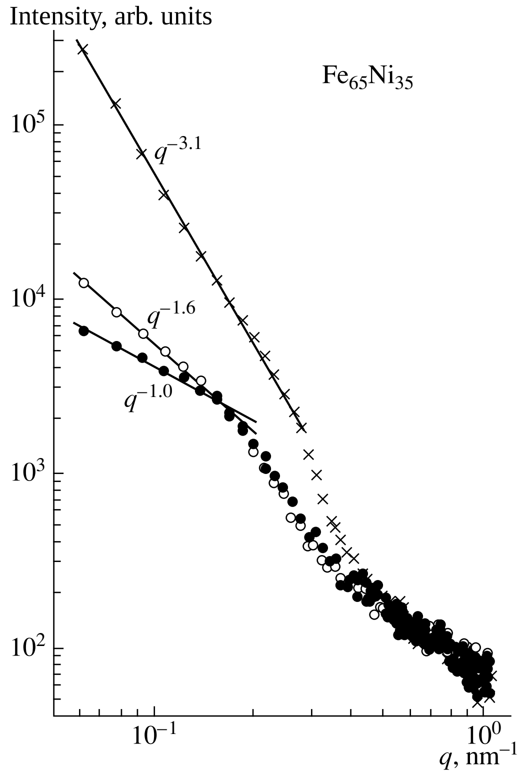


Fig. 11. Experimental dependences of the neutron scattering intensity on the transferred momentum at the asymptote for $Fe_{65}Ni_{35}$ alloy in a zero magnetic field (×) and in a field of 4 kOe parallel (●) and perpendicular (○) to the scattering vector. Sample thickness L = 0.7 cm.

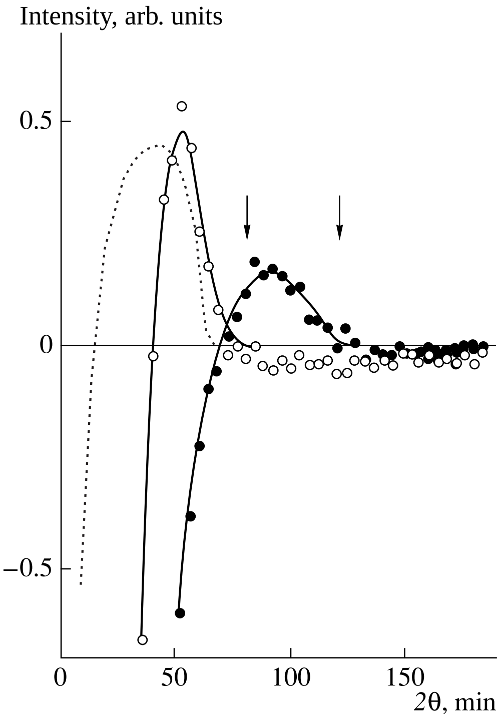


Fig. 12. Angular dependences of the difference intensity of neutron scattering $I_{\mathbf{H} \parallel \mathbf{q}}(q)$ and $I_{\mathbf{H} \perp \mathbf{q}}(q)$ measured in a magnetic field of 4 kOe for Invar alloys $Fe_{65}Ni_{35}$ (○) and Fe_3Pt (●). The arrows indicate the critical angles θ_c for the existence of long-wavelength spin waves. The dashed curve gives the difference intensity for $Fe_{50}Ni_{50}$ alloy calculated theoretically in [19].

reaching a critical value, becomes zero. The dashed curve in this figure gives the similar difference intensity for spin-wave neutron scattering from [19] calculated for $Fe_{50}Ni_{50}$ alloy allowing for the influence of the external magnetic field and dipole–dipole interaction. Also plotted are the results of studying magnetic small-angle neutron scattering in a sample of ordered Fe_3Pt alloy. The arrows indicate the angular positions corresponding to the experimental values of the critical angle. These are 80 and 120 min for $Fe_{65}Ni_{35}$ and Fe_3Pt, respectively. The exchange hardness coefficients for the spin waves D_0 calculated from these data are 120 and 90 meV Å^2, respectively, which satisfactorily agree with the published data [20].

Thus, in measurements of the angular dependence of magnetic small-angle neutron scattering for unmagnetized samples of ferromagnetic materials inelastic spin-wave scattering effects are observed but their intensity is fractions of percent of the total intensity of multiple neutron scattering at the domain structure. Nevertheless, we can assume that spin wave scattering distorts the behavior of the dependence $I(q) \propto q^{-4}$ when the angle θ_c becomes appreciable as a result of the smallness of the exchange hardness coefficient D. For example, this effect is observed in $Fe_{65}Ni_{35}$ (Fig. 11) and Fe_3Pt alloys for which the angular dependence at the asymptote obeys the law $I(q) \propto q^{-3}$ rather than $I(q) \propto q^{-4}$. In addition the angular dependence of the small-angle scattering intensity reveals characteristics which are observed as steps in a log–log plot. However, these aspects require further study.

5. CONCLUSION

These numerical calculations of the general Moliére formula for the multiple neutron scattering intensity show that its analytic analogs obtained in the approximations $\alpha \ll 1$ [9] and $\alpha \gg 1$ [10] can be used jointly to analyze the experimental data on small-angle magnetic neutron scattering. Multiple scattering is observed as a broadening of the primary beam whose thickness dependence is expressed by the power function $w \propto L^\beta$, where β = 0.5–0.6. Using the corresponding experimental dependences of the broadening on the sample thickness, we can make approximate estimates of the effective size of the magnetic domains.

In measurements of the angular dependence of the small-angle magnetic neutron scattering intensity for unmagnetized samples, we must bear in mind that there is always a range of angles where multiple neutron scattering is observed at the domain structure. This generally corresponds to the range of transferred momenta $q < 1$ nm^{-1} where the small-angle scattering intensity is extremely sensitive to the external magnetic field and is approximately two orders of magnitude higher than the intensity of inelastic scattering at spin waves which takes place in this range of angles.

The question naturally arises as to whether it is possible to carry out small-angle neutron scattering experiments to identify spin wave fluctuations without applying an external magnetic field to the sample. This need frequently arises because of the design characteristics of a device where it is difficult to make temperature measurements in an external magnetic field. The answer to this question may be positive if we only take into account the range of small transferred momenta in which multiple neutron scattering effects are small. Information on spin fluctuations is usually only concentrated in that part of the angular dependence where the exponent in the law $I \propto q^{-n}$ is close to two.

The authors are grateful to É. Z. Valiev and Yu. N. Skryabin for numerous useful discussions of the aspects considered here.

This work was supported by the State Scientific-Technical Program "Topical Directions in the Physics of Condensed Media", "Neutron Research" (project no. 4).

REFERENCES

1. R. Von Nardroff, Phys. Rev. **28**, 240 (1927).
2. D. L. Dexter and W. W. Beeman, Phys. Rev. **76**, 1782 (1949).
3. D. J. Hughes, M. T. Burgy, R. B. Heller, *et al.*, Phys. Rev. **75**, 565 (1949).
4. S. Sh. Shil'shtein, V. A. Somenkov, M. Kalanov, *et al.*, Fiz. Tverd. Tela **18**, 3231 (1976).
5. S. G. Bogdanov, E. Z. Valiev, and A. Z. Menshikov, Solid State Commun. **76**, 809 (1990).
6. Yu. I. Smirnov, P. Yu. Peksheev, A. O. Éĭdlin, *et al.*, Fiz. Tverd. Tela. **33**, 2273 (1991); Yu. G. Abov, Yu. I. Smirnov, D. S. Denisov, *et al.*, Fiz. Tverd. Tela **34**, 1408 (1992).
7. N. F. Berk and K. A. Hardman-Rhyne, J. Appl. Crystallogr. **18**, 467 (1985); J. Appl. Crystallogr. **18**, 473 (1985); Physica B **136**, 223 (1988); J. Appl. Crystallogr. **21**, 645 (1988); A. Z. Allen and N. F. Berk, Neutron News **9**, 13 (1988).
8. N. F. Mott and G. Massey, *The Theory of Atomic Collisions* (Clarendon, Oxford, 1965; Mir, Moscow, 1965).
9. S. V. Maleev and B. P. Toperverg, Zh. Éksp. Teor. Fiz. **78**, 315 (1980).
10. S. V. Maleev, R. V. Pomortsev, and Yu. N. Skryabin, Phys. Rev. B **50**, 7133 (1994).
11. A. Z. Menshikov, S. G. Bogdanov, and Yu. N. Skryabin, Physica B **234–236**, 584 (1997).
12. Yu. N. Skryabin, Solid State Commun. **88**, 747 (1993).
13. R. J. Weiss, Phys. Rev. **83**, 379 (1951).
14. A. Z. Men'shikov, S. K. Sidorov, and V. E. Arkhipov, Zh. Éksp. Teor. Fiz. **61**, 311 (1971).
15. M. Hatherly, K. Hirakawa, R. D. Lowde, *et al.*, Proc. Phys. Soc. **84**, 55 (1964).
16. D. I. Svergun and L. A. Feĭgin, *X-ray and Neutron Small-Angle Scattering* (Nauka, Moscow, 1986).
17. Yu. G. Abov, D. S. Denisov, F. S. Dzheparov, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 2194 (1998).
18. A. Z. Men'shikov and V. A. Shestakov, Fiz. Metall. Metalloved. **43**, 722 (1977).
19. M. W. Stringfellow, Report of Atomic Energy Research Establishment (Harwell, UK, 1966), No. AERE-R 4535.
20. I. Ishikawa, S. Onodara, and R. Tajima, JMMM **10**, 183 (1979).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 121–128.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 136–144.
Original Russian Text Copyright © 2000 by Bykovskiĭ, Karavanskiĭ, Kotkovskiĭ, Kuznetsov, Chistyakov, Lomov, Gavrilov.

SOLIDS

Photophysical Processes Stimulated in Nanoporous Silicon by High-Power Laser Radiation

Yu. A. Bykovskiĭ[1], V. A. Karavanskiĭ*[2], G. E. Kotkovskiĭ[1],
M. B. Kuznetsov[1], A. A. Chistyakov[1], A. A. Lomov[3], and S. A. Gavrilov[4]
[1] *Moscow State Engineering-Physics Institute (Technical University), Moscow, 115409 Russia*
[2] *Institute of General Physics, Russian Academy of Sciences,*
Moscow, 117924 Russia
[3] *Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, 177333 Russia*
[4] *Moscow Institute of Electronics, Moscow, 111250 Russia*
**e-mail: karavan@kapella-1.gpi.ru*

Received July 15, 1999

Abstract—Photoprocesses initiated on the surface of porous silicon irradiated with laser radiation with various wavelengths (λ = 266, 337, and 532 nm) in a wide range of intensities (up to 2×10^7 W/cm^2) were investigated. Laser-induced luminescence and laser mass-spectrometry were used as experimental procedures. X-ray reflection was used to determine the parameters of the porous silicon films. The photoluminescence spectra obtained at different wavelengths and low intensities were analyzed. This analysis showed that for an optically thin layer of porous silicon the luminescence spectrum does not depend on the wavelength of the exciting radiation. This indicates the existence of a separate system of levels in porous silicon that are responsible for the luminescence. The behavior of the photoluminescence spectra as a function of the intensity q of the exciting radiation was investigated. It was shown that the luminescence intensity is a nonlinear function of q. At high intensities of the exciting radiation, the luminescence intensity saturates and a short-wavelength shift of the spectra is observed; this is due to the high concentrations of photoexcited carriers. This increases the probability of the experimentally observed nonequilibrium photodesorption of H_2 and Si from the surface of porous silicon. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The steady interest in the photophysical properties of nanoporous silicon is due primarily to the fact that this material is viewed as a promising material for developing a new generation of optoelectronic devices [1]. A unique photophysical property of nanoporous silicon is its photoluminescence, discovered by Canham in 1990 [2] (bulk silicon is an indirect gap semiconductor, and radiative transitions are strongly suppressed in it). However, even though there are many works devoted to this problem, the reason for and the physical mechanisms of photoluminescence of porous silicon are still not completely understood [3].

As is well known [3–6], porous silicon consists of an aggregate of nanocrystals separated by a characteristic distance which is also of the order of several nanometers. Such a structure possesses a very high specific surface area, reaching of the order of 300 m^2/cm^3. Existing models explaining the intense visible-range photoluminescence from porous silicon can be reduced to two alternatives. The first model asserts that transitions between size-quantization levels in silicon nanocrystals are radiative [7–9], and the surface is treated as a source of a nanoradiative recombination channel, where the presence of dangling bonds suppresses luminescence. Here, the smaller the characteristic size of the nanocrystals, the shorter the wavelength of the observed photoluminescence is. In the second model, silicon compounds formed as a result of anodic etching or localized states on the surface of nanocrystals (surface complexes Si–H$_x$ and oxygen vacancies) play the main role in the photoluminescence [10, 11]. In this case, the size-quantization levels determine the absorption, but they bear no relation to luminescence.

There recently have appeared works endeavoring to combine certain characteristic features of these two models in order to explain the photoluminescence of porous silicon [12, 13]. In this approach it is assumed that the absorption spectrum of porous silicon can indeed be explained by size-quantization effects, and surface complexes of the type Si–H$_x$ passivate the dangling bonds and close the nonradiative relaxation channels, while a radiative transition occurs with participation of localized levels associated with the surface and with the position of the size-quantization levels in the nanocrystals themselves. This makes it possible to unite the size-quantization shift of the photoluminescence band and the high efficiency of the photoluminscence.

Most works on the photoluminescence of porous silicon have been performed for low (up to 3 W/cm^2) power densities of continuous-wave exciting laser radiation But, even in these works [7, 10] have shown that as the radiation power increases, the photoluminescence spectra undergoes a shift and distortion, which seem to be due to a characteristic feature of the energy structure of the nanocrystals and its behavior under the conditions of strong excitation at high intensities (10^3–10^6 W/cm^2). Investigation of the processes giving rise to such evolution of the photoluminescence spectra can give additional information about its mechanism in porous silicon. This is now of great interest. Moreover, these investigations make it possible to understand the characteristic features of the electronic excitation relaxation and transfer processes and to answer questions concerning the possibility of photoprocesses, such as photodesorption, photodissociation, and ablation, occurring on the surfaces of nanocrystals of porous silicon [12, 14].

In summary, our objective in this work was to investigate the photoprocesses occurring on the surface of porous silicon under irradiation with high-power laser radiation.

2. EXPERIMENTAL PROCEDURE AND SAMPLES

Layers of porous silicon were obtained by anodization in an electrolyte (HF(49%) : C_2H_5OH in a 1 : 2 volume ratio) in a two-chamber electrochemical cell. Platinum was used as the cathode. After anodization, the samples were rinsed in deionized water and dried in a stream of dry air.

One of the main steps in the sample preparation process is choosing the film formation regimes. On the one hand, it is known that porous silicon films can possess a layered structure: different layers have different photoluminescence properties. Indeed, the properties of the porous-silicon layer formed can vary during the anodization process [15], and under certain conditions even the etching mechanism changes. On the other hand, it has been shown that the photoluminescence spectra can depend strongly on the wavelength of the exciting radiation. Effects associated with the nonuniformity and different excitation apparently can be distinguished if the film of porous silicon is sufficiently thin so that it can be treated as being uniform and so that the excitation conditions at different wavelengths do not differ much. At the same time, preliminary investigations have shown that immediately after the onset of anodization there exists a certain period during which a stationary state of etching is established (this period decreases with increasing anodization current density). This suggests that a layer with variable characteristics, which is determined by a transition from the etching regime of the initial smooth surface to a stationary regime in which extended pores are formed, should exist on the outer boundary. The thickness of this layer presumably could be several diameters of an average pore. Several samples were prepared using different current density j and anodization duration t with a surface layer in the form of a thin film of porous silicon and one sample was prepared in the form of a thick film. The constant value of the charge $Q = jt = 0.15$ C/cm^2 with different current densities was chosen as the parameter for thin films: $j = 1$, 10, and 50 mA/cm^2. In what follows these samples are designated as nos. 1, 2, and 3. Wafers of standard (100) p-type KDB10 silicon (resistivity $\rho \sim 10\ \Omega$ cm) were used as substrates. To prevent variance of the parameters, due to differences between wafers, from having an effect, the entire series of samples was prepared from the same wafer. The sample no. 4 with a thick film was obtained with $Q = 0.85$ C/cm^2 ($j = 12$ mA/cm^2, $t = 425$ s, KDB12 substrate).

For thin films, the thickness and effective porosity were determined by measuring the angular dependence of the intensity of the mirror X-ray reflection near the angles of total external reflection [16]. The thickness of the thick film was estimated from the image of the cleavage surface in an optical microscope and was ~10 μm.

To study the photoluminescence of porous silicon in a wide range of intensities of the exciting radiation and for various wavelengths, a laser fluorimeter, equipped with a repetitive-pulse YAG-Nd^{3+} laser with the harmonics $\lambda = 532$ and 266 nm and a nitrogen laser ($\lambda = 337$ nm). The pulse duration was 10 ns. A collection of interchangeable filters and a system of focusing lenses made it possible to change the power density of the laser radiation at the sample from 7×10^4 to 10^7 W/cm^2 for $\lambda = 532$ nm and from 10^3 to 4×10^5 W/cm^2 for $\lambda = 266$ nm. The power density of the nitrogen laser was 10^4 W/cm^2.

The possible irreversible photoprocesses that could be initiated on the surface of porous silicon by high-power laser radiation (photodesorption, photodissociation, and ablation) were traced by laser-mass spectrometry [12].

In summary, the combination of methods used made it possible to monitor the development and structure of the porous silicon films and to investigate the photoprocesses occurring in porous silicon irradiated with high-power laser radiation, starting from photoluminescence and ending with photodesorption and ablation.

3. EXPERIMENTAL RESULTS

The results of X-ray reflection measurements of the structural parameters of the films are presented in the table.

Experiments with pulsed pumping were performed using the following wavelengths of the exciting radiation: λ = 532, 337, and 266 nm. Since specially prepared samples with very thin porous layers were used (d = 105, 92, and 85 nm), the radiation with each wavelength could be absorbed over the entire depth of the porous layer. This made it possible to decrease (a) the effect of the structural nonuniformity of the porous silicon over depth on the experimental results [17] and (b) the effect of a difference in the absorption coefficients for the first two wavelengths, which in the case of thick films always results in a mismatch between the depths of the luminescing regions. We also note that by choosing samples with a thin porous layer it is possible, according to our estimates, to eliminate any effect due to the reabsorption of the characteristic luminescence in the porous-silicon layer. Reabsorption of photoluminescence must be taken into account only for thicknesses of such samples greater than 10 μm.

The luminescence spectra for thin films are presented in Fig. 1. It is evident that the positions of the maxima in the photoluminescence spectra of porous silicon films obtained with different current densities are different. A large shift into the short-wavelength region corresponds to a high current density. This agrees with the well-known result that films formed with a high current density have smaller silicon nanocrystals. It is also evident that on switching from λ = 532 nm to λ = 266 nm the maximum of the spectra for samples nos. 1 and 2 shifts into the short-wavelength region, while the position of the spectrum of sample no. 3 remains virtually unchanged. For comparison, the spectra for sample no. 4 with a thick film of porous silicon, for which this shift is greatest, 34 nm, is presented in Fig. 2. The results obtained for the shift agree qualitatively with the results of other investigations for comparatively thick films of porous silicon (d > 1 μm) [9]. Apparently, the observed dependence of the shift on the film thickness and on the conditions of formation (porosity and current density) is more important. As one can see from the plots in Figs. 1 and 2, the shift of the maximum of the photoluminescence spectrum decreases monotonically with thickness. It is 15 and 10 nm for samples nos. 1 and 2, respectively. For the thinnest sample (no. 3, thickness d = 85 nm) this shift is virtually absent, and the spectra obtained at three wavelengths (266, 337, and 532 nm) are identical to within the limits of the experimental error.

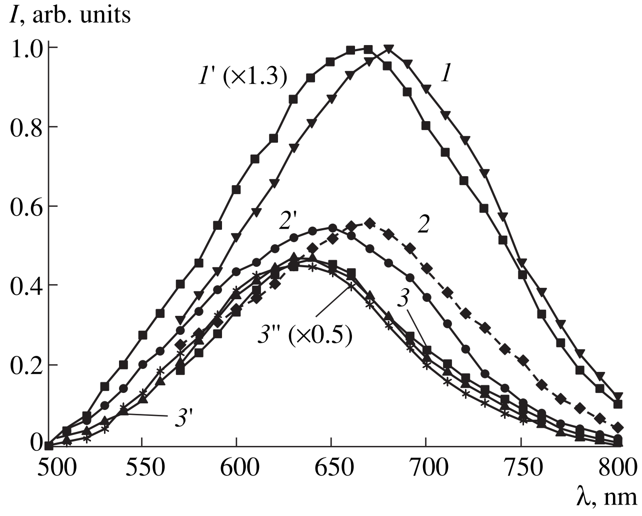


Fig. 1. Photoluminescence spectra of samples of porous silicon. (*1*) sample no. 1, λ = 532 nm, (*1*') sample no. 1, λ = 266 nm; (*2*) sample no. 2, λ = 532 nm, (*2*') sample no. 2, λ = 266 nm; (*3*) sample no. 3, λ = 532 nm, (*3*') sample no. 3, λ = 266 nm, (*3*'') sample no. 3, λ = 337 nm. Laser power flux density $q_{266} = 2.5 \times 10^3$ W/cm^2 and $q_{532} = 2 \times 10^5$ W/cm^2.

The physical reason for the observed shift could be the difference not only in the optical thickness of the sample but also in their porosity. Indeed, samples with a higher porosity typically have smaller nanocrystals and, as is well known, long-wavelength radiation (λ = 532 nm in our case) will be absorbed only by relatively "large" crystallites, whose luminescence spectrum is shifted into the "red" region of the spectrum. But, then, the largest shift of the spectra should be expected for the third sample, for which this shift is absent. To confirm the arguments presented above, a control experiment was performed using thin samples of nanoporous silicon, whose luminescence spectrum with a maximum in the region 580–610 nm indicates that the characteristic size of the crystallites is very small. However, even in this case, good agreement was observed between the luminescence spectra for excitation with radiation with λ = 532 and 266 nm.

Therefore it can be concluded that the shift of the photoluminescence spectra in the thicker samples (nos. 1 and 2 and also no. 4) is due precisely to the different absorption coefficient, more accurately, the different "optical thickness" for different wavelengths. For sample no. 3 the optical thicknesses for all excitation wavelengths are the same.

In thick samples the short-wavelength radiation (with λ = 266 and 337 nm and two to three orders of magnitude larger absorption coefficient than for λ = 532 nm) excites the top layer, where the porosity is greater than the average porosity of the sample and which is more strongly oxidized. For this reason, the spectrum of such samples will be shifted into the short-

Structural parameters P and d of films of porous silicon according to measurements performed by the method of total external reflection of X-rays [21]

Sample no.	P, %	d, nm
1	58 ± 1	105 ± 1
2	38 ± 1	92 ± 1
3	55 ± 1	85.0 ± 0.5

Note: P is the porosity; d is the average film thickness.

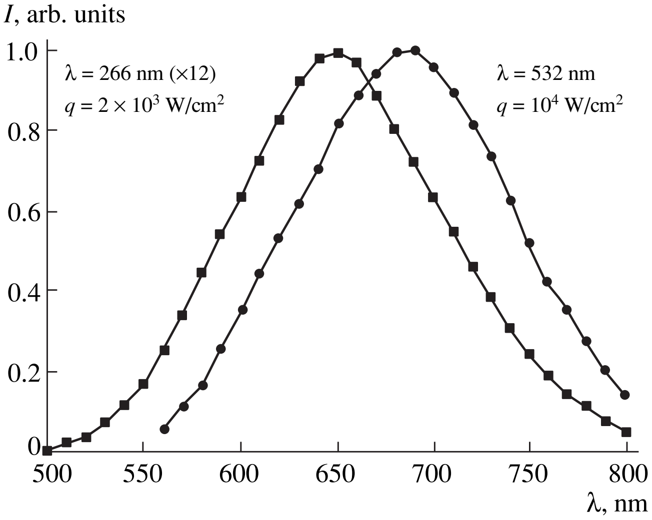


Fig. 2. Photoluminescence spectra of a sample of porous silicon under excitation with radiation with λ = 532 and 266 nm.

wavelength region compared with the spectrum obtained by excitation with λ = 532 nm radiation, for which nanocrystals are excited on the surface and in the interior volume of the porous layer.

Thus, we arrive at the following important conclusion. In samples with the same optical thickness, for the entire excitation spectrum lying above the absorption edge, the photoluminescence spectrum of porous silicon does not depend on the excitation wavelength (in our case for radiation with $\lambda \leq 532$ nm). Under these conditions, irrespective of the photon energy, after excitation the charge carriers relax nonradiatively on a separate system of levels in nanocrystals of porous silicon, and luminescence occurs from this system.

We shall now examine the behavior of the photoluminescence for various intensities of the exciting radiation. Figures 3 and 4 show the dependence of the photoluminescence intensity of the same samples on the radiation power q for wavelengths λ = 266 and 532 nm, respectively. The laser fluorimeter was tuned to the wavelengths corresponding to the maxima of the photoluminescence spectra for small q. The nonlinear dependence of the photoluminescence intensity for $q > 7 \times 10^3$ W/cm² at λ = 266 nm and $q > 2 \times 10^5$ W/cm² at λ = 532 nm with the intensity I subsequently reaching values for which the photoluminescence saturates is especially interesting.

One explanation for the experimentally observed character of the dependence of I on q is saturation of the excited states. Indeed, since the lifetime of the photoex-

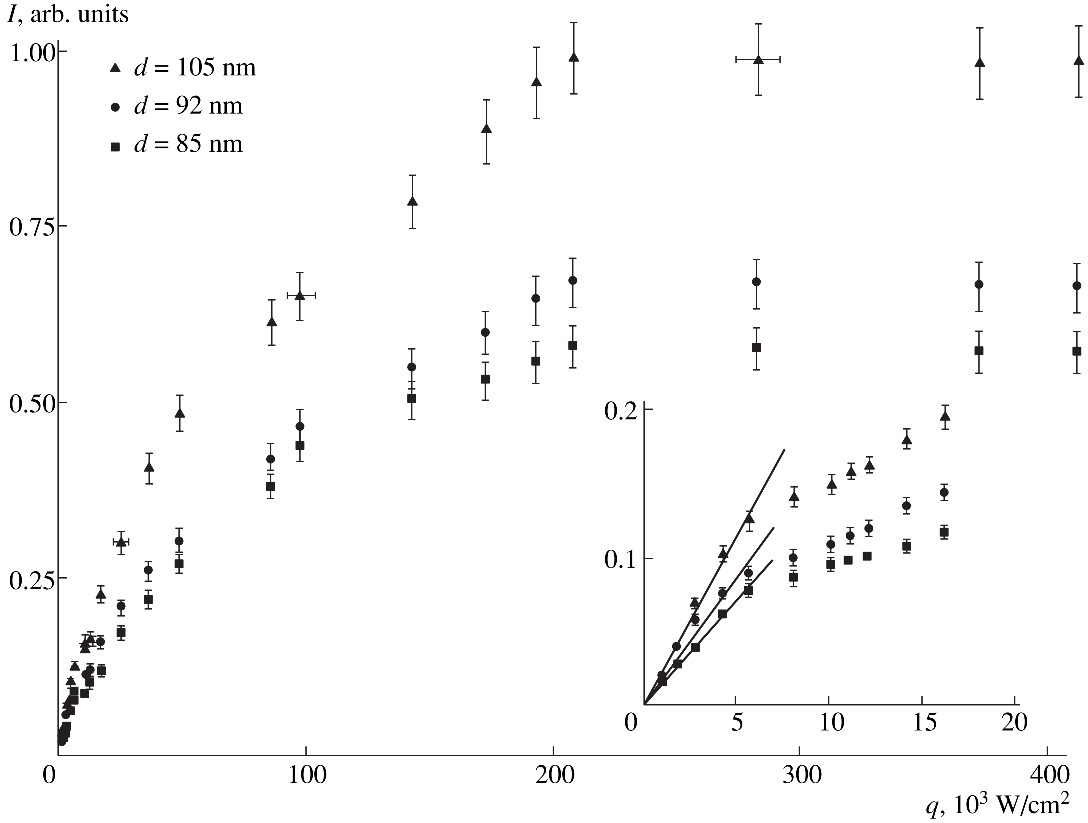


Fig. 3. Photoluminescence intensity versus the power density q of the exciting radiation for samples nos. 1, 2, and 3. The wavelength is λ = 266 nm.

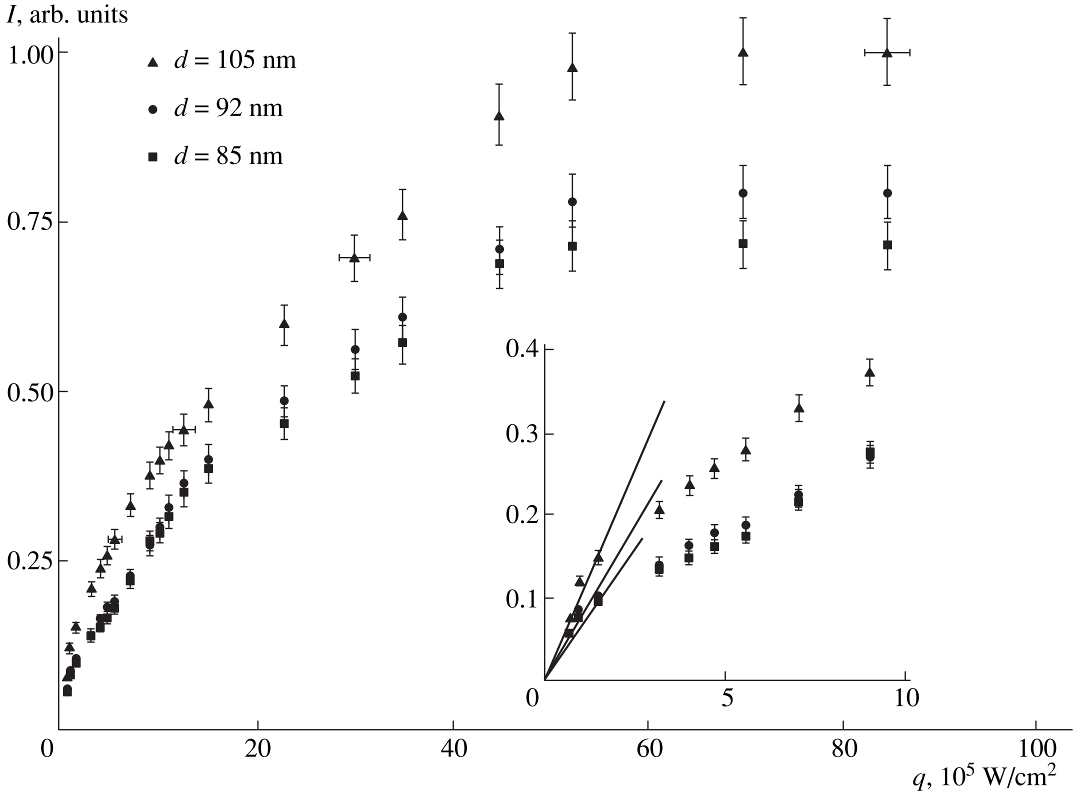


Fig. 4. Photoluminescence intensity versus the power density q of the exciting radiation for samples nos. 1, 2, and 3. The wavelength is $\lambda = 532$ nm.

cited carriers in porous silicon is of the order of $\tau \sim 10$–$100\ \mu$s [3, 4], it can be expected that carriers are effectively transferred into excited states under our experimental conditions. Estimates made on the basis of [18] show that an appreciable deviation from a linear dependence $I(q)$ should be expected for

$$q_0 \geq \frac{\hbar\omega}{\tau\sigma},$$

where σ is the absorption cross section of the radiation exciting the luminescence.

For $\lambda = 532$ nm $q_0 > 10^6$ W/cm^2, and for $\lambda = 266$ nm $q_0 > 2 \times 10^4$ W/cm^2. It is evident that the experimental dependences confirm these estimates.

Thus, high concentrations of photoexcited carriers, for which, for example, nonradiative Auger recombination of carriers can be significant [19], will correspond to the flux densities of the exciting radiation ($q = 10^7$ W/cm^2 for $\lambda = 532$ nm) which were realized under our experimental conditions.

To obtain a more complete picture of the behavior of the photoluminescence for various intensities we must examine the photoluminescence spectra.

The photoluminescence spectra obtained for the same samples with various limits of q and pump wavelengths $\lambda = 266$ and 532 nm, respectively, are displayed in Figs. 5 and 6. It is evident in the figures that as the intensity of the laser radiation increases, the spectra shift into the short-wavelength region. The shift of the maxima of the photoluminescence spectra with excitation by light with wavelength $\lambda = 266$ nm is $\Delta\lambda = 35$,

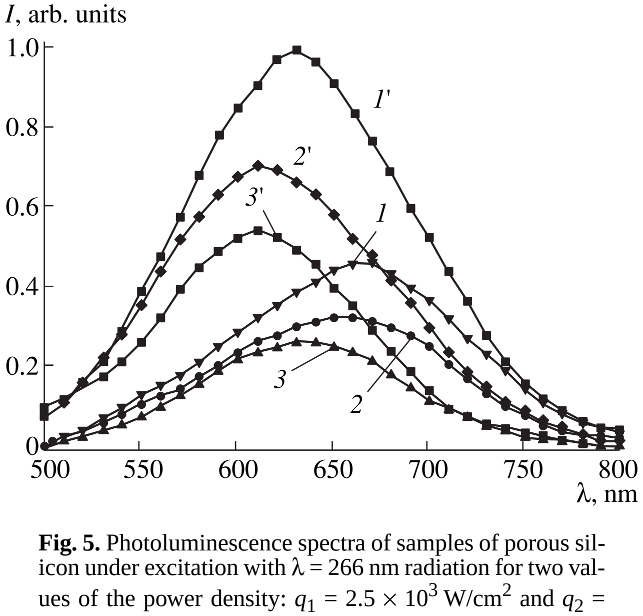


Fig. 5. Photoluminescence spectra of samples of porous silicon under excitation with $\lambda = 266$ nm radiation for two values of the power density: $q_1 = 2.5 \times 10^3$ W/cm^2 and $q_2 = 10^4$ W/cm^2: (*1*) sample no. 1 with q_1, (*1*') sample no. 1 with q_2; (*2*) sample no. 2 with q_1, (*2*') sample no. 2 with q_2; (*3*) sample no. 3 with q_1, (*3*') sample no. 3 with q_2.

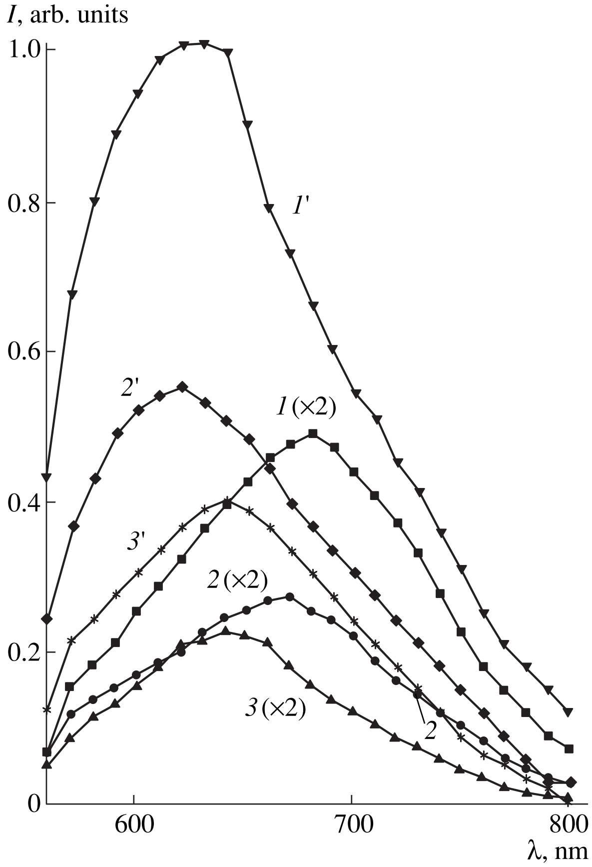


Fig. 6. Photoluminescence spectra of samples of porous silicon under excitation with λ = 532 nm radiation for two values of the power density: $q_1 = 2 \times 10^5$ W/cm^2 and $q_2 = 2.5 \times 10^6$ W/cm^2: (*1*) sample no. 1 with q_1, (*1*') sample no. 1 with q_2; (*2*) sample no. 2 with q_1, (*2*') sample no. 2 with q_2; (*3*) sample no. 3 with q_1, (*3*') sample no. 3 with q_2. The distortion of the spectra in the range 550–570 nm is caused by the OS-23-1 filter, which cuts off λ = 532 nm radiation.

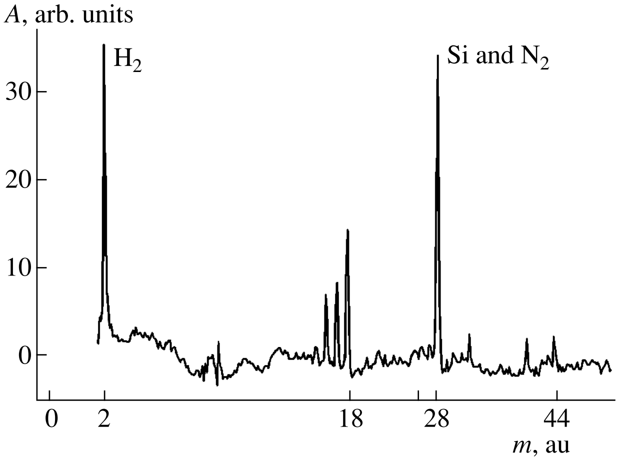


Fig. 7. Characteristic mass spectrum under the action of radiation with $q = 5 \times 10^6$ W/cm^2 on the surface of the porous-silicon sample.

30, and 25 nm, respectively, for samples nos. 1, 2, and 3 (Fig. 5); for excitation with light with wavelength λ = 532 nm the shift is Δλ = 45, 40, and 5 nm, respectively, for samples nos. 1, 2, and 3, respectively (Fig. 6).

Time-resolved spectroscopy [9, 10, 20] records the shift of the maximum of the photoluminescence with time into the long-wavelength region, i.e., “fast” relaxation, with a time of the order of 10 ns; radiative relaxation occurs in the short-wavelength range; “slow” relaxation (τ ~ 10–100 μs) occurs in the long-wavelength range [3, 9]. Some investigators [9] attribute this to the fact that in smaller nanocrystals (and, correspondingly, shorter-wavelength photoluminescence spectrum), the radiative relaxation time is shorter than in nanocrystals with a larger characteristic size.

Thus, the short-wavelength shift of the photoluminescence spectrum of porous silicon with increasing *q* can be explained by the fact that in this case the smaller nanocrystals, for which the dependence of *I* on *q* is still linear, start to make a large contribution to the photoluminescence, while in large nanocrystals photoluminescence starts to saturate.

It should also be noted that the magnitude of the shift of the maximum of the photoluminescence with increasing *q* is directly proportional to the thickness of the sample. This can be explained by the fact that the nanocrystals in a thinner sample are more uniform with respect to their characteristic size and therefore undergo saturation for close values of *q*.

Mass-spectrometric investigations showed that for *q* ranging from 3×10^6 to 8×10^6 W/cm^2 photodesorption of H_2 and Si is observed simultaneously at values for which photoluminescence saturates. The characteristic mass spectrum is shown in Fig. 7, and the interpreted mass spectrum is displayed in Fig. 8. Generally speaking, the high carrier densities realized under our experimental conditions on nanocrystals should lead to a sharp increase in the probability of surface photoprocesses, such as photodesorption, observed in our experiments, and photodissociation. Efficient laser photooxidation is possible. On the other hand, the observed photodesorption, in principle, increases the number of dangling bonds and therefore also the number of nonradiative relaxation channels. This mechanism can lead to an irreversible degradation of luminescence.

The dependence of the luminescence intensity at the maximum of the spectrum on the time or on the number of laser pulses is displayed in Fig. 9 for λ = 532 nm and $q = 8 \times 10^6$ W/cm^2. It is evident that for 1000 pulses the intensity decreases by 40%.

As the intensity of the exciting radiation increases further (above 8×10^6 W/cm^2 for λ = 532 nm), a sharp and irreversible decrease in the photoluminescence intensity occurs. Mass-spectrometric investigations show that in this case the following products of laser

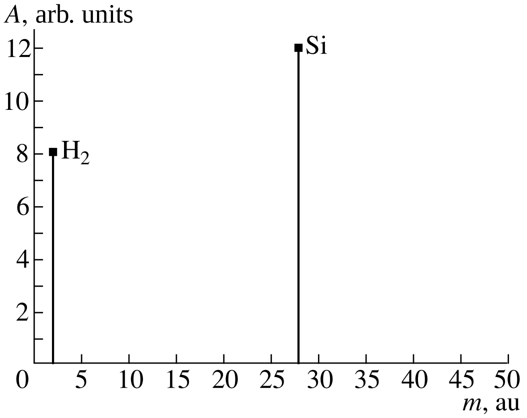


Fig. 8. Interpreted mass spectrum under the action of radiation with $q = 5 \times 10^6$ W/cm^2 on the surface of a porous-silicon sample.

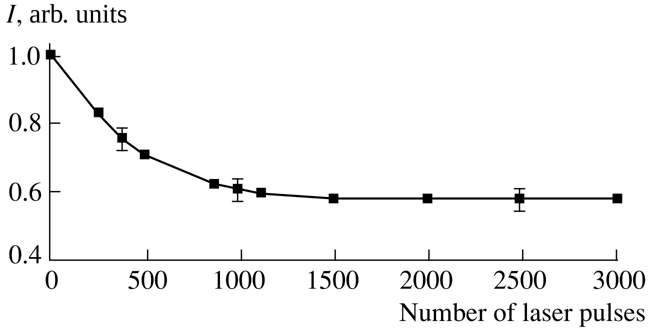


Fig. 9. Peak intensity of the luminescence of sample no. 3 versus the number of pulses of laser radiation with λ = 532 nm for $q = 8 \times 10^6$ W/cm^2.

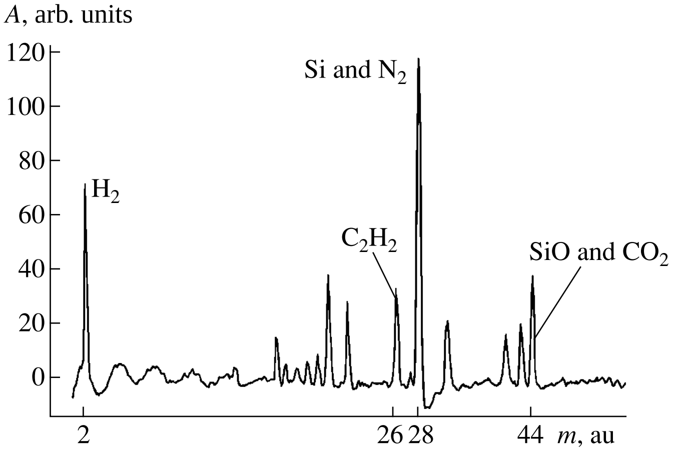


Fig. 10. Characteristic mass spectrum under action of radiation with $q = 2 \times 10^7$ W/cm^2 on the surface of a sample of porous silicon.

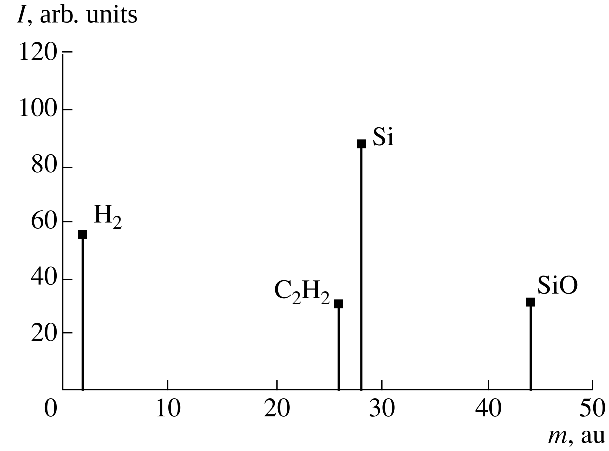


Fig. 11. Interpreted mass spectrum under action of radiation with $q = 2 \times 10^7$ W/cm^2 on the surface of a sample of porous silicon.

action are formed: H_2, Si, C_2H_2, and SiO (Figs. 10, 11), which are characteristic for ablation and destruction of the porous layer [21].

4. CONCLUSIONS

Photoprocesses occurring under the action of laser radiation on the surface of porous silicon were investigated in a wide range of intensities, up to ~10^7 W/cm^2, for wavelengths λ = 532, 337, and 266 nm.

It was shown that samples with the same optical thickness for the entire excitation spectrum lying above the absorption edge, the photoluminescence spectrum of porous silicon does not depend on the excitation wavelength (in our case for radiation with $\lambda \le 532$ nm) in the linear region of the dependence $I(q)$. Under these conditions, irrespective of the photon energy, after excitation the charge carriers relax nonradiatively on a separate system of levels in nanocrystals of porous silicon, and luminescence occurs from this system.

The nonlinear character of the dependence of the photoluminescence intensity on q, due to the high level of excitation of the semiconductor, was observed experimentally.

The shift of the luminescence spectrum as a function of the intensity of the exciting radiation into the short-wavelength region of the spectrum was observed and investigated.

Photodesorption with formation of H_2 and Si was observed for $q \sim 8 \times 10^6$ W/cm^2.

It was shown that nonequilibrium photodesorption gives rise to new nonradiative relaxation channels on account of an increase in the number of dangling bonds. For $q > 9 \times 10^6$ W/cm^2 and λ = 532 nm, a sharp decrease in the photoluminescence intensity on account of an increase in the nonradiative relaxation rate as a result of heating and ablation was observed.

ACKNOWLEDGMENTS

This research was supported in part by the Ministry of Science and Technologies of the Russian Federation (projects no. 08-02-48 and no. 97-10-73) and the program for Integration of Higher Education and Fundamental Science (project no. A0103).

REFERENCES

1. N. Koshida and H. Koyama, Appl. Phys. Lett. **60**, 347 (1992).
2. L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).
3. A. G. Cullis, L. T. Canham, and P. D. J. Calcott, J. Appl. Phys. **82** (3), 909 (1997).
4. B. M. Kostishko, A. M. Orlov, S. N. Mikov, *et al.*, Nonorganic Materials **31**, 444 (1995).
5. S. M. Prokes, O. J. Glembocki, V. M. Bermúdez, *et al.*, Phys. Rev. B **45**, 13 788 (1992).
6. C. Tsai, K. H. Li, J. Sarathy, *et al.*, Appl. Phys. Lett. **59**, 2814 (1991).
7. M. Koós, I. Pócsik, and É. B. Vázsonyi, Appl. Phys. Lett. **62**, 1797 (1993).
8. V. Lehmann and U. Gösele, Appl. Phys. Lett. **58** (8), 856 (1991).
9. M. S. Bresler and I. N. Yassievich, Fiz. Tekh. Poluprovodn. **27**, 871 (1993) [Semiconductors **27**, 475 (1993)].
10. K. L. Narasimhan, S. Banerjee, A. K. Srivastava, *et al.*, Appl. Phys. Lett. **62**, 331 (1993).
11. M. Ohmukai and Y. Tsutsumi, J. Appl. Phys. **84**, 4459 (1998).
12. Yu. A. Bykovskiĭ, G. E. Kotkovskiĭ, M. B. Kuznetsov, *et al.*, in *XVI International Conference on Coherent and Nonlinear Optics Technical Digest* (Moscow, 1998), p. 243.
13. M. Kondo, J. Non-Crystalline Solids **164–166**, 941 (1993).
14. A. Hashimoto, K. Iwata, M. Ohkubo, *et al.*, J. Appl. Phys. **75**, 5447 (1994).
15. S. A. Gavrilov, T. N. Zavaritsakaya, V. A. Karavanskiĭ, *et al.*, Élektrokhimiya **33**, 1064 (1997).
16. V. A. Karavanskiĭ, A. A. Lomov, E. V. Rakova, *et al.*, Poverkhnost, no. 12, 1999 (in press).
17. A. N. Obraztsov, V. A. Karavanskiĭ, H. Okushi, *et al.*, Fiz. Tekh. Poluprovodn. **32**, 1001 (1998) [Semiconductors **32**, 896 (1998)].
18. V. P. Gribkovskiĭ, *The Theory of Absorption and Emission of Light in Semiconductors* (Nauka i Tekhnika, Minsk, 1975).
19. V. P. Gribkovskiĭ and G. N. Yaskevich, Zh. Prikl. Spektrosk. **12**, 231 (1970).
20. R. Laiho, A. Pavlov, O. Hovi, *et al.*, Appl. Phys. Lett. **63**, 275 (1993).
21. Yu. A. Bykovskiĭ, V. A. Karavanskiĭ, G. E. Kotkovskiĭ, *et al.*, Poverkhnost, No. 9, 23 (1999)

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 129–132.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 145–148.
Original English Text Copyright © 2000 by Kageyama, Onizuka, Ueda, Nohara, Suzuki, Takagi.

Low-Temperature Specific Heat Study of $SrCu_2(BO_3)_2$ with an Exactly Solvable Ground State[¶]

H. Kageyama*[1], K. Onizuka[1], Y. Ueda[1], M. Nohara[2],
H. Suzuki[2], and H. Takagi[2]

[1] *Institute for Solid State Physics, University of Tokyo,*
Tokyo 106–8666, Japan
[2] *Graduate School of Frontier Science, University of Tokyo,*
Tokyo 113–8656, Japan
**e-mail: kage@issp.u-tokyo.ac.jp*

Received July 22, 1999

Abstract—The specific heat of a two-dimensional spin gap system $SrCu_2(BO_3)_2$ realizing the Shastry–Sutherland model was measured between 1.3 and 25 K under various magnetic fields up to 12 T. The analysis based on an isolated dimer model in a low temperature region revealed that the value of the spin gap at zero field is $\Delta = 34.4$ K. It turned out that Δ decreases in proportion to H due to the Zeeman splitting of the excited triplet levels. This simplest model, however, fails to reproduce the result in a high-temperature region, suggesting rather strong spin–spin correlation of the system. © *2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Exactly solvable models have been extensively studied in the area of strongly correlated electron systems for the purpose of elucidating various exotic physical phenomena because some rigorous results can be derived from them, sometimes providing us a crucial key to solve underlying problems of the phenomena. Such models, even if being far from realistic, can remain tantalizing theoretical subjects owing to the beauty of the solutions. For example, Majumdar and Ghosh first proved that an exact dimer ground state for *z* one-dimensional spin chain imposed a stringent condition on the first and second nearest neighbor interactions [1]. Stimulated by this discovery, a number of systems with the identical exact wave function have been explored from the theoretical point of view for one-, two,- and three-dimensions (see, for example [2] and references therein). However, in spite of extensive efforts by chemists to tailor experimental examples, no material had been discovered for a long time.

Recently, we reported the magnetic properties of an inorganic compound $SrCu_2(BO_3)_2$, which consists of a two-dimensional orthogonal dimer lattice, concluding that this material verifies the Shastry–Sutherland model, which has the exact dimer ground state [3–5]. Although an imaginary lattice Shastry and Sutherland considered—i.e., a two-dimensional square lattice with some additional diagonal bonds—differs from the real one of $SrCu_2(BO_3)_2$, these two are equivalent from a topological point view. The value of the spin gap was estimated from various measurements like measurements of the temperature variation of the magnetic susceptibility (34 K) [6] and electron spin resonance (ESR; 34.7 K) [7]. It was also found that the spin system for $SrCu_2(BO_3)_2$ is fairly frustrated, located very close to the critical point $(J'/J)_c = 0.70$ between the exact dimer state and the Néel-ordered state [3, 5]: the ratio of intradimer and interdimer interactions, respectively, $J = 100$ K and $J' = 68$ K, is 0.68. Furthermore, several quantized plateaux were observed in the magnetization [3, 6, 8], which originates from the extremely localized triplet excitations [5].

In the present paper, we performed the specific heat measurement of $SrCu_2(BO_3)_2$ under magnetic fields H in order to obtain more information on the exchange interactions as well as the effect of the spin-gapped behavior upon H. The data were analyzed in terms of an isolated dimer model, and the spin gap in the absence of the field was evaluated to be 34.4 K. Furthermore, it was found that application of magnetic fields causes the Zeeman splitting of the excited triplet states, leading to a H-linear decrease in the value of the spin gap.

2. EXPERIMENT

The specific heat measurement was performed by a heat-relaxation method [9] in a temperature range between 1.3 and 25 K under magnetic fields between 0 and 12 T. A bulk single crystal of $SrCu_2(BO_3)_2$ was used, which was grown by the traveling solvent floating zone (TSFZ) method with an image furnace using a solvent, $LiBO_2$ under flowing O_2 gas ($P_{O_2} = 1$ atm, 99.99%). For a detailed procedure of the crystal growth, see [10]. A piece of the crystal with the dimensions of $2 \times 2 \times 1$ mm was attached to a sapphire

---

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

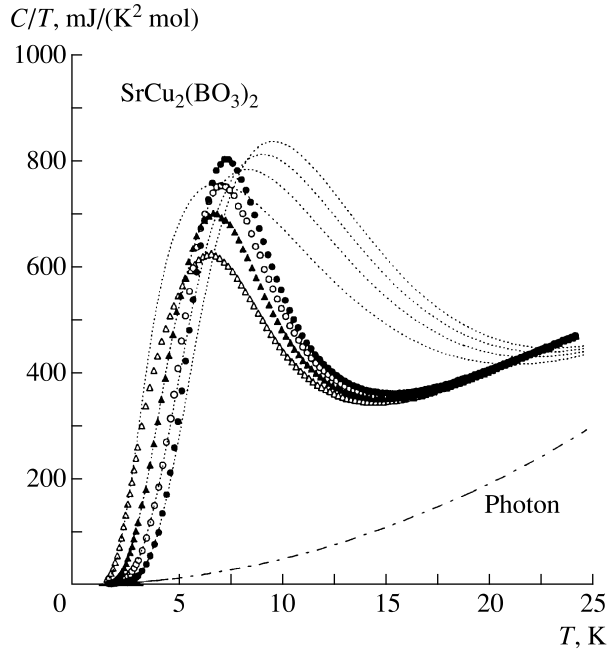


Fig. 1. C/T versus T measured at $H = 0$ (●), 6 (○), 9 (▲), and 12 T (△). Dotted curves are the calculations based on the isolated dimer model for $\Delta(0) = 34.4$ K. Dot-dashed curve represents the phonon term, βT^3 ($\beta = 0.460$ mJ/K^4 mol).

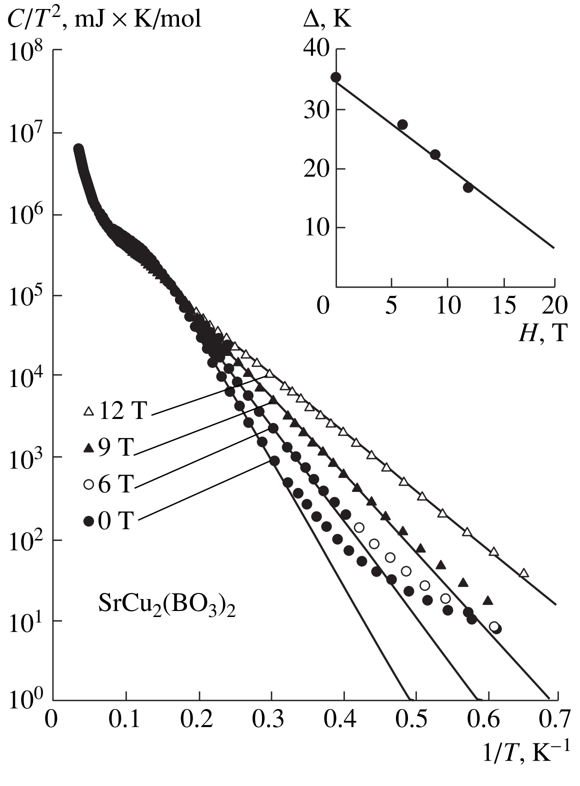


Fig. 2. Logarithmic plot of CT^2 as a function of $1/T$. Solid lines denote the fit to equation (2). Inset shows the magnetic field variation of $\Delta(H)$.

substrate by a small amount of Apiezon *N* grease. The magnetic fields were applied perpendicular to the ab plane, i.e., the Shastry–Sutherland lattice. The substrate was weakly coupled by tungsten wires to a copper heat sink. A bare chip of Cernox resistance sensor (Lake Shore) was used as a thermometer to minimize the addenda heat capacity. The magnetic field dependence of the thermometer was calibrated using a capacitance thermometer. The heat capacity of the sample was obtained by subtracting the addenda heat capacity, which was determined in a separate run without the sample. No appreciable magnetic field dependence was observed for the addenda heat capacity. The resolution of the measurement was about 0.5%, and the absolute accuracy determined from the measurement of a Cu standard was better than 5%. The measurements were performed with increasing temperature.

3. RESULTS AND DISCUSSION

A total specific heat divided by T, C/T, measured in the absence of a magnetic field is plotted as a function of T by closed circles in Fig. 1. With decreasing T from 15 K, C/T rises, reaches a round maximum at 7.5 K, and then falls rapidly, approaching naught. These behaviors, that is to say, the so-called Schottky anomalies are typical of spin-singlet system with a finite spin gap to a lowest exited state. A gradual increase in C/T with T above 15 K comes from the phonon term, with is in general known to vary as $C \propto \beta T^3$. As also shown in Fig. 1, qualitatively similar features described above appear even when magnetic fields are applied, indicating that the system still has a spin–gapped ground state at least for $H < 12$ T. A prominent difference is that a peak of C/T shifts to lower temperature with rising H: the temperature at which C/T–T curve reaches a maximum ($=T_{\max}$) for $H = 6.9$ and 12 T is, respectively, 7.3, 6.9 and 6.8 K, implying a reduction in the actual size of the spin gap $\Delta(H)$ with H. This is quantitatively discussed below.

Because of a lack of an appropriate theory for the specific heat from the standpoint of the Shastry–Sutherland model, we will analyze the experimental data utilizing the isolated dimer model, where J' is neglected and only J is taken into consideration. Let us define the magnetic specific heat under a certain magnetic field H as $C(H)$. Take the example of $H = 0$, $C(0)$ is given by the following formula,

$$C(0) = \frac{3R(\Delta(0)/T)^2\exp(\Delta(0)/T)}{[1+3\exp(\Delta(0)/T)]^2}, \qquad (1)$$

where R is 8.30 J/(K mol) (see, for example, [11]). Likewise, $C(H)$ for a finite magnetic fields is easily calculated. In the low temperature limit, the magnetic specific heat the isolated dimer model can be reduced to the following expression as long as the system is in a gapful state:

$$C(H) \propto T^{-2}\exp(-\Delta(H)/T). \qquad (2)$$

Thus CT^2 is plotted against $1/T$ in a logarithmic scale as shown in Fig. 2. One can see all data roughly follows a linear reversal-temperature dependency. Using the reduced expression of equation (2), we obtained $\Delta(0)$ = 35.9 K, Δ(6 T) = 27.5 K, Δ(9 T) = 22.5 K, and Δ(12 T) = 16.8 K. The deviation from the calculations (the solid lines in Fig. 2) in lower temperature region, more prominent in case of lower field, is for a most part due to the phonon contribution which is neglected here and will be included later. The obtained values of $\Delta(H)$ are plotted in the inset of Fig. 2 as a function of H. It is clear that $\Delta(H)$ decreases nearly in proportion to H. The origin of the decrease should be the Zeeman splitting of the excited states. Namely, a three-fold degeneracy of the lowest excited triplet states (S = 1) in the absence of the magnetic field is lifted up by applied magnetic field. $\Delta(0)$ was estimated to be 35.0 K using the following relation: $\Delta(H) = \Delta(0) - g\mu_B H$, where g is the g-factor of the Cu^{2+} electron spin and μ_B is the Bohr magneton. An isotropic g-value, i.e., g = 2.0 was assumed. The obtained value of $\Delta(0)$ is consistent with that obtained in other measurements using a single crystalline $SrCu_2(BO_3)_2$ such as the magnetic susceptibility (34 K) [6], ESR (34.7 K) [7] and Boron nuclear magnetic resonance (B-NMR; 36 K) [12], Cu-NMR (35 K) [12], and neutron scattering (34 K) [13].

Next, let us take a phonon term into consideration. Then, the total specific heat is given by the sum of the magnetic and phonon terms, $C = C(H) + \beta T^3$. Dotted curves in Fig. 1 denote the results of the global least-square fit in the T range well below the spin-gap size, namely, 2.6 K < T < 4.8 K for 0 T, 2.4 K < T < 4.1 K for 6 T, 2.1 K < T < 3.5 K for 9 T, and 1.5 K < T < 2.8 K for 12 T, from which we obtained once again a reasonable value of $\Delta(0)$ = 34.4 K together with β = 0.460 mJ/(K^4 mol) and g =2.03. The phonon contribution is independently shown by the dot-dashed curve in Fig. 1, which also seems to reproduce the temperature dependence of the experiment above 15 K.

As demonstrated above, it seems that the isolated dimer model nicely reproduces the experimental data, providing a consistent value of $\Delta(0)$. In a higher temperature region, however, the deviation between the experiment and the theory is appreciable. One can notice from Fig. 1 that experimental $T_{\max}$ is lower as compared with the theoretical one in any magnetic field, and above $T_{\max}$ the value of experimental C/T is much suppressed. In Fig. 3, we show the T variation of the magnetic entrope of the system for H = 0, which should reach $2R\ln 2$ ideally in the high-T limit. For comparison, a theoretical curve for the isolated dimer model for $\Delta(0)$ = 34.4 K is shown by the solid line. The experimental entropy starts to deviate largely from the theoretical one at around 10 K. For example, the magnetic entrope at 25 K is still about and 74% of that for the isolated dimer model and 62% of the full entropy. This indicated that the spin system of $SrCu_2(BO_3)_2$ is effectively correlated over much higher temperatures, and thus consistent with the estimation of exchange constants by Miyahara and Ueda; J = 100 K and J' = 68 K [5]. It is noteworthy that the value of J is identical with that of $\Delta(0)$ for the isolated dimer model, and J = 35 K (=$\Delta(0)$) derived from the isolated dimer model is too much smaller.

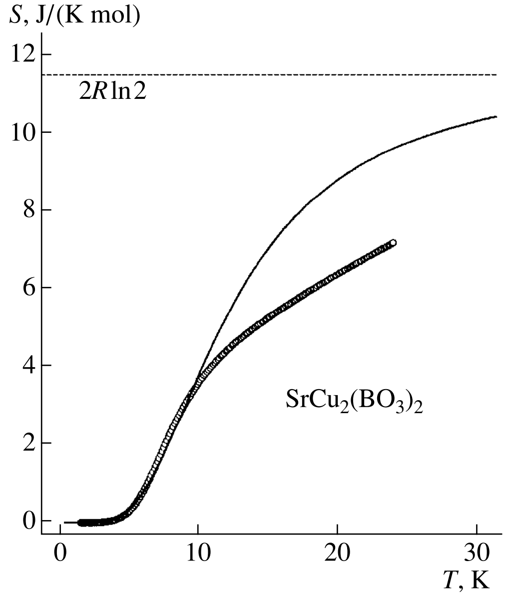


Fig. 3. Magnetic entropy of $SrCu_2(BO_3)_2$ at H = 0 (circles). Solid curve represents the magnetic entrope for the isolated dimer model for $\Delta(0)$ = 34.4 K.

To summarize, we have measured the specific heat of $SrCu_2(BO_3)_2$ under various magnetic fields. From the fitting based on the isolated dimer model, the gap was estimated to be 34.4 K, which is in good agreement with the values determined from other physical measurements. With increasing H, the gap decreases in proportion with H. The simple dimer model, however, can not explain the data at all in higher-temperature, region, suggesting rather stronger correlation of the spin system. We are looking forward to a theory based on the Shastry–Sutherland model with J and J' to reproduce our specific heat data over the whole T range.

REFERENCES

1. C. K. Majumdar and D. K. Ghosh, J. Math. Phys. **10**, 1399 (1969).
2. M. P. Gelfand, Phys. Rev. B **43**, 8644 (1991); T. Oguchi and H. Kitatani, J. Phys. Soc. Jpn. **64**, 1403 (1994).

3. H. Kageyama, K. Yashimura, R. Stern, *et al.*, Phys. Rev. Lett. **82**, 3168 (1999).
4. B. S. Shastry and B. Sutherland, Physica B **108**, 1069 (1981).
5. S. Miyahara and K. Ueda, Phys. Rev. Lett. **82**, 3701 (1999).
6. H. Kageyama, K. Onizuka, T. Yamauchi, *et al.*, J. Phys. Soc. Jpn. **68**, 1821 (1999).
7. H. Nojiri, H. Kageyama, K. Onizuka, *et al.* (submitted to J. Phys. Soc. Jpn.); cond-mat/9906072.
8. H. Kageyama, K. Onizuka, Y. Ueda, *et al.*, J. Phys. Soc. Jpn. **67**, 4304 (1998).
9. R. Bachmann, E. J. DiSalvo, Jr., T. H. Geballe, *et al.*, Rev. Sci. Instrum. **43**, 205 (1972).
10. H. Kageyama, K. Onizuka, T. Yamauchi, *et al.* (submitted to J. Crystal Growth).
11. R. L. Carlin, *Magnetochemistry* (Springer-Verlag, Berlin, 1986).
12. M. Takigawa, private communication.
13. H. Kageyama *et al.* (in preparation).

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 133–143.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 149–161.
Original Russian Text Copyright © 2000 by Babushkin, Logvin, Loĭko.

SOLIDS

Interrelation of Spatial and Temporal Instabilities in a System of Two Nonlinear Thin Films

I. V. Babushkin*, Yu. A. Logvin, and N. A. Loĭko*****
Institute of Physics, Belarussian National Academy of Sciences, Minsk, 220072 Belarus
**e-mail: babush@dragon.bas-net.by*
***e-mail: logvin@dragon.bas-net.by*
****e-mail: loiko@dragon.bas-net.by*

Received May 26, 1999

Abstract—The spatiotemporal dynamics of a system of two thin films possessing a resonance nonlinearity and irradiated on both sides with spatially uniform monochromatic light with the same intensity is investigated. The conditions under which bistability and symmetry breaking occur in the system are obtained. It is shown that self-pulsations can arise in the system as a result of the retardation of the light between the films, if the aperture of the incident beam is sufficiently small, and the dynamical regimes arising in the process are investigated numerically. As the beam aperture increases, the pulsations break down and a stationary spatially nonuniform field distribution is established. The transverse structures arising in this case are studied, and the relation between the symmetry breaking, bistability, self-pulsations, and spatial structures in the system investigated is established. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, a great deal of attention has been devoted to the development of optical methods for transferring and processing information. One of the main advantages of optical over traditional methods is the possibility of using spatially distributed signals, which is impossible in ordinary electronics because of the extremely long wavelength of the carrier radiation. In this connection, it is now urgent to study the mechanisms leading to the formation of spatial light structures (patterns) in laser and nonlinear-optical systems [1, 2]. The investigations of the spontaneous formation of patterns in optics is also important from the standpoint of the general theory of nonequilibrium systems, since it is possible to determine the mechanisms leading to self-organization, which are common to optical and hydrodynamic, chemical, and biological systems [3].

Our objective in the present work is to investigate the spatiotemporal dynamics of the transmission of light through a system of two thin films of a nonlinear medium in a more general form than in preceding works. It has been shown previously that, together with bistability, in such a system symmetry breaking, where for identical fields incident on a system from both sides the reflected fields have different amplitudes [4], and the appearance of asymmetric structures [5], instability, and chaos [6] can be observed. We shall show, taking into consideration the spatial degrees of freedom and retardation effects, that the spontaneous formation of spatial structures and temporal pulsations are occur by the same feedback mechanism in which the phase relations in the feedback circuit play a key role.

As is well known, the appearance of spatial structures is closely related with the existence of bistability in the system [7, 8]. This bistability is even observed in the presence of a single thin film of two-level atoms [9–12]. When external feedback, accomplished using a mirror, is introduced into the system, more complicated dynamical regimes, such as the appearance of self-pulsations [13–15] and formation of transverse static and traveling spatial structures [16–19], can appear. A system consisting of two thin bistable films is an elementary object for studying the effect of feedback in multifilm systems.

In the present paper symmetry breaking in a system of two films is investigated and the relation of symmetry breaking with bistability is determined in a more general case than in [4], specifically, without imposing any conditions on the phase relations in the films. This paper is organized as follows. The model describing the system under study is presented in Section 2. In Sections 3 and 4 the stationary states are determined analytically and their stability is analyzed on the symmetric and asymmetric branches of the solutions, and the regions of spatial and temporal instabilities are determined. In Section 5 the temporal dynamics generated by instabilities on the symmetric and asymmetric branches are compared. It is found that stable periodic pulsations can appear only on the asymmetric branch of the solution, while on the symmetric branch we observed only quasistable pulsations, which are due to switching from one bistable branch to another and whose decay time increases exponentially as the distance between the films increases. Quasipulsations of

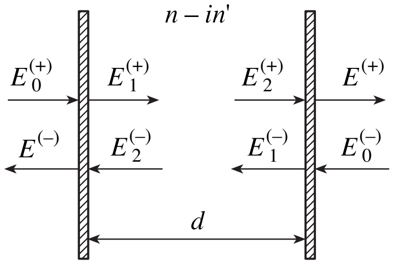


Fig. 1. A system consisting of two films separated by a distance d; $E_0^{(\pm)}$ are the amplitudes of the fields incident on the first (+) and second (–) films, $E_1^{(\pm)}$ are the amplitudes of the fields transmitted through the films, and $E_2^{(\pm)}$ are the amplitudes of the fields which reach the opposite film.

this kind have been investigated in the general theory of differential equations with retardation [20].

In Section 6 it is shown that the temporal dynamics of the system becomes qualitatively different when the transverse spatial coordinates are taken into account. In this case the oscillatory solutions break down and the system passes into a stationary spatially nonuniform state, and even an attempt to stabilize the most unstable static spatial harmonics does not lead to the appearance of oscillatory solutions, but rather only new classes of stationary spatially nonuniform solutions arise. The general conclusions are given in Section 7.

2. MODEL

Let us consider a system of two thin nonlinear films located at a distance d from one another and separated by a linear medium with a complex refractive index $n + in'$ (Fig. 1). The system is illuminated from both sides by monochromatic spatially uniform light fields with amplitudes $E_0^{(+)}$ and $E_0^{(-)}$, respectively. The amplitudes of the fields transmitted through the first and second films are $E_1^{(+)}$ and $E_1^{(-)}$, respectively, and $E_2^{(-)}$ and $E_2^{(+)}$ are the amplitudes of the fields reaching the opposite films (see Fig. 1).

We shall use the Bloch equations for the polarization and the population difference between the levels of two-level atoms to describe the interaction of light with the films:

$$\frac{dR_j}{dt} = \left(i(\omega - \omega_0) - \frac{1}{T_2}\right) - i\frac{\mu W_j E_j}{\hbar}, \tag{1}$$

$$\frac{dW_j}{dt} = -\frac{W_j + 1}{T_1} + i\frac{\mu}{2\hbar}(E_j^* R_j + R_j^* E_j), \tag{2}$$

where R_j, $j = 1, 2$, is the slowly varying part of the off-diagonal element of the density matrix of a two-level atom; μ is the transition dipole moment; T_1 and T_2 are the longitudinal and transverse relaxation times; ω is the frequency of the incident radiation; and, ω_0 is the resonance transition frequency of the atoms in the film. The polarization in the film is $P = \mu NR$, where N is the density of atoms in the film. Taking account the normalization relations

$$\gamma = \frac{T_2}{T_1}, \quad \Delta = \frac{\omega - \omega_0}{\sqrt{\gamma}}, \quad w_j = W_j, \quad r_j = \frac{R_j}{\sqrt{\gamma}},$$
$$e_j = E_j\frac{\mu\sqrt{T_1 T_2}}{\hbar}, \quad j = 1, 2, \tag{3}$$

equations (1) and (2) transform into the following equations for the normalized polarization and population difference of atoms in the films:

$$\dot{r}_j = \gamma(-1 + i\Delta)r_j + i\gamma e_j w_j, \tag{4}$$

$$\dot{w}_j = -(w_j + 1) + \frac{i}{2}(e_j^* r_j - r_j^* e_j). \tag{5}$$

It was shown [6, 11] that the effective field in a film can be represented in the form

$$e_1 = e_0^{(+)} + e_2^{(-)} - i\alpha r_1, \tag{6}$$

$$e_2 = e_0^{(-)} + e_2^{(+)} - i\alpha r_2, \tag{7}$$

where α is the nonlinearity parameter in the film and is given by

$$\alpha = \frac{2\pi n N L \omega \mu^2 T_2}{\hbar c}, \tag{8}$$

where L is the film thickness. The relations (6) and (7) were obtained in the approximation where the film thickness is small compared with the wavelength of the incident radiation.

The propagation of the field in the linear medium between the films satisfies the diffraction equation, which can be written in the following operator form as

$$e_2^{(\pm)}(r_\perp, \tau) = \rho \exp(is)\exp\left(-i\frac{d}{k}\Delta_\perp\right)e_1^{(\pm)}(r_\perp, \tau), \tag{9}$$

where, respectively, $\rho = \exp(kn'd)$ are the losses, $s = knd$ is the phase shift, τ is the propagation time of the light between the fields, $r_\perp = (x, y)$ is the transverse component of the radius vector of the point (x, y, z), $k_\perp = (k_x, k_y)$ is the transverse component of the total wave vector $\mathbf{k}$ of the light field, and $\Delta_\perp$ is the transverse part of the Laplacian. The amplitudes of the fields transmitted through the first and second films, according to [11], are $e_1^{(+)} = e_0^{(+)} - i\alpha r_1$ and $e_1^{(-)} = e_0^{(-)} - i\alpha r_2$, respectively.

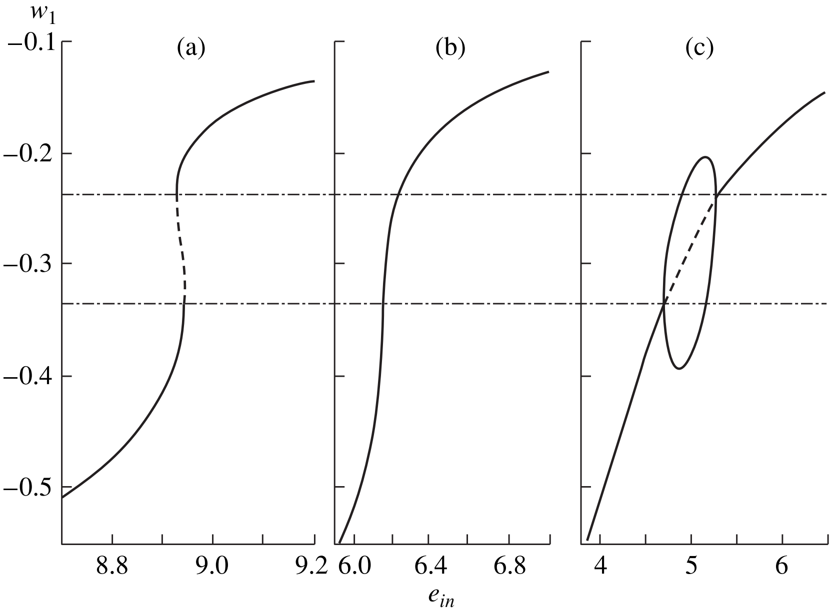


Fig. 2. Stationary states for the parameters $\alpha = 10$, $\Delta = 2$, $\rho = 0.5$, and $\gamma = 0.1$. The solid lines mark stable states and dashed lines mark states which are unstable with respect to perturbations with $\theta = 0$. (a) $s = 0$: bistability on the symmetric branch of the solution; (b) $s = 0.5\pi$: no symmetry breaking or bistability; (c) $s = \pi$: symmetry breaking.

3. STATIONARY UNIFORM STATES OF THE SYSTEM

It has been demonstrated in [6, 4] in the particular case of no offset of the frequency from resonance and s a multiple of π that bistability and symmetry breaking are possible in a system of two films. In what follows, we shall study the relation between these two phenomena in a general form.

We obtain the following relations for the stationary values of the fields in the films:

$$e_{01} = e_1 + i\alpha r_1 + i\alpha\rho\exp(is)r_2, \tag{10}$$

$$e_{02} = e_2 + i\alpha r_2 + i\alpha\rho\exp(is)r_1, \tag{11}$$

$$e_{01} = e_0^{(+)} + \rho\exp(is)e_0^{(-)}, \tag{12}$$

$$e_{02} = e_0^{(-)} + \rho\exp(is)e_0^{(+)}. \tag{13}$$

Here the stationary values of the polarization and population difference have the form

$$r_j = \frac{\eta e_j}{1 + \beta|e_j|^2}, \tag{14}$$

$$w_j = -\frac{1}{1 + \beta|e_j|^2}, \tag{15}$$

where

$$\beta = \frac{1}{1 + \Delta^2}, \quad \eta = \frac{i - \Delta}{1 + \Delta^2} \quad j = 1, 2.$$

When the incident fields are the same ($e_{01} = e_{02}$), the stationary symmetric state $e = e_1 = e_2$ can be found from the equation

$$e_{01} = e_{02} = e + \frac{i\alpha\eta[1 + \rho\exp(is)]e}{1 + \beta|e|^2}, \tag{16}$$

which is well known from the theory of bistability [7, 8]. To analyze symmetry breaking we substitute the expressions (14) and (15) into equations (10) and (11), and after simple algebraic transformations we obtain

$$\begin{aligned} & e_1 + \frac{i\alpha\eta[1 - \rho\exp(is)]e_1}{1 + \beta|e_1|^2} \\ & = e_2 + \frac{i\alpha\eta[1 - \rho\exp(is)]e_2}{1 + \beta|e_2|^2}. \end{aligned} \tag{17}$$

It is evident from the expression (17) that both parts contain the same bistable function as the first part of equation (16) with one difference: s is replaced by $s + \pi$. It is obvious from the S shape of this function that the same value of the right(left)-hand hand part of the expression (17) can be obtained for different values of e_1 (e_2). This situation corresponds to the equality (17) with different values of e_1 on the left-hand side and e_2 on the right-hand side, i.e., symmetry breaking with the same incident fields. It is found that bistability of the symmetric solution (16) is related with symmetry breaking, described by the expression (17) by replacing s by $s + \pi$. This can be seen in Fig. 2, which shows the dependence of the stationary population difference in the films on the magnitude of the incident field e_{in}.

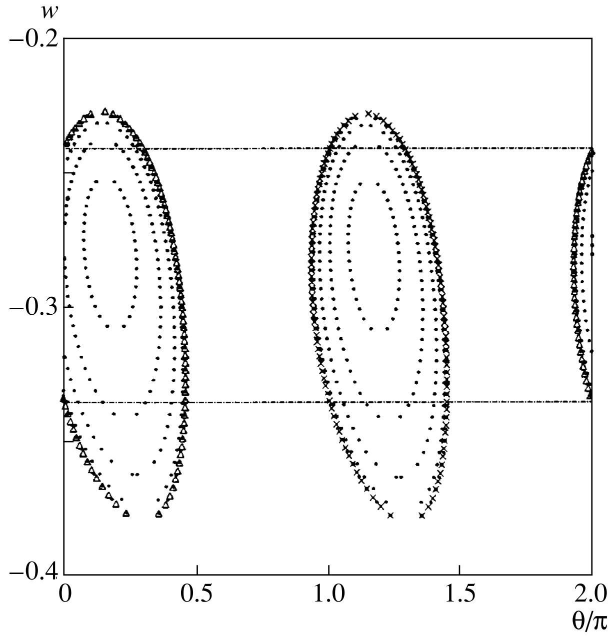


Fig. 3. Regions of static instability for the parameters in Fig. 2a, $\tau = 50$. The dots show the boundaries of the regions where instability with respect to AH perturbations occurs; triangles and crosses show the boundaries of regions with instability with respect to symmetric ($\delta w_1 = \delta w_2$) and asymmetric ($\delta w_1 = -\delta w_2$) static disturbances, respectively.

It is evident that as s changes from $s = 0$ (Fig. 2a) to $s = \pi$ (Fig. 2c), bistability is replaced by symmetry breaking for the same values of the population difference (or light intensity) in the films. For an intermediate value of the phase s (see Fig. 2b) neither bistability nor symmetry breaking occur in the system. The condition for the existence of bistability (symmetry breaking) is determined by the inequality

$$[2\cos(\arg\Lambda) - |\Lambda|]^3 + 27|\Lambda| < 0, \qquad (18)$$

where $\Lambda = i\alpha\eta[1 - \rho\exp(is)]$. It transforms into a similar condition obtained in [4] under the corresponding simplifying assumptions.

Figure 2 demonstrates the important role of phase effects in the transmission of light between the films. These effects determine the development of spatial and temporal instabilities, as will be shown below. We note that an equality similar to equation (17) has been investigated in [13] in a study of 2τ pulsations in a system consisting of a film deposited on a dielectric substrate. In this work e_1 and e_2 represent the values of the field in successive half-periods of the pulsations, and the transition from a regime with pulsations to a regime where only bistability is present, just as in our system a transition from symmetry breaking to bistability, can be accomplished only by changing the phase relations in the feedback circuit. Thus, the same algebraic relations describe the mutually complementary effects observed in nonlinear systems: bistability, symmetry breaking, and 2τ pulsations.

4. STABILITY OF STATIONARY STATES

To investigate the stability of a state of equilibrium with respect to spatiotemporal perturbations (δr_j, δw_j), we shall linearize the system (4)–(5). Linearizing relative to disturbances proportional to $\exp(\lambda t + r_\perp k_\perp)$ gives a system of equations for λ:

$$\begin{pmatrix} D_1 & U_1 \\ U_2 & D_2 \end{pmatrix}\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \lambda\begin{pmatrix} F_1 \\ F_2 \end{pmatrix}, \qquad (19)$$

$$D_j = \begin{pmatrix} -1 + i\Delta + \alpha w_j & 0 & ie_j \\ 0 & -1 - i\Delta + \alpha w_j & -ie_j^* \\ \frac{\gamma}{2}(ie_j^* - \alpha r_j^*) & \frac{\gamma}{2}(-ie_j - \alpha r_j) & \gamma \end{pmatrix}, \qquad (20)$$

$$U_j = \begin{pmatrix} \alpha w_j \exp(i\Theta) & 0 & 0 \\ 0 & \alpha w_j \exp(-i\Theta) & 0 \\ -\frac{\gamma\alpha}{2} e_j^* \exp(i\Theta) & -\frac{\gamma}{2} e_j \exp(-i\Theta) & 0 \end{pmatrix}, \qquad (21)$$

where F_j is the perturbation vector: $F_j = (\delta r_j, \delta r_j^*, \delta w_j)^T$, $\Theta = \theta + s + \lambda\tau$, and $\theta = k_\perp^2 d/k$ characterizes the transverse perturbation with wave number $k_\perp$. Since θ and λ appear in the expression for Θ in the same way, the spatial and temporal instabilities manifest themselves in the same way.

The equation (19) is a characteristic quasipolynomial, in which the unknown quantity λ appears in the form of powers λ^n and in the form of terms of the form $\exp(-\lambda\tau)$. The boundaries of stability are obtained from equation (19) by making the substitution $\lambda = i\Omega$, where Ω is a real quantity. If $\Omega = 0$, we have a region of instability relative to "static" disturbances (with zero temporal frequency), and in the opposite case we have regions of instability relative to disturbances of the Andronov–Hopf (AH) type. Curves of neutral stability of the system with respect to static (marked by triangles and crosses) and AH (marked by dots) instabilities on the symmetric branch for parameters corresponding to Fig. 2a are displayed in Fig. 3. It is evident that the AH zones are embedded in the static zones, so that their presence results only in an increase in the order of the instability of the system. A numerical experiment shows that this has no qualitative effect on the dynamics of the system. As the retardation time τ increases, the number of such zones increases, as all new roots intersect the imaginary axis, while existing zones approach the boundary of static instability.

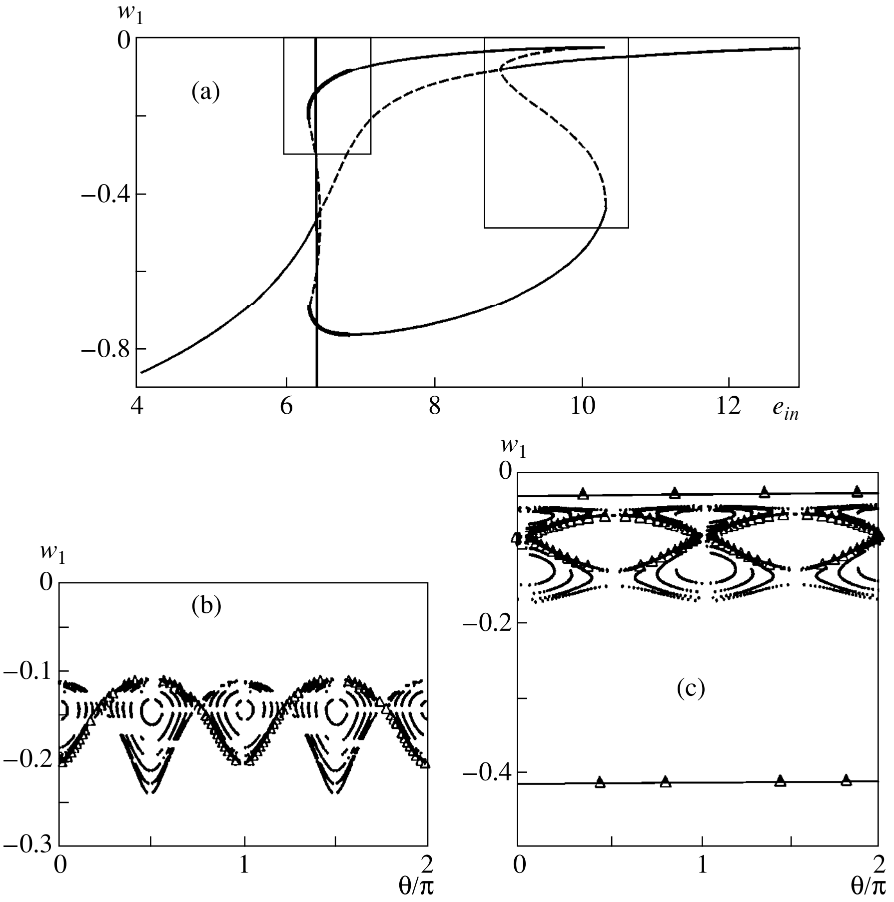


Fig. 4. (a) Stationary curve for the parameters $\alpha = 20$, $\Delta = 2$, $\rho = 0.5$, $s = \pi$, $\gamma = 0.1$, and $\tau = 10$. The solid lines mark stable states, the dashed lines mark states whose instability is determined primarily by static perturbations with $\theta = 0$, and the thick lines show states which are unstable with respect to static perturbations with $\theta \neq 0$ and with respect to AH perturbations. (b) Boundaries of the neutral stability of states corresponding to the part of the asymmetric branch in the left-hand rectangle in Fig. 4a. (c) Boundaries of neutral stability of the states lying on the part of the asymmetric branch in the right-hand rectangle in Fig. 4a. The triangles show the boundaries of stability with respect to static perturbation; the dots show the boundaries of stability with respect to AH perturbations.

As one can see from the expression for Θ, a displacement along s in it is compensated by an opposite displacement along θ. If for such a change in s the quantities $e_j(w_j, r_j)$, satisfying equation (19), once again correspond to stationary states of the system, then the instability zones merely shift along the θ axis without a change in form. Such a situation occurs in the case of the symmetric branch considered above, since the values of e_j, determined by equation (19), remain stationary values of the system as a function of s because the incident field is correspondingly adjusted according to the expression (16).

For parameters corresponding to the case in Fig. 2a, the regions of instability marked by triangles in Fig. 3 are responsible for the development of symmetric perturbations ($\delta w_1 = \delta w_2$), while the region marked by the crosses corresponds to the development of asymmetric perturbations ($\delta w_1 = -\delta w_2$).

We shall now investigate the instability on the asymmetric branch of the solution. Analysis shows that a spatiotemporal instability on the asymmetric branch is possible only if a fork-type bifurcation, corresponding to the appearance of asymmetric solutions, becomes subcritical, which is achieved by increasing the nonlinearity parameter α. An example of this is presented in Fig. 4. Figure 4a displays the stationary dependences of w_1 on the incident field e_{in}. Regions of instability, corresponding to the stationary states on the asymmetric branch located in the left- and right-hand rectangles in Fig. 4a are presented in Fig. 4b. To save space, we show only the zones corresponding to the upper part of the asymmetric branch, since the zone on the bottom part of the branch has a similar form. It follows from Fig. 4b that the states marked by the thick line in Fig. 4a are unstable with respect to static perturbations with $\theta \neq 0$, i.e., here spatial structures can arise. Moreover, these states are unstable with respect to AH perturbations, and zones of this instability do not overlap with regions of static instability for $0 \leq \theta \leq 0.25$, $0.75 \leq \theta \leq 1.25$, and so on, which indicates the possibility of the excitation of pulsations. In contrast to the situation on the symmetric branch, the regions of instability on the asymmetric branch do not shift along the θ axis as s varies, but rather they are compressed or stretched in the vertical direction, since the relations between the stationary values of $e_j(w_j, r_j)$ in the two films change. One of the minima of the curve of neutral stability with respect to the static perturbations always lies on the $\theta = 0$ axis, just as in Fig. 4b.

The stationary states on the part of the asymmetric branch located in the right-hand rectangle in Fig. 4a are unstable for w_1 ranging from –0.41 to –0.03 with respect to static disturbances with any value of θ, including $\theta = 0$. As a result of this, spatially uniform disturbances, transferring the system into stable stationary states existing for given parameters, will develop. In turn, the regions of static instability also contain regions of AH instability and a secondary static instability, which only increase the order of the instability of the solutions, and just as in the case of embedded AH zones on the symmetric branch, they have no qualitative effect on the dynamics of the system.

5. TEMPORAL DYNAMICS

We shall consider first the case where the aperture of the incident light beam is comparatively small, so that only the zero spatial harmonic can be excited. In this case, only the instability on the $\theta = 0$ axis in Figs. 3 and 4 need be considered.

As already mentioned, the regions of AH instability on the symmetric branch always lie inside the static zones, in contrast to the related system of a thin film with a mirror [16], which makes it difficult to obtain AH pulsations, similarly to [16], since after the transient process a stationary stable state is established in the system. However, in [6, 13], where differential equations with a retarded argument reduced in the limit of very large retardation $\tau \longrightarrow \infty$ to discrete mappings, it was shown that discontinuous pulsations with period 2τ can exist. In [6] more complicated regimes, right up to chaos, were also investigated. In these works the question of whether or not the dynamics of the system reduces for very large but finite τ to the dynamics obtained in the limit $\tau \longrightarrow \infty$, in other words, whether or not a finite τ for which this limit cycle or a set close to it is a stable attractor exists, remained open. It is obvious that for arbitrary finite τ this attractor of a discrete mapping, which is a piecewise-continuous function, will not be an attractor of the system (4)–(5), stable or unstable, since, as follows from the general theory of differential equations with retardation [21], the solutions of the system (4)–(5) smooth out with time (if at time $t = 0$ the solution is discontinuous, then at time $t = n\tau$ it will be differentiable $n - 1$ times) and, thus, only an infinitely differentiable function can be a periodic attractor of the system (4)–(5). The process of smoothing of the solutions reflects the presence of temporal dispersion in the system due to the inertia of the polarization and population of the nonlinear medium in films, which is determined by the relaxation times T_1 and T_2. As the retardation increases compared with these quantities, the effect of the inertia decreases and the oscillatory solutions, because of the continuous dependence of solutions of equations with a retarded argument on the retardation [21, 22], can asymptotically approach a discontinuous solution obtained in the case of discrete mappings.

For sufficiently large τ ($1/\tau \ll 1$) the system (4)–(5) is singularly perturbed with respect to the corresponding discrete mapping. It has been shown for very simple one-dimensional systems of this kind [20] that in the case of a mapping possessing two attractors—a stationary point with a quite large region of attraction and a cycle with period 2τ, the following behavior is typical for the corresponding differential-difference equation: If the system is initially in the region of attraction of a 2τ periodic attractor of the corresponding limiting mapping ($\tau = \infty$), then its oscillations break down after a time $\exp(a\tau)$ has elapsed (where a is a constant), and it passes into a stationary state. This means that if the constant a is sufficiently large, then even for small τ we obtain a trajectory which can be assumed to be periodic for a quite long period of time. A numerical experiment showed that such behavior is also characteristic of the system (4)–(5). Figure 5 shows quasipulsations of this kind for the system parameters presented in [6] and the values of s for which symmetry breaking (Fig. 5a) or bistability (Fig. 5b) occurs. In the first case the stationary characteristic is similar to that shown in Fig. 2c, and in the second case it is similar to that presented in Fig. 2a. For sufficiently large τ these quasipulsations acquire a squared form. Figures 4a and 4b show only several periods before the pulsations break down, which occurs after a comparatively long time has elapsed. The process leading to the breakdown of quasipulsations for small τ is shown in Fig. 5c. Figure 5d shows the dependence of their lifetime on the retardation time τ.

We note that to obtain quasipulsations it is sufficient to prescribe special boundary conditions. If the stationary curve exhibits bistability, the state for the first film must correspond to the lower branch of the bistability curve and the state for the second film must correspond to the upper branch. After a time τ has elapsed, switching occurs (the films change places), after a time 2τ the system returns into the initial state, and so on. This corresponds to a stable 2τ cycle, obtained in the limit mapping ($\tau \longrightarrow \infty$). Such a sequence of switchings results in the asymmetric quasipulsations shown in Fig. 5b. Conversely, for a stationary dependence with broken symmetry, both films must be placed on the same asymmetric branch. After a time τ has elapsed, they switch onto the second branch together; this corresponds to symmetric pulsations (Fig. 5a). Nonetheless, the region of attraction of such quasipulsations is quite large. Even if a state in which the films are located in the

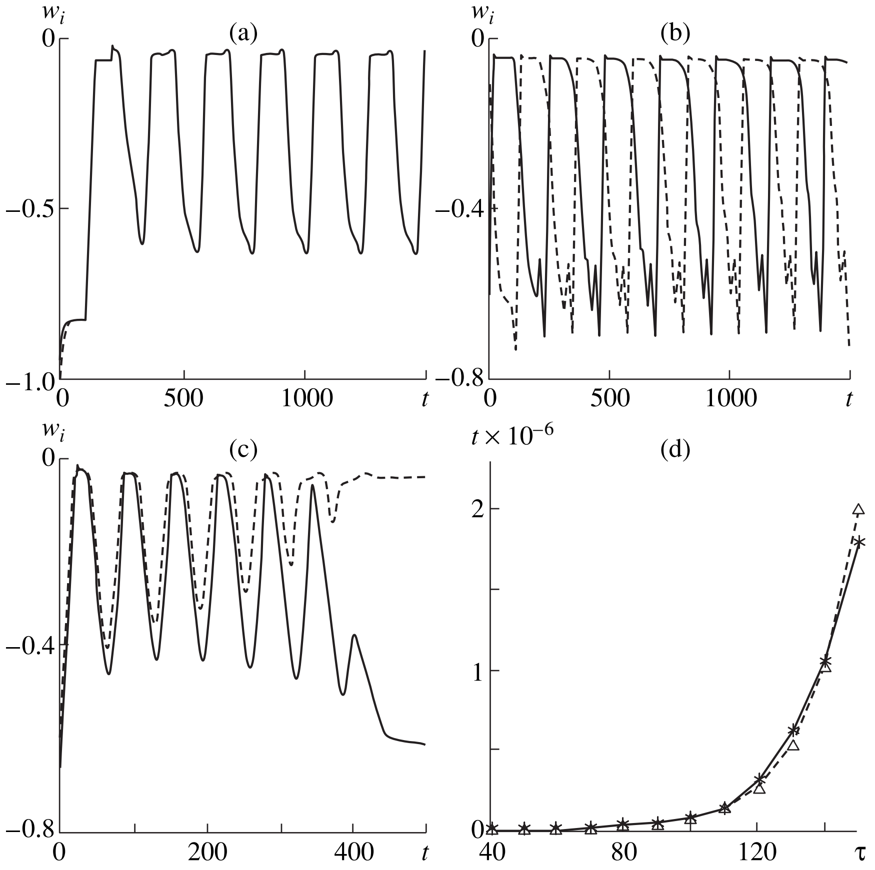


Fig. 5. Quasipulsations generated by switchings from one part of the branch of a stationary curve to another. The values of the parameters are the same as in [6] ($\alpha = 8$, $\Delta = 0$, $\rho = 0.5$, and $\gamma = 0.1$). The solid line shows the time dependence for w_1 and the dashed line shows the saem for w_2. (a) $s = \pi$ and $\tau = 100$: switchings between asymmetric branches (stationary characteristic, just as in Fig. 2c); (b) $s = 0$ and $\tau = 100$: switchings between the upper and lower branches of the bistable curve (stationary characteristic, just as in Fig. 2a); (c) $s = \pi$ and $\tau = 20$: damping of quasipulsations for small values of τ. (d) Lifetime of quasipulsations (dashed curve with triangles) and the best exponential fit to this curve, $C\exp(6.7\tau)$ (the solid curve with the asterisks).

absence of an external field is taken as the initial conditions and the field is switched on instantaneously at time $t = 0$ (as in Fig. 5a), the system nonetheless demonstrates 2τ quasipulsations.

It is evident from Fig. 5d that the lifetime of the quasipulsations actually increases exponentially. In this figure the asterisks connected by a solid line show the data obtained from the numerical experiment; the triangles connected by a dashed line represent the curve $t = C^*\exp(6.7\tau)$, where C^* is a constant. Even for not very large τ the lifetime is several thousandths of periods, and as τ increases, the ratio of the lifetime to the period increases rapidly. Thus, for appropriate retardation times, pulsations exist for such a long time that they are effectively full-fledged periodic oscillations (which is an argument for applicability of switching to limit mappings in [6, 13]).

Returning to Fig. 4b, we shall investigate the temporal dynamics arsing in the system if the initial conditions are chosen on the section of the asymmetric branch that is unstable with respect to temporal perturbations. Figure 6 shows examples of oscillations arising from such stationary states for increasing values of τ. At the initial stage (Fig. 6a, $t < 400$), one can see the development of AH pulsations, but then the system passes into a regime with large-amplitude oscillations that correspond to switching between branches of the stationary characteristic in Fig. 4a. As one can see from Fig. 6b, as τ increases, the pulsations arising on the asymmetric branch can be symmetric. As τ increases further, the pulsations approache a square form with period 4τ (Fig. 5d), but, in contrast to pulsations on the symmetric branch, they are stable. Figure 6c shows oscillations for an intermediate value of τ, which explain the transition from Fig. 6b to Fig. 6d. As one can see by comparing Fig. 6d and Fig. 4a, where the straight vertical line indicates the value of the incident field (the initial conditions of the system are found as the point of intersection of the upper part of the asymmetric branch with this line for the first film and the lower part of the asymmetric branch for the second

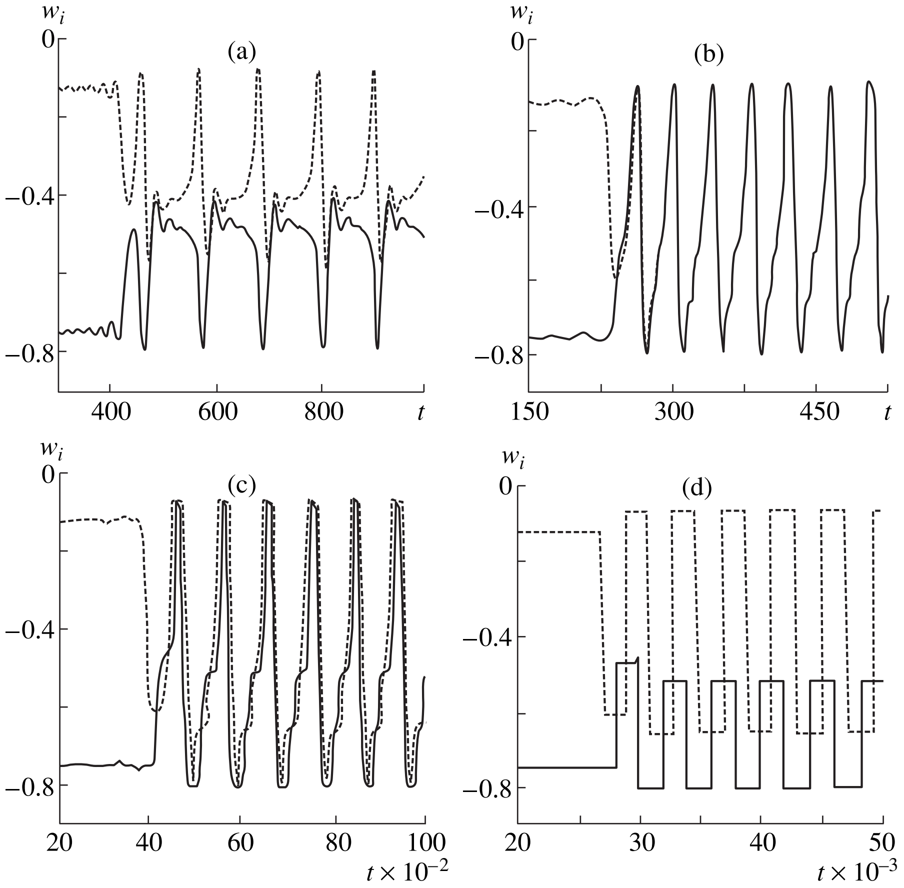


Fig. 6. Stable pulsations for values of the incident field marked by the straight line in Fig. 4a: $\alpha = 20$, $\Delta = 2$, $\rho = 0.5$, $\gamma = 0.1$, $s = \pi$ and $\tau = 4$ (a), 100 (b), 200 (c), and 1000 (d).

film), after each time interval τ a switching occurs between the stationary states of the system, and four switchings correspond to the period 4τ.

6. TAKING ACCOUNT OF THE TRANSVERSE DEGREE OF FREEDOM

We shall now consider the effect of the transverse degrees of freedom on the dynamics of the system. This corresponds to increasing the aperture of the incident light beam. In this case, the harmonics with nonzero $k_\perp$ can be most unstable.

Let us consider first the situation that arises for parameters corresponding to the symmetric branch of the solution in Fig. 2b. A numerical experiment, whose result is presented in Fig. 7, shows that spatial structures can arise in this case (Figs. 7a, 7b). As they form, the unstable harmonics belonging to the regions responsible for the symmetric perturbations and regions responsible for the asymmetric perturbations can interact with one another, as a result of which rhombic structures arise. Figure 7c shows how the interaction of the spatial modes occurs. Here $\mathbf{k}_3$ belongs to the region of instability responsible for the symmetric disturbances, and k_1 and k_2 belong to the regions responsible for the asymmetric disturbances. As one can see from Figs. 7a and 7b, the maxima in the first film occur at the minima in the second film and vice versa.

It follows from Fig. 4b that transverse spatial structures can appear on the asymmetric branch as a result of the instability of static modes with nonzero θ. It has been shown in [5] that in the absence of retardation asymmetric hexagonal structures arise, which in contradistinction to structures on the symmetric branch have a different contrast in the first and second films. There arises the question of what happens when retardation is taken into account and the temporal effects start to influence the spatial dynamics of the system. To answer this question, a numerical experiment was performed with parameters and initial conditions corresponding to the case in Fig. 6. It turns out that in contrast to the case of a small-aperture beam, when the pulsations arising remain stable (Fig. 6), in the present case the development of pulsations drives the system into a spatially uniform but stationary state (Fig. 8). It is evident from the figure that as a result of the transient process (Fig. 8d), symmetric hexagonal structures (Figs. 8a, 8b), which are identical in both films, are

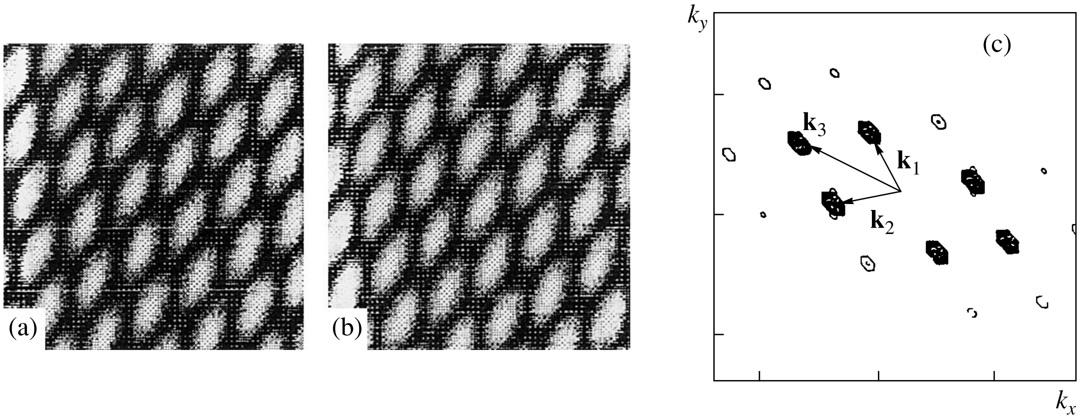


Fig. 7. Asymmetric rhombic structures, generated from a symmetric state for parameters corresponding to the case in Fig. 2c. The population differences in the first (a) and second (b) films and the spatial spectrum of the structure (c): $\mathbf{k}_3$ lies in the region responsible for the development of symmetric disturbances and shown in Fig. 3; $\mathbf{k}_1$ and $\mathbf{k}_2$ belong to the region responsible for the development of asymmetric disturbances.

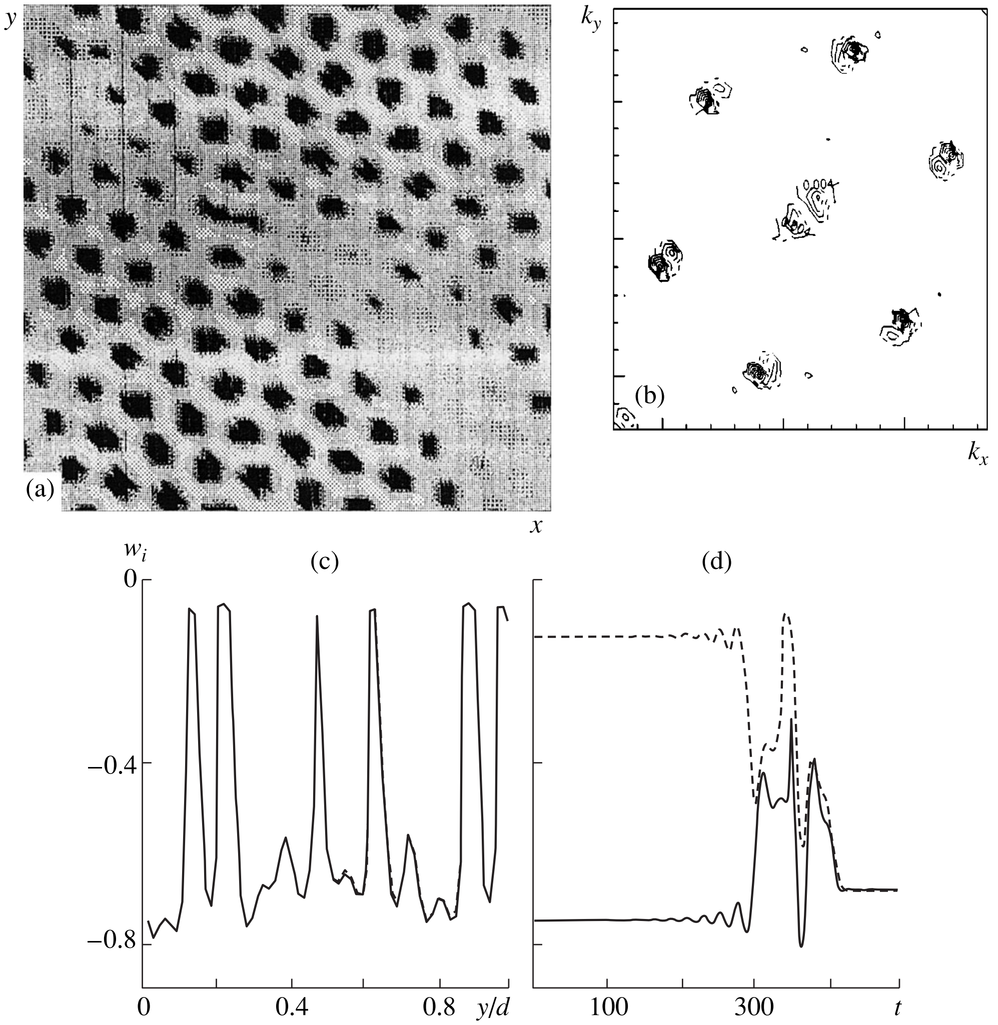


Fig. 8. Structures arising from a uniform state on the asymmetric branch (the value of the incident field is marked by the vertical line in Fig. 4a), for parameters corresponding to Fig. 4 and $\tau = 4$. (a) Spatial distribution of the population difference in the first film. (b) Spatial spectrum of this structure. (c) Section of the spatial distribution in Fig. 8a for $y = 32$. The distribution in the first film is completely identical to the distribution in the second film. (d) Transient process driving the system to the indicated spatial structure. The solid line shows the dynamics at the point (0, 0) of the first film and the dashed line shows the dynamics at the same point of the second film.

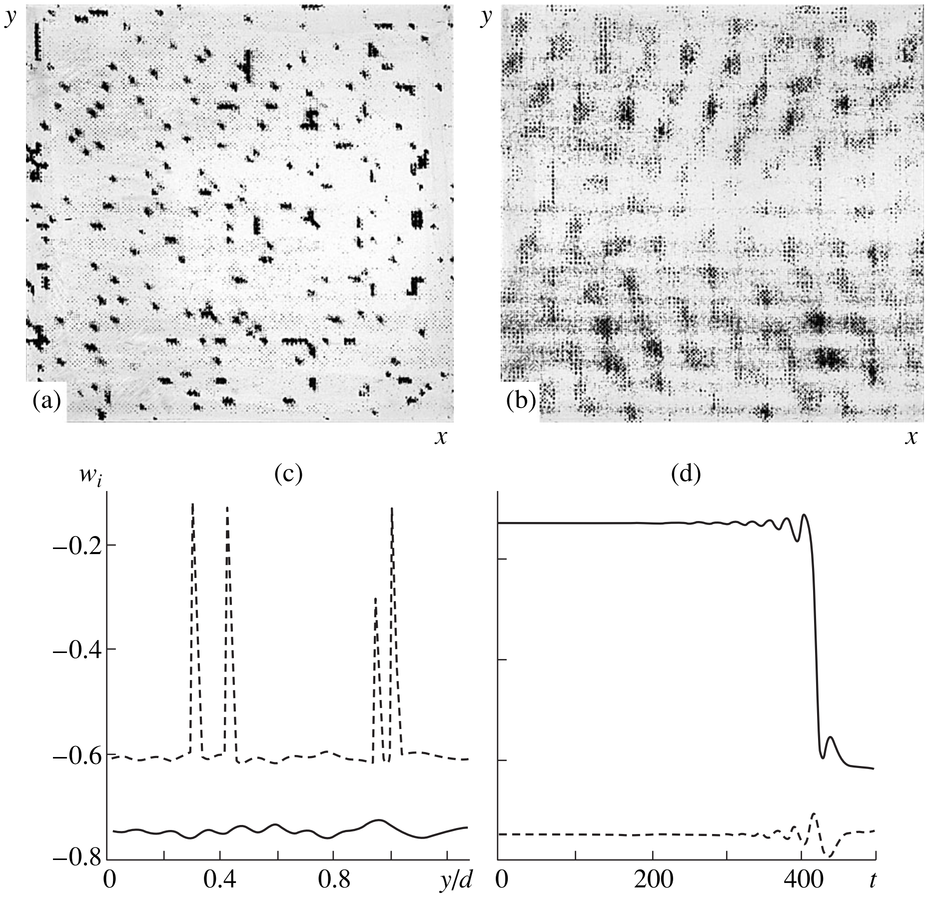


Fig. 9. Spatial structures arising for the same values of the initial parameters as in Fig. 8 but with oscillations on the mode $\theta = 0$ suppressed and the spatial modes which are unstable with respect to static disturbances cut out (see Fig. 4b): (a, b) spatial distributions of the population differences in the first and second films, respectively; (c) section of Fig. 9a (dashed line) and Fig. 9b (solid line) for $y = 32$: (d) temporal dynamics of the transient process for one point (1, 1) in the first (dashed line) and second (solid line) films.

obtained. The maxima and minima of these structures (Fig. 8c) correspond to the upper and lower asymmetric branches of the stationary curve (Fig. 4a). In this case symmetry breaking seems to be distributed in space. This picture is also qualitatively identical to Fig. 6b, the difference being that in Fig. 6b the time dependence is shown, while in Fig. 8c the spatial dependence is shown. Thus, symmetry breaking, pulsations, and spatial structures in this system are seen to be interrelated. This complements the fact already mentioned above that the factors responsible for the spatial and temporal instability enter in the characteristic quasipolynomial of the system in the same manner.

Hopf instability zones, for which $\theta \neq 0$, are seen in Fig. 4b. There arises the question of how these instability zones influence the dynamics when other competing instabilities are suppressed. To clarify this question we suppressed in the numerical experiment the instability of the main AH mode with $\theta = 0$; this corresponds to introducing into the system an additional apparatus, a stabilizer. Once again, initially, the system was in a uniform asymmetric state, the same as in the case shown in Fig. 6. The restrictions imposed on the system have the effect that the pulsations start to develop only on the spatial harmonics with $\Omega \neq 0$. However, the pulsations arising in the spatial structures decay, and with time the system passes into a spatially nonuniform stationary state. Figure 9 shows the transient process (d) and the steady state (a–c) established in the system. Just as the preceding case, a large part of the transverse distribution of the system is close to one of the stationary states, but here sharp, soliton-like peaks, corresponding to switching between the upper and lower solutions in the region of symmetry breaking in Fig. 4a, appear in one of the films (Fig. 9c).

7. CONCLUSIONS

In summary, the results presented above attest to the fact that a system of two thin bistable films irradiated with monochromatic light demonstrates a rich dynamics. Specifically, bistability can arise in the system as a result of the characteristic nonlinearity of each film. The presence of feedback, which is a result of the interaction of the films, also leads to the appearance of symmetry breaking. An increase in the distance between the films and therefore the retardation time in the feedback circuit results in the appearance of pulsations in the system, when the beam aperture is sufficiently small so that the transverse effects do not influence the dynamics of the system. The period of the pulsations is related with the distance between the films. This makes it possible to use this system to obtain a periodic sequence of Π-shaped pulses with a prescribed period. An increase of the beam aperture, which can be described by taking account of the spatial modes with nonzero $k_\perp$, substantially changes the behavior of the system. Specifically, for a short delay, formation of asymmetric rhombic structures on the symmetric branch is possible in the system as a result of the interaction of the spatial modes from the instability regions responsible for the development of asymmetric perturbations and the regions responsible for the development of symmetric perturbations. As the delay time increases, other stationary states are established, for example, symmetric hexagonal structures on the asymmetric branch. Temporal Hopf oscillations arising on the asymmetric branch of the stationary solution of the system with a small aperture of the incident beam collapse. As a rule, this spatial distribution of the field is irregular and has defects. For example, soliton-like spatial structures appear.

REFERENCES

1. N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems* (Nauka, Moscow, 1997).
2. *Self-Organization in Optical Systems and Applications in Information Technology*, Ed. by M. A. Vorontsov and W. B. Miller (Springer, Berlin, 1995).
3. C. Cross and P. C. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).
4. I. V. Babushkin, Yu. A. Logvin, and N. A. Loĭko, Kvantovaya Élektron. **25**, 110 (1998).
5. Yu. A. Logvin, Phys. Rev. A **57**, 1219 (1998).
6. Yu. A. Logvin and A. M. Samson, Zh. Éksp. Teor. Fiz. **102**, 472 (1992) [Sov. Phys. JETP **75**, 250 (1992)].
7. L. A. Lugiato, Prog. Opt. **21** (1984).
8. H. M. Gibbs, *Optical Bistability: Controlling Light with Light* (Mir, Moscow, 1988; Academic Press, Florida, Orlanddo, 1985).
9. S. M. Zakharov, Zh. Éksp. Teor. Fiz. **108**, 829 (1995) [JETP **81**, 452 (1995)].
10. S. M. Zakharov and É. A. Manykin, Poverkhnost, No. 2, 137 (1988).
11. M. Benedict, V. A. Malyshev, E. D. Trifonov, *et al.*, Phys. Rev. A **43**, 3845 (1991).
12. L. N. Oraevsky, D. J. Jones, and D. K. Bandy, Opt. Commun. **111**, 163 (1994).
13. A. M. Basharov, Zh. Éksp. Teor. Fiz. **94**, 12 (1988) [Sov. Phys. JETP **67**, 1741 (1988)].
14. N. A. Loĭko, Yu. A. Logvin, and A. M. Samson, Kvantovaya Élektron. **22**, 389 (1995).
15. A. N. Loĭko, Yu. A. Logvin, and A. M. Samson, Opt. Commun. **124**, 383 (1996).
16. Yu. A. Logvin and N. A. Loĭko, Phys. Rev. E **56**, 3803 (1997).
17. Yu. A. Logvin, B. A. Samson, A. A. Afanas'ev, *et al.*, Phys. Rev. E **54**, R4548 (1996).
18. N. A. Loĭko and Yu. A. Logvin, Laser Phys. **8**, 322 (1998).
19. A. A. Afanas'ev, Yu. A. Logvin, A. M. Samson, *et al.*, Opt. Commun. **115**, 559 (1995).
20. A. M. Sharkovskiĭ, Yu. L. Maĭstrenko, and E. Yu. Romanenko, *Difference Equations and Their Applications* (Naukova Dumka, Kiev, 1986), p. 203.
21. L. É. Éksgol'ts and S. B. Norkin, *Introduction to the Theory of Differential Equations with Deviating Arguments* (Nauka, Moscow, 1971, p. 19; Academic Press, New York, 1973).
22. A. D. Myshkis, Usp. Mat. Nauk **4**, 33 (1949).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 144–152.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 162–174.
Original Russian Text Copyright © 2000 by Ovchinnikov, Dyugaev.

Breakdown of Renormalizability in the Kondo Problem in the High-Temperature Expansion of the Free Energy

Yu. N. Ovchinnikov and A. M. Dyugaev

Max-Planck-Institute for the Physics of Complex Systems,
Dresden, D-01187 Germany
Landau Institute of Theoretical Physics, Russian Academy of Sciences,
Chernogolovka, Moscow oblast, 142432 Russia

Received July 15, 1999

Abstract—It is shown that there are two energy scales in the Kondo problem: T_k and T_0, one of which (T_k) is exponentially small in the coupling constant g. The second scale T_0 is proportional to the squared coupling constant. Perturbation theory is valid only in the region $T \gg T_0$. The point T_0 is apparently the crossover from weak to strong coupling. The first indications of the breakdown of the hypothesis of only one energy scale in the Kondo problem appear in fourth order of perturbation theory. *© 2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

In investigations of the thermodynamics of the Kondo model, it is ordinarily assumed that there is only one energy scale and that the theory is renormalizable [1–3]. Specifically, it is assumed that the dependence of the energy $\mathcal{F}$ on the cutoff parameter D has the form

$$\mathcal{F} = D\sum_n g^n C_n + T_k f\left(\frac{\mu H}{T_k}, \frac{T}{T_k}\right), \tag{1}$$

where g is a dimensionless coupling constant, which will be determined below; $T_k \sim D\exp(-1/2g)$ is the Kondo temperature; C_n are numerical coefficients; μ is the Bohr magneton; and, H is the external magnetic field.

The first term in equation (1) is assumed to be temperature-independent. Physically, it is the shift of the ground state of the system as a function of the coupling constant g. Investigation of the dependence of the average electron spin of an impurity atom on the magnetic field strength $T = 0$ [4] has shown that perturbation theory is inapplicable even for strong magnetic fields, $\mu H \gg T_k$. As a result, in strong magnetic fields, instead of very slowly decreasing logarithmic corrections to the saturated value of the impurity spin, power-law corrections arise [4], which agrees much better with the experimental data [5].

In this connection, it makes sense to check the hypothesis that only one energy scale exists and the renormalizability of the exchange interaction on the basis of finite-temperature perturbation theory. The problem is complicated by the fact that in the Kondo problem the free energy is not a sum of irreducible diagrams, and the perturbation series for the partition function Z must be investigated in order to calculate the free energy. As a result, we shall show that the Kondo model is nonrenormalizable and the expression (1) for the free energy is incorrect. In the Kondo model there exists a second energy scale $T_0 \sim g^2 D \gg T_k$. This energy scale was obtained in [6] in the self-consistent field approximation. Perturbation theory is valid only in the temperature range $T \gg T_0$. In the range $T \ll T_0$ strong coupling is realized.

2. MODEL

We shall assume that the interaction of an electron localized on an impurity atom with the conduction electrons can be described by the exchange Hamiltonian

$$\hat{H} = \hat{H}_0 + \int d^3r_1 d^3r_2 V(r_1 - r_2)\chi^+_\alpha(r_1)\varphi^+_\beta(r_2)\chi_\beta(r_2)\varphi_\alpha(r_1) - \frac{\mu H}{2}\int(\varphi^+_\uparrow(r_1)\varphi_\uparrow(r_1) - \varphi^+_\downarrow(r_1)\varphi_\downarrow(r_1))d^3r_1. \tag{2}$$

In equation (2) the operators φ^+_α and χ^+_α create an electron in a state localized on an impurity and in the continuous spectrum, respectively.

At finite temperature the partition function Z can be represented in the form [7]

$$Z = \operatorname{Tr}\exp(-\hat{H}/T) = \operatorname{Tr}\left\{e^{-\hat{H}_0/T}\left[1 - \int_0^{1/T} d\tau_1 \hat{V}(\tau_1) + \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \hat{V}(\tau_1)\hat{V}(\tau_2) - \ldots\right]\right\}, \tag{3}$$

where the operator $\hat{V}(\tau)$ is determined by the expression

$$\hat{V}(\tau) = e^{\hat{H}_0\tau}V(r)e^{-\hat{H}_0\tau}. \tag{4}$$

1063-7761/00/9001-0144$20.00 © 2000 MAIK “Nauka/Interperiodica”

In what follows, we shall use the symmetric model for cutting off the matrix elements with cutoff energy D.

In the calculation of the perturbation-theory series in the expression (3) for the partition function there arises the quantity

$$\Phi(\tau) = \sum_k (1-n_k)e^{-\tau\varepsilon_k} = \int_{-D}^{D} d\xi e^{-\tau\xi}\left(1 - \frac{1}{1+\exp(\xi/T)}\right), \tag{5}$$

where n_k is the family distribution function, and the electron energy ε_k is measured from the Fermi surface.

Simple calculations show that the following two relations hold:

$$\sum_k n_k e^{\tau\varepsilon_k} = \int_{-D}^{D} d\xi e^{\tau\xi}\left(1+\exp\frac{\xi}{T}\right)^{-1} = \Phi(\tau), \tag{6}$$

$$\Phi(\tau) = \Phi(1/T-\tau). \tag{7}$$

The relations (6) and (7) greatly simplify the calculation of the perturbation-theory series in the expression (3) for the partition function.

The main results—the nonrenormalizability of the Kondo model and the existence of a second energy scale T_0—do not depend on the specific form of the cutoff of the matrix elements.

From equations (5) and (6) we find an explicit expression for the function $\Phi(\tau)$:

$$\Phi(\tau) = \frac{1}{\tau}(1-e^{-D\tau}) + \left(\frac{1-\exp(-D(1/T-\tau))}{1/T-\tau} - \frac{1}{1/T+\tau}\right) - 2\tau\sum_{n=2}^{\infty}\frac{(-1)^n}{(n/T)^2-\tau^2}. \tag{8}$$

Near the singularity ($\tau T \ll 1$) the function $\Phi(\tau)$ can be represented in the form

$$\Phi(\tau) = \frac{1}{\tau}(1-e^{-D\tau}) + \frac{\pi^2}{6}T^2\tau + O(T(T\tau)^3). \tag{9}$$

An important property of the function $\Phi(\tau)$ is that there is no term of the form const · T on the right-hand side of equation (9).

3. PERTURBATION THEORY

In the second order of perturbation theory, we find from equations (2) and (3)

$$Z_2 = Z_0^e Z_0^i g^2 (J_0 + J_H), \tag{10}$$

where

$$J_H = \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \Phi^2(\tau_1-\tau_2)\mathcal{N}(\tau_1-\tau_2), \tag{11}$$

and $J_0 = J_{H=0}$, $Z_0^i = 2\cosh(\mu H/2T)$, and Z_0^e is the partition function of a free electron gas. The function $\mathcal{N}(\tau)$ and the dimensionless coupling constant g are defined as follows:

$$\mathcal{N}(\tau) = (e^{-\mu H\tau} + e^{-\mu H/T + \mu H\tau})/(1+e^{-\mu H/T}), \tag{12}$$

$$\sum_k \langle V\rangle(\ldots) \longrightarrow g\int_{-D}^{D} d\xi(\ldots). \tag{13}$$

The expression (10) is identical to the results obtained in [3]. We shall not require the explicit form of the function J_H, since in the free energy the higher-order and subsequent terms in powers of (D/T) cancel for any function $\Phi(\tau)$ of the form (7) and (9).

In third order of perturbation theory the correction Z_3 to the partition function is

$$Z_3 = -Z_0^e Z_0^i g^3 \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_3) \times \Phi(\tau_1-\tau_3)\{\mathcal{N}(\tau_1-\tau_2) + \mathcal{N}(\tau_2-\tau_3) + \mathcal{N}(\tau_1-\tau_3)\}. \tag{14}$$

We note that the function $\mathcal{N}(\tau)$ can appear in the expressions for the corrections to the partition function in order no higher than linear. This is because in a localized state there is always only one electron.

In the fourth-order correction it is convenient to single out the large term $0.5(J_0 + J_H)^2$, which is proportional to $(D/T)^2$. Then, the quantity Z_4 can be represented in the form

$$\begin{aligned} Z_4 = -Z_0^e Z_0^i g^4 \Bigg\{ & \frac{1}{2}(J_0 + J_H)^2 \\ & + \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 [\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_3-\tau_4) \\ & \times (\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4)) - \mathcal{N}(\tau_1-\tau_2)\mathcal{N}(\tau_3-\tau_4)) \\ & + \Phi^2(\tau_1-\tau_4)\Phi^2(\tau_2-\tau_3) \\ & \times (\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4)) - \mathcal{N}(\tau_1-\tau_4)\mathcal{N}(\tau_2-\tau_3)) \\ & - \Phi^2(\tau_1-\tau_3)\Phi^2(\tau_2-\tau_4) \\ & \times (\mathcal{N}(\tau_1-\tau_3) + \mathcal{N}(\tau_2-\tau_4) + \mathcal{N}(\tau_1-\tau_3)\mathcal{N}(\tau_2-\tau_4))] \\ & + \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 [\Phi(\tau_1-\tau_4)\Phi(\tau_3-\tau_4) \end{aligned} \tag{15}$$

$$\begin{gathered}\times\Phi(\tau_2-\tau_3)\Phi(\tau_1-\tau_2)(2+2\mathcal{N}(\tau_1-\tau_2)+N(\tau_3-\tau_4))\\ +\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_1-\tau_4)+\mathcal{N}(\tau_2-\tau_4)\\ +\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}(\tau_1-\tau_3)+\mathcal{N}(\tau_2-\tau_3)\,)\\ -\Phi(\tau_2-\tau_4)\Phi(\tau_3-\tau_4)\Phi(\tau_1-\tau_3)\Phi(\tau_1-\tau_2)\\ \times(2+\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_3-\tau_4))\\ -\Phi(\tau_1-\tau_4)\Phi(\tau_2-\tau_4)\Phi(\tau_2-\tau_3)\Phi(\tau_1-\tau_3)\\ \times(2+\mathcal{N}(\tau_1-\tau_4)+\mathcal{N}(\tau_2-\tau_3))\,]\Bigg\}.\end{gathered}$$

To establish the existence of a second energy scale in the Kondo problem, it is sufficient to known the partition function Z up to terms of fourth order. For this reason, we shall present in Appendix 1 an expression for the fifth-order correction Z_5 to the partition function.

The expressions (10), (14), and (15) make it possible to find the correction arising in the free energy because of the interaction up to terms of fourth order in g, inclusively. The correction to the free energy contains a large term that is proportional to the cutoff energy D and depends on the interaction constant g. This contribution is ordinarily treated as a shift δE of the ground-state energy and is assumed to be temperature-independent to all orders in the coupling constant. Then, the quantity δE so determined should be identical to all orders in g (for the same cutoff of the matrix elements) to the quantity δE, which is the shift of the ground state and is found at $T = 0$.

We shall show that this agreement occurs only in second and third orders in g. In fourth order in the coupling constant g, there arises a difference that leads to the second energy scale in the Kondo problem. We shall prove this assertion. From equations (10), (14), and (15) we find the following expression for δE:

$$\begin{aligned}-\delta E = D\Bigg\{&g^2 4\ln 2\\ &-3g^3\int_0^\infty\frac{dx}{x}\int_0^x\frac{dy(1-e^{-x})(1-e^{-y})(1-e^{-(x-y)})}{y(x-y)}\\ &+g^4\Bigg[6\int_0^\infty dxdydz\frac{(1-e^{-x})(1-e^{-y})(1-e^{-z})(1-e^{-(x+y+z)})}{xyz(x+y+z)}\\ &-3\int_0^\infty dxdydz\frac{(1-e^{-(x+y)})^2(1-e^{-(y+z)})^2}{(x+y)^2(y+z)^2}\Bigg]\Bigg\}.\end{aligned}\tag{16}$$

The shift $\delta\tilde{E}$ of the ground state are $T = 0$ was found in [4] (see Appendix 2). Using the results of [4], we write the expression for $\delta\tilde{E}$ in the form

$$\begin{aligned}-\delta\tilde{E} = D\Bigg\{&g^2 4\ln 2-3g^3\int_0^1\frac{d\xi_1 d\xi_2 d\xi_3}{(\xi_1+\xi_2)(\xi_2+\xi_3)}\\ &+g^4\int_0^1 d\xi_1 d\xi_2 d\xi_3 d\xi_4\Bigg[\frac{10}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_1+\xi_4)}\\ &-\frac{4}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_3+\xi_4)}\\ &-\frac{4}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_1+\xi_2+\xi_3+\xi_4)}\\ &+\frac{1}{(\xi_1+\xi_2)(\xi_3+\xi_4)(\xi_1+\xi_2+\xi_3+\xi_4)}\Bigg]\Bigg\}.\end{aligned}\tag{17}$$

Calculating the integrals in the third-order terms in equations (16) and (17), we obtain

$$\begin{gathered}\int_0^1\frac{d\xi_1 d\xi_2 d\xi_3}{(\xi_1+\xi_2)(\xi_1+\xi_3)} = 2\ln^2 2+\frac{\pi^2}{6},\\ \int_0^\infty\frac{dx}{x}\int_0^x\frac{dy}{y}\frac{(1-e^{-x})(1-e^{-y})(1-e^{-(x-y)})}{x-y}\\ = \frac{\pi^2}{6}+\left(\frac{\pi^2}{3}-4\sum_{n=1}^\infty\frac{1}{2^n n^2}\right).\end{gathered}\tag{18}$$

Since

$$\frac{\pi^2}{3}-4\sum_{n=1}^\infty\frac{1}{2^n n^2} = 2\ln^2 2,\tag{19}$$

the expressions for δE and $\delta\tilde{E}$ are identical in the second and third orders of perturbation theory. In fourth order we find from equations (16) and (17) that

$$\begin{aligned}-(\delta\tilde{E}-\delta E) &= 4g^4 D\\ \times\int_0^1 d\xi_1 d\xi_2 d\xi_3 d\xi_4\Bigg\{&\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_1+\xi_4)}\\ &-\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_3+\xi_4)}\\ &-\frac{1}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_1+\xi_2+\xi_3+\xi_4)}\\ &+\frac{1}{(\xi_1+\xi_2)(\xi_3+\xi_4)(\xi_1+\xi_2+\xi_3+\xi_4)}\Bigg\}.\end{aligned}\tag{20}$$

The second and third integrals in equation (20) are related by the simple relation

$$\int_0^1 \frac{d\xi_1 d\xi_2 d\xi_3 d\xi_4}{(\xi_1+\xi_2)(\xi_1+\xi_3)(\xi_1+\xi_2+\xi_3+\xi_4)} = \frac{1}{2}\int_0^1 \frac{d\xi_1 d\xi_2 d\xi_3 d\xi_4}{(\xi_1+\xi_3)(\xi_1+\xi_2)(\xi_3+\xi_4)}. \tag{21}$$

Using the relation (21) and simple transformations we write equation (20) in the form

$$-(\delta\tilde{E}-\delta E) = 4g^4 D\left\{\int_0^1 dx \ln^3\left(\frac{1+x}{x}\right) - \frac{3}{2}\int_0^1 dx dy \frac{\ln((1+x)/x)\ln((1+y)/y)}{x+y} + I\right\}. \tag{22}$$

The integral I in equation (22) is determined by the expression

$$I = \int_0^1 \frac{d\xi_1 d\xi_2 d\xi_3 d\xi_4}{(\xi_1+\xi_2)(\xi_3+\xi_4)(\xi_1+\xi_2+\xi_3+\xi_4)} = 40\ln 2 - 24\ln 3 + 2\sum_{n=1}^{\infty}\frac{1}{n^2 2^n}. \tag{23}$$

It is convenient to calculate the first two integrals in equation (22) together:

$$\int_0^1 dx \ln^3\left(\frac{1+x}{x}\right) - \frac{3}{2}\int_0^1 \frac{dx dy}{x+y}\ln\left(\frac{1+x}{x}\right)\ln\left(\frac{1+y}{y}\right) = \ln^3 2 - 6\int_2^{\infty}\frac{dx \ln x \ln(x/2)}{x(x-1)(x-2)} = -2\ln^3 2 - 6\left(\zeta(3) - 2\sum_{n=1}^{\infty}\frac{1}{2^n n^3}\right), \tag{24}$$

where $\zeta(x)$ is the Riemann zeta function. From equations (22)–(24) we find

$$-(\delta\tilde{E}-\delta E) = 4g^4 D\left\{40\ln 2 - 24\ln 3 + 2\sum_{n=1}^{\infty}\frac{1}{n^2 2^{2n}} - 2\ln^3 2 - 6\left(\zeta(3) - 2\sum_{n=1}^{\infty}\frac{1}{2^n n^3}\right)\right\} = g^4 D \times 3.8506. \tag{25}$$

The fact that in equation (25) the right-hand side is different from zero means that in the Kondo problem there exists a second characteristic energy scale T_0 such that $T_k \ll T_0 \ll \varepsilon_F$. To determine this scale it is necessary to investigate the higher-order terms in the perturbation theory. We shall show below that in the presence or absence of a magnetic field the anomalous terms appear only in eighth order of perturbation theory [8].

4. ANOMALOUS TERMS IN THE FREE ENERGY

Anomalous terms arise only in eighth order of perturbation theory, irrespective of whether or not an external magnetic field is present. On the other hand, it is much easier to investigate the perturbation-theory series for $H = 0$. For this reason, since we intend to investigate the eight-order terms of the perturbation theory, we shall confine our attention in this section to the case $H = 0$.

For $H = 0$ the expression for the partition function Z, up to terms of ninth order in g, can be written in the form

$$Z = Z_0^e Z_0^i\{2g^2 J_0 - 3g^3 f_3 + g^4 f_4 + g^5 f_5 + g^6 f_6 + g^7 f_7 + g^8 f_8 + g^9 f_9\} = Z_0^e Z_0^i \exp\{2g^2 J_0 - 3g^3 f_3 + g^4 F_4 + g^5 F_5 + g^6 F_6 + g^7 F_7 + g^8 F_8 + g^9 F_9\}. \tag{26}$$

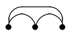
From equation (26) we find a relation between the functions F_i and f_i:

$$\begin{aligned}
F_4 &= f_4 - 2J_0^2,\\
F_5 &= f_5 + 6f_2 f_3,\\
F_6 &= f_6 - 2J_0 f_4 - \frac{9}{2}f_3^2 + \frac{8}{3}f_2^3,\\
F_7 &= f_7 - 2J_0 f_5 + 3f_3 f_4 - 12J_0^2 f_3,\\
F_8 &= f_8 - 2J_0 f_6 - \frac{1}{2}f_4^2 + 3f_3 f_5 + 4f_2^2 f_4 + 18J_0 f_3^2 - 4J_0^4,\\
F_9 &= f_9 - 2J_0 f_7 + 3f_3 f_6 + 4J_0^2 f_5 - 12J_0 f_3 f_4 - 9f_3^2 + 24J_0^3 f_3 - f_4 f_5.
\end{aligned} \tag{27}$$

In equations (27) for the functions F_6 and F_8 we express the functions $f_{4,5}$ in terms of $F_{4,5}$. The result is

$$\begin{aligned}
F_6 &= f_6 - 2J_0 F_4 - \frac{9}{2}f_3^2 - \frac{4}{5}f_2^3,\\
F_8 &= f_8 - 2J_0 F_6 - 9J_0 f_3^2 - \frac{2}{3}J_0^4 + 3f_3 F_5 - \frac{1}{2}F_4^2 - 2J_0^2 F_4.
\end{aligned} \tag{28}$$

To simplify the formulas and make them easier to interpret, we switch to a diagrammatic representation of the equations for F_i. From equations (14) and (26) we find

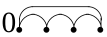
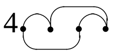
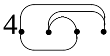
$f_3 =$ 

$$= \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_3)\Phi(\tau_1-\tau_3). \quad (29)$$

Each line in a diagram represents $\Phi(\tau)$, each vertex corresponds to the time τ_i, and all times τ_i are ordered in accordance with the diagram. Two lines always meet at each point. From equations (15) and (27) we find

$F_4 = \square - 3$ . (30)

In equation (30) a square □ represents a sum of all connected diagrams in equation (15) which are constructed from single lines:

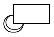
$\square = 10$  $- 4$  $- 4$ . (31)

The second term in equation (30) consists of double lines. As an example, we give its analytic expression:

 $= \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 \Phi^2(\tau_1-\tau_3)\Phi^2(\tau_2-\tau_4).$ (32)

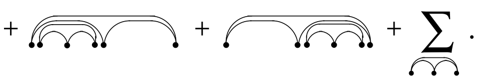
We shall also require the diagrammatic representation of the quantities J_0^2, J_0^3, and J_0^4. As an example, we have

$J_0^2 \equiv 2[$   $+$  $+$  $]$. (33)

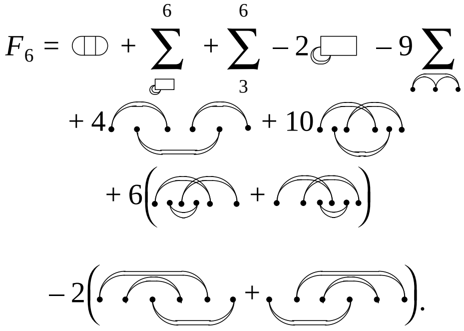
We shall also introduce the following notations:  is the set of all irreducible diagrams in f_6 that are constructed only with single lines; $\sum_3^6$ is the set of all irreducible diagrams in f_6 that are constructed only with double lines; $\sum^6$ is the set of all irreducible diagrams in f_6 that are constructed with one double and four single lines; and,  is the set of all diagrams in equation (31) in which one or several vertices are encompassed by a double line (without all vertices being encompassed).

We also introduce the quantity $\sum$ using the relation

$\frac{1}{2} f_3^2 =$ 

 (34)

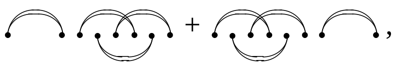
Using equations (28)–(34) we find the following representation for the function F_6:

 (35)

All diagrams in equation (35) are irreducible. Therefore, no anomalous terms appear in the free energy in sixth order in the coupling constant. The equations (28) and (35) make it possible to find the anomalous contribution to the function F_8. The irreducible diagrams do not contain an anomalous contribution. For this reason, the right-hand side of equation (28) for F_8 should be reprojected on all possible reducible diagrams contained in f_8 or in any other term on the right-hand side of equation (28).

To investigate the function F_8 it is necessary to find a representation for the function F_5 that is similar to equation (35) for the function F_6. Such a representation is obtained in Appendix 1 (equation (A.4)).

Analysis shows that all reducible diagrams in F_8, with the exception of diagrams of the form

 (36)

cancel. As a result, we find the following expression for the anomalous contribution to the free energy $\delta\mathcal{F}_{an}$:

$$\delta\mathcal{F}_{an} = 24 J_0 T g^8 \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3$$
$$\times \int_0^{\tau_3} d\tau_4 \int_0^{\tau_4} d\tau_5 \int_0^{\tau_5} d\tau_6 \Phi^2(\tau_1-\tau_4)\Phi^2(\tau_2-\tau_5)\Phi^2(\tau_3-\tau_6). \quad (37)$$

Substituting the explicit expression (8) for the function $\Phi(\tau)$ into equation (37), we obtain [8]

$$\delta\mathcal{F}_{an} = 24 J_0 D g^8$$
$$\times \int_0^{\infty} dx dy dz \frac{I_1(x+y) I_1(y+z)(1-e^{-(x+y+z)})^2}{(x+y+z)^2}, \quad (38)$$

where

$$I_1(a) = \int_a^{\infty} \frac{dx}{x^2}(1-e^{-x})^2. \quad (39)$$

In order of magnitude, the anomalous contribution to the free energy is

$$\delta\mathcal{F}_{an} \sim g^8 D^2/T. \tag{40}$$

Comparing the corrections arising in the specific heat from the fourth-order terms (in equation (15)) and eighth-order terms (equation (40)) shows that the second characteristic energy scale T_0 in the Kondo problem is, in order of magnitude,

$$T_0 \sim g^2 D. \tag{41}$$

This value of T_0 is identical to the expression obtained in [6] in the self-consistent field approximation. The point T_0 seems to be the crossover from perturbation theory to strong coupling. Perturbation theory is valid only in the region $T \gg T_0$. Strong coupling is realized in the region $T \ll T_0$.

In Appendix 2 we show that a magnetic field does not lead to the appearance of anomalous terms in sixth order of perturbation theory.

ACKNOWLEDGMENTS

We are grateful to P. Fulde for helpful remarks and for encouragement. The bulk of this work was performed at the Max-Planck-Institute for the Physics of Complex Systems (Dresden).

APPENDIX 1

We present an expression for the fifth-order correction to the partition function (equation (26)) in an external magnetic field:

$$\begin{aligned}
f_5 = {} & \int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 \int_0^{\tau_4} d\tau_5 \\
& \times \{\Phi(\tau_1-\tau_5)\Phi(\tau_1-\tau_4)\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_4)\Phi(\tau_2-\tau_5) \\
& \times [\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_2-\tau_4)+\mathcal{N}(\tau_3-\tau_4)-\mathcal{N}(\tau_1-\tau_5) \\
& -\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_5)\Phi(\tau_2-\tau_5)\Phi(\tau_2-\tau_4)\Phi(\tau_1-\tau_3)\Phi(\tau_3-\tau_4) \\
& \times [\mathcal{N}(\tau_1-\tau_5)+\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_5)\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_4)\Phi(\tau_3-\tau_4)\Phi(\tau_3-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_5)+\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_2-\tau_5)-\mathcal{N}(\tau_3-\tau_4) \\
& -\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))-\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))] \\
& +\Phi(\tau_1-\tau_5)\Phi(\tau_1-\tau_4)\Phi(\tau_2-\tau_4)\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_5)+\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))] \\
& -\Phi(\tau_1-\tau_5)\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_4)\Phi(\tau_4-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}(\tau_4-\tau_5) \\
& +\mathcal{N}(\tau_1-\tau_5)+\mathcal{N}(\tau_1-\tau_3)+\mathcal{N}(\tau_1-\tau_4)+\mathcal{N}(\tau_2-\tau_4) \\
& +\mathcal{N}(\tau_2-\tau_5)+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))+\mathcal{N}(\tau_3-\tau_5) \\
& +\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))+\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5)) \\
& +\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))+\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_5)\Phi(\tau_1-\tau_3)\Phi(\tau_2-\tau_3)\Phi(\tau_2-\tau_4)\Phi(\tau_4-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_5)+\mathcal{N}(\tau_1-\tau_4)+\mathcal{N}(\tau_4-\tau_5)-\mathcal{N}(\tau_2-\tau_3) \\
& -(\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))-\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5)))] \\
& +\Phi(\tau_1-\tau_2)\Phi(\tau_1-\tau_3)\Phi(\tau_3-\tau_5)\Phi(\tau_2-\tau_4)\Phi(\tau_4-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_4-\tau_5)+\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_2)\Phi(\tau_1-\tau_3)\Phi(\tau_3-\tau_4)\Phi(\tau_4-\tau_5)\Phi(\tau_2-\tau_5) \\
& \times [\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}(\tau_4-\tau_5)+\mathcal{N}(\tau_3-\tau_5)-\mathcal{N}(\tau_1-\tau_2) \\
& -\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))-\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_5)\Phi(\tau_4-\tau_5)\Phi(\tau_1-\tau_4) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_1-\tau_3)-\mathcal{N}(\tau_4-\tau_5) \\
& -\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))] \\
& +\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_5)\Phi(\tau_1-\tau_4)\Phi(\tau_3-\tau_4)\Phi(\tau_3-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))] \\
& +\Phi(\tau_1-\tau_3)\Phi(\tau_2-\tau_3)\Phi(\tau_2-\tau_5)\Phi(\tau_1-\tau_4)\Phi(\tau_4-\tau_5) \\
& \times [\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_4-\tau_5) \qquad \text{(A.1)} \\
& +\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))]\} \\
& -\int_0^{1/T} d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 \int_0^{\tau_3} d\tau_4 \int_0^{\tau_4} d\tau_5 \\
& \times \{\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_5)\Phi(\tau_1-\tau_5)\Phi^2(\tau_3-\tau_4) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_2-\tau_5)+\mathcal{N}(\tau_1-\tau_5) \\
& +\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4)) \\
& +\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_3)\Phi(\tau_1-\tau_5)\Phi(\tau_3-\tau_5)\Phi^2(\tau_2-\tau_4) \\
& \times [\mathcal{N}(\tau_1-\tau_5)-\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))] \\
& +\Phi(\tau_1-\tau_2)\Phi(\tau_2-\tau_3)\Phi(\tau_1-\tau_3)\Phi^2(\tau_4-\tau_5) \\
& \times [\mathcal{N}(\tau_1-\tau_2)+\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_1-\tau_3) \\
& +\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))+\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))
\end{aligned}$$

$$+\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))]$$

$$+\Phi(\tau_1-\tau_2)\Phi(\tau_1-\tau_4)\Phi(\tau_2-\tau_4)\Phi^2(\tau_3-\tau_5)$$

$$\times[\mathcal{N}(\tau_1-\tau_2)-\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))]$$

$$+\Phi(\tau_3-\tau_4)\Phi(\tau_4-\tau_5)\Phi(\tau_3-\tau_5)\Phi^2(\tau_1-\tau_2)$$

$$\times[\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}(\tau_4-\tau_5)+\mathcal{N}(\tau_3-\tau_5)$$

$$+\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))$$

$$+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))]$$

$$+\Phi(\tau_2-\tau_4)\Phi(\tau_4-\tau_5)\Phi(\tau_2-\tau_5)\Phi^2(\tau_1-\tau_3)$$

$$\times[\mathcal{N}(\tau_4-\tau_5)-\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))]$$

$$+\Phi(\tau_1-\tau_4)\Phi(\tau_4-\tau_5)\Phi(\tau_1-\tau_5)\Phi^2(\tau_2-\tau_3)$$

$$\times[\mathcal{N}(\tau_1-\tau_4)+\mathcal{N}(\tau_4-\tau_5)+\mathcal{N}(\tau_1-\tau_5)$$

$$+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))$$

$$+\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))]$$

$$+\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_4)\Phi(\tau_2-\tau_4)\Phi^2(\tau_1-\tau_5)$$

$$\times[\mathcal{N}(\tau_2-\tau_3)+\mathcal{N}(\tau_3-\tau_4)+\mathcal{N}(\tau_2-\tau_4)$$

$$+\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))+\mathcal{N}((\tau_1-\tau_3)+(\tau_4-\tau_5))$$

$$+\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_5))]$$

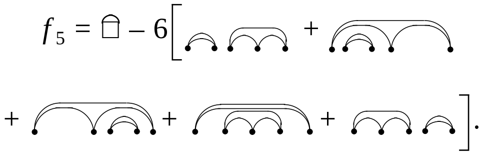
$$+\Phi(\tau_1-\tau_3)\Phi(\tau_1-\tau_4)\Phi(\tau_3-\tau_4)\Phi^2(\tau_2-\tau_5)$$

$$\times[\mathcal{N}(\tau_3-\tau_4)-\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))]$$

$$+\Phi(\tau_2-\tau_3)\Phi(\tau_3-\tau_5)\Phi(\tau_2-\tau_5)\Phi^2(\tau_1-\tau_4)$$

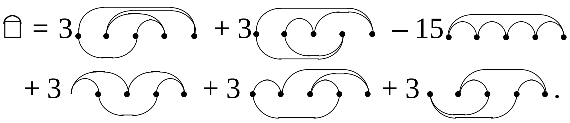
$$\times[\mathcal{N}(\tau_2-\tau_3)-\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))]\}.$$

In a zero magnetic field the expression for the function f_5 simplifies substantially and its diagrammatic representation is presented below:



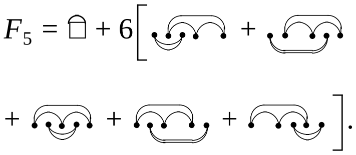
(A.2)

In equation (A.2) the symbol ⌂ represents the set of irreducible diagrams appearing in f_5 (equation (A.1)) and consisting only of single lines. From equation (A.1) we find its diagrammatic representation in a zero magnetic field:



(A.3)

From equations (27) and (A.2) we also find an expression for the function F_5 in a zero magnetic field:



(A.4)

APPENDIX 2

In Appendix 2 we show that a magnetic field does not lead to the appearance of anomalous terms in the free energy in sixth order in the coupling constant g. To check this assertion, we write the sixth-order correction $\delta\mathcal{F}_6$ to the free energy in the form

$$\delta\mathcal{F}_6 = -Tg^6F_6(H), \tag{A.5}$$

where the function $F_6(H)$ is determined by an equation of the form (27) in a magnetic field

$$F_6 = -\frac{1}{6}(2J_0+\delta J_H)^3-\frac{1}{2}J_3^2(H)-(2J_0+\delta J(H))\square$$

$$-(2J_0+\delta J_H)\int\limits_0^{1/T}d\tau_1\int\limits_0^{\tau_1}d\tau_2\int\limits_0^{\tau_2}d\tau_3\int\limits_0^{\tau_3}d\tau_4$$

$$\times\{[\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\mathcal{N}(\tau_1-\tau_2)\mathcal{N}(\tau_3-\tau_4)]$$

$$\times\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_3-\tau_4) \tag{A.6}$$

$$+[\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\mathcal{N}(\tau_1-\tau_4)\mathcal{N}(\tau_2-\tau_3)]$$

$$\times\Phi^2(\tau_1-\tau_4)\Phi^2(\tau_2-\tau_3)$$

$$+[\mathcal{N}(\tau_1-\tau_3)+\mathcal{N}(\tau_2-\tau_4)+\mathcal{N}(\tau_1-\tau_3)\mathcal{N}(\tau_2-\tau_4)]$$

$$\times\Phi^2(\tau_1-\tau_3)\Phi^2(\tau_2-\tau_4)\}+f_6(H).$$

In equation (A.6) the function δJ_H is $\delta J(H) = J_H - J_0$, and the functions $J_3(H)$ and $f_6(H)$ are related with the third- and sixth-order corrections to the partition function by the relations

$$Z_3 = -Z_0^eZ_0^ig^3J_3(H), \quad Z_6 = Z_0^eZ_0^ig^6f_6(H). \tag{A.7}$$

The square in equation (A.6) must be removed in a magnetic field. Its expression is determined by the last term in equation (15).

We have shown above (see equation (35)) that in a zero magnetic field the function $F_6(0)$ does not contain anomalous terms. We shall show that there are also no anomalous terms in $F_6(H)$. Retaining only the anomalous terms on the right-hand side of equation (A.6), we put the expression for $F_6(H)$ in the form

$$\begin{aligned}F_6(H) \sim & -2J_0^2\delta J_H - 2J_0\int_0^{1/T}d\tau_1\int_0^{\tau_1}d\tau_2\int_0^{\tau_2}d\tau_3\int_0^{\tau_3}d\tau_4\\ & \times\{\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_3-\tau_4)\\ & \times[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\delta\mathcal{N}(\tau_1-\tau_2)\\ & -\delta\mathcal{N}(\tau_3-\tau_4)]+\Phi^2(\tau_1-\tau_4)\Phi^2(\tau_2-\tau_3)\\ & \times[\delta\mathcal{N}((\tau_1-\tau_4)-(\tau_2-\tau_3))-\delta\mathcal{N}(\tau_1-\tau_4)\\ & -\delta\mathcal{N}(\tau_2-\tau_3)]\}+4J_0\int_0^{1/T}d\tau_1\int_0^{\tau_1}d\tau_2\int_0^{\tau_2}d\tau_3\int_0^{\tau_3}d\tau_4\\ & \times\Phi^2(\tau_1-\tau_3)\Phi^2(\tau_2-\tau_4)[\delta\mathcal{N}(\tau_1-\tau_3)+\delta\mathcal{N}(\tau_2-\tau_4)]\\ & +3\delta J_H\int_0^{1/T}d\tau_1\int_0^{\tau_1}d\tau_2\int_0^{\tau_2}d\tau_3\int_0^{\tau_3}d\tau_4\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_2-\tau_4)\\ & +(f_6(H)-f_6(0))-\frac{1}{2}(J_3^2(H)-J_3^2(0))\\ & -\delta[(2J_0+\delta J(H))\square].\end{aligned}\tag{A.8}$$

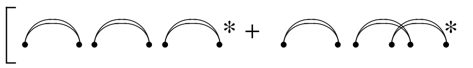
In equation (A.8) the function $\delta\mathcal{N}(\tau)$ is given by

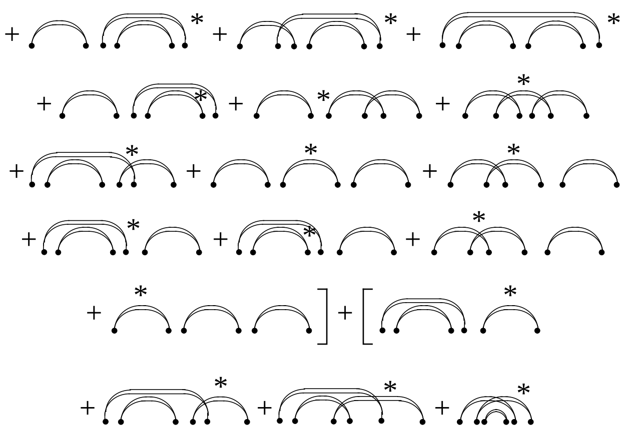
$$\delta\mathcal{N}(\tau) = \mathcal{N}(\tau)-1. \tag{A.9}$$

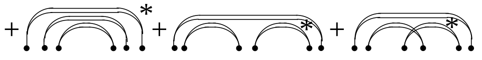
The symbol δ in front of the last term in equation (A.9) means that its value for $H = 0$ should be subtracted from the expression in parentheses.

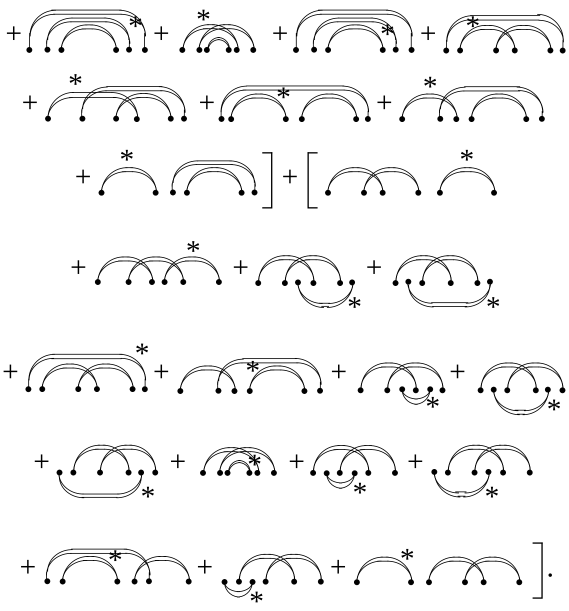
In the subsequent analysis each term in equation (A.8) is represented as a sum of time-ordered diagrams. For example, we have

$$\frac{1}{2}J_0^2\delta J_H =$$









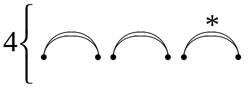
(A.10)

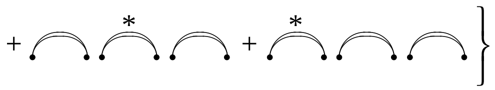
In equation (A.10) the symbol  represents a double line with the factor $\delta\mathcal{N}(\tau)$

$$= \Phi^2(\tau)\delta\mathcal{N}(\tau). \tag{A.11}$$

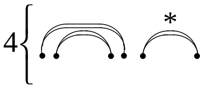
We also present an expression for the anomalous part of the function $f_6(H)$ consisting of blocks only with double lines:

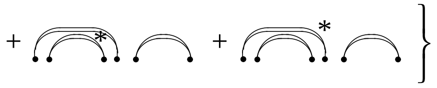
$$(f_6(H)-f_6(0))_{\text{an}} = 4\Big\{$$





$$\begin{aligned}& +\int d\tau\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_3-\tau_4)\Phi^2(\tau_5-\tau_6)\\ & \times\{[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4)+(\tau_5-\tau_6))\\ & -\delta\mathcal{N}(\tau_1-\tau_2)-\delta\mathcal{N}(\tau_3-\tau_4)-\delta\mathcal{N}(\tau_5-\tau_6)]\\ & +[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_4))-\delta\mathcal{N}(\tau_1-\tau_2)\\ & -\delta\mathcal{N}(\tau_3-\tau_4)]+[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_5-\tau_6))\\ & -\delta\mathcal{N}(\tau_1-\tau_2)-\delta\mathcal{N}(\tau_5-\tau_6)]\\ & +[\delta\mathcal{N}((\tau_3-\tau_4)+(\tau_5-\tau_6))-\delta\mathcal{N}(\tau_3-\tau_4)\\ & -\delta\mathcal{N}(\tau_5-\tau_6)]\}+4\Big\{\end{aligned}$$





$$+\int d\tau\,\Phi^2(\tau_2-\tau_3)\Phi^2(\tau_1-\tau_4)\Phi^2(\tau_5-\tau_6)$$

$$\times\{[\delta\mathcal{N}((\tau_1-\tau_4)-(\tau_2-\tau_3)+(\tau_5-\tau_6))$$

$$-\delta\mathcal{N}(\tau_1-\tau_4)-\delta\mathcal{N}(\tau_2-\tau_3)-\delta\mathcal{N}(\tau_5-\tau_6)]$$

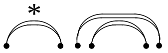
$$+[\delta\mathcal{N}((\tau_1-\tau_4)-(\tau_2-\tau_3))-\delta\mathcal{N}(\tau_1-\tau_4)$$

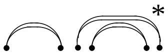
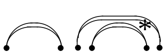
$$-\delta\mathcal{N}(\tau_2-\tau_3)]+[\delta\mathcal{N}((\tau_1-\tau_4)+(\tau_5-\tau_6))$$

$$-\delta\mathcal{N}(\tau_1-\tau_4)-\delta\mathcal{N}(\tau_5-\tau_6)]$$

$$+[\delta\mathcal{N}((\tau_2-\tau_3)+(\tau_5-\tau_6))-\delta\mathcal{N}(\tau_2-\tau_3)$$

$$-\delta\mathcal{N}(\tau_5-\tau_6)]\}+4\{$$



$+$  $+$  $\}$

$$+\int d\tau\,\Phi^2(\tau_1-\tau_2)\Phi^2(\tau_3-\tau_6)\Phi^2(\tau_4-\tau_5)$$

$$\times\{[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_6)-(\tau_4-\tau_5))$$

$$-\delta\mathcal{N}(\tau_1-\tau_2)-\delta\mathcal{N}(\tau_3-\tau_6)-\delta\mathcal{N}(\tau_4-\tau_5)] \tag{A.12}$$

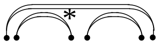
$$+[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_3-\tau_6))-\delta\mathcal{N}(\tau_1-\tau_2)$$

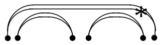
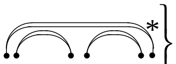
$$-\delta\mathcal{N}(\tau_3-\tau_6)]+[\delta\mathcal{N}((\tau_1-\tau_2)+(\tau_4-\tau_5))$$

$$-\delta\mathcal{N}(\tau_1-\tau_2)-\delta\mathcal{N}(\tau_4-\tau_5)]$$

$$+[\delta\mathcal{N}((\tau_3-\tau_6)-(\tau_4-\tau_5))-\delta\mathcal{N}(\tau_3-\tau_6)$$

$$-\delta\mathcal{N}(\tau_4-\tau_5)]\}+4\{$$



$+$  $+$  $\}$

$$+\int d\tau\,\Phi^2(\tau_1-\tau_6)\Phi^2(\tau_2-\tau_3)\Phi^2(\tau_4-\tau_5)$$

$$\times\{[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_2-\tau_3)-(\tau_4-\tau_5))$$

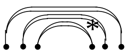
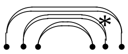
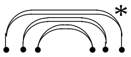
$$-\delta\mathcal{N}(\tau_1-\tau_6)-\delta\mathcal{N}(\tau_2-\tau_3)-\delta\mathcal{N}(\tau_4-\tau_5)]$$

$$+[\delta\mathcal{N}((\tau_2-\tau_3)+(\tau_4-\tau_5))-\delta\mathcal{N}(\tau_2-\tau_3)$$

$$-\delta\mathcal{N}(\tau_4-\tau_5)]+[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_2-\tau_3))$$

$$-\delta\mathcal{N}(\tau_1-\tau_6)-\delta\mathcal{N}(\tau_2-\tau_3)]+[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_4-\tau_5))$$

$$-\delta\mathcal{N}(\tau_1-\tau_6)-\delta\mathcal{N}(\tau_4-\tau_5)]\}$$

$+4\{$  $+$  $+$  $\}$

$$+\int d\tau\,\Phi^2(\tau_1-\tau_6)\Phi^2(\tau_2-\tau_5)\Phi^2(\tau_3-\tau_4)$$

$$\times\{[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_2-\tau_5)+(\tau_3-\tau_4))$$

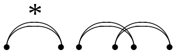
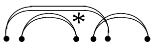
$$-\delta\mathcal{N}(\tau_1-\tau_6)-\delta\mathcal{N}(\tau_2-\tau_5)-\delta\mathcal{N}(\tau_3-\tau_4)]$$

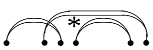
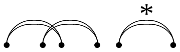
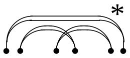
$$+[\delta\mathcal{N}((\tau_2-\tau_5)-(\tau_3-\tau_4))-\delta\mathcal{N}(\tau_2-\tau_5)-\delta\mathcal{N}(\tau_3-\tau_4)]$$

$$+[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_2-\tau_5))-\delta\mathcal{N}(\tau_1-\tau_6)$$

$$-\delta\mathcal{N}(\tau_2-\tau_5)]+[\delta\mathcal{N}((\tau_1-\tau_6)-(\tau_3-\tau_4))$$

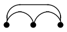
$$-\delta\mathcal{N}(\tau_1-\tau_6)-\delta\mathcal{N}(\tau_3-\tau_4)]\}$$

$+\{$  $+$  $+$ 

$+$  $+$  $+$  $\}.$

In equation (A.12) the symbol $\int d\tau$ represents a time-ordered integral

$$\int d\tau=\int_0^{1/T}d\tau_1\int_0^{\tau_1}d\tau_2\int_0^{\tau_2}d\tau_3\int_0^{\tau_3}d\tau_4\int_0^{\tau_4}d\tau_5\int_0^{\tau_5}d\tau_6.$$

The remaining anomalous blocks consisting of double lines in equation (A.8) are quite simple, and we shall not write them out explicitly. It follows from equations (A.8), (A.10), and (A.12) that all anomalous terms in the function $\delta\mathcal{F}_6(H)$ that contain only double lines cancel. The anomalous terms containing a square □ or  can be analyzed quite simply. These terms in equation (A.8) also all cancel. We have thereby shown that there are no anomalous terms in the function $F_6(H)$.

REFERENCES

1. A. A. Abrikosov and A. A. Migdal, J. Low Temp. Phys. **3**, 519 (1970).
2. A. M. Tsvelick and P. B. Wigmann, Adv. Phys. **32**, 453 (1983).
3. N. Andrei and K. Furuya, Physics **55**, 331 (1983).
4. Yu. N. Ovchinnikov and A. M. Dyugaev, Zh. Éksp. Teor. Fiz. **115**, 1263 (1999) [JETP **88**, 696 (1999)].
5. W. Felsch, Zeitschrift fur Physik B **29**, 212 (1978).
6. Yu. N. Ovchinnikov, A. M. Dyugaev, P. Fulde, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 184 (1997) [JETP Lett. **66**, 195 (1997)].
7. A. A. Abrikosov, A. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, NJ, 1963).
8. Yu. N. Ovchinnikov and A. M. Dyugaev, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 106 (1999) [JETP Lett. **70**, 111 (1999)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 153–159.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 175–181.
Original Russian Text Copyright © 2000 by Mamaluĭ, Chupis.

Surface Polaritons in a Dielectric at a Boundary with a Metal in Crossed Electric and Magnetic Fields

D. A. Mamaluĭ* and I. E. Chupis**
Verkin Physicotechnical Institute for Low Temperatures,
Ukrainian National Academy of Sciences,
Khar'kov, 310164 Ukraine
**e-mail: mamaluy@ilt.kharkov.ua*
***e-mail: chupis@ilt.kharkov.ua*

Received May 7, 1999

Abstract—The spectrum of surface polaritons in a dielectric at a boundary with an ideal metal or a superconductor in crossed constant electric and magnetic fields is studied. It is shown that the polariton spectrum possesses strong nonreciprocity (polaritons with fixed frequency propagate only in one direction; this is the rectification effect) and depends strongly on the directions of the external fields and their ratios H_0/E_0. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Surface polaritons, i.e., electromagnetic waves on surfaces and interfaces between different media, are being actively studied theoretically and experimentally (see, for example, [1]). Coupled electromagnetic and electric polarization waves in a dielectric are called phonon polaritons. In metals and semiconductors, plasmon polaritons result from the action of an electromagnetic wave on free charge carriers. In magnets, the coupled states of electromagnetic and spin excitations are called magnetic polaritons (see the review in [2]). The linear magnetoelectric effect in magnetic crystals with a definite symmetry introduced its own characteristic features into the spectrum of polariton excitations. Specifically, it leads to nonreciprocity of the spectrum, so that $\omega(-\mathbf{k}) \neq \omega(\mathbf{k})$, where $\mathbf{k}$ is the wave vector [3–5].

It is known that nonreciprocity in the spectrum exists not only in magnets but also in nonmagnetic crystals in the presence of an external magnetic field. The effect of an external magnetic field on surface polaritons has been studied with respect to its action on an electron plasma (magnetoplasmon polaritons), i.e., in metals and semiconductors (see references in [1]).

In nonmagnetic dielectrics, the interaction of the electric polarization $\mathbf{P}$ with a magnetic field $\mathbf{H}$ can be described [6] by a universal scalar magnetoelectric energy of dynamical origin:

$$W_{int} = \frac{V_0}{mc}\mathbf{P} \cdot [\mathbf{\Pi} \times \mathbf{H}].$$

This is the interaction energy of the electric polarization $\mathbf{P}$ with the effective electric field $\mathbf{E}_{eff} = -(1/c)\mathbf{v} \times \mathbf{H}$ arising when a charge moves with velocity $\mathbf{v}$ in a magnetic field (c is the speed of light). The vector $\mathbf{\Pi} = (m/V_0)\mathbf{v}$ is the momentum density, m is the mass of the charge, and V_0 is the volume of the unit cell. In our case, the electric polarization $\mathbf{P}$ consists of electronic and ionic parts. The indicated dynamical energy is a scalar, i.e., it is present in crystals with any symmetry.

Taking account of this dynamical magnetoelectric energy in the analysis of polaritons in a dielectric at a boundary with vacuum in the presence of a constant electric (magnetic) field has shown the existence of interesting effects. Thus, in the presence of a constant electric field oriented in a direction normal to the surface of the dielectric, the spectrum of phonon polaritons is different for opposite directions of this field (or in 180° domains of a ferroelectric), and birefringence is possible in certain frequency ranges [7]. A magnetic field makes possible the conversion of virtual phonon polaritons into real polaritons [8]. However, these new effects are weak because the relativistic magnetostatic energy W_{int} is small.

Nonetheless, the effects caused by the influence of the dynamical magnetoelectric interaction are no longer weak at the contact surface of a dielectric with an ideal metal or superconductor.

It is known that surface polaritons do not exist in a dielectric at a boundary with a metal because of "metallic quenching" (the requirement that the tangential components of the electric field be zero) [1]. As shown in [9], in the presence of a constant electric field the indicated dynamical magnetoelectric energy W_{int} lifts this forbiddenness. It turns out that at a contact with an ideal metal the penetration depth of phonon polaritons in the dielectric is inversely proportional to W_{int}. Since the energy W_{int} is proportional to the constant electric field, the penetration depth is all the smaller, the stronger this field is. While W_{int} substantially determines the penetration depth of polaritons in a dielectric, the dispersion relation for polaritons is essentially independent of the energy W_{int}, because this energy is so small. The dispersion law for surface polaritons has the same analytic form as for volume polaritons. The frequency ranges of phonon polaritons in an electric field oriented in the direction of the outer and inner normals to the interface are not small and differ substantially, so that the switching of the direction of the electric field indicates that surface polaritons with this frequency are "switched on" or "switched off" [9]. This large effect arises because of the contact of the dielectric with an ideal metal even though the magnetoelectric interaction is weak. However, if a constant magnetic field **H** is applied in the plane of contact of the dielectric with an ideal metal or superconductor, then the existence of surface polaritons is also found to be possible, and their frequency ranges (with wave vector $\mathbf{k} \perp \mathbf{H}$) are substantially different not only for opposite directions of the magnetic field but also for opposite directions of wave propagation. The spectrum exhibits strong nonreciprocity (in contrast to the ordinarily weak nonreciprocity [1])—phonon polaritons with a fixed frequency propagate only in one direction, i.e., rectification occurs.

In the present paper, phonon polaritons of a dielectric at its boundary with an ideal metal or superconductor in the presence of a constant magnetic field applied in the contact plane ($\mathbf{H}_0 \parallel \mathbf{y}$) together with a constant electric field oriented in a direction normal to the contact boundary ($\mathbf{E}_0 \parallel \mathbf{z}$), i.e., in crossed electric and magnetic fields, are studied. The spectrum obtained depends strongly not only on the orientations of the electric and magnetic fields but also on their ratio H_0/E_0. The spectrum is characterized by strong nonreciprocity and consists of three well-separated frequency ranges. Breaks appear in the dispersion curves, and a gap appears in the lower nonactivational branch. The widths of the gap, the breaks, and the frequency ranges characterized by strong nonreciprocity depend on the ratio H_0/E_0 and can be regulated over wide limits.

2. LINEAR RESPONSE OF A DIELECTRIC IN CROSSED ELECTRIC AND MAGNETIC FIELDS

Let us consider a dielectric occupying the half-space $z > 0$ and bounding a metal ($z < 0$). For definiteness, we assume the dielectric to be uniaxial (z is the "easy axis"), though the results obtained below are general.

We write the energy density W of the dielectric in the form

$$\begin{aligned} W = {} & \frac{c_1}{2}P_z^2 + \frac{c_2}{2}(P_x^2 + P_y^2) + \frac{\Pi^2}{2\rho} \\ & - \mathbf{E}\cdot\mathbf{P} + \xi\mathbf{P}\cdot[\mathbf{\Pi}\times\mathbf{H}], \end{aligned} \tag{1}$$

where **P** is the electric polarization, **Π** is the momentum density, $\mathbf{H} = \mathbf{H}_0 + \mathbf{h}$, $\mathbf{E} = \mathbf{E}_0 + \mathbf{e}$, and **e** and **h** are the ac electric and magnetic fields. The constant magnetic field $\mathbf{H}_0$ and the constant electric field $\mathbf{E}_0$ are applied along the y and z axes, respectively. The term in the expression for W with the coefficient ξ is the scalar magnetoelectric energy, mentioned above, of the dynamical origin, $\xi = V_0/mc$ and $\rho = m/V_0$. Generally speaking, the polarization **P** contains both ionic and electronic contributions. In the infrared region of the spectrum, especially near the characteristic ionic frequencies, the ionic polarizability prevails, the ions make the largest contribution to the dynamical magnetoelectric energy, and therefore m and **Π** are the ion mass and momentum, respectively. In the optical region of the spectrum, where the electronic polarizability is much larger than the ionic polarizability, m is the electron mass and **Π** is the electron momentum.

If the dielectric is not a ferroelectric, then $c_1 > 0$ and the equilibrium value of the electric polarization $\mathbf{P}_0$ in a constant electric field $\mathbf{E}_0 \parallel \mathbf{z}$ is, as follows from equation (1),

$$P_0 = P_{0z} = \frac{E_0}{c_1 - \xi^2\rho H_0^2}.$$

For a ferroelectric, the constant $c_1 < 0$, the anharmonic term $\delta P_z^4/4$ must be taken into account in the expression (1), and the spontaneous polarization vector is

$$P_0 = P_{0z} = \pm\sqrt{-\frac{c_1 - \xi^2\rho H_0^2}{\delta}}.$$

If the dielectric is a ferroelectric, we assume that there is no constant electric field. A characteristic feature of the crossed-field case is the existence of an equilibrium value of the momentum density

$$\Pi_0 = \Pi_{0x} = -\rho\xi H_{0y}P_{0z} = -\frac{1}{c}H_0P_0,$$

which means that toroidal ordering with moment density $\mathbf{T} \propto \mathbf{\Pi}_0$ is induced in the dielectric by the external fields.

The linear response of a nonmagnetic dielectric with magnetic permeability $\mu = 1$ in the field of an electromagnetic wave in the absence of damping and neglecting spatial dispersion in the case of crossed fields can be obtained similarly to [9]. In so doing, the existence of an equilibrium value Π_0 must be taken into account. Then, the nonzero components of the electric χ_{ik}^e and magnetoelectric $\chi_{ik}^{em} = \partial P_i/\partial h_k = (\chi_{ki}^{me})^*$ susceptibilities are as follows:

$$\varepsilon_1 = 1 + 4\pi\chi_{xx}^e = 1 + \frac{4\pi\bar{\omega}_0^2(\omega_e^2 - \omega^2)}{(\omega_1^2 - \omega^2)(\omega_2^2 - \omega^2)},$$

$$\varepsilon_2 = 1 + 4\pi\chi_{yy}^e = 1 + \frac{4\pi\bar{\omega}_0^2}{(\omega_0^2 - \omega^2)},$$

$$\begin{aligned}
\varepsilon_3 &= 1 + 4\pi\chi^{e}_{zz} = \frac{(\Omega_1^2-\omega^2)(\Omega_2^2-\omega^2)}{(\omega_1^2-\omega^2)(\omega_2^2-\omega^2)},\\
\varepsilon' &= 4\pi i\chi^{e}_{xz} = -4\pi i\chi^{e}_{zx} = \frac{8\pi\bar{\omega}_0^2\omega g H_0}{(\omega_1^2-\omega^2)(\omega_2^2-\omega^2)},\\
\gamma_1 &= 4\pi i\chi^{em}_{xy} = \frac{4\pi\omega g P_0(\omega_e^2+3g^2H_0^2-\omega^2)}{(\omega_1^2-\omega^2)(\omega_2^2-\omega^2)},\\
\gamma_2 &= 4\pi i\chi^{em}_{yx} = \frac{4\pi\omega g P_0}{(\omega^2-\omega_0^2)},\\
\gamma_3 &= 4\pi\chi^{em}_{zy} = \frac{8\pi g^2 P_0 H_0(\omega_0^2-g^2H_0^2)}{(\omega_1^2-\omega^2)(\omega_2^2-\omega^2)},\\
\gamma_4 &= 4\pi\chi^{em}_{yz} = \frac{4\pi g^2 P_0 H_0}{(\omega^2-\omega_0^2)},
\end{aligned} \tag{2}$$

where

$$\begin{aligned}
\omega_{1,2}^2 &= \frac{1}{2}[\omega_0^2+\omega_e^2+2g^2H_0^2\\
&\mp\sqrt{(\omega_0^2-\omega_e^2)^2+8g^2H_0^2(\omega_0^2+\omega_e^2)}\,],\\
\Omega_{1,2}^2 &= \frac{1}{2}[\omega_0^2+\Omega_e^2+2g^2H_0^2\\
&\mp\sqrt{(\omega_0^2-\Omega_e^2)^2+8g^2H_0^2(\omega_0^2+\Omega_e^2)}\,],\\
g &= \frac{e}{mc},\quad \bar{\omega}_0^2 = \frac{e^2}{mV_0},\quad \omega_0^2 = \bar{\omega}_0^2 c_2,\\
\omega_e^2 &= \bar{\omega}_0^2 c_1,\quad \Omega_e^2 = \omega_e^2 + 4\pi\bar{\omega}_0^2.
\end{aligned} \tag{3}$$

Here ω_e is the excitation frequency of P_z (for a ferroelectric $\omega_e^2 = -2\bar{\omega}_0^2 c_1$), and ω_0 is the excitation frequency for the transverse components P_x and P_y of the polarization in the absence of a magnetic field. For a uniaxial dielectric, which is the case considered here, $\omega_0 > \omega_e$, and a natural condition is $\Omega_e > \omega_0$, i.e., $\chi^{-1}_{zz}(0) + 4\pi > \chi^{-1}_{xx}(0)$, where $\chi_{zz}(0) = c_1^{-1}$ and $\chi_{xx}(0) = c_2^{-1}$ are the static dielectric susceptibilities.

3. SPECTRUM OF SURFACE POLARITONS

3.1. Dispersion Law. Penetration Depth

We seek the solution of Maxwell's equations for waves propagating along the boundary with a metal in the direction of the x ($k = k_x$) axis in the form

$$\mathbf{e}, \mathbf{h} \propto \exp[i(k_x x - \omega t) - k_0 z], \tag{4}$$

where k_0^{-1} is the penetration depth of surface phonon polaritons in the dielectric.

The following conditions must be satisfied at the interface with an ideally conducting metal:

$$b_z = \tilde{b}_z, \quad \mathbf{h}_t = \tilde{\mathbf{h}}_t, \quad d_z = \tilde{d}_z, \quad \mathbf{e}_t = 0, \tag{5}$$

where **d** is the electric induction, **b** is the magnetic induction, and **e** and **h** are, respectively, the electric and magnetic field intensities. In equation (5) the letters with an overtilde designate quantities referring to the metal and the quantities labeled with an index t are the tangential components of the fields. For a superconductor, in equation (5) $b_z = 0$, but in the case at hand this equality does not introduce additional conditions in Maxwell's equations, so that all results obtained also hold for a dielectric in contact with a superconductor.

Taking account of the equalities (2), (4), and (5), we can write Maxwell's equations in the form

$$\begin{aligned}
e_z\frac{\omega}{c}\varepsilon' + h_y\left(k_0+\frac{\omega}{c}\gamma_1\right) &= 0,\\
e_z\left(k_x+\frac{\omega}{c}\gamma_3\right) + h_y\frac{\omega}{c} &= 0,\\
e_z\frac{\omega}{c}\varepsilon_3 + h_y\left(k_x+\frac{\omega}{c}\gamma_3\right) &= 0.
\end{aligned} \tag{6}$$

As one can see from equations (6), only the components e_z and h_y of the electromagnetic field are different from zero. The dispersion law and the penetration depth of the electromagnetic field are as follows:

$$k_x = \frac{\omega}{c}(\pm\sqrt{\varepsilon_3}-\gamma_3), \quad k_0 = \frac{\omega}{c}\left(\pm\frac{\varepsilon'}{\sqrt{\varepsilon_3}}-\gamma_1\right). \tag{7}$$

Analysis of equations (7), the condition for k_0 to be positive, and taking account of the frequency dependences (2) makes it necessary to distinguish the four possible cases of orientation of the external fields $H_0 = H_y$ and $E_0 = E_z$. In what follows, it is convenient for us to use instead of E_0 the value P_0 proportional to it and the following notations for possible orientations of the external fields:

I $gP_0 > 0, \quad gH_0 > 0;$

II $gP_0 > 0, \quad gH_0 < 0;$

III $gP_0 < 0, \quad gH_0 > 0;$

IV $gP_0 < 0, \quad gH_0 < 0.$

In addition, it should be kept in mind that the gyromagnetic ratio $g = e/mc$ is positive for ionic excitations and negative for electronic excitations.

The dispersion laws for surface phonon polaritons for various orientations of the external fields are shown in Fig. 1. Figure 1a corresponds to the case I, and its

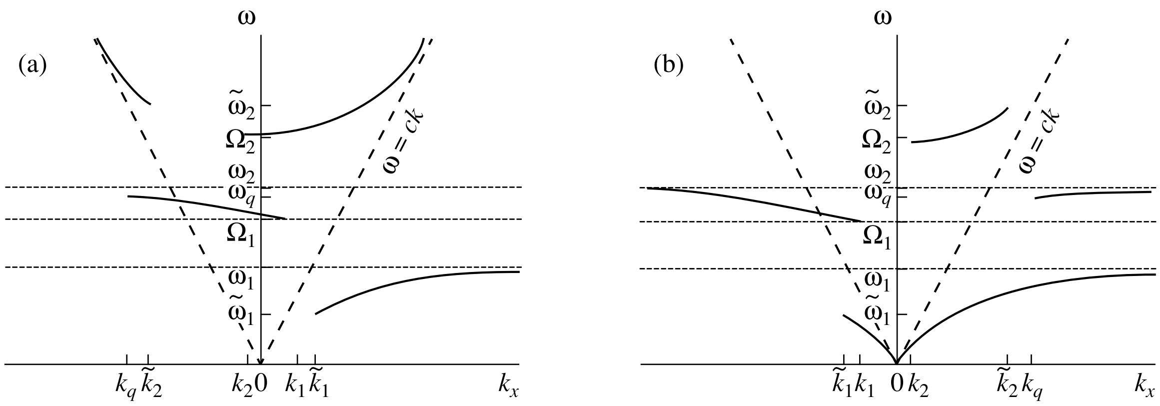


Fig. 1. Spectra of surface polaritons in the cases I ($gP_0 > 0$, $gH_0 > 0$) (a) and III ($gP_0 < 0$, $gH_0 > 0$) (b). In the cases II ($gP_0 > 0$, $gH_0 < 0$) and IV ($gP_0 < 0$, $gH_0 < 0$) the spectra are mirror images with respect to the frequency axis of the spectra in the cases I and III, respectively.

mirror image with respect to the frequency axis corresponds to the case II. Figure 1b refers to the case III, and its mirror image with respect to the frequency axis corresponds to the case IV. The spectrum of surface polaritons consists of three well-separated frequency ranges: $\omega < \omega_1 \approx \omega_e$, $\Omega_1 < \omega < \omega_2$ (Ω_1 and ω_2 near ω_0) and $\omega > \Omega_2 \approx \Omega_e$. In the IR range, for ionic excitations ($g > 0$) $\omega_e \sim 10^{13}$ rad/s. In the optical range, where electrons make the main contribution to the electric polarization ($g < 0$), these frequencies are at least an order of magnitude higher. For certain values of the frequencies ω_q and $\tilde{\omega}_{1,2}$ and the wave vectors $k_q = k(\omega_q)$, $\tilde{k}_{1,2} = k(\tilde{\omega}_{1,2})$, and $k_{1,2} = k(\Omega_{1,2})$ we employed the smallness of the ratio of the cyclotron frequency gH_0 to the optical frequency $\omega_{1,2}$: $gH_0/\omega_{1,2} \ll 1$. We have approximately

$$\omega_q^2 = \frac{\omega_2^2 q^2 + \Omega_1^2}{1 + q^2}, \quad k_q = -\frac{\omega_q}{c}\frac{(1 + 2q^2)}{q}\sqrt{\frac{\Omega_e^2 - \omega_0^2}{\omega_0^2 - \omega_e^2}},$$

$$q = \frac{2H_0}{P_0}\frac{\bar{\omega}_0^2}{\sqrt{(\omega_0^2 - \omega_e^2)(\Omega_e^2 - \omega_0^2)}},$$

$$\tilde{\omega}_{1,2}^2 = \omega_e^2 + 2\pi\bar{\omega}_0^2\left(1 \mp \sqrt{1 + \left(\frac{H_0}{\pi P_0}\right)^2}\right), \tag{8}$$

$$k_1 = \frac{\omega_0 P_0 (\Omega_e^2 - \omega_0^2)}{2cH_0 \bar{\omega}_0^2}, \quad k_2 = \frac{2\Omega_e}{c}\frac{g^2 P_0 H_0 \omega_0^2}{(\Omega_e^2 - \omega_0^2)\bar{\omega}_0^2}.$$

The expressions for $\tilde{k}_{1,2} = k(\tilde{\omega}_{1,2})$ are too complicated to present here. Then can be calculated from equation (7) using equations (2), (3), and (8). The sequence of the arrangement of $k_{1,2}$, k_q, and $\tilde{k}_{1,2}$ is not necessarily as shown in the figures (though $|\tilde{k}_2| > |k_2|$); it depends on the ratio (H_0/P_0) of the magnetic and electric fields.

3.2. Lower Branch of the Spectrum

The lower branch of the spectrum lies in the frequency range $\omega < \omega_1$, where, because the ratio gH_0/ω_0 is small, the approximate value of ω_1, as follows from equation (3), is

$$\omega_1 \approx \omega_e - \frac{1}{2\omega_e}\frac{\omega_0^2 + 3\omega_0^2}{\omega_0^2 - \omega_e^2} g^2 H_0^2.$$

As follows from equation (8), the frequency $\tilde{\omega}_1$ is real, if

$$\left|\frac{H_0}{P_0}\right| < \nu = \frac{\sqrt{4\pi\chi + 1}}{2\chi}, \quad \chi = \chi_{zz}(0). \tag{9}$$

Far from the ferroelectric transition temperature, the static dielectric susceptibility χ ordinarily is essentially 1, so that $\nu \sim 1$. If $|H_0/P_0| > \nu$, then the lower polariton branch exists in the range $[0, \omega_1]$, characterized by strong nonreciprocity: In the cases I and III this mode exists only for $k_x > 0$, and in the cases II and IV it exists only for $k_x < 0$ (see Fig. 1), i.e., rectification of the polaritons occurs: They propagate only in one direction. The lower polariton branch also has a similar form in the absence of an electric field ($P_0 = 0$). The variation of the lower polariton mode in an electric field is a threshold process, since $\tilde{\omega}_1 \propto \sqrt{E_0 - E_\nu}$ arises only in a field $|E_0| > E_\nu = |H_0|\chi^{-1}\nu^{-1}$. The character of the changes in the spectrum for $E_0 > E_\nu$ depends on the direction of the electric field. If $gP_0 > 0$ (the cases I and II), then a gap $\tilde{\omega}_1$, whose width increases with the electric field, appears in the lower branch, while the allowed frequency range $[\tilde{\omega}_1, \omega_1]$ decreases. However, if $gP_0 < 0$ (the cases III and IV), then the electric field "symmetrizes" the lower branch, adjusting the frequency interval for $k_x < 0$ (III) and $k_x > 0$ (IV). In the absence of a magnetic field ($H_0 = 0$), the lower branch is symmetric, the spectrum is reciprocal (III and IV) or there are no surface polaritons (I and II) [9]. A magnetic field in the cases I and II creates a nonreciprocal gap mode at locations where there was no mode for $H_0 = 0$, and in the cases III and IV it “cuts off” the symmetric spectrum earlier, creating frequency ranges $[\tilde{\omega}_1, \omega_1]$ where waves propagate in only one direction. The magnitude of the frequency interval $[\tilde{\omega}_1, \omega_1]$ is

$$\omega_1^2 - \tilde{\omega}_1^2 = 2\pi\bar{\omega}_0^2\left(\sqrt{1+\left(\frac{H_0}{\pi P_0}\right)^2} - 1\right). \quad (10)$$

As is evident from equation (10), it depends on the ratio of the electric and magnetic fields and therefore can vary over wide limits: from zero (for $H_0/P_0 = 0$) up to ω_1 (for $|H_0/P_0| \geq \nu$).

3.3. Radiation Branch

The frequency range $[\Omega_1, \omega_2]$ is shown in Fig. 1 as being disproportionately large. Actually, as follows from equation (3), because the ratio $gH_0/\omega_{1,2}$ is small the frequencies Ω_1 and ω_2 lie near ω_0, and the range $[\Omega_1, \omega_2]$ is small:

$$\frac{\omega_2 - \Omega_1}{\omega_0} \approx \frac{8\pi g^2 H_0^2 \bar{\omega}_0^2}{(\Omega_e^2 - \omega_0^2)(\omega_0^2 - \omega_e^2)} \sim \left(\frac{gH_0}{\omega_0}\right)^2. \quad (11)$$

This frequency range arises only in the presence of a magnetic field, since for $H_0 = 0$ the necessary condition $\varepsilon_3 > 0$ is not satisfied (see equation (7)). It lies in a previously forbidden (for $H_0 = 0$) zone, where strong damping occurs.

However, this frequency branch is interesting because, in contrast to the lower $[0, \omega_1]$ and upper $[\Omega_2, \infty]$ branches, it is a radiation branch: It can intersect with the electromagnetic mode $\omega = ck$. This means that, in principle, resonance excitation of such a polariton mode of an electromagnetic wave is possible. In the absence of an electric field ($P_0 = 0$), the radiation point always exists, and its frequency is $\omega_r^2 = \omega_0^2 - g^2 H_0^2$. In crossed fields, the point of intersection of the electromagnetic and polariton modes (the radiation point) can occur for both positive and negative values of k_x depending on the ratio P_0/H_0 of the fields. Thus, in the case I the position of the radiation point ω_r depends on the relation $\sqrt{\varepsilon_3} + \gamma_3 = -k_x/|k_x|$, which, for example, for the frequency ω_0 (i.e., when the radiation point $\omega_r = \omega_0$) means the following condition for P_0/H_0:

$$\frac{P_0}{H_0} \approx \frac{\sqrt{3\omega_0^2 + \omega_e^2}}{8\pi\omega_0^2}\left(\sqrt{3\omega_0^2 + \Omega_e^2} \pm \sqrt{3\omega_0^2 + \omega_e^2}\right). \quad (12)$$

The plus sign in equation (12) corresponds to the radiation point lying on the right-hand side ($k_x > 0$) and a minus sign corresponds to the left-hand side ($k_x < 0$). The radiation point can be moved into or out of the range $[\Omega_1, \omega_2]$ by varying the ratio of the electric and magnetic fields.

As one can see from Fig. 1 , reversing the direction of the magnetic field is equivalent to reversing the direction of wave propagation. In the absence of an electric field ($P_0 = 0$, $\omega_q = \omega_2$, and $k_1 = 0$), the polariton mode under study exists in the cases I and III only on the left-hand side ($k_x < 0$), and in the cases II and IV it exists only on the right-hand side ($k_x > 0$). The spectrum is characterized by strong nonreciprocity: polaritons propagate only in one direction (rectification effect). In the cases I and II (Fig. 1a), an electric field deforms the polariton branch, leaving it completely nonreciprocal. In the cases III and IV (Fig. 1b) it symmetrizes the spectrum, additionally adjusting the ranges $[\omega_q, \omega_2]$ and, therefore, decreasing the frequency range $[\Omega_1, \omega_q]$ where rectification is possible. The position of the frequency ω_q in the range $[\Omega_1, \omega_2]$ depends on the value of q (8), i.e., on the ratio H_0/P_0 of the fields.

Near the frequency Ω_1, i.e., a zero of the permittivity ε_3, as follows from equations (6) and (7), the electric field $h_y = \pm\sqrt{\varepsilon_3}\,e_z$ predominates in the wave. Conversely, near the characteristic frequencies $\omega_{1,2}$ the wave is predominantly magnetic, and for $\omega = \omega_q$ the ratio of the fields in the wave depends on the ratio $h_y/e_z = (2H_0/P_0)(c_2 - c_1)^{-1}$ of the constant fields.

3.4. Upper Branch of the Spectrum

The upper branch of the polariton spectrum,

$$\omega > \Omega_2 \approx \Omega_e + \frac{1}{\Omega_e}\frac{\omega_0^2 + 3\Omega_e^2}{\Omega_e^2 - \omega_0^2} g^2 H_0^2,$$

in a magnetic field and in the absence of an electric field is characterized by strong nonreciprocity and exists only for $k_x > 0$ in the cases I and III and for $k_x < 0$ in the cases II and IV, and $k_2 = 0$, $\tilde{\omega}_2 \longrightarrow \infty$. The electric field limits the spectrum (cases III, IV, Fig. 1b) or symmetrizes it (cases I, II, Fig. 1a). When only an electric field ($H_0 = 0$, $\tilde{\omega}_2 = \Omega_2 = \Omega_e$) is present, the polariton spectrum is symmetric and exists only in the cases I and II. The addition of a magnetic field covers the spectrum: in the cases I and II (Fig. 1a) it cuts out sections of the curve in the range $[\Omega_2, \tilde{\omega}_2]$ (the width of this range is determined by the same expression as in equation (10)) and adds these same sections in the cases III and IV (Fig. 1b).

Table

	I	II	III	IV
$[0, \tilde{\omega}_1]$	0	0	1	1
$[\tilde{\omega}_1, \omega_1]$	1	0	1	0
$[\Omega_2, \tilde{\omega}_2]$	1	0	1	0
$[\tilde{\omega}_2, \infty]$	1	1	0	0

4. DISCUSSION

We note first the following properties of the spectra presented in Fig. 1.

1. The substitution $H_0 \longrightarrow -H_0$ is equivalent to $k_x \longrightarrow -k_x$, so that the plots for the cases I and II, III and IV are mirror images of one another.

2. In the cases differing by the signs of the electric and magnetic fields simultaneously (I and IV, II and III), the spectra mutually complement one another up to a complete symmetric spectrum. For example, the spectrum of the upper branch in the case IV is a “cutout” piece of the spectrum of this branch in the case I and so on. For a fixed orientation of the external fields the transitions I $\longleftrightarrow$ IV and II $\longleftrightarrow$ III correspond to a transition from the IR region of the spectrum, where the ionic contribution to the polarizability predominates and therefore $g > 0$, into the optical region where the electronic polarizability predominates ($g < 0$). Here, one should bear in mind the change in the order of the characteristic frequencies.

The penetration depth of the electromagnetic field in the dielectric, k_0^{-1} (see equations (7) and (2)), can be finite only in the presence of at least one field (electric or magnetic). If there is no field, then $k_0 = 0$, i.e., the waves are volume waves. In our case of crossed fields, the electric and magnetic fields make additive contributions to k_0, since $\varepsilon' \propto H_0$ and $\gamma_1 \propto P_0$. If these contributions have the same sign, then both fields together displace the electromagnetic wave out of the dielectric, converting it into a surface wave. The penetration depth decreases with increasing P_0 and H_0. However, if the signs of the terms in k_0 (7) are different, then the electric and magnetic fields act oppositely and k_0 can vanish, i.e., the wave becomes a volume wave. Such a situation ($k_0^{-1} \longrightarrow \infty$) occurs for frequencies ω_q, $\tilde{\omega}_1$, and $\tilde{\omega}_2$ in Fig. 1. Even though the penetration depth of the field is inversely proportional to the small magnetoelectric interaction energy W_{int}, the penetration depth can be small because of the frequency dispersion of ε_3, ε', and γ_1 (7). Thus, the penetration depth of an electromagnetic wave in the dielectric approaches zero near the poles of k_0 (7), i.e., near the frequencies ω_1, Ω_1, ω_2, and Ω_2.

Although the dynamic magnetoelectric interaction W_{int} is weak, it produces effects which are not small for phonon polaritons of a dielectric at the surface of contact with an ideal metal or a superconductor in external electric and magnetic fields. If only an electric field is present, the spectrum of surface phonon polaritons has only one (lower or upper) branch, depending on the direction of the external electric field $\mathbf{E}_0$, so that when the direction of $\mathbf{E}_0$ is switched, the polaritons with a given frequency are switched on (off) [9]. If only a magnetic field is applied to the system, then the polariton spectrum possesses three branches characterized by strong nonreciprocity: a polariton on any of these branches propagates only in one direction.

When electric and magnetic fields are present together, the spectrum consists of three frequency ranges and depends not only on the direction of the fields but also on their ratio H_0/P_0.

When electric and magnetic fields are present at the same time, breaks appear in the frequency branches and a gap $\tilde{\omega}_1$ appears in the previously nonactivational lower branch. The spectrum is characterized by strong nonreciprocity in the wide frequency ranges $[\tilde{\omega}_1, \omega_1]$ and $[\Omega_2, \tilde{\omega}_2]$. The width of these ranges depends on the ratio H_0/P_0. For example, for the tetragonal crystal MgF_2 we have $\varepsilon_{zz}(0) \approx 5$ [11] (i.e., $c_1 \approx \pi$), while for fields $H_0 = 1$ T and $E_0 = 1$ mV/m the relative width of the gap in the upper branch is $(\tilde{\omega}_2 - \Omega_2)/\Omega_2 \approx 2.6$. Since the widths of the gap and the breaks in the spectrum and the nonreciprocity ranges depend on the ratio H_0/P_0, they can be regulated by varying the intensities of the fields. This property could be helpful for producing filters, whose frequency can be tuned by means of external fields, and rectifiers for surface electromagnetic waves.

The middle branch of the polariton spectrum in the range $[\Omega_1, \omega_2]$ is interesting in that it can interact resonantly with an electromagnetic wave, i.e., it is a radiation branch, a fact that is very important for studying a polariton spectrum in practice. The position of the “radiation point” (resonance point) in crossed fields can be controlled by these external fields, since it depends on their ratio P_0/H_0 (see (12)). Unfortunately, however, this radiation mode lies in the region of strong absorption.

We note one other possibility for using the strong dependence of the polariton spectrum on the external fields that arises in crossed electric and magnetic fields. In the table the number 1 (0) denotes the existence (absence) of surface polaritons for various orientations of the fields I–IV in wide frequency ranges. The dependence of the spectrum on the direction of the external fields means that the polariton modes in 180° domains of ferroelectrics and magnets are different. Spectral investigations of surface polaritons in the system under study make it possible to determine the relative orientation of the electric polarization and the magnetization. It is evident from the table that if the two fixed frequencies are chosen from the pair of ranges $[0, \tilde{\omega}_1]$ and $[\tilde{\omega}_1, \omega_1]$, or $[\tilde{\omega}_1, \omega_1]$ and $[\tilde{\omega}_2, \infty]$, or $[0, \tilde{\omega}_1]$ and $[\Omega_2, \tilde{\omega}_2]$, or $[\Omega_2, \tilde{\omega}_2]$ and $[\tilde{\omega}_2, \infty]$, then one of the four possible methods of orientation of the polarization and magnetization can be determined on the basis of the presence or absence of a surface wave at these two frequencies. Indeed, in this case a binary code, uniquely determining this orientation, corresponds to each of the orientations I–IV (see table). This property could be helpful in computer technology, for example, when ferromagnets are used as elements with binary (electric and magnetic) memory.

REFERENCES

1. *Surface Polaritons*, Ed. by V. M. Agranovich and D. L. Mills (Nauka, Moscow, 1985; North-Holland, Amsterdam, 1982).
2. M. I. Kaganov, N. B. Pustyl’nik, and T. I. Shalaeva, Usp. Fiz. Nauk **167**, 191 (1997).
3. V. N. Lyubimov, Dokl. Akad. Nauk SSSR **181**, 858 (1968) [Sov. Phys. Dokl. **13**, 739 (1969)].
4. V. A. Markelov, M. A. Novikov, and A. A. Turkhin, Pis’ma Zh. Éksp. Teor. Fiz. **25**, 404 (1977) [JETP Lett. **25**, 378 (1977)].
5. V. D. Buchel’nikov and V. G. Shavrov, Zh. Éksp. Teor. Fiz. **109**, 706 (1996) [JETP **82**, 380 (1996)].
6. I. E. Chupis, Ferroelectrics **204**, 173 (1997).
7. I. E. Chupis and D. A. Mamaluĭ, Fiz. Nizk. Temp. **24**, 1010 (1998) [Low Temp. Phys. **24**, 762 (1998)].
8. I. E. Chupis and N. Ya. Alexandrova, J. Korean Phys. Soc. **32**, 51 134 (1998).
9. I. E. Chupis and D. A. Mamaluĭ, Pis’ma Zh. Éksp. Teor. Fiz. **68**, 876 (1998) [JETP Lett. **68**, 922 (1998)].
10. I. E. Chupis and D. A. Mamaluĭ, Fiz. Nizk. Temp. **25**, 1112 (1999) [Low Temp. Phys. **25**, 833 (1999)].
11. V. V. Bryskin, D. N. Mirlin, and I. I. Reshina, Fiz. Tver. Tela **15**, 1118 (1973) [Sov. Phys. Solid State **15**, 760 (1973)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 160–172.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 182–195.
Original Russian Text Copyright © 2000 by Agafonov, Manykin.

SOLIDS

Insulator–Superconductor–Metal Transitions Induced by Spin Fluctuations in the Paramagnetic Phase of Doped Insulators

A. I. Agafonov* and É. A. Manykin**
Institute of Superconductivity and Solid-State Physics,
Kurchatov Institute of Atomic Energy (All-Russia Scientific Center), Moscow, 123182 Russia
**e-mail: aai@isssph.kiae.ru*
***e-mail: edmany@isssph.kiae.ru*

Received May 31, 1999

Abstract—The band structure of cuprates as a doped 2*D* insulator is modeled assuming that the excess charge carriers are associated with the corresponding substitution atoms, and the phase diagram of the paramagnetic states as a function of the degree x of doping at zero temperature is studied. The Hamiltonian contains electronic correlations on impurity orbitals and hybridization between them and the initial band states of the insulator. It is shown that the change in the electronic structure of a doped compound includes the formation of impurity bands of distributed and localized electronic states in the initial insulator gap. It is established that in the case of one excess electron per substitution atom the spin fluctuations (1) give rise to an insulator state of the doped compound for $x < x_{thr,\,1}$, (2) lead to a superconducting state for $x_{thr,\,1} < x < x_{thr,\,2}$, and (3) decay as $x > x_{thr,\,2}$ increases further, and the doped compound transforms into a paramagnetic state of a "poor" metal with a high density of localized electronic states at the Fermi level. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Phase transitions in doped cuprates are now being intensively studied. As a rule, the initial compounds possess an antiferromagnetic insulator state. It is known that the magnetic state vanishes for low concentrations of the substitution impurity [1, 2]. As the impurity concentration increases further, insulator–superconductor–metal phase transitions occur in the paramagnetic phase. However, this paramagnetic phase exhibits unusual spin-dynamic properties, which can contain substantial information about the orgin of the normal state and about the mechanism of superconductivity [2–9].

The construction of the theory of normal and superconducting states is largely based on an assumption about the nature of the excess charge carriers introduced into the electronic system by the impurity ensemble. At present, most works assume that these excess charge carriers (1) are free, (2) leave the intermediate layers of the cuprates, which are regarded as reservoirs for charge carriers, and (3) migrate in the CuO_2 plane [1, 2, 6, 7, 10]. The most common direction of investigations of HTSCs is based on this and consists in the fact that HTSC is due to processes occurring in the regular 2*D* lattice of the CuO_2 planes with variable filling [1–8, 11–14]. To describe these processes, models of strongly correlated electrons, based on the Hubbard model in the site representation, are used.

After it was found that the mobility of charge carriers in cuprates is comparable to the Mott–Ioffe–Regel' criterion, it became obvious that localization effects play an important role in these systems [15, 16]. As a result of this, a different direction arose in the investigation of HTSC, in which HTSC materials are regarded as doped semiconductors [15, 17–20]. This approach is based on the assumption that the excess charge carriers are initially associated with the corresponding impurity atoms. An important circumstance for such an analysis is that in the paramagnetic insulator phase, which is followed immediately by the superconducting phase, the electronic properties (specifically, the conductivity) of doped cuprates are determined by activated hops of charge carriers between localized states [17, 21–24], which is typical for doped semiconductors.

The ARPES data for $Bi_2Sr_2CaCu_2O_{8+y}$ showed that in the normal and in the superconducting states the electronic structure near the Fermi level is formed by two types of bands (see [25] and the references cited there). The first type corresponds to narrow bands lying in the range ±40 meV at the Fermi level and possessing a well-determined electronic dispersion, i.e., these bands correspond to distributed states. Bands of the second type have no dispersion in a wide range of values of the wave vectors. Smoothing of the dispersion dependence in certain regions of the spectrum can be obtained on the basis of Emery's model [2, 13, 25]. However, considering the properties of cuprates in the paramagnetic insulator phase, it can be inferred [26]

1063-7761/00/9001-0160$20.00 © 2000 MAIK "Nauka/Interperiodica"

that these bands belong to localized states, which coexist with the distributed states near the Fermi level and contribute to the formation of the superconducting state.

At present there is no single theory of HTSC. This is due primarily to the fact that the theory must describe adequately all unusual properties of cuprates in the insulator, metallic, and superconducting phases. For this reason, the approach which we have developed provides one possibility for describing the insulator, superconducting, and metallic states.

2. MODEL

The initial cuprates belong to systems of strongly correlated electrons and are insulators, in which primarily the O 2*p* and Cu 3*d* orbitals form an important part of the electronic structure near the insulator gap, due to charge transfer [2]. The unit cell of the compound La_2CuO_4 contains a single CuO_2 plane, lying between two LaO planes. When a substitution impurity is introduced into $La_{2-x}Sr_xCuO_4$, the La^{3+} atoms randomly replace the Sr^{2+} atoms. Both valence electrons of Sr go to the formation of valence bonds, and therefore it can be expected that a singly filled impurity orbital of acceptor type will arise near the cuprate plane [17, 19, 27].

Since the basis of initial band states is determined by the CuO_2 layer, which does not change on doping, the interaction of the impurity orbitals with the band states of the CuO_2 plane can be described by the hybridization between them. Here, for simplicity, we confine our attention to single-band approximation (valence band) for the density of states in the CuO_2 plane. Taking account of the strong anisotropy of the physical characteristics of cuprates, the Hamiltonian for one structural cell, not coupled with other cells by charge transfer along the $c(z)$ axis, can be represented in the form [27]

$$\begin{gathered} H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} a^{+}_{\mathbf{k}\sigma} a_{\mathbf{k}\sigma} + \sum_{j\sigma} \varepsilon_0 a^{+}_{j\sigma} a_{j\sigma} \\ + \frac{1}{2} \sum_{j\sigma} U n_{j\sigma} n_{j,-\sigma} + \sum_{j,\mathbf{k}\sigma} \{ V_{\mathbf{k}j} a^{+}_{\mathbf{k}\sigma} a_{j\sigma} + \text{h.c.} \}, \end{gathered} \tag{1}$$

where $a_{\mathbf{k}\sigma}$ and $a_{j\sigma}$ are the standard annihilation operators for the initial band state of the insulator and for the impurity orbital with number *j*, respectively; $\sigma = \pm 1/2$ is the spin index; $\mathbf{k}$ is the 2*D* wave vector of an electron in a band state with energy $\varepsilon_{\mathbf{k}}$; ε_0 is the energy of the impurity orbital; $V_{\mathbf{k}j}$ is the hybridization matrix element; and, *U* is a seed parameter for the intracenter interaction of electrons in an impurity orbital ($U > 0$).

The structure of Nd_2CuO_4 is similar to La_2CuO_4 with the exception of the positions of the O atoms out of the CuO_2 planes. For “electronic doping,” the replacement of Nd^{3+} atoms by Ce^{4+} atoms should result in the appearance of a singly filled donor level in $Nd_{2-x}Ce_xCuO_4$. For La_2CuO_{4+y}, the excess oxygen atoms can occupy interstitial positions near the CuO_2 planes. Then all valence electrons of the O atoms can participate in the formation of the impurity bands.

In equation (1) the impurity orbital is characterized by two parameters, ε_0 and *U*. The latter parameter is an important parameter of the model, and it can be estimated as

$$U \simeq \frac{e^2}{\varepsilon_{opt} a_{eff}}, \tag{2}$$

where a_{eff} is the effective Bohr radius of the impurity orbital and ε_{opt} is the high-frequency permittivity of the compound. The estimated value of the effective radius for doped cuprates in an insulator state is well known: $a_{eff} = 4\text{–}8$ Å [17, 21–24]. Substituting into equation (2) the typical value $\varepsilon_{opt} = 3$ [28], we obtain $U \simeq 0.6$ eV. At the same time, according to the ARPES data, the typical width of the main peak of the valence band, determined by the spectrum of single-electron states of the cuprate plane, is $2D_b \simeq 3$ eV [25].

It was noted above that the change in the electronic structure of doped cuprates should include the formation of impurity bands of both distributed and localized states near the insulator gap. Formally, Hamiltonian (1) is the Anderson Hamiltonian [29] extended to an ensemble of substitution atoms randomly distributed in the initial lattice. As is well known, the Anderson Hamiltonian for a single impurity atom possesses a solution only for localized states in the gap. When the model is extended to an ensemble of impurities, Hamiltonian (1) contains solutions corresponding to the formation of impurity bands of both localized and distributed electronic states in the gap. This was demonstrated by solving equation (1) in the Hartree–Fock approximation [30]. The formation of narrow impurity bands of distributed states with a high density is due to hybridization, which results in virtual single-electron transitions: initial impurity site $\longrightarrow$ $\mathbf{k}$-band state $\longrightarrow$ difference site $\longrightarrow$ $\mathbf{k}_1$-band state, and so on.

As is well known, spin fluctuations are neglected in the Hartree–Fock approximation. It was found [30] that in the case of a single excess electron on an impurity the paramagnetic solution (1) in this approximation always corresponds to a metallic state, in which the Fermi level lies in two overlapping bands of localized and distributed states, i.e., the insulator–metal paramagnetic transition accompanying a change in the degree of doping cannot be described in the Hartree–Fock approximation.

However, the situation changes radically when the spin fluctuations are taken into account. In this case it is possible to obtain insulator, superconducting, and metallic states in the paramagnetic phase of the doped insulator. This will be presented below.

3. PARAMAGNETIC SOLUTION

We introduce the normal Green's functions

$$G_{\eta\eta_1}^{\sigma\sigma_1}(t, t_1) = -i\left\langle N_{\sigma_1-\sigma}\middle|T\tilde{a}_{\eta\sigma}(t)\tilde{a}^{+}_{\eta_1\sigma_1}(t_1)\middle|N_0\right\rangle,$$

and the anomalous Green's functions

$$F_{\eta\eta_1}^{(+)\sigma\sigma_1}(t, t_1)$$

$$= \exp(-2i\mu t)\left\langle (N+2)_{\sigma+\sigma_1}\middle|T\tilde{a}^{+}_{\eta\sigma}(t)\tilde{a}^{+}_{\eta_1\sigma_1}(t_1)\middle|N_0\right\rangle,$$

$$F_{\eta\eta_1}^{\sigma\sigma_1}(t, t_1)$$

$$= \exp(2i\mu t)\left\langle N_0\middle|T\tilde{a}_{\eta,\sigma}(t)\tilde{a}_{\eta_1\sigma_1}(t_1)\middle|(N+2)_{\sigma+\sigma_1}\right\rangle,$$

where the operators in the Heisenberg representation with Hamiltonian (1) are used; η assumes the values $\mathbf{k}$ or j; μ is the chemical potential of the system. Averaging is performed over the spin states of the system with the total number of particles N and $N + 2$ [31]. The bottom index in the ground states $|N_{\sigma_1-\sigma}\rangle$ and $|(N+2)_{\sigma+\sigma_1}\rangle$ indicates the presence of excess spins with a fixed total number of particles compared with the spin state in $|N_0\rangle$ and $|(N+2)_0\rangle$.

Using equation (1) it is easy to obtain a closed system of equations for the Green's functions. The paramagnetic solution of this system for the Green's functions, averaged over a random distribution of substitution atoms, was found in [32]. This solution will be presented below, but we shall first discuss the order parameters in the model under study. Here the fixed particle number representation is used.

The first parameter is determined by the average filling number of the localized orbitals:

$$G_{jj}^{\sigma\sigma}(-0^{+}) = -i\frac{1}{\pi}\int\limits_{-\infty}^{\mu} d\omega \operatorname{Im} G_{jj}^{\sigma\sigma}(\omega) = iA_{j,\sigma}. \qquad (3)$$

In addition, the system contains two different types of order parameters, associated with the spin fluctuations [32]. One of them is determined by the spin-off-diagonal normal Green's function for localized states:

$$G_{jj}^{\sigma,-\sigma}(0^{+}) = -i\frac{1}{\pi}\int\limits_{\mu}^{+\infty} d\omega \operatorname{Im} G_{jj}^{\sigma,-\sigma}(\omega) = i\lambda_j^{\sigma,-\sigma}. \qquad (4)$$

It is easy to see that for the normal state of the system the effect of the spin fluctuations is determined by λ.

The possibility of a superconducting state in the system is due to the anomalous Green's functions. Here, the order parameter determined by the spin-nondiagonal variable site Green's functions $F_{jj}^{(+)\sigma,-\sigma}(\omega)$ and $F_{jj}^{\sigma,-\sigma}(\omega)$ is important. These functions describe localized bosons with zero spin. We introduce the notation

$$F_{jj}^{\sigma,-\sigma}(0^{+}) = \lim_{t\to 0^{+}}\int\frac{d\omega}{2\pi}F_{jj}^{\sigma,-\sigma}(\omega)\exp(-i\omega t) = i\beta_j^{\sigma,-\sigma}. \quad (5)$$

The integral equations (3)–(5) must be supplemented by an equation for the Fermi energy μ:

$$(1+x)N_t = -\frac{1}{\pi}\int\limits_{-\infty}^{\mu} d\omega \operatorname{Im}\operatorname{Tr} G(\omega), \qquad (6)$$

where N_t is the total number of electrons in the initial band of the insulator (for definiteness, the valence band of the CuO_2 plane, hole doping) and $N_{im} = xN_t$ is the density of substitution atoms. Here it is assumed that the impurity level is singly filled. The quantity x will be interpreted as the degree of doping.

We assume that the impurity atoms occupy randomly the equivalent sites of the crystal lattice of the material. In this case the order parameters (3)–(5) do not depend on j. For the paramagnetic phase, the spin indices for the parameters A, λ, and β can be dropped [32]. The quantity A is the filling factor of a localized orbital and it is therefore real and positive. Since the Hamiltonian (1) does not change as a result of the following transformation in the two spin subspaces

$$a_{\eta,\uparrow} \longrightarrow a_{\eta,\uparrow}\exp(i\phi), \quad a_{\eta,\downarrow} \longrightarrow a_{\eta,\downarrow}\exp(-i\phi)$$

with constant phase ϕ, one of the quantities λ (4) or β (5) can be chosen to be real. For definiteness, we choose λ to be real. The characteristics of the superconducting state are determined only by the modulus of the complex quantity β.

For the impurity ensemble under consideration, the hybridization matrix element has the form

$$V_{\mathbf{k}j} = V_{\mathbf{k}l}\exp(i\mathbf{k}\mathbf{R}_j),$$

where $\mathbf{R}_j$ is a two-dimensional radius vector of the jth substitution atom.

Taking account of equation (1), the solution of the closed system of equations for the Green's functions can be represented in the form of series whose terms possess a similar structure:

$$\sum_{\{j_{m+1}\},\{\mathbf{k}_m\}} V_{\mathbf{k}j_1}V_{j_1\mathbf{k}_1}V_{\mathbf{k}_1 j_2}V_{j_2\mathbf{k}_2}\ldots V_{\mathbf{k}_m j_{m+1}}V_{j_{m+1}\mathbf{k}}\prod_m F(\mathbf{k}_m),$$

where $F(\mathbf{k}_m)$ are known functions that depend on the wave vectors.

Taking account of $V_{\mathbf{k}j}$, in performing the summation over $\{j\}$ the result obtained must be averaged over the random distribution of the impurity atoms $\{\mathbf{R}_j\}$. In so doing, it is necessary to calculate the *s*th moments of the following type [33]:

$$M_s(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_s) = \left\langle \sum_{\{n_1\}} \sum_{\{n_2\}} \ldots \sum_{\{n_s\}} \exp\left(-i \sum_m \mathbf{k}_m \mathbf{R}_{n_m}\right) \right\rangle_{av},$$

where $\langle\ldots\rangle_{av}$ denotes averaging over all possible configurations of the impurity ensemble. Then, the calculation of the configuration-averaged Green's functions reduces to summation of the series arising, whose *s*th term is a polynomial of degree *s* in the impurity concentration [30, 33].

It is convenient to represent this solution in matrix form, in which the Green's functions which are not diagonal in the bottom indices are expressed in terms of the diagonal Green's functions [32]. There arise residual terms proportional to double sums of the following type:

$$\sum_{\mathbf{k}_1 \neq \mathbf{k},\, j_1 \neq j} V_{\mathbf{k} j_1} V_{j_1 \mathbf{k}_1} G^{\sigma\sigma_1}_{\mathbf{k}_1 j}(\omega).$$

The first correction due to such terms to the Green's functions which are nondiagonal with respect to the bottom indices is zero, since the summation over j_1 and averaging over the random distribution of impurity atoms give

$$\left\langle \sum_{\mathbf{k}_1 \neq \mathbf{k},\, j_1 \neq j} \exp(i(\mathbf{k} - \mathbf{k}_1)\mathbf{R}_{j_1}) \right\rangle_{av} = N_{im} \sum_{\mathbf{k}_1 \neq \mathbf{k}} \delta(\mathbf{k}_1 - \mathbf{k}) = 0.$$

In second order, corrections $\propto x^2$ to the Green's functions which are nondiagonal in the bottom indices arise. Since, in reality $x \ll 1$, such double sums were neglected in [32], and diagrams $\propto x^n$ with $n \geq 2$ were summed only partially. The computational details can be found in [32].

We introduce the notation

$$G^{(0)}_{\mathbf{kk}}(\omega) = (\omega - \varepsilon_{\mathbf{k}})^{-1},$$

$$F^{(0)}_{\mathbf{kk}}(\omega) = (\omega + \varepsilon_{\mathbf{k}} - 2\mu)^{-1},$$

$$\tilde{G}_{ll}(\omega, A, \lambda) = (\omega - \varepsilon_0 - U(A - \lambda))^{-1},$$

$$\tilde{F}_{ll}(\omega, A, \lambda, \mu) = (\omega + \varepsilon_0 - 2\mu + U(A + \lambda))^{-1}.$$

Now, we present the expressions for the site Green's functions for the paramagnetic phase [32]:

$$\begin{aligned} G^{\sigma\sigma_1}_{jj} &= \frac{1}{2}(G_{loc}(\omega; \mu, A, \lambda, \beta) \\ &+ (-1)^{\sigma - \sigma_1} G_{loc}(\omega; \mu, A, -\lambda, \beta)), \end{aligned} \tag{7}$$

where

$$G_{loc} = \frac{\omega + \varepsilon_0 - 2\mu - S_2(\lambda)}{(\omega + \varepsilon_0 - 2\mu - S_2(\lambda))(\omega - \varepsilon_0 - S_1(\lambda)) - |\beta|^2 U^2 Z^2(\lambda)} \tag{8}$$

and

$$\begin{aligned} F^{(+)\sigma\sigma_1}_{jj} &= \frac{1}{2}((-1)^{\sigma - \sigma_1} F_{loc}(\omega; \mu, A, \lambda, \beta) \\ &- F_{loc}(\omega; \mu, A, -\lambda, \beta)), \end{aligned} \tag{9}$$

where

$$F_{loc} = \frac{\beta^* U Z(\omega; A, \mu, \lambda; \beta)}{(\omega + \varepsilon_0 - 2\mu - S_2(\lambda))(\omega - \varepsilon_0 - S_1(\lambda)) - |\beta|^2 U^2 Z^2(\lambda)}. \tag{10}$$

In the expressions (8) and (10) S_1 and S_2 are the self-energy parts, which have the form

$$S_1(\omega; A, \lambda, \mu, \beta) = U(A - \lambda) + \sum_{\mathbf{k}} \frac{V^2_{\mathbf{k}l}[F^{(0)^{-1}}_{-\mathbf{k}-\mathbf{k}} - x N_t T^{-1}_{ll}(\omega) \tilde{G}^{-1}_{ll} V^2_{\mathbf{k}l}]}{G^{(0)^{-1}}_{\mathbf{kk}} F^{(0)^{-1}}_{-\mathbf{k}-\mathbf{k}} - x N_t V^2_{\mathbf{k}l} T^{-1}_{ll}(\omega)[\tilde{G}^{-1}_{ll} G^{(0)^{-1}}_{\mathbf{kk}} + \tilde{F}^{-1}_{ll} F^{(0)^{-1}}_{-\mathbf{k}-\mathbf{k}} - x N_t V^2_{\mathbf{k}l}]} \tag{11}$$

and

$$S_2(\omega; A, \lambda, \mu, \beta) = -U(A + \lambda) + \sum_{\mathbf{k}} \frac{V^2_{\mathbf{k}l}[G^{(0)^{-1}}_{\mathbf{kk}} - x N_t T^{-1}_{ll}(\omega) \tilde{F}^{-1}_{ll} V^2_{\mathbf{k}l}]}{G^{(0)^{-1}}_{\mathbf{kk}} F^{(0)^{-1}}_{-\mathbf{k}-\mathbf{k}} - x N_t V^2_{\mathbf{k}l} T^{-1}_{ll}(\omega)[\tilde{G}^{-1}_{ll} G^{(0)^{-1}}_{\mathbf{kk}} + \tilde{F}^{-1}_{ll} F^{(0)^{-1}}_{-\mathbf{k}-\mathbf{k}} - x N_t V^2_{\mathbf{k}l}]}, \tag{12}$$

where

$$T_{ll}(\omega) = \tilde{G}_{ll}^{-1}\tilde{F}_{ll}^{-1} - |\beta|^2 U^2. \qquad (13)$$

The function $Z(\omega; A, \lambda, \mu, \beta)$ has the following form:

$$Z(\omega; A, \lambda, \mu, \beta) = 1 + \frac{xN_t}{T_{ll}(\omega)}\sum_{\mathbf{k}} \frac{V_{\mathbf{k}l}^4}{G_{\mathbf{kk}}^{(0)^{-1}} F_{-\mathbf{k}-\mathbf{k}}^{(0)^{-1}} - xN_t V_{\mathbf{k}l}^2 T_{ll}^{-1}(\omega)[\tilde{G}_{ll}^{-1} G_{\mathbf{kk}}^{(0)^{-1}} + \tilde{F}_{ll}^{-1} F_{-\mathbf{k}-\mathbf{k}}^{(0)^{-1}} - xN_t V_{\mathbf{k}l}^2]}. \qquad (14)$$

The one- and two-particle distributed states are determined by the following Green's functions [32]:

$$\begin{aligned} G_{\mathbf{kk}}^{\sigma,\sigma_1} = {} & \frac{1}{2}(G_{ext}(\mathbf{k}, \omega; \mu, A, \lambda, \beta) \\ & + (-1)^{\sigma-\sigma_1} G_{ext}(\mathbf{k}, \omega; \mu, A, -\lambda, \beta)), \end{aligned} \qquad (15)$$

where

$$G_{ext} = \frac{F_{-\mathbf{k}-\mathbf{k}}^{(0)^{-1}} - b_1(\lambda)W(\lambda)}{(G_{\mathbf{kk}}^{(0)^{-1}} - b_2(\lambda)W(\lambda))(F_{-\mathbf{k}-\mathbf{k}}^{(0)^{-1}} - b_1(\lambda)W(\lambda)) - |\beta|^2 U^2 W^2(\lambda)}, \qquad (16)$$

and

$$\begin{aligned} F_{-\mathbf{k}\,\mathbf{k}}^{(+)\sigma,\sigma_1} = {} & \frac{1}{2}((-1)^{\sigma+\sigma_1} F_{ext}(\mathbf{k}, \omega; \mu, A, \lambda, \beta) \\ & + F_{ext}(\mathbf{k}, \omega; \mu, A, -\lambda, \beta)), \end{aligned} \qquad (17)$$

where

$$F_{ext} = \frac{\beta^* U W(\lambda)}{(G_{\mathbf{kk}}^{(0)^{-1}} - b_2(\lambda)W(\lambda))(F_{-\mathbf{k}-\mathbf{k}}^{(0)^{-1}} - b_1(\lambda)W(\lambda)) - |\beta|^2 U^2 W^2(\lambda)}. \qquad (18)$$

The following notation is used in equations (15)–(18):

$$b_1(\lambda) = \tilde{G}_{ll}^{-1} - \sum_{\mathbf{k}_1} V_{\mathbf{k}_1 l}^2 G_{\mathbf{k}_1\mathbf{k}_1}^{(0)}, \qquad (19)$$

$$b_2(\lambda) = \tilde{F}_{ll}^{-1} - \sum_{\mathbf{k}_1} V_{\mathbf{k}_1 l}^2 F_{\mathbf{k}_1\mathbf{k}_1}^{(0)} \qquad (20)$$

and

$$W(\lambda) = \frac{N_{im} V_{\mathbf{k}_1 l}^2}{b_1(\lambda) b_2(\lambda) - |\beta|^2 U^2}. \qquad (21)$$

We shall now analyze the solution of equations (3)–(21). This solution is determined by four self-consistent parameters, μ, A, λ, and β, which are found from the four integral equations (3)–(6). If equation (5) possesses only the trivial solution $\beta = 0$, then, as one can easily see, the state of the doped material is normal. The electronic structure is determined only by the single-particle normal Green's functions for localized (7) and distributed (15) electronic states. These functions have poles in the region of the initial band and in the region of the initial insulator gap, where impurity bands of localized and distributed states are formed. It was noted above that the formation of the impurity bands of distributed single-particle states is due to hybridization, which results in single-electron transitions: initial impurity site ⟶ band state ⟶ other sites ⟶ band state, and so on.

The appearance of a superconducting state is determined only by the existence of a nontrivial solution $\beta \neq 0$ of equation (5) for the anomalous singlet Green's function $F_{jj}^{(+)\sigma,-\sigma}$. Such a solution indicates the formation of localized bosons with spin 0. In addition, localized spin-1 bosons should also be present, as is evident from equations (9) and (10). This spin-fluctuation parameter β gives the function F_{ext} (18), in terms of which the superconducting condensate is determined in the singlet and in the triplet quasiparticle pairing channels (17).

Thus, in this theory of spin-fluctuation superconductivity in doped insulators equations of the BCS type for the energy gap do not arise. Instead, the key equation for the superconducting state is equation (5), describing localized spin-zero bosons. The mechanism leading to the formation of a superconducting condensate is due to two-quasi-particle transitions along an impurity ensemble: localized boson on an impurity site ⟶ pair of coupled quasiparticles (–**kk**) in the impurity band of the distributed states ⟶ localized boson on another impurity site, and so on. The singlet and triplet quasiparticle pairing channels are intercoupled (see equation (17)). We note that in this model superconducting bosons arise only if the center of mass is at rest. This is due to the absence of translational symmetry in a doped insulator and is a consequence of averaging over an ensemble of substitution atoms.

Using equations (9) and (10), we rewrite equation (5) in the form

$$\beta = \beta\sum_{n=0,1}\int_{\mu}^{\infty}\frac{d\omega}{\pi}\mathrm{Im}\frac{(U/2)Z(\omega;\,A,\mu,(-1)^n\lambda;\,\beta)}{(\omega+\varepsilon_0-2\mu-S_2((-1)^n\lambda))(\omega-\varepsilon_0-S_1((-1)^n\lambda))-|\beta|^2U^2Z^2+i0^+}. \tag{22}$$

The existence of a solution $\beta \neq 0$ of equation (22) is determined by the values of the complex function Z (14) at the poles or their regions of the anomalous singlet Green's function (9), which lie near the Fermi level on the $\mathrm{Re}\,\omega \geq \mu$ semiaxis in the lower part of the complex plane ($\mathrm{Im}\,\omega < 0$). The function UZ can be regarded as an intrasite interaction between localized electrons, renormalized by spin fluctuations and hybridization in the quasiparticle pairing channel in the distributed states. If the second term on the right-hand side of equation (14) is less than 1, which arises here from the seed correlation energy, then, as one can easily see, equation (22) will have only the trivial solution $\beta = 0$ and, correspondingly, the system will be in the normal state (or metallic, or insulator). The solution $\beta \neq 0$ can exist if there exists at least a region of poles of the Green's function (9) near the Fermi level, where $\mathrm{Re}\,Z < 0$. This would mean that in this region of the spectrum the correlation energy UZ changed its initial positive sign and started to correspond to attraction.

4. COMPUTATIONAL DETAILS

For the calculations, we chose a model of a symmetric valence band with density of **k** states per spin

$$\rho^{(0)}(\varepsilon) = \begin{cases}\dfrac{N_t}{\pi D_b^2}[D_b^2-\varepsilon^2]^{1/2}, & |\varepsilon| \leq D_b\\ 0, \quad |\varepsilon| > D_b. & \end{cases} \tag{23}$$

Here $2D_b$ is the width of the band and N_t is the total number of states in the band. Doping will be described in terms of the relative density of substitution atoms: $x = N_{im}/N_t$.

The poles of the Green's function for localized and distributed electronic states lie in the region of the initial band and in the gap. According to equations (7)–(10) and (15)–(18), each of these functions is represented as a sum of two terms. The difference between these terms is that the spin-fluctuation parameter λ occurs in them with different signs ±. For this reason, the electronic structure divides into pairs of bands for which the substitution $\lambda \longrightarrow -\lambda$ transforms one band into another. For such pairs we shall use the same notations, which differ only by the upper index ±.

To show clearly the influence of doping on the density of distributed states in the valence band, we introduce the change in the density of states per substitution atom

$$\Delta\rho_{ext}(\omega) = -\frac{1}{\pi x N_t}\sum_{\mathbf{k}}\mathrm{Im}(G_{\mathbf{kk}}^{\sigma\sigma} - G_{\mathbf{kk}}^{(0)}). \tag{24}$$

In the region of the initial gap it is convenient to normalize the density of the localized and distributed states to the density of substitution atoms:

$$\rho_{ext}(\omega) = -\frac{1}{\pi x N_t}\sum_{\mathbf{k}}\mathrm{Im}G_{\mathbf{kk}}^{\sigma\sigma} \tag{25}$$

and

$$\rho_{loc}(\omega) = -\frac{1}{\pi}\mathrm{Im}G_{jj}^{\sigma\sigma}. \tag{26}$$

The superconducting condensate was calculated as follows. The **k** distribution of the singlet and triplet bosons has the form

$$\rho^{\sigma,\sigma_1}(\mathbf{k}) = \left|\int_{-\infty}^{\mu}\frac{d\omega}{\pi}\mathrm{Im}F_{-\mathbf{kk}}^{(+)\sigma,\sigma_1}(\omega)\right|^2. \tag{27}$$

We denote the branches of the poles of the anomalous Green's function $F_{-\mathbf{kk}}^{(+)\sigma,\sigma_1}$ as $\omega_{\mathbf{k}} = f_n(\varepsilon_{\mathbf{k}})$, where n is the branch number. Then, introducing the inverse function $\varepsilon_{\mathbf{k}} = f_n^{-1}(\omega_{\mathbf{k}})$, we obtain from equation (27) the energy distribution of the superconducting condensate, normalized to the density of substitution atoms:

$$\begin{aligned}\rho_n^{\sigma,\sigma_1}(\omega_{\mathbf{k}}) &= \frac{1}{xN_t}\rho^{\sigma,\sigma_1}(f_n^{-1}(\omega_{\mathbf{k}}))\\ &\times\rho^{(0)}(f_n^{-1}(\omega_{\mathbf{k}}))\frac{df_n^{-1}(\omega_{\mathbf{k}})}{d\omega_{\mathbf{k}}},\end{aligned} \tag{28}$$

where the density of states $\rho^{(0)}$ is given by equation (23).

The substitution

$$V_{\mathbf{k}l} = V_k \times N_t^{-1/2},$$

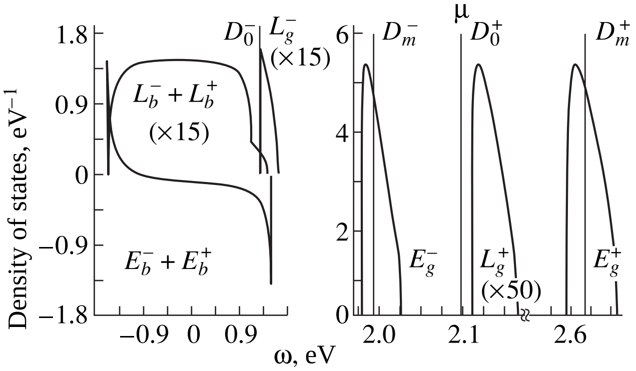


Fig. 1. Electronic structure of the insulator state for $x = 0.02$. The Fermi energy $\mu = 2.1$ eV and coincides with the position of the peak D_0^+. The curve $E_b^- + E_b^+$ represents the change in the density of states in the initial band. Parameters: $D_b = 1.5$ eV, $\varepsilon_0 = D_b - 0.4$ eV, $V_k = 1.5$ eV, and $U = 1$ eV.

where V_k has the dimension of energy, is used in the expressions obtained for the Green's functions. In the calculations, V_k is assumed to be independent of **k**.

Thus, the four constants with the dimension of energy D_b, U, ε_0, and V_k completely determine the model (1). In the calculations the values of the constants D_b and U which agree with the existing experimental data (see Section 2) were used. The values of the two remaining constants ε_0 and V_k were chosen so that (1) the Fermi level in the paramagnetic insulator phase lies above the valence-band top by $\simeq 0.4$–0.6 eV, which corresponds to the position of the known spectral feature (MIR band) in the doped cuprates (see [2] and references cited there) and (2) the insulator–superconductor phase transition occurs at a reasonable impurity concentration, $x \simeq 0.05$ [2].

The solution of (1) depends on the four self-consistent parameters A, μ, λ, and β. These parameters were found from the four integral equations (3)–(6). An iteration procedure was used to solve these equations. Taking into account the fact that the exact single-particle Green's function must satisfy Levinson's theorem, which means that the total number of electronic states is conserved, the accuracy of our calculations using the Green's functions obtained can be estimated as follows:

$$\gamma = \frac{1}{\pi N_{im}} \mathrm{Im} \int d\omega \mathrm{Tr}(G(\omega) - G^{(0)}(\omega)),$$

where $G^{(0)}$ is the unperturned Green's function, which is determined by the first and second terms on the right-hand side of the Hamiltonian (1). The computational accuracy decreases as the degree of doping x increases. For the results presented below $\gamma \leq 0.7 \times 10^{-3}$.

5. RESULTS AND DISCUSSION

The model presented makes it possible to describe at the same time insulator–superconductor–metal transitions in a doped compound. We begin with a discussion of the normal paramagnetic phase. In this state the spin fluctuations are represented by the parameter λ, which leads to the above-mentioned doubling of the spectral features of the electronic structure. We shall show that the spin fluctuations (1) lead to an insulator state of the material for low values of x and (2) give rise to an insulator–metal phase transition at some threshold value of the degree of doping.

5.1. Insulator–Metal Transition

The electronic structure of the paramagnetic insulator state with $x = 0.02$ is shown in Fig. 1. The curve $E_b^- + E_b^+$ represents the change in the density of distributed states in the valence band. A common feature for various values of the parameters of the model is that the doping-induced change in the total number of band states is negative. For the result presented, this change is −0.2531 per spin per substitution atom. This deficit of states is exactly compensated by the distributed states, which form in the region of the insulator gap. Spin fluctuations lead to splitting of these states into two narrow bands, E_g^- and E_g^+, shown in Fig. 1.

An important circumstance is that the maximum density of distributed states in each band $E_g^{\pm}$,

$$xN_t\rho_{g,ext}^{max} = \frac{N_t}{4.712\ \mathrm{eV}}, \tag{29}$$

is equal to half the maximum density of states in the valence band, while the total number of states in these bands is small compared with the density xN_t of substitution atoms. In the insulator state, the completely filled E_g^- band lies below the Fermi energy $\mu = 2.1$ eV, and the E_g^+ band is unfilled.

In the notations used above, the density of localized states is normalized to 1. The curve $L_b^- + L_b^+$ in Fig. 1 corresponds to the density of localized states near the initial band. The total number of these states is 0.2414 per spin per substitution atom. This band is due to hybridization, and it arises even when the Anderson Hamiltonian is solved with a single impurity atom [29].

In addition, the energy distribution of localized states contains four δ functions, which correspond to the simple poles of $G_{jj}^{\sigma\sigma}(\omega)$ (7). Two of these poles, $\omega_0^{\pm} = \varepsilon_0 + U(A \pm \lambda)$, and the corresponding δ peaks are labelled in Fig. 1 as $D_0^{\pm}$. In the insulator state the position of the partially filled peak D_0^+ in the gap always coincides with the Fermi energy. The number of states in each of these peaks is $x/(1 + 2x)$ per spin per substitution atom. Taking account of the normalization of the density to 1, these peaks are low-concentration peaks for $x \ll 1$.

The two other simple poles correspond to high-concentration δ peaks, which are denoted as $D_m^{\pm}$. The number of localized states per spin per substitution atom is 0.3052 for the D_m^{-} peak and 0.3868 for the D_m^{+} peak. A general feature of the model for various values of the parameters is that the $D_m^{\pm}$ peaks lie inside the band of distributed states $E_g^{\pm}$, as shown in Fig. 1. The electronic structure also includes two bands $L_g^{\pm}$ of localized states. The number of states in these bands is usually relatively small ($\simeq 10^{-2}$ in the units adopted).

We note that in the region of the gap all bands of localized states with upper index "+" lie above all bands of localized states with upper index "–" (see Fig. 1). Then, using the expressions for the Green's function (7) and (8) and the definitions (3) and (4), it can be shown that in the insulator state the following relation holds between the filling factor A and the spin-fluctuation parameter λ:

$$A + \lambda = 1. \tag{30}$$

For the electronic structure presented in Fig. 1, $A = 0.5904$ and $\lambda = 0.4096$. Therefore the spin fluctuations in the model (1) cannot be neglected.

The change in the electronic structure in the region of the valence band with increasing doping x is qualitatively similar to the result presented in Fig. 1, and it will not be shown below.

For the normal phase, the behavior of the electronic structure in the region of the initial gap with increasing x gives rise to an insulator–metal transition in the doped compound. In the insulator state the Fermi level coincides with the position of the D_0^{+} peak at $\omega_0^{+} = \varepsilon_0 + U(A + \lambda)$. Because of the relation (30), the value of ω_0^{+} does not change with increasing x. However, the position of the E_g^{-} band of distributed states depends on x, and this band shifts with increasing x in the direction of the partially filled peak of localized D_0^{+} states. Figure 2 shows the computed electronic structure near the Fermi level for $x = 0.0524$. It is evident that the top of the E_g^{-} band reached the D_0^{+} peak. Here the relation (30) is still satisfied: $\lambda = 0.4135$ and $A = 0.5865$. The number of localized states per spin per substitution atom is 0.2755 for the D_m^{-} peak and 0.3572 for D_m^{+}. The number of distributed states per spin per substitution atom is 0.1484 for the E_g^{-} band and 0.0865 for E_g^{+}.

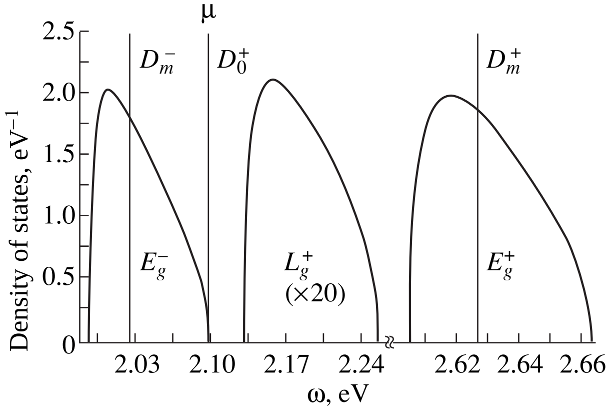


Fig. 2. Density of states in the initial insulator gap for $x = 0.0524$. This doping level lies near the insulator–metal boundary $x_{thr,\,1}$ in the phase diagram of normal states. The parameters are the same as in Fig. 1.

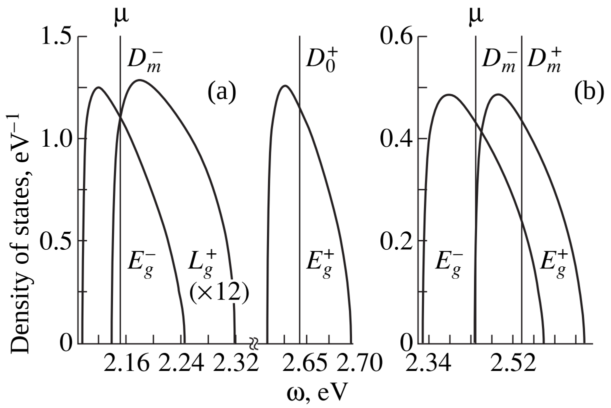


Fig. 3. Electronic structure of the metallic state in the region of the initial insulator gap for various degrees of doping x: (a) $x = 0.085$; (b) $x = 0.22$. The parameters are the same as in Fig. 1.

For $x = 0.085$ the D_0^{+} peak lies below the E_g^{-} band, and the Fermi level now coincides with the position of the high-concentration peak D_m^{-} of localized states (Fig. 3a). This partially filled peak lies in two overlapping bands, one of which is the band E_g^{-} of distributed states and the other is the band L_g^{+} of localized states. Thus, the electronic structure presented here in the region of the initial insulator gap corresponds to the state of a poor metal, in which the distributed and localized states coexist near the Fermi level. Moreover, the total number of localized D_m^{-} states is 2.12 times greater than the number of states in the E_g^{-} band, which is 0.1265 per spin per substitution atom. The number of

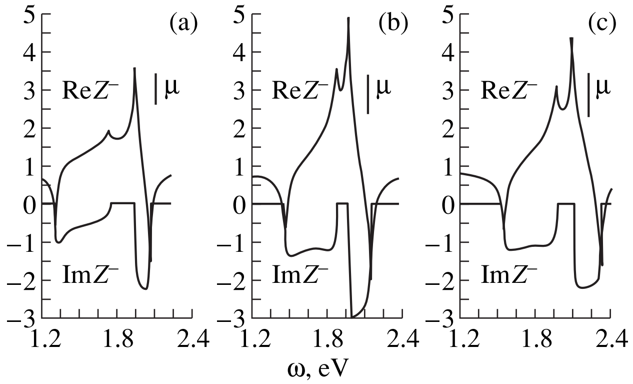


Fig. 4. The function Z, representing the renormalization of the intrasite interelectronic interaction in the quasiparticle pairing channel in the distributed states, versus the energy ω for various degrees of doping x: (a) $x = 0.0524$; (b) $x = 0.085$; (c) $x = 0.107$. The parameters are the same as for Fig. 1.

states in the L_g^+ band is relatively small, 0.0148 in the units adopted. It was found that $A = 0.6494$ and $\lambda = 0.3368$. Thus, equation (30) is not satisfied in the metallic state.

As x increases further, the L_g^+ band also shifts below the Fermi level. Then we obtain from equations (4) and (7), (8) the following expression for the spin-fluctuation parameter λ:

$$\lambda = N_m^+ - N_m^-(1 - A_m^-), \qquad (31)$$

where N_m^+ and N_m^- are the number of states per spin per substitution atom in the D_m^+ and D_m^- bands, respectively, and A_m^- is the filling factor of the D_m^- peak. All bands of localized states, with the exception of D_m^+ and D_m^-, lie below the Fermi level, and there total number of states increases with x. Then, the contribution from these bands to the filling factor A also increases and, correspondingly, A_m decreases as x increases. Since the values of N_m^+ and N_m^- are close, $\lambda \longrightarrow 0$ in the metallic phase as x increases. In consequence, there is a tendency for any pair of bands whose designations differ only by the upper index ± to merge, and the electronic structure has its own asymptotic solution (1) in the Hartree–Fock approximation ($\lambda = 0$). This tendency for the electronic structure to become modified is represented in Fig. 3b. For $x = 0.22$ we obtain $\lambda = 0.07115$ and $A = 0.6254$. Both bands of distributed states $E_g^\pm$ overlap with one another and are partially filled. The peak of localized states D_m^+ is still not filled, but it lies closer to D_m^-. The number of localized states per spin per substitution atom is 0.2103 for the D_m^- peak and 0.2229 for the D_m^+ peak. Here the filling $A_m^- = 0.2785$, while for $x = 0.085$ we have $A_m^- = 0.9743$.

It should be noted that the Fermi energy is essentially independent of x. According to Figs. 1–3 μvaries in a ≃10% range, while the doping changes by a factor of 11. It also follows from Figs. 1–3 that, although the width of the $E_g^\pm$ bands increases with increasing x, the maximum density of states in these bands does not depend on doping and is given by equation (29).

5.2. Renormalization of the Intrasite Interelectron Interaction in the Quasiparticle Pairing Channel in Distributed States

It was shown above that the possibility of a nontrivial solution of equation (22) for the spin-fluctuation parameter $\beta \neq 0$ is determined by the function $Z(\omega)$ (14) on the semiaxis $\omega \geq \mu$. If such a solution exists, then the doped material is a superconductor. The quantity UZ can be interpreted as the effective intrasite interaction. Since the other spin fluctuation parameter λ leads to doubling of the spectral features of the electronic structure, the solution (22) depends on the two functions $Z(\omega; A, \mu, \pm\lambda, \beta)$. However, it is easy to show from equation (14) that $Z^+(\mu + \omega; A, \mu, \lambda, \beta) = Z^-(\mu - \omega; A, \mu, -\lambda, \beta)$. For this reason, in what follows, we shall present the computational results only for one of them.

We have shown that in the normal phase, near the Fermi level, there is always a density of localized single-particle states that is determined by the normal Green's function (7), and therefore there exists a region of poles of the anomalous Green's function (9) for localized bosons. For this reason, the solution $\beta \neq 0$ can exist, if there exists a finite energy range near the Fermi level where $\mathrm{Re}Z^\pm < 0$. We shall present below the function $Z^-(\omega)$ for the normal states ($\beta = 0$) for various values of x. Actually, in the superconducting state the β is much less than A and λ, and it does not substantially affect $Z(\omega)$.

Figure 4a shows the functions $\mathrm{Re}Z^-(\omega)$ and $\mathrm{Im}Z^-(\omega)$ for $x = 0.0524$ close to the threshold doping level $x_{thr,\,1}$ for the insulator–metal transition (see Fig. 2). Here $\mu = 2.1$ eV and $\mathrm{Re}Z^-(\omega = \mu) = 0.0771$, $\mathrm{Im}Z^-(\omega = \mu) = 0$. It is evident that there exists a region where $\mathrm{Re}Z^- < 0$. The width of this region is 49 meV, but it lies below the Fermi level. A region where $\mathrm{Im}Z^- < 0$ also lies below μ. For smaller values of x these regions become narrower and shift to the left from $\mu = 2.1$ eV. We have established that for $x < 0.0524$ there exists only the trivial solution (22), and the normal state of the doped compound is an insulator (see Figs. 1, 2).

As x increases, the region $\mathrm{Re}Z^- < 0$ shifts to the right with respect to μ. For $x = 0.085$ the function $Z^-(\omega)$ also is presented in Fig. 4b. Here $\mu = 2.1513$ eV lies in the region of negative values of $\mathrm{Re}Z^-$ and $\mathrm{Im}Z^-$, and $\mathrm{Re}Z^-(\omega = \mu) = -1.2675$ and $\mathrm{Im}Z^-(\omega = \mu) = -1.8937$. The width of the region where $\mathrm{Re}Z^- < 0$ is 84 meV. Here, one should note also another region, $\mathrm{Re}Z^- < 0$, lying below 1.5 eV. This region does not contribute to the integral of the expression containing Z^- on the right-hand side of equation (22). However, because of the relation between $Z^\pm$, the function $\mathrm{Re}Z^+$, together with the region of negative values of $\mathrm{Re}Z^+$ near μ, acquires a second region where $\mathrm{Re}Z^+ < 0$, located near the peak of localized states D_m^+. For this reason, for fixed x it can be expected that there exist at least three positive contributions to the right-hand side of equation (22), one of which is related with Z^- and the other two are related with Z^+. If $\beta = 0$, the doped material would be in a metallic state, whose electronic structure is presented in Fig. 3. However, for this degree of doping, there exists a solution $\beta \neq 0$ of equation (22) and, correspondingly, the doped compound is a superconductor.

As x increases further, this region passes through the Fermi surface and for $x = 0.107$ it lies to the right of $\mu = 2.2173$ eV (Fig. 4c). Here $\mathrm{Re}Z^-(\omega = \mu) = 1.0671$ and $\mathrm{Im}Z^-(\omega = \mu) = -2.1392$. For $x > 0.107$ we obtained only the trivial solution (22). Correspondingly, the normal state of the doped compound is metallic (see Fig. 3).

Thus, superconductivity can exist in a bounded range of doping. Comparing the results presented in Figs. 2 and 4, it can be concluded that the onset of superconductivity is associated with the first threshold doping level $x_{thr,\,1}$. For the model parameters used, $x_{thr,\,1} \simeq 0.0524$. As the normal phase is approached, near $x_{thr,\,1}$, an insulator–metal transition occurs and for the same doping the region with $\mathrm{Re}Z^- < 0$ approaches μ. We were not able to determine whether a transition from the insulator into the superconducting state occurs immediately or there exists an intermediate metallic phase. The solution for the superconducting state for $x = 0.085$, i.e., quite far from the threshold value, will be presented below. We have shown that the model is also characterized by a second threshold doping level $x_{thr,\,2}$, for which a superconductor–metal transition occurs.

We note that $x_{thr,\,1}$ increases with D_b. For a width of the initial band $2D_b = 6$ eV and the same values of the other model parameters, we obtain $x_{thr,\,1} = 0.134$. For this reason, the superconducting state induced by the spin fluctuations actually can arise in doped insulators with initially narrow allowed bands.

5.3. Superconducting State

Assuming $\beta = 0$, the normal state of the doped compound is metallic with $x = 0.085$ (see Fig. 3). However, equation (22) possesses the solution $|\beta| = 0.00395$. We note that here no cutoff of the integral on the right-hand side of equation (22) was used, and the integration was performed along the entire semiaxis $\omega \geq \mu$.

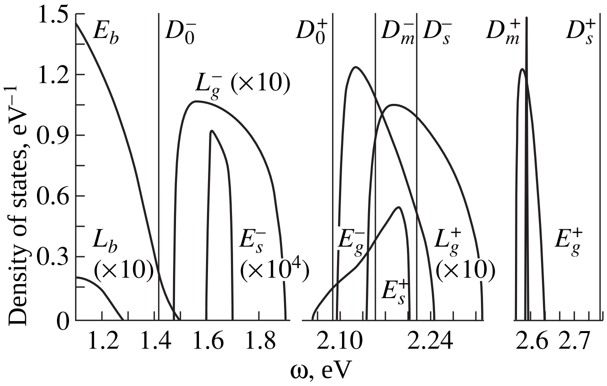


Fig. 5. Density of single-particle electronic states in the superconducting state of a doped compound with $x = 0.085$. The parameters are the same as for Fig. 1.

The following changes in the parameters were obtained in the superconducting state: $\Delta A = -2.7 \times 10^{-4}$, $\Delta\lambda = 3.2 \times 10^{-4}$, and $\Delta\mu = -0.28$ meV with respect to their values in the normal state. The electronic structure of the single-particle states is shown in Fig. 5. Here the curves $E_b^\pm$ and $L_b^\pm$ are fragments of the distribution of distributed and localized states in the region of the initial band near its boundary with the gap at 1.5 eV. Comparing this result with the results shown in Fig. 3, it is evident that in the superconducting state, two new bands $E_s^\pm$ of distributed states and two new peaks $D_s^\pm$ of localized states near the gap have appeared. The number of states per spin per substitution atom is 0.7377×10^{-5} for the E_s^- band and 0.4497×10^{-5} for the E_s^+ band. The number of states in the $D_s^\pm$ peaks is the same and is equal to 1.755×10^{-6}. We note that small but finite widths have appeared for the high-concentration peaks of the localized states $D_m^\pm$. Near the Fermi level there are six overlapping bands. Two of them are bands of the distributed states E_g^- and E_s^+, and the others are bands of localized states.

Now, we shall discuss the solution of equation (22). This equation contains the anomalous Green's function $F_{jj}^{(+)\sigma,\,-\sigma}$. Its poles coincide with the poles of the normal Green's function $G_{jj}^{(+)\sigma\sigma}$, as one can see from equations (7), (8) and (9), (10). Therefore, for this solution with $\beta \neq 0$ we have

$$1 = \sum_m Q_m(A, \mu, \lambda;\, \beta), \tag{32}$$

where Q_m is the contribution of the mth band of single-particle localized states, which lies no lower than the

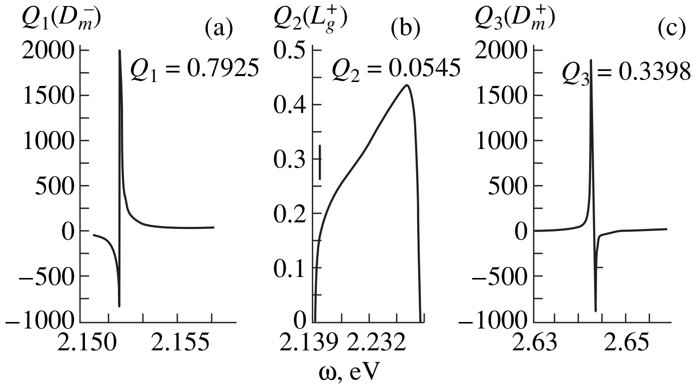


Fig. 6. The solution $|\beta| = 0.00395$ of equation (22) with $x = 0.085$: (a) contribution from the D_m^- band; (b) contribution from L_g^+; (c) contribution from D_m^+. The contributions from $D_s^\pm$ are the same and are equal to -0.0904. The corresponding bands are shown in Fig. 5.

Fermi level $\mu = 2.1511$ eV, to the integral on the right-hand side of equation (22).

It follows from Fig. 5 that there are five such bands (or regions of integration in equation (22)). Three of them are the bands D_m^-, D_m^+, and L_g^+, which arise in the solution for the normal phase and which are also shown in Fig. 3a. The contribution of the peak D_m^-, which coincides with the Fermi level μ and now has a finite width, in equation (32) is determined by the function Z^-. As shown in Fig. 4, $\mathrm{Re}Z^- < 0$ near the Fermi level. Here also $\mathrm{Im}Z^- < 0$, which is due to the damping of the quasiparticles. The integrand for this band is shown in Fig. 6a and its contribution is $Q_1(D_m^-) = 0.7925$ for $|\beta| = 0.00395$.

The contribution of D_m^+ and L_g^+ in equation (32) is determined by the values of the function Z^+ in the regions of these bands. Considering the relation between Z^+ and Z^- (Fig. 4), it is easy to see that Z^+ possesses two regions where $\mathrm{Re}Z^+ < 0$. The first region lies near the Fermi energy, where the band L_g^+ lies. The integrand for this band is shown in Fig. 6b and its contribution is relatively small $Q_2(L_g^+) = 0.0545$ for $\beta = 0.00395$. This is due to the fact that the amplitude of the poles for the single-particle states in this region is also small, as shown in Figs. 3a and 5. The second region with $\mathrm{Re}Z^+ < 0$ lies near the peak D_m^+ and corresponds to the region $\mathrm{Re}Z^- < 0$, lying below 1.5 eV (Fig. 4). For this band, the integrand is shown in Fig. 6c and its contribution is $Q_3(D_m^+) = 0.3338$ for $\beta = 0.00395$.

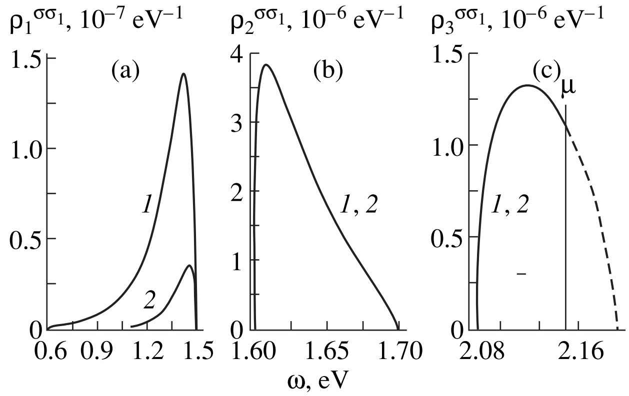


Fig. 7. Energy distribution of the distributed bosons in the region of the initial band of the insulator (a), in the region of the band E_s^- (b), and in the region of the band E_s^+ (c). The corresponding bands are shown in Fig. 5: curve *1*, spin zero; curve *2*, spin 1.

The two remaining regions of integration in equation (22) are related with the two new peaks of localized states $D_s^\pm$, which have appeared, in the region of the gap. Their contributions are the same and negative: $Q_{4,5}(D_s^\pm) = -0.0904$.

Thus, we have obtained the solution of equation (32), and therefore the doped compound is a superconductor. The energy distribution of the distributed bosons in the singlet and triplet pairing channels were calculated according to equation (28). The results are presented in Fig. 7. An unexpected result is that there is a distribution of bosons in the region of the initial valence band far from the Fermi surface (Fig. 7a). Here the number of pairs per spin per substitution atom is relatively small and is equal to 3.46×10^{-8} in the singlet channel and 5.57×10^{-9} in the triplet channel.

In the region of the initial gap there are two regions of distribution of the condensate. The first region also lies far from μ, and the corresponding distribution of the superconducting bosons coincides with the band of single-particle distributed states E_s^- (Fig. 7b). Here the number of pairs per spin per substitution atom in the singlet and triplet channels is the same and is equal to 1.821×10^{-7}. The second region in the gap lies near μ, and its position coincides with the band of single-particle distributed states E_s^+ (Fig. 7c). The energy distribution of the pairs in this region is cut off by the Fermi energy. The number of pairs per spin per substitution atom in the singlet and triplet channels is the same and is equal to 1.053×10^{-7}.

This model of superconductivity is also characterized by the number of localized bosons with spin 0 and 1. Using equations (9) and (10), this number per spin per substitution atom is

$$N_{LB} = \sum_{\sigma, \sigma_1} \left| \int_{-\infty}^{\mu} \frac{d\omega}{\pi} \mathrm{Im} F_{jj}^{(+)\sigma\sigma_1}(\omega) \right|^2.$$

For $\beta = 0.00395$ we obtain $N_{LB} = 4.50 \times 10^{-5}$.

We now obtain the following relation between the quasiparticle concentrations in the superconducting state. Summing over the three bands shown in Fig. 7, we obtain that the total number of pairs per substitution atom is 0.64×10^{-6} in the singlet pairing channel and 0.58×10^{-6} in the triplet channel. Therefore the concentration of distributed bosons is much lower than the concentration of localized bosons. The latter concentration, in turn, is much less than the concentration of filled single-particle states near the Fermi energy. The total number of single-particle distributed states near the Fermi energy is largely determined by the E_g^- band and is 0.253 per substitution atom. This value is 2.12 times less than the total number of localized states, which is mainly related with the peak D_m^-.

In conclusion, we note the following. The phase diagrams and properties in the normal state of cuprates and the compounds $R_{1-x}M_xTiO_3$ and $R_{1-x}M_xVO_3$ (R^{3+} is a rare-earth element and M^{2+} is an alkaline-earth element) are similar, the difference being that the latter are not superconductors. It has been established [34, 35] that introducing a substitution impurity in a compound based on Ti and V leads to the appearance of a spectral density in the region of the initial insulator gap, and therefore difficulties appear in the explanation of the results obtained on the basis of the Hubbard model [34, 35]. Our approach makes it possible to describe the insulator–metal phase transition in the case where equation (5) possesses only the trivial solution $\beta = 0$.

The mechanism leading to the formation of a superconducting state is electronic and is related with spin fluctuations in the doped compound. However, in this model the electron–phonon interaction should have a very large effect on the properties of the compounds. For example, this interaction can lead to the formation of a pseudogap in the spectrum of distributed states near the energy μ in the normal state, if the interaction with phonons in the form

$$H_{el-ph} = \sum_{\mathbf{q}j\sigma} \kappa_{\mathbf{q}j} \hat{n}_{j\sigma} \phi_{\mathbf{q}},$$

where $\phi_{\mathbf{q}}$ is the field operator for phonons, is introduced into equation (1). This is due to the fact that in the metallic state there is a high density of localized electronic states D_m^- at the Fermi surface. This interaction results in dephasing of the site electrons. Then the localized states D_m^- will be characterized by damping, and the peak will be broadened. As a result of hybridization, a corresponding imaginary part will also appear in the energies of electrons in the band of distributed states E_g^-. This will lead to a dip in the density of states in the E_g^- band near μ, the depth of the depth being determined by the strength $\kappa_{\mathbf{q}j}$ of the interaction.

REFERENCES

1. G. M. Éliashberg, in *Physical Properties of High-Temperature Superconductors*, Ed. by D. M. Ginzberg (Mir, Moscow, 1990), p. 505.
2. E. Dagotto, Rev. Mod. Phys. **66**, 673 (1994).
3. N. Bulut and D. J. Scalapino, Phys. Rev. B **50**, 16078 (1994).
4. Q. Chen, I. Kosztin, B. Jankó, *et al.*, Phys. Rev. Lett. **81**, 4708 (1998).
5. Yu. A. Uzyumov, Usp. Fiz. Nauk **161**, 2 (1991) [Sov. Phys. Usp. **34**, 361 (1991)]; Usp. Fiz. Nauk **165**, 403 (1995) [Sov. Phys. Usp. **38**, 385 (1995)].
6. N. M. Plakida, *High T_c Superconductivity* (Springer, Berlin, 1995).
7. P. Monthoux and D. Pines, Phys. Rev. B **49**, 4261 (1994).
8. A. Chubukov and D. K. Morr, Phys. Rev. Lett. **81**, 4716 (1998).
9. T. J. Smith, K. H. Andersen, U. Beck, *et al.*, J. Magn. Magn. Mater. **177–181**, 543 (1998).
10. H. A. Blackstead, J. D. Dow, and D. B. Pulling, Physica C **265**, 143 (1996).
11. M. Sigrist and T. M. Rice, Rev. Mod. Phys. **64**, 503 (1995).
12. D. J. Scalapino and S. R. White, Phys. Rev. B **58**, 8222 (1998).
13. V. J. Emery and S. A. Kivelson, Phys. Rev. Lett. **74**, 3253 (1995).
14. Q. Si, J. Phys. Cond. Matt. **8**, 9953 (1996).
15. A. S. Alexandrov, Physica C **274**, 237 (1997).
16. M. V. Sadovskiĭ, Sverkhprovodimost': Fiz., Khim., Tekh. **3**, 337 (1995).
17. P. P. Edwards, N. F. Mott, and A. S. Alexandrov, J. Supercond. **11**, 151 (1998).
18. H. A. Blackstead, J. D. Dow, and D. B. Pulling, Physica C **265**, 143 (1996).

19. A. I. Agafonov and É. A. Manykin, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 419 (1997) [JETP Lett. **65**, 439 (1997)].
20. M. I. Ivanov, V. M. Loktev, and Yu. G. Pogorelov, Fiz. Nizk. Temp. **24**, 615 (1998) [Low Temp. Phys. **24**, 463 (1998)].
21. N. F. Mott, *Metal-Insulator Transition* (Taylor and Francis, London, 1990).
22. P. P. Edwards, T. V. Ramakrishnan, and C. N. R. Rao, J. Phys. Chem. **99**, 5228 (1995).
23. C. Y. Chen, R. J. Birgeneau, M. A. Kastner, *et al.*, Phys. Rev. B **43**, 392 (1991).
24. G. A. Thomas, in *High Temperature Superconductivity*, Ed. by D. P. Tunstall and W. Barford (Adam Hilger, Bristol, 1991), p. 169.
25. Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
26. C. Quitmann, J. Ma, R. J. Kelly, *et al.*, Physica C **235-240**, 1019 (1994).
27. A. I. Agafonov and E. A. Manykin, Physica B **259–261**, 458 (1999).
28. T. Timusk, S. L. Herr, K. Kamaras, *et al.*, Phys. Rev. B **38**, 6683 (1988).
29. F. D. M. Haldane and P. W Anderson, Phys. Rev. B **13**, 2553 (1976).
30. A. I. Agafonov and E. A. Manykin, Phys. Rev. B **52**, 14571 (1995); A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **109**, 1405 (1996) [JETP **82**, 758 (1996)].
31. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, New Jersey, 1963).
32. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **114**, 1765 (1998) [JETP **87**, 956 (1998)].
33. F. Yonezawa and T. Matsubara, Prog. Theor. Phys. **35**, 357 (1966).
34. S. W. Robey, L. T. Hudson, C. Eylem, *et al.*, Phys. Rev. B **48**, 562 (1993).
35. T. Katsufuji, Y. Okimoto, and Y. Tokura, Phys. Rev. Lett. **75**, 3497 (1995); M. Kasuya, Y. Tokura, T. Arima, *et al.*, Phys. Rev. B **47**, 6197 (1993); Y. Taguchi, Y. Tokura, T. Arima, *et al.*, Phys. Rev. B **48**, 511 (1993).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 17–23.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 22–28.
Original Russian Text Copyright © 2000 by Achasov, Beloborodov, Berdyugin, Bozhenok, A. Bukin, D. Bukin, Burdin, Vasil'ev, Ganyushin, Gaponenko, Golubev, Dimova, Dolinskiĭ, Druzhinin, Dubrovin, Zakharov, Ivanchenko, Ivanov, Korol', Koshuba, Kukhartsev, Mamutkin, Otboev, Pakhtusova, Sal'nikov, Serednyakov, Sidorov, Silagadze, Sharyĭ, Shatunov.

Measurement of the Relative Probability of $\phi \longrightarrow \eta\gamma$ Decay in the $\eta \longrightarrow \pi^+\pi^-\pi^0$ Channel

**M. N. Achasov, K. I. Beloborodov, A. V. Berdyugin, A. V. Bozhenok, A. D. Bukin*,
D. A. Bukin, S. V. Burdin, A. V. Vasil'ev, D. I. Ganyushin, I. A. Gaponenko,
V. B. Golubev, T. V. Dimova, S. I. Dolinskiĭ, V. P. Druzhinin, M. S. Dubrovin,
A. S. Zakharov, V. N. Ivanchenko, P. M. Ivanov, A. A. Korol', S. V. Koshuba,
G. A. Kukhartsev, A. A. Mamutkin, A. V. Otboev, E. V. Pakhtusova, A. A. Sal'nikov,
S. I. Serednyakov, V. A. Sidorov, Z. K. Silagadze, V. V. Sharyĭ, and Yu. M. Shatunov**

*Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
Novosibirsk State University, Novosibirsk, 630090 Russia
e-mail: bukin@inp.nsk.su

Received July 28, 1999

Abstract—The relative decay probability $B(\phi \longrightarrow \eta\gamma) = (1.259 \pm 0.030 \pm 0.059)\%$ in the decay channel $\eta \longrightarrow \pi^+\pi^-\pi^0$ has been measured in an experiment using a spherical neutral detector in the VEPP-2M electron-positron storage ring. The result agrees with the tabulated value and with measurements of this probability using the spherical neutral detector in other η-meson decay channels. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The $\phi \longrightarrow \eta\gamma$ magnetic-dipole transition has been studied in several experiments, mainly using counter-propagating electron-positron beams [1]. In these experiments the systematic error is comparable to or greater than the statistical error. Since it is more difficult to estimate systematic errors, in order to improve the experimental accuracy of the averaged result it is important to make independent measurements of this value for different detectors and in different η-meson decay modes. So far the neutral η-meson decay modes: $\eta \longrightarrow \gamma\gamma$ [2, 3] and $\eta \longrightarrow 3\pi^0$ have mainly been used although the $\eta \longrightarrow \pi^+\pi^-\pi^0$ decay channel was used in a recent experiment using the KMD-2 detector [5]. For the spherical neutral detector (SND) measurements of the decay probability $B(\phi \longrightarrow \eta\gamma)$ in the $\eta \longrightarrow \pi^+\pi^-\pi^0$ channel are of additional value for studying systematic errors when recording events not only with neutral particles but also with charged ones. In addition, $\phi \longrightarrow \eta\gamma$ decay has now been measured using the SND in all significant η-meson decay channels [3,4] so that errors in measurements of η-meson decay probabilities in different channels can be almost completely eliminated.

2. EXPERIMENT

Experiments were carried out in 1996 [6] and 1998 [7] using the VEPP-2M electron-positron storage ring [8] with the SND detector [9]. The photon energy resolution in the detector calorimeter is given by

$$\frac{\sigma_E}{E} = \frac{4.2\%}{\sqrt[4]{E\,[\mathrm{GeV}]}},$$

and the angular resolution as a function of the photon energy is

$$\sigma_\varphi = \sqrt{\frac{(0.82^\circ)^2}{E\,[\mathrm{GeV}]} + (0.63^\circ)^2}.$$

The coordinate resolution of the drift chamber for the charged particles determines the angular resolution with respect to the azimuthal angle $\sigma_\varphi = 0.5^\circ$ and the polar angle $\sigma_\theta = 1.8^\circ$ (the polar axis is directed along the beam axis in the collider),

Seven sweeps of the ϕ-meson energy range were made in 1996 and two sweeps were made in 1998 in the range $2E = 980$–1060 MeV. The total integrated luminosity is 13.2 pb^{-1} and the total number of created ϕ-mesons was 2.1×10^7. For conciseness the sweeps will be arbitrarily designated PHI9601, PHI9602, …, PHI9802. The luminosity was measured for large-angle electron and positron elastic scattering processes and also for two-quantum annihilation. The difference between the results of the measurements by these two methods does not exceed 1%. In the present study we used the luminosity measured using elastic scattering, as this process is closer to that being studied. The accuracy of the theoretical formulas used to model the elastic scattering allow us to estimate the accuracy of the luminosity measurements at around 2%.

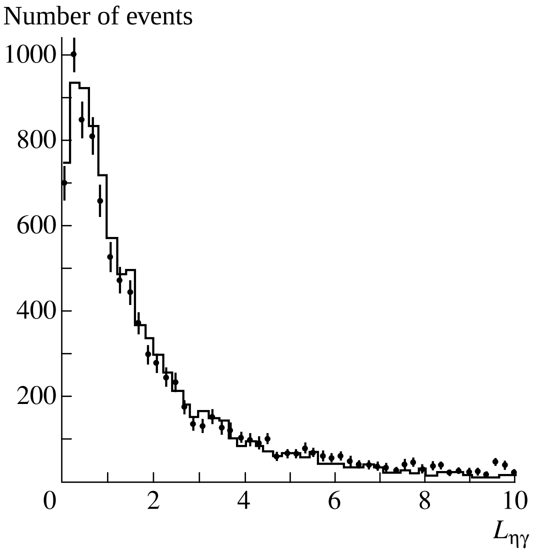


Fig. 1. Distribution of modeled events (histogram) and experiment (circles with error bars) over the parameter $L_{\eta\gamma}$.

3. EVENT SELECTION

The events of the process being studied $\phi \longrightarrow \eta\gamma$, $\eta \longrightarrow \pi^+\pi^-\pi^0$, and $\pi^0 \longrightarrow \gamma\gamma$ have two charged particles (pions) and three photons in the final state. For the analysis we selected events with two charged noncollinear particles and three or more γ-quanta. The main background process is $\pi^+\pi^-\pi^0$ creation when an "excess" photon is formed in the event by means of one of several possible mechanisms: "splitting" of photons in the calorimeter, "noise" triggering of crystals, large-angle photon emission by initial particles, and the "superposition" of particles from other events which are simulated with poor accuracy. However, because of erroneous reconstruction there is a small spurious contribution of events from $\phi \longrightarrow K_S^0 K_L^0$, $\phi \longrightarrow K^+K^-$, and $ee \longrightarrow \omega\pi^0$.

For the selected events we made a kinematic reconstruction which uses the measured angles of all particles and the photon energies taking two hypotheses:

(1) The event is a decay process

$$\phi \longrightarrow X\gamma \longrightarrow \pi^+\pi^-\pi^0\gamma \longrightarrow \pi^+\pi^-\gamma\gamma\gamma,$$

(2) The event is a background event

$$\phi \longrightarrow \pi^+\pi^-\pi^0.$$

In both cases, from all the recorded photons in the event we selected those which best satisfy the hypothesis and neglected the remaining photons (in the selected events of the effect in the experiment 27% of the events have an excess photon whereas in the $\phi \longrightarrow \eta\gamma$ modeling this figure is only 13%). In both hypotheses the π^0 mass was fixed whereas in the first hypothesis the η-meson mass (the invariant mass of the $\pi^+\pi^-\pi^0$ system) was a free parameter. As a result of kinematic reconstruction we obtained three parameters which were used for the ensuing selection: the two plausibility functions $L_{\eta\gamma}$ and $L_{3\pi}$, and the invariant mass m_η of the π^+ system. In order to simplify the profile of the plausibility function in multidimensional space and to improve the results of the numerical minimization, the minimum with respect to three angular variables corresponding to the rotation of the event as a whole, was determined analytically [10]. We give a complete list of the selection criteria:

$$\begin{gathered} R_1 < 0.5\,\text{cm}, \quad R_2 < 0.5\,\text{cm}, \quad \alpha_{12} < 130°, \quad L_{\eta\gamma} < 7, \\ L_{\eta\gamma} - L_{3\pi} < -1, \quad |m_\eta - 550| < 50\,\text{MeV}, \\ 27° < \theta_i < 153°. \end{gathered} \tag{1}$$

Here $R_{1,\,2}$ is the minimum distance between the charged particle trajectory and the beam axis (the half-height width of the distribution is approximately 0.1 cm and 50% of the events have $R_{1,\,2} < 0.06$ cm) and α_{12} is the spatial angle between the charged particles (this condition can suppress the background from collinear events, mainly from K^+K^- and also from $K_S^0K_L^0$ events). The kinematic reconstruction of events in the fundamental hypothesis $\phi \longrightarrow X\gamma$ was carried out using only particles having the polar angle $27° < \theta_i < 153°$ and the other particles were neglected. This selection was carried out to isolate small angles for background particles and also because of the lower reliability in modeling the detector edges.

These sampling conditions are weak for the process being studied. The recording probability was obtained as 16–21% (the fraction of the solid angle to the fifth power is $0.891^5 = 56\%$).

Although the distributions of the events over the parameters specified in the list of selection criteria show fairly good agreement with the experimental distributions, for each parameter we can identify a selection limit of the modeling events which is more consistent with the experimental selection limit. For these refined selection boundaries we obtain a certain event recording efficiency. A change in the efficiency when the standard selection conditions were replaced by the refined ones served as the basis for introducing a correction to the efficiency and the measure of systematic error, which is determined by the difference between the distributions for the modeling and the experiment. These corrections to the selection limits can be selected most easily for the distribution over the mass m_η where, as a result of fitting the distribution, we can obtain its full width at half-maximum and the coordinate of the distribution maximum in the form of peak parameters.

The greatest efficacy in suppressing the background is obtained by sampling over the parameter $L_{\eta\gamma}$ (Fig. 1), which requires the energy and momentum conservation laws to be satisfied and the invariant mass of two photons to be equal to the pion mass. The background from the process $e^+e^- \longrightarrow \pi^+\pi^-\pi^0(\gamma)$ is mainly suppressed by sampling over $L_{\eta\gamma}$ and $L_{\eta\gamma} - L_{3\pi}$ and for the remainder (around 10%) we made a correction based on the m_η distribution (Fig. 2). This figure gives a histogram showing the $\eta\gamma$ modeling and the circles with the error bars give the experimental results (sweep PHI9801). The full width at half-maximum of the m_η distribution is $h_\eta = 29.5 \pm 0.6$ MeV for the modeling and for the experimental distribution of the events for all the 1996 sweeps we have $h_\eta = 34.7 \pm 0.9$ MeV, for sweep PHI9801 we have $h_\eta = 29.3 \pm 1.0$ MeV, and for PHI9802 we have $h_\eta = 29.8 \pm 1.4$ MeV.

By approximating the experimental m_η distribution and assuming that the distribution of background events is close to a straight line near $m_\eta \sim 550$ MeV, we can obtain an estimate of the total number of background events in the range 500 MeV $< m_\eta <$ 600 MeV. By using the results of fitting the distributions of these experimental events over the energy $2E$ (Fig. 3), we can obtain the number of nonresonant background events. Now the number of resonant background events can be determined as the difference between the total number of background events and the number of nonresonant events. The final error in calculations of this correction in our case is largely determined by the statistical error of the nonresonant cross section. The corrections to the resonant residual background were determined separately for each sweep. Statistically these all agree with their average value (9.4 ± 0.7)%. Finally for all the sweeps we used this average correction to the residual resonant background.

The coefficient of suppression of the main background processes associated with the ϕ-meson resonance is extremely large (the recording probability is $\varepsilon_{K^+K^-} \sim 4 \times 10^{-5}$, $\varepsilon_{K_S^0K_L^0} \sim 3 \times 10^{-5}$, $\varepsilon_{3\pi} \sim 2.4 \times 10^{-4}$) and provides no basis for subtracting the remaining background according to the modeling data. The nonresonant background was subtracted automatically when fitting the ϕ-meson resonance curve with the free background.

4. RESONANCE CURVE FITTING

The detector state (the presence of nonoperating calorimeter channels, the detector triggering conditions, and so on) changed significantly from one sweep to another. Hence, separate results were obtained for each sweep so that it was ultimately possible to obtain not only an average result but also to estimate any systematic error of unknown origin which could be manifest in the instability of the result from one sweep to another. The cross section was parametrized in accordance with standard formulas for the phase volumes and the dependence of the resonance width on the energy $\Gamma(s)$ from [11] taking into account the contribution of ω- and ρ-mesons:

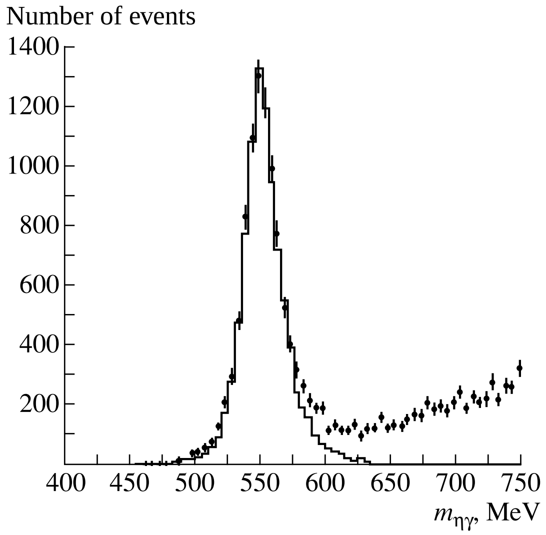


Fig. 2. Distribution of $\phi \longrightarrow \eta\gamma$ modeling events (histogram) and experiment (circles with error bars) over the parameter m_η.

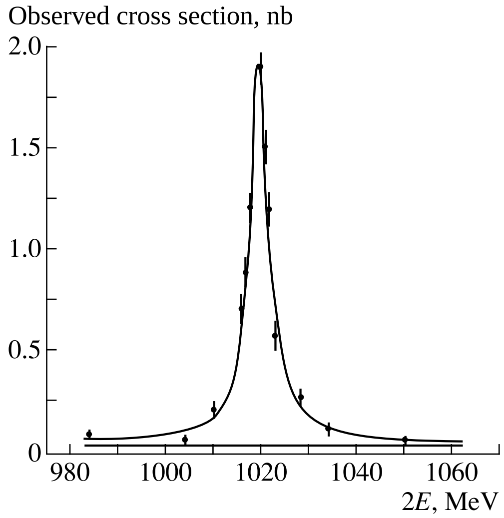


Fig. 3. Result of fitting ϕ-meson resonance excitation curve for sweep PHI9801.

$$\sigma(s) = \frac{12\pi}{\sqrt{s}}\left(\frac{m_e r_e}{\alpha}\right)^2 \times \left|\sqrt{\frac{F(s)}{s}}b + \sum_{V=\rho,\omega,\phi}\frac{\sqrt{\Gamma_{V\to ee}\Gamma_{V\to\eta\gamma}(s)m_V}\,e^{i\varphi_V}}{s - m_V^2 + im_V\Gamma_V(s)}\right|^2. \qquad (2)$$

Here the possible nonresonant contribution, in accordance with [13], is written with an arbitrary complex constant b. The phase volume $F(s)$ is given by

$$F(s) = [(s - m_\eta^2)/2\sqrt{s}]^3.$$

Table summarizing results of analysis of each sweep of the ϕ-meson energy range (statistical errors)

Sweep	L, pb^{-1}	χ^2/n_D	m_ϕ, MeV	ε, %	$B(\phi \longrightarrow \eta\gamma)$, %	$\sigma(\phi \longrightarrow \eta\gamma)$, nb
PHI9601	0.459	6.4/7	1018.96 ± 0.23	3.877	1.380 ± 0.098	58.4 ± 4.1
PHI9602	0.494	6.4/7	1019.420 ± 0.030	4.044	1.247 ± 0.074	52.6 ± 3.1
PHI9603	0.807	20.8/9	1019.366 ± 0.023	3.652	1.379 ± 0.050	58.2 ± 2.1
PHI9604	0.707	7.0/10	1019.130 ± 0.070	3.630	1.180 ± 0.068	49.8 ± 2.9
PHI9605	1.037	5.0/10	1019.053 ± 0.050	3.666	1.333 ± 0.060	56.4 ± 2.5
PHI9606	0.666	8.0/8	1019.08 ± 0.26	3.641	1.165 ± 0.083	49.3 ± 3.5
PHI9608	0.222		1019.413	3.770	1.294 ± 0.074	54.7 ± 3.1
PHI9801	4.226	16.0/13	1019.830 ± 0.108	4.750	1.179 ± 0.032	49.8 ± 1.4
PHI9802	4.541	17.3/13	1019.848 ± 0.091	4.709	1.251 ± 0.032	52.8 ± 1.4

The observed reaction cross section was fitted taking into account the radiation corrections [12] and the particle energy spread in the beam. The energy dependence of the recording efficiency was neglected: near a ϕ-meson ±20 MeV estimates using modeling events give

$$d\varepsilon/d\sqrt{s} = (6.5 \pm 8.5) \times 10^{-5}\,\mathrm{MeV}^{-1},$$

which is of the order of zero. In addition, if the efficiency has a weak, first-order dependence on energy, the contribution to the cross section is zero.

Fitting the resonance curve, even in sweeps PHI9801 and PHI9802 which have the largest number of energy points and a large integrated luminosity, yielded values of the constant b of the order of zero:

$$\begin{aligned} \text{PHI9801:}\quad \mathrm{Re}(b) &= (-0.9 \pm 1.9) \times 10^{-7}\,\mathrm{MeV}^{-1},\\ \mathrm{Im}(b) &= (0.4 \pm 5.5) \times 10^{-7}\,\mathrm{MeV}^{-1},\\ \text{PHI9802:}\quad \mathrm{Re}(b) &= (-0.1 \pm 2.4) \times 10^{-7}\,\mathrm{MeV}^{-1},\\ \mathrm{Im}(b) &= (0.3 \pm 14.8) \times 10^{-7}\,\mathrm{MeV}^{-1}. \end{aligned} \tag{3}$$

Thus, the value of this constant was subsequently fixed at zero and the cross section was parametrized using the simple model of vector dominance. Quite clearly more accurate experiments and possibly a wider energy range are needed to study this anomalous contribution.

In addition to these fitting parameters, we can also construct another parameter

$$\begin{aligned} \sigma(\phi \longrightarrow \eta\gamma) &= 12\pi\left(\frac{m_e r_e}{m_\phi \alpha}\right)^2 \\ &\times B(\phi \longrightarrow ee) B(\phi \longrightarrow \eta\gamma), \end{aligned} \tag{4}$$

which is the reaction cross section at the energy $\sqrt{s} = m_\phi$ assuming that $\Gamma(\rho \longrightarrow \eta\gamma)$ and $\Gamma(\omega \longrightarrow \eta\gamma)$ are zero.

The table gives the results of fitting for each sweep. A graphical example of the fitting result is plotted in Fig. 3.

The free parameters used to minimize the plausibility function were the ϕ-meson mass, the decay probability $B(\phi \longrightarrow \eta\gamma)$, and the nonresonant background level. If the width of the ϕ-resonance is left as a free parameter, its statistical error for sweeps PHI9801 and PHI9802 is found to be of the order of 0.3 MeV (even larger for the 1996 sweeps) whereas the tabulated value of the width is considerably more accurate. The distribution of the optimum values of the width in different sweeps around the tabulated value does not contradict the assumption that these fluctuations are of a statistical nature and thus the width, together with the two other fitting parameters, was determined taking into account the experimental accuracy of the measurements [1]:

$$\Gamma_\phi = 4.43 \pm 0.05\ \mathrm{MeV},$$

$$B(\rho \longrightarrow \eta\gamma) = (2.4 \pm 0.9) \times 10^{-4},$$

$$B(\omega \longrightarrow \eta\gamma) = (6.5 \pm 1.0) \times 10^{-4}.$$

Allowance for the tabulated accuracy of these parameters was standard: the parameter was assumed to be free but half the square of the deviation of the parameter from the tabulated value in units of the measurement error was added to the logarithmic plausibility function.

The optimum values of the mass in different sweeps have a spread considerably greater than the statistical error. This can be attributed to the inaccuracy in reproducing the energy after tuning the magnetic fields in the collider before the beginning of sweeping. The observed spread of mass values is consistent with this additional error of the order of 0.3 MeV (allowance for this error gives $\chi^2/n_D = 1$). If this error is added when averaging the mass values over different sweeps, we obtain the result

$$m_\phi = 1019.36 \pm 0.12\ \mathrm{MeV},$$

which agrees with the tabular value.

The phases of the resonances in (2) were selected as follows: $\varphi_\omega = \varphi_\rho = 0$, $\varphi_\phi = \pi$. The graph of the plausibility function as a function of the phase φ_ϕ plotted in Fig. 4 shows that the interference with ρ and ω must definitely be taken into account, but the statistics at this level are insufficient to make the phase a free parameter.

The recording efficiency is given in Table 1 allowing for $B(\eta \longrightarrow \pi^+\pi^-\pi^0) = 0.231 \pm 0.005$. Different values of the efficiency for different sweeps are obtained because of changes in the trigger and differences in the list of inoperative calorimeter crystals.

In sweep PHI9608 measurements were only made at the resonance maximum so that the mass and background level were fixed. In view of the difficulties involved in making a correct choice of values of these parameters and the small fraction of statistics in it, sweep PHI9608 was not included in the final result.

Averaging $\sigma(\phi \longrightarrow \eta\gamma)$ and $B(\phi \longrightarrow \eta\gamma)$ over all sweeps (except PHI9608), we obtain

$$\sigma(\phi \longrightarrow \eta\gamma) = 52.77 \pm 0.75 \text{ nb},$$

$$B(\phi \longrightarrow \eta\gamma) = (1.248 \pm 0.017)\%$$

with $\chi^2/n_D = 17/7$. The level of statistical agreement according to the χ^2 criterion ($P(\chi^2) = 1.7\%$) is satisfactory, although the observed appreciable spread of the results from one sweep to another can be taken into account to some extent in the statistical error. Let us assume that there is an unknown random shift of the order of 5%. If the unknown statistical error of this quantity is added quadratically to the intrinsic statistical error of each sweep, the ratio $\chi^2/n_D = 7/7$ is obtained and, as a result of averaging, we obtain

$$\sigma(\phi \longrightarrow \eta\gamma) = 53.2 \pm 1.2 \text{ nb},$$

$$B(\phi \longrightarrow \eta\gamma) = (1.259 \pm 0.030)\%.$$

In order to obtain the final result, we need to take into account various systematic errors. The statistical modeling error (1%), the systematic error in modeling the distributions used to sample events (1%), the systematic error in the luminosity measurements (2%), the reconstruction errors and the inefficiency of the chambers (1%), the trigger (1%), the indeterminacy of the phase φ_ϕ (1%), and the corrections for the residual resonance background (0.5%) together give 3.0%. Taking this into account we have

$$\sigma(\phi \longrightarrow \eta\gamma) = 53.2 \pm 1.2 \pm 1.6 \text{ nb},$$

$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma)B(\eta \longrightarrow \pi^+\pi^-\pi^0)$$

$$= (0.870 \pm 0.021 \pm 0.027) \times 10^{-6}.$$

Using the tabulated value $B(\eta \longrightarrow \pi^+\pi^-\pi^0) = 0.231 \pm 0.005$, we obtain

$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma)$$

$$= (3.765 \pm 0.092 \pm 0.143) \times 10^{-6}.$$

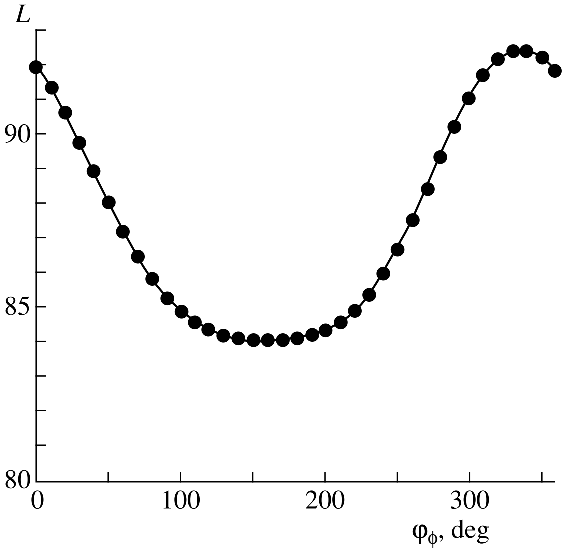


Fig. 4. Graph of plausibility function as a function of the phase φ_ϕ for sweep PHI9801.

Finally, taking from the tables [1] $B(\phi \longrightarrow ee) = (2.99 \pm 0.08) \times 10^{-4}$, we obtain

$$B(\phi \longrightarrow \eta\gamma) = (1.259 \pm 0.030 \pm 0.059)\%.$$

5. DISCUSSION OF RESULTS

We shall analyze several of the most accurate measurements of the decay probability $B(\phi \longrightarrow \eta\gamma)$:

$$B(\phi \longrightarrow \eta\gamma)$$

$$= \begin{cases} (1.246 \pm 0.025 \pm 0.057)\%, \\ \text{SND}(\eta \longrightarrow 3\pi^0)\ [4], \\ (1.338 \pm 0.012 \pm 0.052)\%, \\ \text{SND}(\eta \longrightarrow 2\gamma)\ [3], \\ (1.259 \pm 0.030 \pm 0.059)\%, \\ \text{SND}(\eta \longrightarrow \pi^+\pi^-\pi^0), \\ (1.18 \pm 0.03 \pm 0.06)\%, \\ \text{CMD-2}(\eta \longrightarrow \pi^+\pi^-\pi^0)\ [5], \\ (1.26 \pm 0.06)\%, \ \text{World aver.}\ [1]. \end{cases}$$

When results obtained in the same $\pi^+\pi^-\pi^0$ channel using the SND and CMD-2 detectors are compared, it must be borne in mind that most of the systematic error (2.7% $B(\phi \longrightarrow ee)$ and 2.2% $B(\eta \longrightarrow \pi^+\pi^-\pi^0)$) makes no contribution to the ratio of these two results:

$$\frac{B(\phi \longrightarrow \eta\gamma)[\text{SND}]}{B(\phi \longrightarrow \eta\gamma)[\text{CMD-2}]} = 1.067 \pm 0.037 \pm 0.052. \quad (5)$$

Quite clearly the agreement between these two results is good.

Similarly, the ratio of the results of measurements for different η-meson decay channels using the same SND detector contains no systematic error from $B(\phi \longrightarrow ee)$ and provides information on the ratio of the decay probabilities η in different channels:

$$\frac{B(\eta \longrightarrow \pi^+\pi^-\pi^0)}{B(\eta \longrightarrow 2\gamma) + B(\eta \longrightarrow 3\pi^0)} = 0.3141 \pm 0.0081 \pm 0.0058 \tag{6}$$

(the contributions to the systematic error are: 1% statistical modeling error, 1% reconstruction and inefficiency of the chambers, 0.5% residual resonance background, and 1% trigger), which should be compared with the tabulated data [1]:

$$\left[\frac{\Gamma(\eta \longrightarrow 2\gamma)}{\Gamma(\eta \longrightarrow \pi^+\pi^-\pi^0)} + \frac{\Gamma(\eta \longrightarrow 3\pi^0)}{\Gamma(\eta \longrightarrow \pi^+\pi^-\pi^0)}\right]^{-1} \tag{7}$$
$$= [1.70 \pm 0.05 + 1.39 \pm 0.05]^{-1} = 0.324 \pm 0.007.$$

The accuracy of the result was slightly worse than the world average, but the statistical agreement is good.

The fact that the probability $B(\phi \longrightarrow \eta\gamma)$ in all significant η-meson decay modes is measured using the same detector can also be utilized to improve the accuracy:

$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma) = (3.848 \pm 0.036 \pm 0.070) \times 10^{-6} \tag{8}$$

(here the systematic error includes: 1% modeling, 1% trigger, 1% luminosity measurements and 0.5% indeterminacy of $B(\eta \longrightarrow 2\gamma) + B(\eta \longrightarrow 3\pi^0) + B(\eta \longrightarrow \pi^+\pi^-\pi^0)$). The error in the luminosity was reduced because the luminosity in the neutral channels was measured using the two-quantum annihilation process whereas that in the charged channel was measured using large-angle elastic scattering.

Now, from the sum of the three measurements of $B(\phi \longrightarrow \eta\gamma)$ using the SND detector we can obtain

$$B(\phi \longrightarrow \eta\gamma) = (1.287 \pm 0.012 \pm 0.042)\%, \tag{9}$$

which is the most accurate measurement made using a single detector. The remaining indeterminacy is mainly determined by the error in $B(\phi \longrightarrow ee)$.

6. CONCLUSIONS

The probability $B(\phi \longrightarrow \eta\gamma)$ in the $\eta \longrightarrow \pi^+\pi^-\pi^0$ channel has been measured in experiments on the VEPP-2M electron-positron storage ring [8] using an SND detector [9], carried out over the period 1996–1998:

$$\sigma(\phi \longrightarrow \eta\gamma) = 53.2 \pm 1.2 \pm 1.6 \text{ nb},$$
$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma)B(\eta \longrightarrow \pi^+\pi^-\pi^0)$$
$$= (0.870 \pm 0.021 \pm 0.027) \times 10^{-6},$$
$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma)$$
$$= (3.765 \pm 0.092 \pm 0.143) \times 10^{-6},$$
$$B(\phi \longrightarrow \eta\gamma) = (1.259 \pm 0.030 \pm 0.059)\%.$$

Measurements of the probability $B(\phi \longrightarrow \eta\gamma)$ using the same detector simultaneously in the three most significant η-meson decay channels can be used to determine various quantities more accurately. The value of the ratio

$$\frac{B(\eta \longrightarrow \pi^+\pi^-\pi^0)}{B(\eta \longrightarrow 2\gamma) + B(\eta \longrightarrow 3\pi^0)} = 0.3141 \pm 0.0081 \pm 0.0058$$

agrees with the world average of 0.324 ± 0.007 and is of comparable accuracy.

From the sum of three measurements using the SND detector in different η-meson decay channels we can also obtain

$$B(\phi \longrightarrow ee)B(\phi \longrightarrow \eta\gamma)$$
$$= (3.848 \pm 0.036 \pm 0.070) \times 10^{-6},$$
$$B(\phi \longrightarrow \eta\gamma) = (1.287 \pm 0.012 \pm 0.042)\%,$$

which is the most accurate measurement made using a single detector.

This work was supported by the STP Integration Fund (Grant no. 274) and by the Russian Foundation for Basic Research (Grants nos. 96-15-96327, 97-02-18563, and 99-02-17155).

REFERENCES

1. Review of Particle Physics, Eur. Phys. J. C **3** (1998).
2. V. P. Druzhinin, V. N. Golubev, V. N. Ivanchenko, *et al.*, Phys. Lett. B **144**, 136 (1984).
3. M. N. Achasov, A. V. Berdyugin, A. V. Boshenok, *et al.*, Preprint No. 99-39 (Novosibirsk, Budker Institute of Nuclear Physics, 1999).
4. M. N. Achasov, S. E. Baru, A. V. Berdyugin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 573 (1998).
5. R. R. Akhmetshin, G. A. Aksenov, E. V. Anashkin, *et al.*, Preprint No. 99-11 (Novosibirsk, Budker Institute of Nuclear Physics, 1999); T. A. Purlatz, in *Proc. Intern. Workshop on* e^+e^- *Collisions from* φ *to* J/Ψ (Novosibirsk, 1999); R. R. Akhmetshin, E. V. Anashkin, V. S. Banzarov, *et al.*, E-printes Archive no. hep-ex/9907003, submitted for publication in Phys. Lett. B.
6. M. N. Achasov, M. G. Beck, K. I. Beloborodov, *et al.*, Preprint No. 97-78 (Novosibirsk, Budker Institute of Nuclear Physics,1997); M. N. Achasov, K. I. Beloborodov, A. V. Berdyugin, *et al.*, in *Proc. Conf. HADRON-97* (Brookhaven, 1997), p. 26.

7. M. N. Achasov, V. M. Aulchenko, S. E. Baru, *et al*., Preprint No. 98-65 (Budker Institute of Nuclear Physics, Novosibirsk, 1998).
8. G. M. Tumaĭkin, in *Proc. 10th Int. Conf. on High-Energy Particle Accelerators* (Serpukhov, Russia, 1977), vol. 1, p. 443.
9. V. M. Aul'chenko, B. O. Baĭbusinov, T. V. Baĭer, *et al*., *Proc. Working Meeting on Physics and Detectors for DAPHNE* (Frascati, Italy, 1991); M. N. Achasov, V. M. Aul'chenko, S. E. Baru, in *Proc. Intern. Workshop on* e^+e^- *Collisions from* ϕ *to* J/Ψ (Novosibirsk, 1999).
10. A. D. Bukin, Preprint No. 97-50 (Novosibirsk, Budker Institute of Nuclear Physics, 1997); A. D. Bukin, in *Proc. Computing in High-Energy Physics, CHEP-97* (Berlin, 1997).
11. N. N. Achasov, A. A. Kozhevnikov, M. S. Dubrovin, *et al.*, Int. J. Mod. Phys. A **7**, 3187 (1992).
12. E. A. Kuraev and V. S. Fadin, Yad. Fiz. **41**, 733 (1985).
13. M. Benayoun, S. I. Eidelman, and V. N. Ivanchenko, Z. Phys. C **72**, 221 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 173–182.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 196–206.
Original Russian Text Copyright © 2000 by Kashurnikov, Rudnev, Gracheva, Nikitenko.

SOLIDS

Phase Transitions in a Two-Dimensional Vortex System with Defects: Monte Carlo Simulation

V. A. Kashurnikov, I. A. Rudnev*, M. E. Gracheva, and O. A. Nikitenko

Moscow State Engineering-Physics Institute,
Moscow, 115409 Russia
**e-mail: rudnev@supercon.mephi.ru*

Received June 4, 1999

Abstract—The phase states and phase transitions in a system consisting of a two-dimensional vortex lattice with defects are studied by the Monte Carlo method. It is shown that a "rotating lattice" phase, which is an intermediate phase between the vortex crystal and vortex liquid phases, is present. The dependence of the temperature of the transition from the rotating lattice phase into a vortex liquid on the strength of the defect potential is determined. The current-voltage characteristics of the system are calculated at various temperatures for point, square, and linear defects. It is shown that the phase state of the system strongly affects its transport properties. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, a great deal of attention has been devoted to investigations of the dynamics and phase transformations in the vortex lattice of high-temperature superconductors (HTSCs) [1]. Experimental and theoretical investigations have been performed. The complexity of the problems arising often makes it necessary to use numerical simulation (see review [2]). It can now be positively stated that the Abrikosov lattice [3] is not the only state of a vortex system in HTSCs. Actually, as the temperature increases, the vortex lattice melts and a vortex lattice–vortex liquid phase transition is observed. A convincing proof of the existence of a phase transition has been given, for example, in [4], where the short-range correlations in the vortex lattice were analyzed by numerical simulation and the melting temperature of the vortex lattice was calculated. The melting temperature can be found from Lindemann's criterion, i.e., by comparing the standard deviation of the vortices with the period of the vortex lattice [5, 6]. A convenient criterion of a phase transition has also been investigated in [5], where the melting temperature was determined according to the features in the temperature dependence of the specific heat $C(T)$ of the vortex system.

Investigations of phase transformations in a defect field are of special interest. Thus, phase transitions from a triangular lattice of vortices into a vortex glass in the presence of a periodic system of pinning centers have been investigated [7, 8], the temperature dependence of the pinning force has been taken into account [9], and the temperature of the transition into a vortex glass has been estimated using various criteria [10–12]. One of the new and important results obtained is that the picture of phase transformations changes in the presence of defects. Thus, the question of how melting of a vortex structure occurs in the presence of pinning centers has been discussed in our previous works [13–15]. It has been shown that there exist three stages of melting. First, at low temperature, a triangular lattice starts to break up far from the pinning centers. When strict long-range order is completely absent, "islands" of the lattice form around the defects. The islands start to "rotate" around the secured pinning centers, forming a "rotating lattice" phase. At the third stage, complete melting of the lattice occurs (vortex liquid phase) [11].

The questions of the dynamic interaction of a vortex lattice with pinning centers in the presence of a transport current and the current-voltage characteristics (IVCs) obtained are important for practical applications of superconducting materials. The first Monte Carlo calculations of the IVCs appeared only very recently [16–18]. The IVCs were calculated in the presence of a large number of defects with various potential energies (relative number of vortices) [16], but defects with different activation energies and temperature dependences of the current–voltage characteristics were not investigated.

The data, obtained by analyzing the IVCs of HTSCs, on various phase regimes of current flow [9] should be especially noted (see also [19–22]). Thus, pinned vortex glass regimes, a plastic flow regime of a vortex liquid, and a moving vortex glass regime have been observed. These phase states of an Abrikosov lattice and the transitions between them are close to the phase transitions, which we have studied, between the states of a rotating lattice and a vortex liquid only in the current state. Computational results for the IVCs and the critical current of two-dimensional HTSCs in various phase states of the vortex structure are presented in [23].

In the present paper, we present a systematic picture of the phase transitions and the dynamics of a vortex structure in a two-dimensional system, simulating a layer of the high-temperature superconductor $Bi_2Sr_2CaCu_2O_x$. A situation with a low density of point defects, which act as pinning centers, is studied. Equilibrium configurations of the vortex system that correspond to various phase states will be demonstrated; the phase transition temperatures and their dependence on the defect density and the magnitude of the defect potential are calculated. Computational results are also presented for the IVCs of model layered superconductors, and they are compared with the experimental IVCs. It will be shown how the IVC is modified as a result of a change in temperature or an increase in the number of defects, how the IVC depends on the depth of the defects, and how the form of the defects influences the IVC.

All results will be obtained by Monte Carlo simulation of the vortex system. A detailed description of this method in application to calculations of vortex structures is given in Secs. 2 and 3.

2. DESCRIPTION OF THE MODEL

In most cases, the Monte Carlo (MC) method is used to investigate the behavior of the lattice of Abrikosov vortices in HTSCs.

By choosing an ensemble, such as, for example, the canonical ensemble, the MC method makes it possible to calculate the observed physical quantities with a fixed number of particles, volume, and temperature. The main advantage of the MC method is that it does not require any approximations, and the interaction is taken into account exactly. The system is described by a model Hamiltonian that is most convenient for the problem [2]. In the present work, the standard MC method together with Metropolis' algorithm for a canonical ensemble was used [24].

A two-dimensional system of Abrikosov vortices in the form of model classical particles with a long-range potential, which are arranged on a periodic square grid, is studied. The discreteness of the spatial grid is chosen so that its period is much smaller than the period of a triangular lattice. In the present model, the contribution of the interaction of vortices with one another and with defects to the energy is taken into account. In the calculations, the shortest distance between the vortices, taking account of the periodicity, is always chosen. On this basis, the model Hamiltonian has the form

$$H = \frac{1}{2}\sum_{i \neq j}^{N} H(r_i, r_j) n_i n_j + \sum_{i=1}^{N} U_p(r_i) n_i, \qquad (1)$$

where

$$\begin{aligned} H(r_i, r_j) &= \frac{\Phi_0^2 d}{2\pi\lambda^2(T)\mu_0} K_0\left(\frac{|r_i - r_j|}{\lambda(T)}\right) \\ &= U_0(T) K_0\left(\frac{|r_i - r_j|}{\lambda(T)}\right), \end{aligned} \qquad (2)$$

$U_p(r_i)$ is the vortex–defect interaction energy at the site i, n_i is the occupation number of the vortices (0 or 1) at the site i of the spatial lattice, $\Phi_0 = hc/2e$ is the magnetic flux quantum, K_0 is a modified Bessel function of order zero, d is the thickness of the superconducting layer, $\mu_0 = 4\pi \times 10^{-7}$ H m, and $\lambda(T) = \lambda(0)/\sqrt{1 - (T/T_c)^{3.3}}$ is the penetration depth of the magnetic field.

3. COMPUTATIONAL PROCEDURE

For a specific calculation, parameters close to the characteristics of the layered $Bi_2Sr_2CaCu_2O_x$ HTSC were chosen: $d = 2.7$ Å, $\lambda(T = 0) = 1800$ Å and $T_c = 84$ K. The external field was chosen in the range $B \approx 0.1$ T, which corresponds to the real scale of the magnetic induction for which melting of a vortex lattice is observed in bismuth HTSCs. The two-dimensional model is applicable in this case because the volume vortices in $Bi_2Sr_2CaCu_2O_x$ decompose in a wide range of fields (including at $B = 0.1$ T) and temperatures into flat two-dimensional "vortex pancakes" with weak coupling between the planes ($3D$–$2D$ transition). For this reason, in the present case we can talk about the quasi-two- dimensionality of the vortex structure and consider an approximate two-dimensional model.

The calculations were performed on a flat 200 × 200 spatial grid with periodic boundary conditions. The basic data were obtained for $N_v = 150$ vortices and $N_d = 5$–200 defects (the two-dimensional defect density was $\sim 10^{10}$–4×10^{11} cm^{-2}). The defects were arranged randomly on the grid. The real vortex density corresponding to a given field B was reproduced by varying the value of a division of the spatial cell so that the period a_v of a triangular lattice of vortices would satisfy the relation $a_v = \sqrt{2\Phi_0/B\sqrt{3}}$, i.e., as the external magnetic field varied, the vortex density varied. This is reflected in the model: As the magnetic field varies, the area of the region under study with a constant simulation parameter (the number of vortices) varies. Therefore, there was no need to include a direct interaction with an external field in the model (1), since the number of vortices is fixed.

We note that we are justified in treating the two-dimensional vortices as point quasiparticles because the characteristic distance between the vortices (the lattice period a_v) in a field $B = 0.1$ T is approximately $a_v \approx 2000$ Å, which is two orders of magnitude greater than the characteristic size of a vortex core $\zeta \approx 20$ Å. The linear size of the system is approximately 20000 Å.

The defects were simulated in the form of point-like wells with the same potentials, and the interaction energy of a vortex with a pinning center was taken to be of the form

$$U_p(r, T) = -\frac{U_0(T)}{8}\delta_{r, ri}. \quad (3)$$

It should be underscored that the temperature dependence of the depth of the potential well (3) was chosen to be the same as for the intervortex interaction potential. This is not always valid for a real situation, but the dependence (3), as a rule, is used in computational works (see, for example, [9, 18]).

The real picture is such that the density of vortex filaments for the magnetic fields investigated is much greater than the defect density. In addition, in the calculations we employ the defect densities which are close to realistic values [1].

The specific heat as a function of temperature was reproduced from the numerical results. The expression for the specific heat in terms of the fluctuation of the internal energy has the form ($k_B = 1$)

$$C(T) = \frac{|\langle E^2\rangle - \langle E\rangle^2|}{T^2}. \quad (4)$$

Figure 1 shows examples of the calculation of $C(T)$ and the determination of the melting temperature T_m for two values of the defect potential: $U_p = 1$ meV and $U_p = 100$ meV. The calculations were performed for defect potentials ranging from 1 to 100 meV with $N_v = 150$ vortices.

To show that the number of vortices is macroscopically large for this problem, a calculation of the specific heat was performed for the same defect potential but with a different number of vortices (for example, $N_v = 50$, 100, and 150). The results are presented in the inset in Fig. 1. It is evident that even for $N_v = 100$ and $N_v = 150$ the temperatures T_m are the same to within the limits of error, and as the number of vortices increases from 50 to 100, the peak in the specific heat shifts in the direction of high temperatures and increases in height. For $N_v = 150$ vortices the temperature of the peak no longer shifts. From this it can be concluded that the number of vortex points $N_v = 150$ is macroscopically large for our problem. Test calculations of the specific heat for a different number of MC steps were performed in the same manner.

The standard MC procedure consisted of the following. First, the vortices were distributed arbitrarily on a grid in phase space, simulating the plane of the HTSC. It is obvious that the initial arbitrary arrangement of vortices does not correspond to a minimum of the energy, so that the system is forced to evolve in order to sort the energetically most favorable configurations according to the so-called Metropolis algorithm, which realizes a Gibbs distribution. The typical total number of MC steps for the calculation was chosen to be 60 000. The number of steps required for thermalization of the system (when the system reaches stable configurations with minimum energy) is 30 000, so that there was enough time for the system to reach equilibrium in the chosen number of MC steps. The minimum change in an instantaneous configuration of the system is an elementary displacement of one vortex. By MC step we mean, as usual, a single displacement of all vortices in the system, i.e., in a MC step each vortex was shifted on the average once. Estimates showed that this number of MC steps is sufficient to calculate the specific heat with the minimum statistical errors.

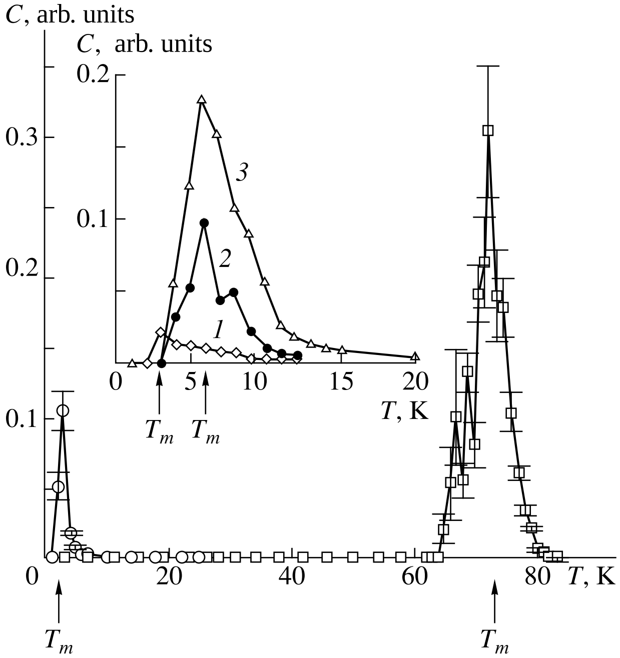


Fig. 1. Temperature dependence of the specific heat in a defect field: (○) $U_p = 1$ meV; (□) $U_p = 100$ meV. Inset: Effect of the vortex density N_V on the melting temperature: (*1*) $N_V = 50$; (*2*) $N_V = 100$; (*3*) $N_V = 150$, $U_p = 70$ meV.

To investigate dynamical processes, a transport current was introduced into the system. In this case, a term due to the action of the Lorentz force on each vortex was added to the Hamiltonian describing the behavior of the entire system. For an elementary motion of the vortex along the x axis, a term of the form $\delta U = \Phi_0 J\Delta x$ was subtracted from the total energy, if the direction of motion of the vortex coincided with the Lorentz force acting on it, or added if the vortex moved in a direction opposite to the Lorentz force. The transport current was directed along the y axis. For a nonzero current the vortex filaments were allowed to move randomly over any distance within the system. The voltage arising on the boundaries of the system was calculated from the relation $V = Bv_{dr}$, where B is the external field and

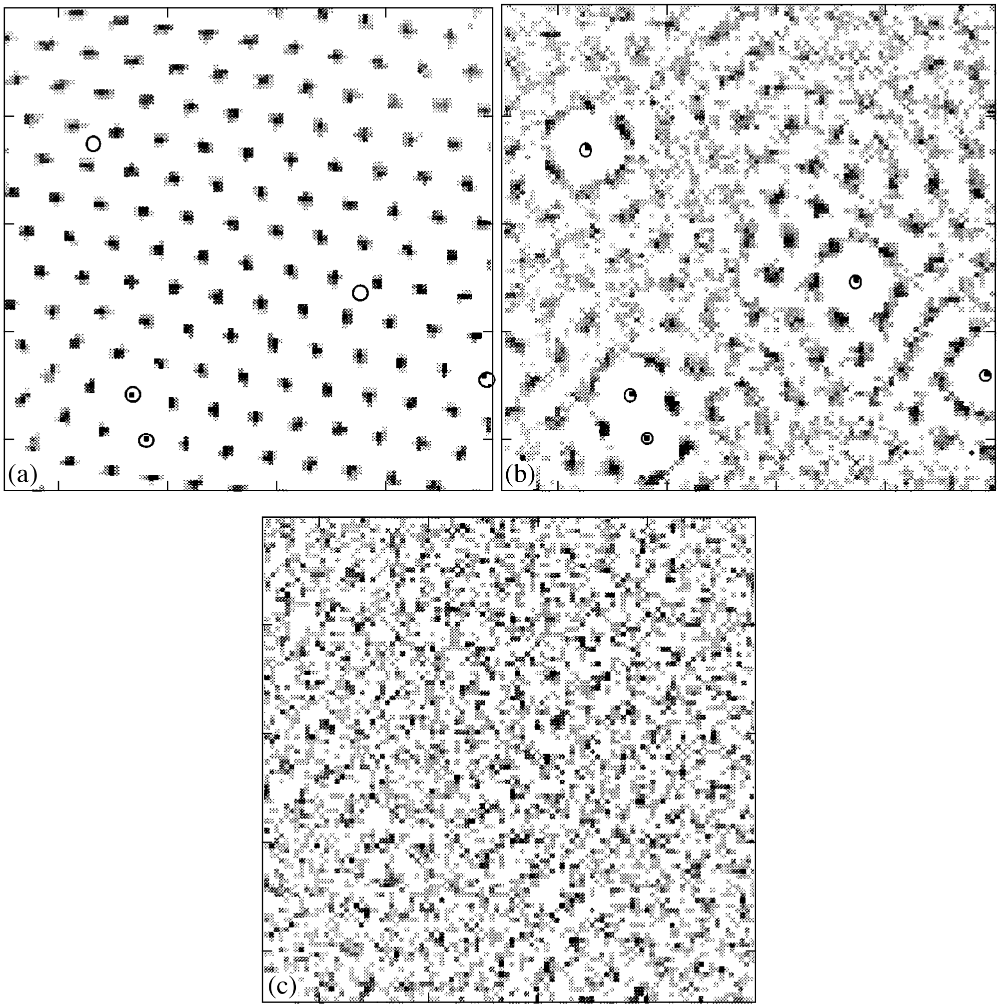


Fig. 2. Form of the vortex system for weak pinning at T = (a) 1, (b) 5, and (c) 35 K. The total vertical and horizontal linear dimensions of the system are 2 μm.

$v_{dr} = X_s/\tau$ is the vortex drift velocity, i.e., the displacement X_s of the center of mass of a vortex filament in unit time τ. Nominally, one elementary MC step is chosen as a unit of time, i.e., there is an uncertainty in the time scale that is reflected in the arbitrariness in the choice of the dimension of the voltage scale. This uncertainty can later be removed by normalizing to a real IVC.

In order to take account of the boundary conditions correctly, vortex filaments were not allowed to reach the edge of the system by a single random displacement under the action of the current. This condition actually determined the maximum possible current for the calculation and was always checked. Thus, in [22] a typical distribution of displacements of a vortex filament relative to the center of mass (actually, the distribution of drift velocities) was presented for T = 70 K and $J/J_p = 5$ ($J_p = 5 \times 10^{10}$ A/m^2). The distribution is ellipsoidal, and it is extended perpendicular to the direction of the current and does not reach the boundaries of the system.

4. COMPUTATIONAL RESULTS AND DISCUSSION

4.1. Phase Transitions in the Vortex System

We shall consider the equilibrium configurations of the vortex system at various temperatures.

A virtually ideal triangular lattice is reproduced in Fig. 2a (T = 1 K). We note that when the field was increased to 1 T, we observed a transition from a triangular into a square lattice at T = 2 K [15]. As one can see, not all defects are occupied by vortices; this attests to the stiffness of the lattice [1] at such temperatures. The defects would be occupied by vortices if the correspondence between the arrangement of the defects and the centers of the triangular lattice was ideal. At a temperature of 3 K, all defects are occupied by vortices, which rigidly hold the lattice around themselves. As a result of the irregularity of their arrangement, the defects seemingly drag around pieces of the lattice, breaking up the lattice far away from themselves. The boundaries between the coherent regions arising are melted, and the lattice loses its stiffness. We note that a similar "orientational melting" was observed in [24] in a model planar problem. As the temperature increases further (up to 5 K, Fig. 2b), it is evident that the "islands" of the triangular lattice, which are confined around the defects and possess mobility relative to a defect as a rotation axis, seemingly rotate, smearing the vortex density into concentric circles with maxima on the coodination spheres. A region depleted of vortices is formed near the defects themselves at a distance of one coordination sphere, corresponding to the period of the ideal triangular lattice, since an immobile pinned vortex does not allow other vortices to approach closer.

In this new phase state (the transition temperature is estimated to be $T_{m_1} = 3$ K), which we shall call a rotating lattice by analogy to a floating lattice [7], there is still a long-range order in the coherent regions much larger than the average distance between the vortices and rigidly pinned on the pinning centers. As the temperature increases further (right up to $T = T_{m_2} = 7.5$ K), the vortices begin to detach from the defects, the coherent regions break up, and, for example, at $T = 35$ K (already far from the transition point), a completely melted vortex liquid is observed (Fig. 2c).

Thus, the melting of the lattice in the presence of pinning centers is a three-phase process: triangular lattice–rotating lattice–vortex liquid.

We also investigated melting in the presence of defects with a 30 times higher intervortex attraction energy (i.e., U_p ($T = 2$ K) = 0.115 eV). The basic differences of this situation from the melting examined above in weaker pinning are as follows.

(a) The transition point from a triangular lattice to a rotating lattice shifted in the direction of lower temperatures ($T_{m_1} = 2$ K). Thus, displacement of the lattice already starts at $T = 2$ K, while for weak pinning and the same temperature we see an immobile triangular lattice. Pinning centers with strong attraction and minimum thermal activity of the vortices immediately break up the regular lattice into parts which then drift behind them.

(b) Strong pinning results in an expansion of the temperature range of the rotating lattice phase. Almost ideal twisted regions at intermediate temperatures are observed right up to $T = 70$–80 K.

For a quantitative investigation of the points of phase transitions with melting of the vortex lattice, ordinarily either the structure factor $S(q)$ is calculated to estimate the degree of long-range order or the hexagonal parameter S_6 [5] is calculated in order to analyze the short-range correlations (equivalent to the Lindemann criterion for the standard deviations).

As already mentioned, an alternative method is to calculate the specific heat $C(T)$ of the system taking account of the fact that, according to the fluctuation-dissipation theorem, this quantity is related with the fluctuation $\langle E \rangle$ of the total energy by the relation (4). We reproduce the function $C(T)$ for the case of weak pinning also in Fig. 1. A feature can be seen at $T_{m1} = 3$ K, corresponding to the point of the transition into a rotating lattice state. In addition, a quite sharp "front" is observed, proving that this transition is of a thermodynamic character. The function $C(T)$ is divided sharply into three stages: at temperature $T = 3$ K growth of the specific heat corresponding to T_{m1} starts (Fig. 2a); next, after this weakly expressed maximum, a sharp jump is observed at $T = 5$ K; and, finally, the final transition into the vortex liquid state corresponding to a sharp peak at $T = T_{m2} = 7.5$ K occurs. A calculation of the specific heat for strong pinning gives two clear peaks in the function $C(T)$ that correspond to the temperatures $T_{m1} = 2.5$ K and $T_{m2} = 73$ K, which corresponds completely to the previously indicated differences of the weak and strong pinning regimes.

4.2. Dependence of the Melting Temperature on the Strength of the Defect Potential

We have already examined two limiting cases: strong and weak pinning. To analyze the dependence of the phase-transition temperature on the magnitude of the defects, we calculated the temperature dependence of the specific heat of the system for various values of the defect potential. The transition temperature at which the vortex lattice, divided into islands pinned on defects, transforms into a vortex liquid phase can be determined accurately according to the characteristic features in the function $C(T)$. This is the temperature T_m (or T_{m2} according to the terminology already used above).

Having determined T_m, we obtained the dependence of the transition temperature on the defect potential (Fig. 3). It is evident that as the defect potential increases, the melting temperature increases monotonically. The results obtained can be interpreted using a simple physical picture of the processes occurring with a vortex lattice in a HTSC in the presence of defects. Consider a vortex lattice with a certain defect potential. Three interactions compete in it: the interaction of vortices with one another, which determines the stiffness of the lattice; thermal fluctuations; and, the interaction of vortices with defects. At low temperature the structure of the lattice is preserved. A vortex pinned in the potential well of the defects is localized on the well with a high probability. As the temperature increases, the lattice starts to lose its stiffness, at first far from the defects. A vortex pinned on defects holds near it an island of a coherent region of the lattice of vortices (rotating lattice phase [13]). Far from a defect the lat-

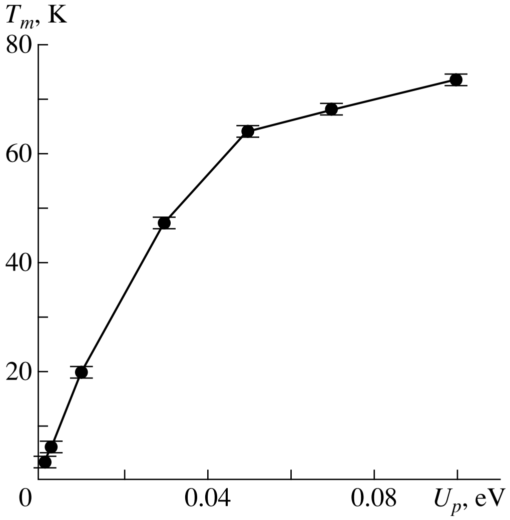


Fig. 3. Melting temperature versus the strength of the defect potential.

tice starts to break up. As temperature increases further, thermal fluctuations grow and the size of the "confined" region decreases. For this reason, the probability that a vortex (together with the coherent region that is confined) leaves the potential well increases. Finally, at even higher temperatures, the islands break up, the vortices detach from the defects, and complete melting of the vortex system occurs. It is obvious that for high values of the defect potentials a defect will confine vortices near itself at higher temperatures; this is what explains the increase in T_m with increasing depth of the potential well of defects in a complicated pattern of the rotating lattice phase. A separate question is why slower growth of T_m is observed near the critical region (Fig. 3). We believe that this is explained simply by the temperature dependence of the depth of the potential well of the defects that appears in the expression (3). Moreover, the calculations show that renormalization of the depth of the potential well, taking account of the temperature, "rectifies" the dependence $T_m(U_p)$.

It should be noted especially that experiments have repeatedly shown the existence of an intermediate phase, investigated in our model, with dynamics different from that of a triangular lattice or a vortex liquid. For example, in the experimental work [25] performed on $TmBa_2Cu_3O_{6.9}$ crystals, a scaling behavior of the normalized relaxation rate of the screening currents was obtained:

$$\frac{S}{T} = f_s\frac{H}{H_{\max}},$$

where

$$S = \frac{d\lg M_{irr}}{d\lg v}, \quad v = \frac{dH}{dt},$$

M_{irr} is the width of the hysteresis loop and v is the scan rate of the magnetic field. Three regions with different magnetic flux dynamics clearly stood out in the curve of $S(T)$ versus $H/H_{\max}$. The first region, in the opinion of the authors of [25], was qualitatively consistent with the elastic theory of a collective creep for small vortex bundles; the second was attributed by the authors to the appearance of plastic deformations in the vortex lattice; and, the third corresponded to the appearance of a vortex liquid.

The experimental proof of the presence of an intermediate phase in the phase diagram of bismuth HTSCs was given in [26], where it was stated that the solid phase below the line of melting "is obviously divided into two phases." In the first phase (I) the vortices are immobile, in the second phase (II) they begin to move, and as temperature increases further, a transition into the vortex liquid occurs. The phase I was determined to be a "weakly disordered Bragg glass," and the phase II was determined to be a "strongly disordered Josephson glass." The effect of the point pinning on the system of vortices and the formation of a reverse line in the phase diagram were qualitatively investigated in [27]. It was shown there that in the two-dimensional case the point pinning influences the vortex liquid in a definite region below the line of melting. In our case, this region corresponds to the rotating lattice phase (orientational melting)—the phase II.

In works on the experimental observation of the specific heat of a vortex system, peaks corresponding to a transition into a vortex liquid are also clearly seen in the temperature dependences of the specific heat of $YBa_2Cu_3O_{7-\delta}$ single crystals [28, 29], corresponding to a transition into a vortex liquid. The question of the nature of the transition remains open. However, knowing that experimentally, together with a peak in the specific heat, a jump is observed in the magnetization of the system, it can only be inferred that this is a first-order thermodynamic transition [28, 29].

4.3. Current-voltage Characteristics of a Vortex System

The current–voltage characteristics for systems containing from 1 to 100 defects were calculated at temperatures of 10, 20, 30 K, and so on (up to 83 K). Near the critical temperature the characteristics were calculated with an interval of 1 K right up to the critical region. The typical IVCs are presented in Fig. 4 for $T = 20$ K and different defect density.

These data were compared with the experimental IVCs (for low currents) for $Bi_2Sr_2CaCu_2O_x$ films irradiated with high-energy ions [30]. Comparing the experimental and theoretical IVCs permits determining the real scale of the electric field intensity (i.e., the quantity E_p). Thus, in our case we have $E_p = 5 \times 10^{-2}$ V/m, which indeed corresponds to the actually observed values. Moreover, if the magnetic field is known, the relaxation time τ of the vortices can be estimated. In our calculations it is $\tau \simeq 10^{-12}$ s, which also agrees well with physical estimates [1].

There is another area of agreement. According to our estimates, the characteristic critical currents obtained as a result of the calculation differ from the real currents in HTSCs by approximately a factor of 5. But, this numerical factor reflects the approximate fraction of superconducting layers (of thickness 2.7 Å) in the total volume of the unit cell of the compound $Bi_2Sr_2CaCu_2O_x$!

Let us analyze the computational data. It is evident that as the defect density increases, the IVC rectifies, increasingly demonstrating simple ohmic behavior of the system. This also occurs as the temperature approaches the critical value (see Fig. 5 below). Analysis of the initial part of the IVC (at low currents) leads to the following result: The IVC in double-logarithmic coordinates with the current and voltage (see inset in Fig. 4) is strictly linear, which confirms the presence of magnetic flux creep in this regime.

Figure 5 shows the IVCs of a system with a fixed defect density but at different temperatures.

Our computed data on the motion of a flux system in a defect field in the presence of a current demonstrate a different behavior of the IVC depending on the phase state of the system. Thus, in Fig. 5 we see two groups of curves, nominally separated by the temperature boundary $T_{m2} \simeq 70$ K. Visual analysis of the distribution density of the vortices (as in Fig. 2) shows that a rotating lattice occurs for $T < T_{m2}$, while a vortex glass occurs for $T > T_{m2}$. In the rotating lattice phase, the IVCs change little with increasing temperature. This is explained by the still strong interaction with the pinning centers. Conversely, at temperatures $T > T_{m2}$ we see a strong effect of the temperature (virtually equidistant increase of the voltage with the temperature increasing by only 2 degrees) right up to the critical region. Thus, the observed differences in the temperature behavior of the IVCs of a real HTSC can be explained by the different phase states of the vortex system which are described in Sec. 4.1.

We calculated the IVCs of compounds not only with point-like defects with the same potentials but also with defects with different potentials. The potential was chosen in the range from 20 to 50 meV. For such defects the qualitative form of the IVC did not change. In this case two more important dependences were obtained: the dependence of the critical current on the temperature and the dependence of the differential resistance $\rho = \partial E/\partial J$ on the temperature. The critical current J_c was chosen just as it is chosen in an experiment – according to a prescribed voltage threshold (in our case $E/E_p = 0.1$). The dependence $J_c(T)$ is presented in Fig. 6. A kink can be seen in the dependence. This kink corresponds to a transition from the rotating lattice phase into a vortex liquid phase. It is helpful to analyze the dependence $\rho(T)$, shown in Fig. 7, on the basis of the results of the well-known works [31, 32], where the pinning of a vortex liquid in an HTSC was studied. It was shown analytically that in a vortex liquid in a HTSC there exist two temperature ranges with different temperature

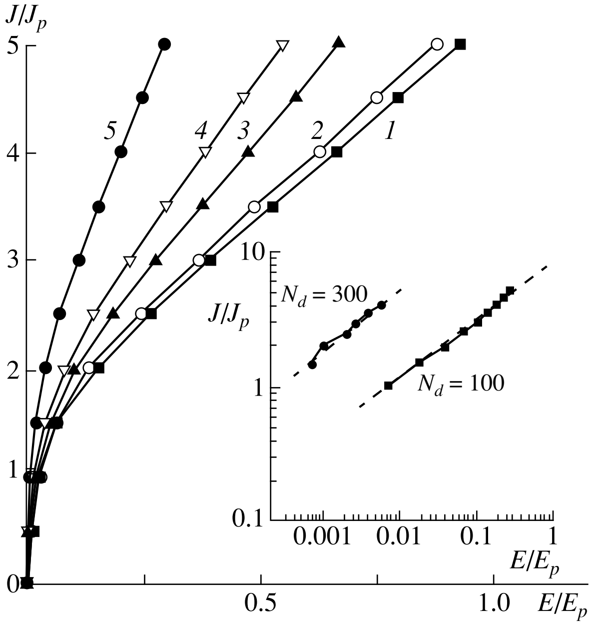


Fig. 4. Typical IVCs for $T = 20$ K for various numbers of defects N_d: (*1*) 1, (*2*) 10, (*3*) 40, (*4*) 60, (*5*) 100. Inset: IVCs for $N_d = 100$ and $N_d = 300$ in double-logarithmic coordinates.

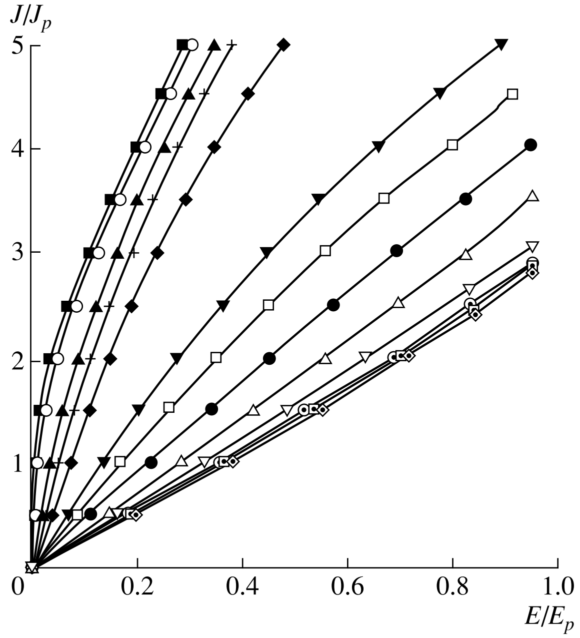


Fig. 5. IVCs for $N_d = 100$ and various temperatures (in degrees Kelvin): T = (■) 10, (○) 20, (▲) 40, (+) 50, (◆) 60, (▼) 70, (□) 72, (●) 74, (△) 76, (▽) 78, (⊙) 80, (⊡) 80, (◈) 83.

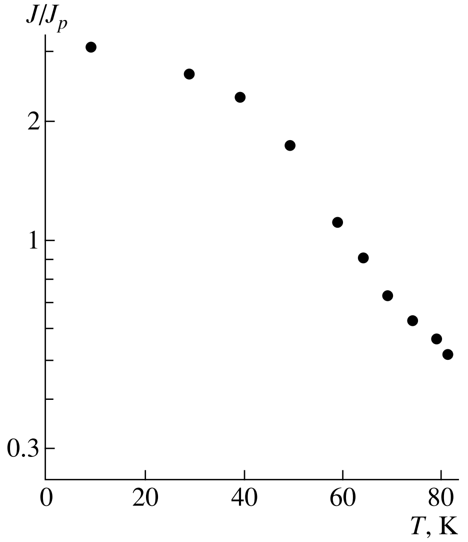


Fig. 6. Critical current versus the temperature for point defects, $N_d = 100$.

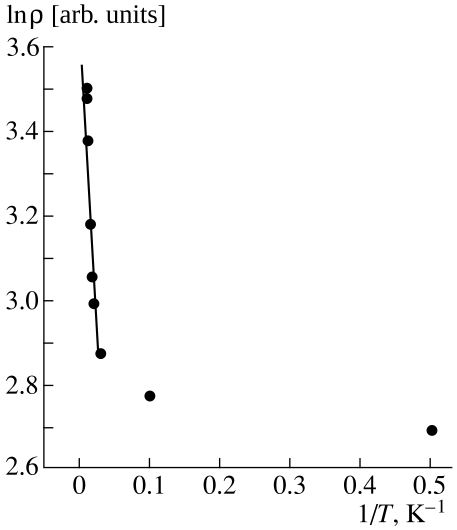


Fig. 7. Resistivity versus the temperature.

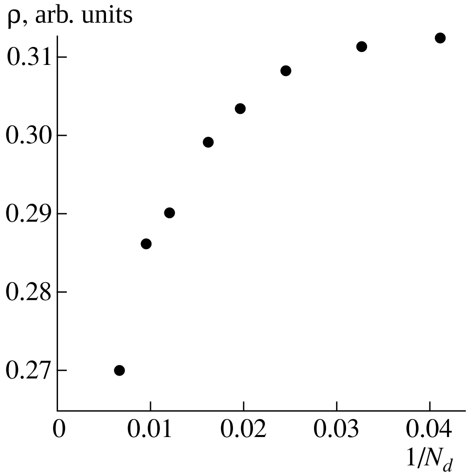


Fig. 8. Resistivity versus the number of defects.

dependences of the resistance. At high temperatures the resistivity is equal to some value ρ_{flow} which is characteristic for the flux flow regime. As the temperature decreases, a transition occurs into a partially pinned, thermally activated, flux flow regime with a linear resistivity much less than ρ_{flow} and exponentially decreasing with temperature. It was found that the resistivity which we calculated from the IVC follows an exponential law at temperature above the melting temperature of the vortex lattice; this is demonstrated in Fig. 7. In addition, the slope of the straight line in this figure is proportional to the activation energy of the vortex system. Therefore it is evident that the dependences constructed agree qualitatively with the theory of thermally activated magnetic flux flow [31, 32].

The effect of the degree of disorder γ on ρ was also examined in [31, 32], and it was stated that in the region of thermally activated flux flow the resistivity decreases with increasing γ according to the law $\rho \propto \gamma^{-1}$. We attempted to construct this dependence, taking as the degree of disorder the defect density N_d. Figure 8 shows the dependence $\rho(N_d^{-1})$. It is evident that, qualitatively, the resistivity decreases with increasing defect density, but nonlinearly. The latter behavior seems to reflect the fact that the degree of disorder is not linearly related with the defect density.

In concluding this section, we shall examine the effect of the form and size of the defects on the IVCs and the critical current of our vortex system, simulating various technological and natural defects occurring in real conductors. Any defects, differing from point defects, can be represented as a set of a definite number of point defects. There arises the question of how point defects arbitrarily distributed throughout the system and collected in clusters in the form of, for example, squares or lines affect the IVCs. For this, we introduce into the system 90 point defects by three different methods:

(1) ninety randomly distributed point defects with the same potentials (which can simulate vacancies and interstices);

(2) ten square defects consisting of nine point defects (which corresponds to pores or clusters);

(3) ten linear defects consisting of nine point defects (dislocations or twinning boundaries).

Thus, we have the same density of point defects (90), but with 10 macrodefects in the second and third cases. Likewise, to compare the same number of macrodefects and point defects, we add a fourth case: 10 point defects.

For each situation, the IVCs at different temperatures were constructed. From these curves, the four temperature dependences of the critical current shown in Fig. 9 were obtained.

Let us see how the effectiveness of defects varies with temperature. At low temperature there is virtually no difference in the square and linear defects, and point

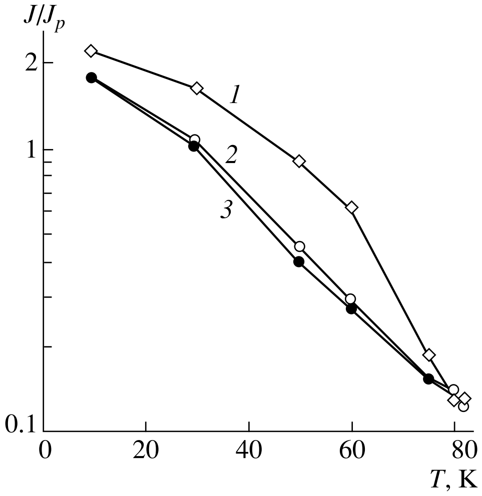


Fig. 9. Critical current versus the temperature for various defects: (*1*) 90 point; (*2*) 10 linear; (*3*) 10 square.

defects give an approximately 20% higher critical current. In the temperature range 20–60 K the effectiveness of the point defects increases rapidly. This is expressed as an increase in the critical current of a system with point defects by almost a factor of 2 for square and linear defects. At the same time, a difference appears in the effectiveness of linear and square defects. Finally, near the critical temperature the effectiveness of all defects decreases, and the difference in the type of defects completely disappears.

The result obtained can be understood on the basis of the previously considered phase states of a vortex system. Indeed, at low temperatures the vortex system is close to a state of a vortex crystal, collective pinning is effective, and therefore the number and type of defects play no substantial role. Conversely, at high temperatures the vortex system is melted, pinning is of an individual character, and the type of defects once again is not a determining factor. Differences in the types of defects are most noticeable only in the intermediate region corresponding to the rotating lattice phase. Point defects arbitrarily distributed through the system form centers of islands of a rotating lattice, on the whole confining the maximum number of vortices. This stabilizes the entire lattice. The total number of confined vortices is smaller for square and linear defects, since the possible centers of rotation are localized in macrodefects. We note that this is picture that we observe in the calculation of the visual distribution of vortices at different temperatures for different defects.

5. CONCLUSIONS

We shall summarize the main results of this work. We investigated the phase transition triangular lattice–vortex liquid in a two-dimensional vortex system simulating superconducting layers in an HTSC. In the presence of pinning the system passes through a rotating-lattice intermediate phase, whose existence can be explained by the competition between two processes against the background of an increasing thermal mobility of the vortices: the high stiffness of a triangular lattice competes with the quite strong interaction with pinning centers, striving to fix a vortex together with its surroundings, which results in the lattice breaking up into islands around defects. In layered HTSCs, which possess a lower lattice stiffness because of the weak bond between the superconducting layers, the pinning force competes with the elastic properties of the vortex structure, expanding the region of the rotating-lattice phase up to the range $T \simeq T_c$. In conventional superconductors, the high stiffness of the volume Abrikosov lattice weakens only near T_c [33], so that the temperature range of the phase state mentioned is much narrower. The fact that the temperature of the transition into the rotating phase is low (2–3 K) means that real and, therefore, defective HTSCs in a mixed state virtually immediately are in this state. This results in substantial broadening of the IVCs, even in weak magnetic fields. This could explain sometimes the absence of an Abrikosov lattice, tested according to the broadening of the IVCs, sometimes observed in thin films of conventional superconductors [33].

The defect potential strongly influences the temperature of the transition between the rotating-lattice and vortex-liquid phases. As the depth of the potential well of a defect increases, the temperature of this transition increases and the temperature range where an intermediate rotating-lattice phase occurs increases.

The calculations of the IVCs and the temperature dependences of the critical current for various types of defects showed that the phase state of a vortex system has a determining effect on the transport properties of the system. It is in the rotating-lattice phase that differences are observed in the effectiveness of point, square, and linear defects, which disappear when the vortex lattice melts.

It was also shown, on the basis of a calculation of the resistivity versus the temperature, a thermally activated flux flow regime occurs.

ACKNOWLEDGMENTS

This work was performed with the financial support of the State Scientific and Technical Program “Topical Directions in Condensed-State Physics” (subprogram “Superconductivity,” projects no. 96026 and no. 99011), and the Federal Target Program “Integration” (projects no. A-0133 and no. A-0099).

REFERENCES

1. G. Blatter, M. V. Feigelman, V. B. Geshkenbein, *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).

2. M. E. Gracheva, M. V. Katargin, V. A. Kashurnikov, *et al.*, Fiz. Nizk. Temp. **23**, 1151 (1997) [Low Temp. Phys. **23**, 863 (1997)].
3. A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957) [Sov. Phys. JETP **5**, 1174 (1957)].
4. J. W. Schneider, S. Schafroth, and P. F. Meier, Phys. Rev. B **52**, 3790 (1995).
5. S. Ryu, S. Doniach, G. Deutscher, *et al.*, Phys. Rev. Lett. **68**, 710 (1992).
6. K. Yates, D. J. Newman, and P. A. J. de Groot, Physica C **241**, 111 (1995).
7. C. Reichardt, J. Groth, C. J. Olson, *et al.*, Phys. Rev. B **54**, 16108 (1996).
8. D. Ertas and D. R. Nelson, Physica C **272**, 79 (1996).
9. S. Ryu, M. Hellerqvist, S. Doniach, *et al.*, Phys. Rev. Lett. **77**, 5114 (1996).
10. R. E. Hetzel, A. Sudbo, and D. A. Huse, Phys. Rev. Lett. **69**, 518 (1992).
11. K. Yates, D. J. Newman, and P. A. J. de Groot, Phys. Rev. B **52**, 13149 (1995).
12. R. Sasik and D. Stroud, Phys. Rev. B **52**, 3696 (1995).
13. M. E. Gracheva, V. A. Kashurnikov, and I. A. Rudnev, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 269 (1997) [JETP Lett. **66**, 291 (1997)].
14. M. E. Gracheva, V. A. Kashurnikov, and I. A. Rudnev, Phys. Low-Dim. Struct. **8/9**, 125 (1997).
15. M. E. Gracheva, V. A. Kashurnikov, and I. A. Rudnev, Phys. Low-Dim. Struct. **9/10**, 193 (1998).
16. E. Bonabeau and P. Lederer, Phys. Rev. Lett. **77**, 5122 (1996).
17. R. Sugano, T. Onogi, and Y. Murayama, Physica C **263**, 17 (1996).
18. K. Moon, R. T. Scalettar, and G. T. Zimanyi, Phys. Rev. Lett. **77**, 2778 (1996).
19. A. N. Lykov, C. Attanasio, L. Maritato, *et al.*, Supercond. Sci. Technol. **10**, 119 (1997).
20. B. Khaykovich, M. Konczykowski, E. Zeldov, *et al.*, Phys. Rev. B **56**, R517 (1997).
21. C. Goupil, A. Ruyter, V. Hardy, *et al.*, Physica C **278**, 23 (1997).
22. M. E. Gracheva, V. A. Kashurnikov, and I. A. Rudnev, Fiz. Nizk. Temp. **25**, 148 (1999) [Low Temp. Phys. **25**, 105 (1999)].
23. K. Binder and D. W. Heerman, *Monte Carlo Simulation in Statistical Physics: An Introduction* (Nauka, Moscow, 1995; Springer, Berlin, 1997).
24. Yu. E. Lozovik and E. A. Rakoch, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 268 (1997) [JETP Lett. **65**, 282 (1997)].
25. G. K. Perkins, L. F. Cohen, A. A. Zhukov, *et al.*, Phys. Rev. B **55**, 8110 (1997).
26. D. T. Fuchs, E. Zeldov, T. Tamegai, *et al.*, Phys. Rev. Lett. **80**, 4971 (1998).
27. D. R. Nelson, Physica C **263**, 12 (1996).
28. A. Schilling, R. A. Fisher, N. E. Phillips, *et al.*, Phys. Rev. Lett. **78**, 4833 (1997).
29. M. Roulin, A. Junod, A. Erb, *et al.*, Phys. Rev. Lett. **80**, 1722 (1998).
30. V. F. Elesin, I. A. Esin, I. A. Rudnev, *et al.*, Sverkhprovodimost: Fiz., Khim., Tekh. **6**, 807 (1993).
31. V. M. Vinokur, M. V. Feigelman, V. B. Geshkenbein, *et al.*, Phys. Rev. Lett. **65**, 259 (1990).
32. V. M. Vinokur, V. B. Geshkenbein, A. I. Larkin, *et al.*, Zh. Éksp. Teor. Fiz. **100**, 1104 (1991) [Sov. Phys. JETP **73**, 610 (1991)].
33. A. V. Nikulov, D. Yu. Remisov, and V. A. Oboznov, Phys. Rev. Lett. **75**, 2586 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 183–193.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 207–217.
Original English Text Copyright © 2000 by Nagaev.

Bound Ferromagnetic and Paramagnetic Polarons as an Origin of the Resistivity Peak in Ferromagnetic Semiconductors and Manganites¶

E. L. Nagaev
Institute for High Pressure Physics,
Troitsk, Moscow oblast, 142092 Russia;
e-mail: tsir@elch.chem.msu.ru

Received 16 June 1999

Abstract—A theory of resistivity is developed for ferromagnetic semiconductors, possibly, including manganites. The theory is based on analysis of the interaction of the free and bound charge carriers with the magnetization of the crystal. The temperature dependence of free energy for nonionized donors and free electrons is calculated for the spin-wave and paramagnetic regions. In addition to the trapping by the ferromagnetic fluctuations (the ferromagnetic polarons), the electron trapping by the random magnetization fluctuations as $T \longrightarrow \infty$ is taken into account (the paramagnetic polarons). For the nondegenerate semiconductors, the theory makes it possible to explain a nonmonotonic temperature dependence of the activation energy, with the value for $T = 0$ being lower than that for $T \longrightarrow \infty$. For degenerate semiconductors, the theory explains a metal–insulator transition that occurs with increasing temperature in samples with relatively low charge carrier density. If the density is larger, a reentrant metal–insulator transition should take place, so that the crystal is highly conductive as $T \longrightarrow \infty$. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The present paper deals with ferromagnetic semiconductors, both degenerate and nondegenerate. All these semiconductors display a resistivity peak in the vicinity of the Curie point T_C. The heavily doped semiconductors displaying the metallic conductivity at $T = 0$ can remain in the insulating state up to very high temperatures after passing this peak; i.e., the metal–insulator transition takes place with increasing temperature. Still more heavily doped semiconductors return to a highly conductive state after passing the peak, i.e., the metal–insulator transition is reentrant in them. The nondegenerate semiconductors have a temperature-dependent activation energy for the conductivity. This energy passes a maximum in the vicinity of the Curie point. The high-temperature activation energy exceeds its low-temperature value. All these materials display a colossal negative magnetoresistance [1].

In what follows, we talk about the donors and conduction electrons, although all the results obtained below remain in force for the holes and acceptors. A semiqualitative explanation of the properties of nondegenerate ferromagnetic semiconductors was given in [1–3]. The point is that the electron levels decrease with increasing magnetization. But the local magnetization in the vicinity of a nonionized donor is higher than the average magnetization in the crystal, which was first pointed out in [4, 5]. In fact, since the indirect exchange via free charge carries is exponentially small in nondegenerate semiconductors, the average ferromagnetic ordering is supported only by the superexchange. In contrast, the magnetization near the donor is supported also by the indirect exchange via the donor electron. Hence, at finite temperatures, the donor magnetization is destroyed to a lesser degree than the average magnetization.

The donor overmagnetization means that, with increasing temperature, the donor level depth first increases, since the conduction band bottom rises much more rapidly than the donor level. But, with further increase in the temperature, the local ordering begins to disappear. The level depth will then decrease with increasing temperature. As a result, the charge carrier density is minimal and the resistivity is maximal at a certain temperature (to avoid a misunderstanding, the magnetization dependence of the donor level depth and of the free charge carrier density was not investigated in [4, 5]).

Unfortunately, the calculation [1–3] was carried out under the assumption that the electron dwells only at the magnetic atoms nearest to the donor atom, which was also assumed in [4, 5]. On the other hand, the orbital radius must depend on the magnetization for the same reason as in the antiferromagnetic semiconductors [6]: at finite temperatures, the overmagnetized region close to the donor is a potential well for the donor electron. Hence, the electron is attracted to the donor not only by the Coulomb potential, but also by

---

¶This article was submitted by the author in English.

the magnetic potential well. As a result, the orbital radius must be magnetization-dependent and should be found by a self-consistent calculation similar to that carried out for the antiferromagnetic semiconductors in [6]. In complete analogy with the case of donors in antiferromagnetic semiconductors, where magnetized regions arise close to the nonionized impurities, one can use the term “the bound magnetic polaron”, or “the bound ferron” for the overmagnetized donors.

Calculations of the bound ferrons in the spin-wave region will be carried out below. In this case the overmagnetized region is determined by enhanced ferromagnetic correlations in the vicinity of the nonionized donor. The bound ferron radius and free energy are determined. This allows us to find the free charge carrier density and its activation energy as functions of temperature. In essence, this part of the paper develops ideas set forth for antiferromagnetic semiconductors in [1, 2], although it requires a quite different calculation procedure.

In addition to the already known low-temperature bound ferron, a new type of bound ferron will be considered. It exists in the limit $T \longrightarrow \infty$ and can be called the paramagnetic bound polaron (ferron). While the ferrons investigated so far are related to a self-consistent enhancement of the ferromagnetic correlations in the region of the electron localization, the correlations are absent here, and only the fluctuating magnetization of the region increases with decreasing size; it is of the order of $1/N_I^{1/2}$, where N_I is the number of magnetic atoms over which the donor electron is spread.

Although the mean local magnetization remains zero, the electron spin adjusts to the fluctuating magnetization, fluctuating jointly with it and thus ensuring the gain in the exchange energy between the electron and the magnetic atoms. This means that the tendency arises for the electron to be concentrated inside a region as small as possible in size. This tendency competes with the Coulomb interaction and kinetic energy in determining the equilibrium orbital radius. The shrinking of the electron orbit caused by the magnetization fluctuations can lead to a considerable lowering of the donor level as compared with its depth at $T = 0$. Hence, the low-temperature activation energy for the resistivity is less than the high-temperature activation energy. The uncorrelated fluctuations possibly also can trap the free charge carriers, but the binding energy of the free fluctuation polarons, if it is nonzero, should be very small.

Calculations carried out for nondegenerate ferromagnetic semiconductors are generalized for the degenerate semiconductors in the following way. The free energy of the impurity metal consisting of the ionized donors and electrons is calculated and compared with the free energy of separate donors. If the former energy at $T = 0$ is lower than the latter value, then an increase in T can cause a crossover, which means that there is a transition to the insulating state. As for the reverse transition at elevated temperatures, it can occur only for those systems in which the donor density markedly exceeds the density of the metal–insulator transition at $T = 0$. If this is not the case, the reverse transition is prohibited due to the paramagnetic ferrons. This result agrees with the experimental data on degenerate ferromagnetic semiconductors presented in [1].

It should be pointed out that earlier explanation of the temperature-induced metal–insulator transition in degenerate ferromagnetic semiconductors was given in terms of a modified Mott theory, in which only the stability of the metallic state was investigated [1, 7]. But this approach seems to be less fruitful than that used here. In particular, it does not lead to the insulating state as $T \longrightarrow \infty$, i.e., it does not allow one to explain some essential features of the degenerate ferromagnetic semiconductors. In addition, it can be used only if the donor density is very close to the density at which the Mott transition takes place at $T = 0$. The approach used here is more general, allowing one to overcome these limitations and to prove similarity of the physical mechanisms leading to the resistivity peak in the nondegenerate ferromagnetic semiconductors and to the metal–insulator transitions in degenerate ferromagnetic semiconductors. But some problems treated in [1, 7] remain beyond the scope of our paper (e.g., the charge carrier mobility). Thus, the present approach and that adopted in [1, 7] are complementary.

2. INDIRECT EXCHANGE HAMILTONIAN FOR A NONIONIZED DONOR

To analyze the magnetic properties of the nonionized donors, it is advisable to begin with construction of the magnetic Hamiltonian describing the indirect exchange via the donor electron. It must differ strongly from the RKKY indirect Hamiltonian since the latter assumes that the conduction electrons are completely spin-depolarized to the zero approximation. Certainly, the situation with a single donor electron is quite opposite.

As usual, the *s*–*d* model is used. The Hamiltonian of the system in the coordinate representation is given by

$$\begin{gathered} H = H_s(\mathbf{r}) + H_{sd}(\mathbf{r}) + H_{dd}, \\ H_s = -\frac{\Delta}{2m} - \frac{e^2}{\varepsilon r}, \\ H_{sd} = -A\sum_{\mathbf{g}}(\mathbf{S_g s})D(\mathbf{r}-\mathbf{g}), \\ H_{dd} = -\frac{I}{2}\sum_{\mathbf{g},\Delta}(\mathbf{S_g S_{g+\Delta}}), \end{gathered} \tag{1}$$

where $\mathbf{S_g}$ is the *d*-spin of the atom $\mathbf{g}$, $\mathbf{s}$ is the conduction electron spin, $D(\mathbf{r} - \mathbf{g})$ is equal to unity inside the unit cell $\mathbf{g}$ and zero outside it, m is the electron effective mass, ε is the dielectric constant, Δ is the vector connecting the nearest neighbors, $\hbar = 1$. The s–d exchange integral A is assumed to be positive.

In what follows, the inequality $AS \ll W$ is assumed to be met, where S is the d-spin magnitude and W is the conduction band width. This inequality is certainly met in rare-earth semiconductors (EuO, EuS, and others) [1] and can also be satisfied in transition metal compounds. In particular, it can possibly be met in colossal magnetoresistance manganites, although the experimental situation about them is not yet clear. Many investigators believe that the holes in these systems move not over the Mn ions but over the oxygen ions [8–10]. In this case, the s–d exchange is relatively weak and the band width is relatively large, in contrast to the double-exchange case where holes move over the Mn ions.

As, in the usual theory of indirect exchange, the adiabatic approximation is used when, in dealing with the s-electron, the d-spins are considered as the classical vectors. In the first approximation in AS/W, the wave function of the system cam be separated into the orbital part and the spin part:

$$\Psi(\mathbf{r}, \{S^z\}, \sigma) = \psi(\mathbf{r})\eta(\{S^z\}, \sigma), \tag{2}$$

where the normalized magnetic wave function η of the set of the d-spin variables $\{S^z\}$ and s-electron spin variable σ will be specified below as a functional of the donor ground-state orbital wave function ψ. After constructing the wave equation with Hamiltonian (1) and wave function (2), multiplying it by $\psi(\mathbf{r})$ on the left side, and integrating over $\mathbf{r}$, we obtain the wave equation for the magnetic subsystem

$$\begin{gathered} H_{av}\eta = (E - E_I)\eta, \quad E_I = \int d^3r\psi H_s\psi, \\ H_{av} = -A\sum w(\mathbf{g})(\mathbf{S_g s}), \quad w(\mathbf{g}) = \psi^2(\mathbf{g})a^3, \end{gathered} \tag{3}$$

where E_I is the energy of the s-electron bound to the impurity, and a is the lattice constant.

The eigenfunction of H_{av} is represented in the form

$$\begin{aligned} \eta(\{S^z\}, \sigma) &= \phi(\{S^z\})\delta(\sigma, 1/2) \\ &+ \chi(\{S^z\})\delta(\sigma, -1/2), \end{aligned} \tag{4}$$

where $\delta(\sigma, \pm 1/2)$ is the s-electron spin wave function with $\delta(x, y) = 1$ for $x = y$ and $\delta(x, y) = 0$ for $x \neq y$, (ϕ, χ) is the two-component wave function of the d-system.

Using equations (3) and (4), we can represent the wave equation in the form (E_I is omitted)

$$\begin{gathered} \frac{AL^+}{2}\phi + \left(E - \frac{AL^z}{2}\right)\chi = 0, \\ \frac{AL^-}{2}\chi + \left(E + \frac{AL^z}{2}\right)\phi = 0, \\ \mathbf{L} = \sum w(\mathbf{g})\mathbf{S_g}, \quad L^\pm = L^x \pm iL^y. \end{gathered} \tag{5}$$

In the particular case of $w(\mathbf{g}) = 1/N_I$, the system of equations (5) can be solved exactly. Accordingly, we use the following relations, which are valid for any function $f(S^z)$ of S^z:

$$S^- f(S^z) = f(S^z + 1)S^-, \quad L^- L^+ = L^2 - L^z(L^z + 1).$$

They follow from the definition of the S^- operator and from the commutation rules for the spin operators. The exact result for $w(\mathbf{g}) = \text{const} \equiv 1/N_I$ is

$$\left(E + \frac{A}{2N_I}\right)^2 \phi = \left(\frac{A}{2N_I}\right)^2 \left(\mathbf{L}^2 + \frac{1}{4}\right)\phi,$$

which corresponds to the effective magnetic Hamiltonian

$$H_{mI} = \frac{A}{2N_I}\left(\frac{1}{2} \pm \sqrt{\mathbf{L}^2 + \frac{1}{4}}\right). \tag{6}$$

The double sign in equation (6) corresponds to two possible spin projections of the conduction electron onto the total moment $\mathbf{L}$ of the d-spins. As should be the case, the exact eigenvalues of the Hamiltonian (6) are $(-AL/2)$ and $A(L + 1)/2$.

For an arbitrary $w(\mathbf{g})$, the system of equations (5) can be solved with accuracy of $1/2SN_I$, where N_I is the number of magnetic atoms over which the localized electron is spread. The terms of this order (omitted below) arise as a result of commuting L^- and $(E - AL^z/2)^{-1}$ after exclusion of χ from the second equation in (5). In this case, one obtains the following relation with the accuracy indicated above:

$$H_{mI} = \pm\frac{A}{2}\sqrt{\sum_{\mathbf{g}, \mathbf{f}} w(\mathbf{g})w(\mathbf{f})(\mathbf{S_g S_f})}. \tag{7}$$

In contrast to the RKKY Hamiltonian, Hamiltonians (6) and (7) are linear and not quadratic in A. More importantly, they not only describe the bilinear exchange, but the multispin exchange as well in which up to $N_I(N_I - 1)$ spins take part simultaneously. Hamiltonian (7) contains also the biquadratic terms and terms of higher orders in the scalar product of the spins, as well as terms of still more complicated structure. This is seen from equation (7), if one separates the diagonal terms with $\mathbf{g} = \mathbf{f}$ and then expands equation (7) in terms of the nondiagonal terms with $\mathbf{g} \neq \mathbf{f}$.

The strength of indirect exchange between the spins does not depend on the distance between them, but depends on the distance of each d-spin from the donor atom. Obviously, Hamiltonians (6) and (7) are isotropic, and there is no gap in the spectrum for the uniform spin rotation, as should be the case. At $T = 0$, Hamiltonians (6) and (7) correctly reproduce the s–d exchange energy for the complete ferromagnetic ordering (the latter with accuracy of $1/2SN_I$).

3. DONOR STATES AND THE RESISTIVITY PEAK IN NONDEGENERATE FERROMAGNETIC SEMICONDUCTORS (SPIN-WAVE REGION)

In this section, our first task is to calculate the free energy for a ferromagnetic system containing nonionized donors. This calculations is inapplicable to the manganites, since the undoped manganites are antiferromagnetic and only the heavily doped manganites are ferromagnetic. The problem can be solved by using a variational procedure for the free energy under condition that the donor electron is in the ground state with the wave function

$$\psi(r) = \sqrt{\frac{x^3}{\pi a_B^3}}\exp\left(-\frac{xr}{a_B}\right), \quad a_B = \frac{\varepsilon}{me^2}, \tag{8}$$

where x is the variational parameter.

In addition to the electron energy, the total free energy includes a contribution from the magnons, whose frequencies are renormalized as a result of their interaction with the donor electron, realizing the indirect exchange between the d-spins in the vicinity of the donor atom. The state of the magnon subsystem is determined from the spin-wave Hamiltonian, including the direct d–d exchange from equation (1) and the indirect exchange (7). The Hamiltonian is obtained from these equations after the Holstein–Primakoff transformation

$$S^z_{\mathbf{g}} = S - b^*_{\mathbf{g}} b_{-\mathbf{g}}, \quad S^+_{\mathbf{g}} = \sqrt{2S} b_{\mathbf{g}}, \quad S^-_{\mathbf{g}} = \sqrt{2S} b^*_{\mathbf{g}}, \tag{9}$$

where the electron distribution $w(\mathbf{g})$ corresponds to equation (8):

$$\begin{aligned} H_{mg} &= IS\sum_{\mathbf{g},\Delta}(b^*_{\mathbf{g}} b_{\mathbf{g}} - b^*_{\mathbf{g}} b_{\mathbf{g}+\Delta}) \\ &+ \frac{A}{2}\sum_{\mathbf{g},\mathbf{f}} w(\mathbf{g}) w(\mathbf{f})(b^*_{\mathbf{g}} b_{\mathbf{g}} - b^*_{\mathbf{g}} b_{\mathbf{f}}). \end{aligned} \tag{10}$$

The last term $\sim b^*_{\mathbf{g}} b_{\mathbf{f}}$ in equation (10) is basically important to ensure the absence of the gap in the magnon spectrum. But it does not influence the bulk of the magnon frequencies. For example, if $w(\mathbf{g}) = 1/N_I$, only the $q = 0$ magnon has the zero frequency. In the absence of the d–d exchange, $N_I - 1$ magnons with other wave vectors have the same frequency $A/2N_I$. Hence, in calculating the free energy, we can use the following Hamiltonian for the magnon frequencies:

$$\begin{aligned} H_{mg} &= IS\sum_{\mathbf{g},\Delta}(b^*_{\mathbf{g}} b_{\mathbf{g}} - b^*_{\mathbf{g}} b_{\mathbf{g}+\Delta}) + H(\mathbf{g}) b^*_{\mathbf{g}} b_{\mathbf{g}}, \\ H(\mathbf{g}) &= \frac{A\psi^2(\mathbf{g})a^3}{2}. \end{aligned} \tag{11}$$

However, Hamiltonian (11), written with an allowance for equation (8), is still too complicated to be diagonalized exactly. The perturbation theory also cannot be used here. To carry out an approximate calculation, we must replace the magnon potential hump $H(\mathbf{g})$ of a complicated shape in equation (11) by a rectangular potential hump with the height h and radius ρ equal to the mean height of the hump (11) and the mean radius of the electron wave function, respectively:

$$\begin{aligned} h &= \sum H(\mathbf{g})\psi^2(\mathbf{g}) = \frac{Ax^3}{16\pi b^3}, \\ \rho &= \sum \mathbf{g}\psi^2(\mathbf{g}) = \frac{3a_B}{2x}, \quad b = \frac{a_B}{a}. \end{aligned} \tag{12}$$

This means that the magnon frequency in the region close to a nonionized donor is given by

$$\begin{aligned} \omega^I_{\mathbf{q}}(x) &= \Omega_{\mathbf{q}} + h, \quad \Omega_{\mathbf{q}} = J(1-\gamma_{\mathbf{q}}), \\ J &= zIS, \quad \gamma_{\mathbf{q}} = \frac{1}{z}\sum_{\Delta}\exp(i\mathbf{q}\Delta), \end{aligned} \tag{13}$$

where z is the coordination number.

To calculate the density of the conduction electrons in a nondegenerate semiconductor, it is necessary to write the spin-wave Hamiltonian with an allowance for the conduction electrons. The relative number of the donors ζ is assumed to be small.

This makes it possible to disregard the interaction between s-electrons. We can divide all regular magnetic atoms into those which enter spheres of radius ρ surrounding donors and those which are outside these spheres (the number of the latter spheres greatly exceeds the total number of the former ones). Using the expression for the conduction-electron-magnon Hamiltonian H_{mg} (11), (13), we can represent the total electron-magnon Hamiltonian in the form

$$\begin{aligned} H &= \sum n_{I,i}\Big(E_I + \sum \omega^I_{\mathbf{q}} m_{\mathbf{q},i}\Big) \\ &+ \sum (1-n_{I,i})\Omega_{\mathbf{q}} m_{\mathbf{q},i} + \sum E_{\mathbf{k}} n_{\mathbf{k}} \\ &+ \sum B_{\mathbf{q}} n_{\mathbf{k}} m_{\mathbf{q}} + \sum \Omega_{\mathbf{q}} m_{\mathbf{q}} - \mu\sum(n_{I,i} + n_{\mathbf{k}}), \end{aligned} \tag{14}$$

where $m_{\mathbf{q},i}$ and $m_{\mathbf{q}}$ are the magnon operators for the ith sphere and outside the spheres that surrounds donors, respectively. Since the magnon number operators for different donor regions and outside them are constructed of the magnon operators $b^*_{\mathbf{g}}$ and $b_{\mathbf{g}}$ with different $\mathbf{g}$, all the operators $m_{\mathbf{q},i}$ and $m_{\mathbf{q}}$ are independent.

Further, $n_{I,i}$ and $n_{\mathbf{k}}$ are the operators for an electron in the localized state at the donor i and for the delocalized electrons with the quasimomentum $\mathbf{k}$, respectively. The spin index is absent from the electron operators, since the electrons are completely spin-polarized in the spin-wave region. For the same reason, the s–d exchange energy ($-AS/2$) is the same for all the electron states considered and therefore can be omitted as an additive constant. The quantity μ is the chemical potential.

The energy E_I of an electron at the donor is given by equations (3) and (8). At low temperatures, we can set $x = 1$ in equation (8), so that $E_I = -E_B = -e^2/2\varepsilon a_B$. The quantity $B_{\mathbf{q}}$, which describes the *s–d* interaction of the delocalized electrons with magnons when the electron quasimomenta are small compared to the magnon quasimomenta, has the form [1, 11]

$$B_{\mathbf{q}} = \frac{Aq^2}{2N(p^2+q^2)}, \quad p^2 = 2mAS, \tag{15}$$

where m is the *s*-electron effective mass, and N is the total number of atoms.

With an allowance for the mutual independence of $m_{\mathbf{q},i}$ and $m_{\mathbf{q}}$, the mean number of electrons at the donor is given by the expression (the index of the donor is omitted)

$$\begin{aligned}\langle n_I\rangle &= \sum_m \exp\left[-(E_I-\mu)/T - \sum_{\mathbf{q}} \omega^I_{\mathbf{q}} m_{\mathbf{q}}/T\right] \\ &\times\left(\sum_m \left\{\exp\left[-(E_I-\mu)/T - \sum_{\mathbf{q}} \omega^I_{\mathbf{q}} m_{\mathbf{q}}/T\right]\right.\right. \\ &\left.\left.+ \exp\left[-\sum_{\mathbf{q}} \Omega_{\mathbf{q}} m_{\mathbf{q}}/T\right]\right\}\right)^{-1}.\end{aligned} \tag{16}$$

In equation (16), the summation over m denotes summation over $m_{\mathbf{q}}$. Carrying out the summation, we find

$$\begin{aligned}\langle n_I\rangle &= \left\{1 + \exp\left(\frac{E_I-\mu}{T}\right)\frac{\Pi_{\mathbf{q}}[1-\exp(-\omega^I_{\mathbf{q}}/T)]}{\Pi_{\mathbf{q}}[1-\exp(-\Omega_{\mathbf{q}}/T)]}\right\}^{-1} \\ &= \left[1 + \exp\left(\frac{E_I+\delta F_{mI}-\mu}{T}\right)\right]^{-1}, \\ \delta F_{mI} &= F_{mI} - F^0_{mI} = N_I(f_I - f_0),\end{aligned} \tag{17}$$

where F_{mI} and F^0_{mI} are the magnon free energies for a region of radius ρ containing the nonionized and ionized donor, respectively,

$$\begin{aligned}f_I &= T\frac{a^3}{(2\pi)^3}\int d^3q \ln\left[1-\exp\left(-\frac{\omega^I_{\mathbf{q}}}{T}\right)\right], \\ f_0 &= T\frac{a^3}{(2\pi)^3}\int d^3q \ln\left[1-\exp\left(-\frac{\Omega_{\mathbf{q}}}{T}\right)\right], \\ N_I &= \frac{4\pi\rho^3}{3a^3}.\end{aligned} \tag{18}$$

A similar calculation is carried out for the mean number $\langle n_{\mathbf{k}}\rangle$ of electrons with the quasimomentum $\mathbf{k}$:

$$\begin{aligned}\langle n_{\mathbf{k}}\rangle &= \left\{1 + \exp\left(\frac{E_{\mathbf{k}}-\mu}{T}\right)\right. \\ &\left.\times\frac{\Pi_{\mathbf{q}}[1-\exp(-\Omega_{\mathbf{q}}/T - B_{\mathbf{q}}/T)]}{\Pi_{\mathbf{q}}[1-\exp(-\Omega_{\mathbf{q}}/T)]}\right\}^{-1} \\ &= \left[1 + \exp\left(\frac{E_{\mathbf{k}}+\delta F_{mC}-\mu}{T}\right)\right]^{-1},\end{aligned} \tag{19}$$

$$\delta F_{mC} = F_{mC} - F^0_{mC} = N(f_C - f_0),$$

$$f_C = T\frac{a^3}{(2\pi)^3}\int d^3q \ln\left[1-\exp\left(-\frac{\Omega_{\mathbf{q}}}{T} - \frac{B_{\mathbf{q}}}{T}\right)\right]. \tag{20}$$

Keeping in mind the fact that $B_{\mathbf{q}} \sim 1/N$, we can write

$$\delta F_{mC} = \frac{Aa^3}{2(2\pi)^3}\int d^3q \frac{q^2}{p^2+q^2}\frac{1}{[\exp(\Omega_{\mathbf{q}}/T)-1]}. \tag{21}$$

Equating the number of ionized donors to the total number of the conduction electrons, we find an expression for the charge carrier density n_{cc} for $E_{\mathbf{k}} = k^2/2m$:

$$\begin{aligned}n_{cc} &= \sqrt{nn_{eff}}\exp(-E_B + \delta F_m/2T), \\ n_{eff} &= \frac{(mT)^{3/2}}{2\sqrt{2\pi}} \quad \delta F_m = \delta F_{mI} - \delta F_{mC},\end{aligned} \tag{22}$$

where n_{eff} is the effective density of states in the conduction band, and $n = \zeta/a^3$ is the donor density.

It can be ascertained that the activation energy in equation (22) increases with temperature in the spin-wave region. It is sufficient to consider the case of $J \gg h$, $p^2 \ll 1$. Using equation (17)–(22), we find

$$\delta F_m = -\frac{7Aa^3}{32(2\pi)^3}\int d^3q \frac{1}{[\exp(\Omega_{\mathbf{q}}/T)-1]}. \tag{23}$$

In other words, δF_m is negative, and its absolute value increases with temperature. This conclusion is confirmed by numerical calculations. For example, for $A = 2$, $I = 0.02$ (in the E_B units), $\delta F_m = -0.022$ at $T = 0.01$, but $\delta F_m = -0.214$ at $T = 0.03$. The fact that the activation energy increases with temperature in the spin-wave region suggests that the resistivity peak at elevated temperatures is caused by a minimum in the charge carrier density.

4. TEMPERATURE-INDUCED METAL–INSULATOR TRANSITION (SPIN-WAVE REGION)

In this section we investigate the transition of a degenerate ferromagnetic semiconductor into the insulating state, which occurs with increasing temperature. We will compare the free energy of the highly conducting state with that of the insulating state. First, using equations (1), (8), (12), and (13), we write the total free energy of a separated nonionized donor in the E_B units:

$$F_I(x) = (x^2 - 2x) + \delta F_{mI}(x). \tag{24}$$

If one considers the term $\delta F_{mI}(x)$ in equation (24) as a small perturbation, the optimum value of x is

$$x = 1 - \frac{1}{2}\frac{d\delta F_{mI}}{dx}(1) \tag{25}$$

and, to a first approximation, the optimum free energy is

$$F_I^{opt} = -1 + \delta F_{mI}(1). \tag{26}$$

Since $\delta F_{mI}(x)$ [equations (17) and (18)] decreases with decreasing x and, hence, the last term in equation (25) is negative, the parameter x increases, and, accordingly, the electron radius decreases with the temperature. This is a manifestation of the ferron effect: the electron is dragged in by the region of the enhanced magnetization and simultaneously supports it, realizing the indirect exchange inside this region. The temperature-induced decrease in the electron radius points to the tendency toward the temperature-induced transition from metallic to insulating state if the system at $T = 0$ is metallic. In fact, while the orbit overlapping of neighboring atoms at $T = 0$ is sufficient for metallization, at finite temperatures this overlapping is insufficient and the transition to the insulating state should take place. To prove the possibility of such a transition, one should compare the free energy of separated nonionized donors to that of the impurity metal which consists of ionized donors and delocalized electrons.

Under typical conditions for degenerate ferromagnetic semiconductors, due to a relatively small electron density in them, the condition $\mu < AS$ is met (here μ is the Fermi energy [1]). In other words, the electron gas is completely spin-polarized in the spin-wave region. Using expressions for the energy of the electron gas from [12, 13], we find the following expression for the donor metal energy per donor atom (unlike the "magnetic" index m, index M denotes metal):

$$E_M^F = E(k=0) + \frac{3(6\pi^2 n)^{2/3}}{10m} + E_{ex}(n), \tag{27}$$

where $E(k = 0)$ is the electron energy at the conduction band bottom, E_{ex} is the exchange energy between conduction electrons, and n is the electron (or donor) number density. Under conditions of the complete spin polarization, $E_{ex}(n)$ can be easily obtained by generalization of the corresponding Bloch expression for the completely spin-depolarized electron gas, e. g., in [12]:

$$E_{ex}(n) = -\frac{3}{4}\left(\frac{6n}{\pi}\right)^{1/3}\frac{e^2}{\varepsilon}. \tag{28}$$

To calculate the energy $E(k = 0)$, we will use the Wigner–Seitz procedure (see, e. g., [13]). Each ionized donor is surrounded by a sphere of radius $L = (3/4\pi n)^{1/3}$. Inside each Wigner–Seitz shell, the electron wave function Φ corresponding to $k = 0$ satisfies the wave equation

$$\left(-\frac{\Delta}{2m} - \frac{e^2}{\varepsilon r} - E(k=0)\right)\Phi(r) = 0 \tag{29}$$

with the boundary condition

$$\frac{d\Phi}{dr}(L) = 0. \tag{30}$$

As is well known from the theory of metal adhesion, the wave function Φ should be almost constant with the boundary condition (30). A special analysis shows that, for relative densities ζ between 0.001 and 0.1, the Φ = const approximation ensures an accuracy in energy higher than 1%. With sufficient accuracy, we can therefore set

$$E(k=0) = -3\left(\frac{4\pi n}{3}\right)^{1/3} E_B. \tag{31}$$

With an allowance for equations (28) and (31), the energy (27) in the E_B units takes the form

$$E_M^F = \frac{3}{5}(6\pi^2\zeta)^{2/3} b^2 - (36\pi\zeta)^{1/3} - \frac{3}{2}\left(\frac{6\zeta}{\pi}\right)^{1/3} b, \tag{32}$$

where $\zeta = na^3$, and $b = a_B/a$.

At finite temperatures, the free energy of a donor metal with the volume V is given by the expression

$$\begin{gathered} G^F(n) = nVE_M^F(n) + Nf_M, \\ f_M = T\frac{a^3}{(2\pi)^3}\int d^3q \ln\left[1 - \exp\left(-\frac{\omega_{\mathbf{q}}^M}{T}\right)\right], \end{gathered} \tag{33}$$

where, with an allowance for the non-RKKY indirect exchange in our case (since $\mu < AS$), the magnon frequencies are given by the following expression [1, 11] (see also equation (15)):

$$\omega_{\mathbf{q}}^M = \Omega_{\mathbf{q}} + \frac{Aq^2\zeta}{p^2 + q^2}, \quad p^2 = 2mAS. \tag{34}$$

Equating the energy $E_M^F(n)$ (32) to the donor energy $E_I = -E_B$, we find that the density n_0, at which the electron delocalization takes place at $T = 0$, obeys the relation

$n_0^{1/3} a_B = C$, where $C = 0.208$, which is slightly lower than the value of 0.25 found by Mott.

To find the temperature of the transition from highly conducting to insulating state for a material with n exceeding n_0, one should equate the metal free energy G^F (33) to the free energy of the localized state found with the use of equations (26) and (17):

$$F^I = \zeta N(E_I + N_I f_I) + N f_0(1 - \zeta N_I).$$

For n sufficiently close to n_0, we obtain the following implicit expression for the transition temperature:

$$(\zeta - \zeta_0)\frac{d(nE_M^F)}{dn} = \zeta N_I(f_I - f_0) + (f_0 - f_M)$$
$$\zeta_0 = n_0 a^3. \tag{35}$$

Numerical calculations based on equation (32) show that the quantity $d(nE_M^F)/dn$ is negative for $\zeta < 0.2$. This does not mean that the system is unstable, since this derivative is not the electron Fermi energy. It does mean that the energy of the donor metal varies when the number of the donor atoms changes by unity. The expression on the right-hand side of equation (35) is also negative for x close to unity, which is seen from numerical calculations. The proof of this statement is especially simple in the case $S \gg 1$, if one considers the region $T_C/S \ll T \ll T_C$ and uses equations (12), (13), (15), (18), and (33) (T_C is the Curie point). This means that the equality (35) can be met for ζ that exceeds ζ_0 only moderately, and the transition from metallic to insulating state should take place with increasing temperature. But for large densities, when $\zeta > 0.2$, this transition is prohibited, at least in the spin-wave region, which agrees with experimental data of the works cited in the Introduction.

5. FLUCTUATION TRAPPING IN THE PARAMAGNETIC REGION AND RESISTIVITY OF NONDEGENERATE SEMICONDUCTORS

Calculations carried out in this section and in the next one are possibly also applicable to the manganites. First, expression (7) will be analyzed in the limit $T \longrightarrow \infty$. Although correlations between the d-spins are absent, the s–d exchange energy remains nonzero in the first order in AS/W. We see from equation (6) that in this case

$$E_{mI} = \pm\frac{A}{2}\sqrt{\frac{S(S+1)}{N_I}}. \tag{36}$$

The physical meaning of relation (36) is clear if we recall that, according to the mathematical statistics, a system of N noninteracting spins should possess a total moment on the order of $N^{-1/2}$ of their maximum moment NS. The direction of this moment is not fixed but fluctuates freely, so that its mean value should vanish. But the spin of the s-electron adjusts to the direction of the fluctuating moment and fluctuates jointly with it, providing the maximum gain in the s–d exchange energy for the energetically favored direction of the s-electron spin relative to the total spin of its localization region. This gain should be on the order of the total moment per atom, i.e., $\sim AS/N^{1/2}$, as is the case of equation (36). The term of the order of AS/W is essential only for sufficiently small orbital radii. For larger radii, the terms of the next order in AS/W should be taken into account.

Let us now consider the bound ferron at $T \gg T_C$, taking into account the fluctuation lowering of the energy discussed above. When the correlations between the d-spins are weak, the donor magnetic Hamiltonian (7), jointly with the direct d–d exchange Hamiltonian (1), can be represented in the Heisenberg form

$$H_{mP} = -\frac{A}{2}\sqrt{P} + H_H,$$

$$H_H = -\frac{1}{2}\sum_{\mathbf{g}\neq\mathbf{f}} I_{tI}(\mathbf{g}, \mathbf{f})(\mathbf{S_g S_f}),$$
$$I_{tI}(\mathbf{g}, \mathbf{f}) = \frac{A}{2\sqrt{P}}w(\mathbf{g})w(\mathbf{f}) + I(\mathbf{g} - \mathbf{f}), \tag{37}$$

$$P = S(S+1)\sum w^2(\mathbf{g}).$$

The free energy of the system is obtained by the high-temperature expansion to the first-order terms in $1/T$:

$$F^{PI} = E_I - \frac{A}{2}\sqrt{P} + F_{mP},$$

$$F_{mP} = -T\ln\operatorname{Tr}\exp\left(-\frac{H_H}{T}\right) \tag{38}$$

$$= -NT\ln(2S+1) - \frac{S^2(S+1)^2}{12T}\sum I_{tI}^2(\mathbf{g}, \mathbf{f}).$$

Calculating the electron energy E_I using the Hamiltonian H_s (1) and the trial wave function (8), and keeping in mind that the direct exchange integral $I(\mathbf{g} - \mathbf{f})$ is nonzero only in the nearest-neighbor approximation, we can write the x-dependent portion of the free energy (38) in the form (for $z = 6$)

$$F^{PI}(x) = (x^2 - 2x)E_B - Lx^{3/2} - \frac{Q(x)}{T}, \tag{39}$$

where

$$L = \frac{A}{2}\sqrt{\frac{S(S+1)}{8\pi}}b^{-3/2}, \quad b = \frac{a_B}{a},$$

$$Q(x) = \frac{L^2x^3}{12} + AIS^{3/2}(S+1)^{3/2}\sqrt{\frac{x^3b^3}{32\pi}} \times\left(1 + 2xb + \frac{4x^2b^2}{3}\right)e^{-2xb}.$$

In writing equation (39), we calculate the integral

$$\sum w(\mathbf{g})w(\mathbf{g}+\Delta)$$

in $Q(x)$ in elliptic coordinates. Here and below, the entropy term $-NT\ln(2S+1)$ is omitted from the free energy.

Minimizing the free energy (39) with respect to x, we obtain its optimum value and the inverse orbital radius in the limit $T \longrightarrow \infty$ (in the E_B and $1/a_B$ units, respectively):

$$F_\infty = -\frac{8}{3}l^3[l + \sqrt{1+l^2}] - \frac{8}{3}l\sqrt{1+l^2} - 4l^2 - 1, \quad (40)$$

$$x_\infty = [l + \sqrt{1+l^2}]^2; \quad l = \frac{3L}{8E_B} \sim \frac{AS}{(We^2/\varepsilon a)^{1/2}}. \quad (41)$$

If one sets $a_B = a$, then for AS/E_B varying from 1 to 5, the energy F_∞ varies from –1.104 to –1.659, and the radius x_∞, from 1.077 to 1.1445. Hence, the electron interaction with random (uncorrelated) magnetization fluctuations leads to a marked decrease in the donor ionization energy and in the orbital radius; this applies to any type of magnetic ordering at $T = 0$. The corresponding electron state can be called the bound paramagnetic fluctuation polaron (ferron).

Formally, random fluctuations could cause the trapping of a charge carrier in the absence of the impurity potential (the free paramagnetic ferron). In contrast to the ferron self-trapping, which occurs in the region of the enhanced magnetization, no ferromagnetic correlations between d-spins appear in the region of the electron localization. Mathematically, jointly with the solution $x = 0$ corresponding to a free electron, a solution of equation (39) with $x = 4l^2$ exists. The corresponding free energy of the trapped electron is

$$F_t = -\frac{16l^4}{3} \sim AS\left(\frac{AS}{W}\right)^3. \quad (42)$$

According to equation (42), the depth of the levels corresponding to these trapped states is very small: it is beyond the accuracy in AS/W adopted here. For this reason, the free fluctuation ferrons will not be considered in what follows. In the limit $T \longrightarrow \infty$, we obtain the following equation for the charge carrier density similar to equation (22):

$$n_{cc} = \sqrt{nn_{eff}}\exp\left(\frac{F_\infty}{2T}\right). \quad (43)$$

A comparison of equations (43) and (22) shows that the high-temperature activation energy of the conductivity $(-F_\infty/2)$ exceeds the low-temperature activation energy.

6. TEMPERATURE-INDUCED METAL–INSULATOR TRANSITION IN THE PARAMAGNETIC REGION

At finite temperatures, from equations (38) and (39) we obtain the total free energy of a system of N magnetic atoms and nV donors,

$$F(T) = nF_\infty - \frac{nVQ(x_\infty) + NS^2(S+1)^2I^2/2}{T} - NT\ln(2S+1), \quad (44)$$

and for the donor orbital radius we have

$$x(T) = x_\infty + \frac{1}{2T(1 - lx_\infty^{-1/2})}\frac{dQ}{dx}(x_\infty), \quad (45)$$

where x_∞ is given by equation (41). We can prove that the second term on the right-hand side of equation (45) is positive if the parameter

$$c = \sqrt{32\pi S(S+1)b^3}\frac{I}{A}$$

is in the range between –1 and 40. With $I > 0$, for any real parameter values, the inequality $c < 40$ is guaranteed. On the other hand, it can be satisfied even if $I < 0$, but the indirect exchange dominates ensuring the total ferromagnetic ordering at $T = 0$. In fact, the intensity of the indirect exchange is proportional to $A^2S^2\zeta^{1/3}/W$, which can exceed the intensity of the d–d exchange, zIS^2, provided that the latter quantity is small compared with A^2S^2/W. The fact that the second term in equation (45) is positive means that the radius of the donor orbital state decreases with decreasing temperature. This points to the tendency toward the electron localization at lower temperatures if the electrons are delocalized in the limit $T \longrightarrow \infty$.

Let us now investigate in more detail the temperature-induced transition from metallic to insulating state, which can occur with decreasing temperature. In the high-temperature limit, the total free energy of the donor metal is given by

$$G^{PM} = nVE_M^{PM} + \delta G^{PM}. \quad (46)$$

The energy of a nonmagnetized crystal per donor atom, instead of equation (32), is given by the following expression, which includes the correlation contribution [12]:

$$E_M^{PM} = \frac{3}{5}(3\pi^2\zeta)^{2/3}b^2 - (36\pi\zeta)^{1/3}b - \frac{3}{2}\left(\frac{3\zeta}{\pi}\right)^{1/3}b - \frac{0.113\zeta^{1/3}b}{0.1216 + \zeta^{1/3}b}, \quad (47)$$

where $\zeta = na^3$ and $b = a_B/a$.

In complete analogy with equation (38), the magnetic free energy is given by

$$\delta G^{PM} = -NT\ln(2S+1) - \frac{S^2(S+1)^2\sum I_{tM}^2(\mathbf{q})}{12T}, \quad (48)$$
$$I_{tM}(\mathbf{q}) = I(\mathbf{q}) + I_{in}(\mathbf{q}), \quad I(\mathbf{q}) = I\gamma_{\mathbf{q}}.$$

The structure of the indirect exchange integral $I_{in}(\mathbf{q})$ corresponds to the RKKY theory which can be used because electron gas is fully spin-depolarized in the paramagnetic region:

$$I_{in}(\mathbf{q}) = \frac{3nA^2a^3}{8\mu}\left(1 + \frac{4k_F^2 - q^2}{4k_F q}\ln\frac{2k_F + q}{|2k_F - q|}\right),$$
$$\mu = \frac{(3\pi^2 n)^{2/3}}{2m}, \quad k_F = \sqrt{2m\mu}. \quad (49)$$

First, it will be proved that a sample, which was in the highly conducting state at $T = 0$, can become insulating at an elevated temperature and remain nonmetallic up to arbitrarily high temperatures. This stems from the fact that fluctuations lower the donor level strongly, while no such lowering takes place for delocalized electrons. As a result, according to equation (32), the delocalization of the donor electrons at $T = 0$ occurs at the density n_0 which corresponds to the Mott-like equality $n_0^{1/3}a_B = 0.208$. But, equating the energy (47) to the energy F_∞ (40), we find that the delocalization density n_∞ at $T \longrightarrow \infty$ exceeds the value n_0 at $T = 0$ if the ratio AS/E_B exceeds 1.27. Normally, this ratio is essentially greater, and for $AS/E_B = 5$ the Mott-like relation takes the form $n_\infty^{1/3}a_B = 0.378$. Hence, normally, n_∞ exceeds n_0 considerably.

This fact results in a nontrivial temperature dependence of the electrical properties of a degenerate ferromagnetic semiconductor. For the donor density n in the range between n_0 and n_∞, the system behaves like a metal at low temperatures, but remains insulating up to arbitrarily high temperatures upon the transition from metallic to insulating state. If the density n exceeds n_∞, then the reentrant metal–insulator transition takes place with increasing temperature. This suggests a high resistivity peak at elevated temperatures of the order of the Curie temperature. Using equations (44), (46), and (48), we find the following expression for the temperature at which the temperature-induced metal–insulator transition occurs when the donor density n exceeds n_∞:

$$\frac{1}{T} = -\frac{(\zeta - \zeta_\infty)}{\zeta Q - R}\frac{d(nE_M^{PM})}{dn},$$
$$R = \frac{S^2(S+1)^2}{12(2\pi)^3}\int d^3q I_{in}^2(\mathbf{q}). \quad (50)$$

In writing equation (50), we took into account that $d(nE_M^{PM})/dn$ is negative. This fact was established by numerical calculations showing that, at least to $\zeta = 0.2$, this derivative is about -2 in the E_B units.

Numerical calculations show also that, at $I = 0$ and $\zeta = \zeta_\infty$, the denominator in equation (50) for $1/T$ is positive, which accounts for the positive transition temperature T_{tr}. This temperature decreases with increasing density ζ and depends on the direct exchange integral I. For example, at $AS/E_B = 5$, the difference $\zeta Q - R$ is equal to 0.008 for $I = 0.02$, to 0.005 for $I = 0$, and to 0.001 for $I = -0.02$ (a negative I value corresponds to initially antiferromagnetic systems such as manganites, which means that the transition form metallic to insulating state in them is also possible).

7. DISCUSSION OF THE RESULTS

The main results of the present treatment can be formulated as follows. For the nondegenerate semiconductors, it is established that the activation energy of the conductivity in the spin-wave region is determined not only by the depth of the donor level, but also by the difference in the magnon free energies for delocalized and localized electrons. As this difference increases with temperature, the activation energy E_A increases as well. In the paramagnetic region, the activation energy decreases with temperature. Qualitatively, the activation energy behaves like the difference between the local magnetization in the vicinity of a nonionized donor and the mean magnetization over the crystal: with increasing temperature, it first increases and then decreases, passing through a maximum at a temperature comparable with the Curie temperature. The resistivity peak for the nondegenerate semiconductors is located at the temperature at which $dE_A(T)/dT = 0$.

A very important result is the fact that the high-temperature activation energy exceeds the low-temperature value. This is a consequence of the fluctuation lowering of the donor level, which is caused by the fact that the moment of a region in which the localized electron dwells remains finite even when $T \longrightarrow \infty$. The direction of this moment fluctuates in space so that its mean value vanishes. But the *s*-electron spin adjusts to the direction of the moment of the region and fluctuates jointly with the moment. The gain in the *s*–*d* exchange energy therefore remains finite for the localized electron, although it diminishes with increasing size of the region. For a delocalized electron, such a fluctuation lowering is absent. Obviously, the trapping by random fluctuations is possible not only in ferromagnetic semiconductors but also in all magnetic semiconductors independently of their ground-state magnetic ordering.

Let us now discuss in greater detail the case of more heavily doped ferromagnetic semiconductors, which are in the metallic state at $T = 0$. The same reason as for nondegenerate semiconductors—increase in the stability of the localized states as compared to the delocalized states—leads to their transition from metallic to insulating state with increasing temperature. The high-temperature fluctuation lowering of the donor levels again plays an important part. Because of this circumstance, the low-temperature electron delocalization density n_0 turns out to be smaller than the high-temperature delocalization density n_∞.

There are two possible scenarios of the temperature-induced metal–insulator transition. The first corresponds to the case where the donor density exceeds n_0 but is less than n_∞. Then, with increase in temperature, the system undergoes a transition from metallic to insulating state and remains in the latter state as the temperature is raised arbitrarily high. The second scenario corresponds to the case where the donor density exceeds both n_0 and n_∞. Then, with increase in temperature, first, the transition from the highly conducting state to the insulating state takes place and then the reverse transition occurs. Obviously, the temperature range of the insulating state should decrease with increasing density. Then the reentrant metal–insulator transition should manifest itself as a resistivity peak, whose height decreases with increasing density [14].

The following remark is likely to be appropriate here. Many investigators use the terms “insulating” or “semiconducting” to denote the high-temperature state of heavily doped feromagnetic semiconductors, since the resistivity ρ decreases with increasing temperature. In doing so, they ignore the fact that the resistivity exceeds the values typical of nondegenerate semiconductors by many orders of magnitude; it is on the order of the resistivity typical of degenerate semiconductors. If one accepts this terminology, the state of a nondegenerate semiconductor in the portion of the resistivity peak where $d\rho/dT > 0$ should be called metallic, despite its giant value. Hence, this terminology is misleading.

Strictly speaking, these results correspond to the materials which are ferromagnetic when undoped. A situation is more complicated in the case where the undoped crystal is antiferromagnetic and becomes ferromagnetic as a result of doping (e.g., the manganites). The difficulty in finding n_0 is attributed to the formation of bound magnetic polarons (ferrons) in the vicinity of the donors when electrons are localized, and to the ferromagnetic–antiferromagnetic phase separation in the delocalized state of electrons which occurs at $T = 0$. It is still more difficult to investigate the temperature dependence of the critical density at low temperatures.

A direct analysis of such materials is therefore carried out here only in the high-temperature limit. It is established on the basis of equation (50) that, as the temperature is lowered, such materials can undergo the transition from metallic to insulating state. The fact that the electron is delocalized at $T = 0$ is sufficient to conclude that the resistivity peak should exist at intermadiate temperatures in these materials. Moreover, one can state that, in the materials with the initial antiferromagnetic ordering, the resistivity peak should be very close to T_C, in contrast to the materials with the initial ferromagnetic ordering, where they can be well separated. In fact, in the former case the localization of charge carriers leads to disappearance of the indirect exchange producing the ferromagnetic long-range order. Thus, after the charge carrier localization in the initially ferromagnetic materials, the ferromagnetic order is supported by the *d–d* exchange, and it is destroyed only due to thermal fluctuations of magnetization. In the initially antiferromagnetic materials, the ferromagnetic exchange disappears simultaneously.

These theoretical results, which disregard the polaronic effects and are based only on the *s–d* model, make it possible to explain electric properties of degenerate ferromagnetic semiconductors presented in [1], including doped manganites. Many investigators believe that one should take into account the Jahn–Teller (JT) and lattice polarization effects to describe properties of the manganites adequately. As for the former, it should be pointed out that the resistivity peak in the vicinity of T_C and the colossal magnetoresistance are observed in several tens of the non-JT systems [1], so that the JT effect cannot be the origin of the specific features of ferromagnetic semiconductors. The same mechanisms as in other ferromagnetic semiconductors are operative in the manganites, leading to the same specific features.

Search for additional mechanisms in the manganites would be justified, if their resistivity peaks in the vicinity of T_C were considerably higher than in other ferromagnetic semiconductors. The resistivity peak height in the manganites, however, is many orders of magnitude lower than that in ferromagnetic semiconductors such as EuO, EuS, and others. Therefore, one should formally conclude that the JT and polaronic effects rather hinder the manifestation of the particular properties of these materials. I do not insist on this conclusion but am certain that the specific features of the ferromagnetic semiconductors are not related to the lattice effects. The role of the polaron effects in manganites is discussed in more detail in [15].

ACKNOWLEDGMENTS

This investigation was supported in part by the Grant no. 98-02-16148 of the Russian Foundation for Basic Research.

REFERENCES

1. E. L. Nagaev, *Physics of Magnetic Semiconductors* (Mir, Moscow, 1983).

2. E. L. Nagaev, Fiz. Tverd. Tela **11**, 3428 (1969).
3. A. P. Grigin and E. L. Nagaev, Phys. Status Solidi B **61**, 65 (1974).
4. A. Yanase and T. Kasuya, J. Phys. Soc. Jpn. **25**, 1025 (1968).
5. T. Kasuya and A. Yanase, Rev. Mod. Phys. **40**, 684 (1968).
6. E. L. Nagaev, Zh. Éksp. Teor. Fiz. **54**, 228 (1968) [Sov. Phys. JETP **27**, 122 (1968)].
7. E. L. Nagaev, Zh. Éksp. Teor. Fiz. **92**, 569 (1987) [Sov. Phys. JETP **65**, 322 (1987)]; Phys. Rev. B **54**, 16608 (1996).
8. T. Saitoh, A. Bocquet, T. Mizokawa, *et al.*, Phys. Rev. B **51**, 13942 (1995).
9. H. Ju, H.-C. Sohn, and K. Krishnan, Phys. Rev. Lett. **79**, 3230 (1997).
10. T. Saitoh, A. Sekiyama, K. Kobayashi, *et al.*, Phys. Rev. B **56**, 8836 (1997).
11. E. L. Nagaev, Phys. Rev. B **58**, 827 (1998).
12. P. Gombas, *Theorie und Loesungmethoden des Mehrteilchenproblems der Wellenmechanik* (Basel, 1950).
13. C. Kittel, *Quantum Theory of Solids* (Wiley, New York–London, 1963).
14. E. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys. Usp. **39**, 781 (1996)].
15. E. L. Nagaev, Phys. Lett. A **258**, 65 (1999).

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 194–201.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 218–226.
Original Russian Text Copyright © 2000 by Aplesnin.

Quantum Monte Carlo Investigation of the Magnetic Properties of Weakly Interacting Antiferromagnetic Chains with an Alternating Exchange Interaction with Spin $S = 1/2$

S. S. Aplesnin
Kirenskiĭ Institute of Physics,
Siberian Division, Russian Academy of Sciences,
Krasnoyarsk, 660036 Russia;
e-mail: apl@iph.krasnoyarsk.su

Received June 21, 1999

Abstract—An approximation dependence of the spontaneous magnetic moment at a site, $\sigma/\sigma(0) - 1 = 0.71(6)\delta^{2.5(2)}$, and the antiferromagnet-singlet state phase boundary, $J_2/J_1 = 0.52(3)\delta$, are determined by the quantum Monte Carlo method in the self-consistent sublattice molecular field approximation for weakly interacting (J_2) antiferromagnetic chains with spin $S = 1/2$ and alternating exchange interaction ($J_1 \pm \delta$). The Néél temperature and a number of critical temperatures which could be related with the filling energy of two singlets ($\Delta S^z = 0$) and one triplet ($\Delta S^z = 1$) spin bands, each of which is split by the sublattice field ($h^{x,y} \neq h^z$) into two subbands, are determined on the basis of the computed correlation radii of the two- and four-spin correlation function, the squared total spin $\langle (S^z)^2 \rangle$ with respect to the longitudinal components, the dimerization parameter, and the correlation functions between the nearest neighbors with respect to longitudinal and transverse spin components. On the basis of the Monte Carlo calculations, the critical temperatures and possible energy gaps at the band center are determined for the antiferromagnets $CuWO_4$ and Bi_2CuO_4 and for the singlet compounds $(VO)_2P_2O_7$ and $CuGeO_3$, agreeing satisfactorily with existing results, and new effects are also predicted. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

There exists a wide class of magnetic compounds with spatially anisotropic distribution of exchange interactions and with a strong interaction between the magnetic and elastic subsystems that in certain cases results in a spin-Peierls transition. Ordinarily, a transition from a singlet state into the paraphrase is studied in the Hubbard or Heisenberg models with alternating exchange parameter J and hopping integral t using the mean-field theory or Green's functions together with perturbation theory. As a rule, spinon excitations are neglected in the analysis of these systems, resulting in overestimation of the temperature of the spin-Peierls transition when interchain exchange is taken into account, specifically, logarithmic behavior [1].

Alternating exchange can also be achieved by means of the geometry of the crystal lattice, such as in $CuWO_4$ [2], Bi_2CuO_4 [3, 4], $(VO)_2P_2O_7$ [5], and $(CH_3)CHNH_3CuBr_3$ [6]. These compounds are all three-dimensional magnetic systems with alternating exchange. For most of them, the exchange interactions in three directions of the corresponding crystal axes have been determined. The magnetic properties of the antiferromagnets $CuWO_4$ and Bi_2CuO_4 are interpreted in a two-sublattice Heisenberg model, and the existence of several branches of spin excitations, whose intensity becomes zero at different temperatures, and the presence of an energy gap at band center for $\omega = 1.4$ meV in $CuWO_4$ [2] and at $\omega_i = 0.7$, 1.7, 2.3, 3.4, and 4 meV in Bi_2CuO_4 [4, 7, 8] remain incomprehensible. Nonmonotonic temperature behavior of the susceptibility [9], the antiferromagnetic resonance field, and the linewidth in Bi_2CuO_4, whose temperature derivatives have several maxima [10], and an additional maximum of the specific heat at $T \simeq 17$ K ($T_N = 45$ K) in Bi_2CuO_4 [11], are observed in these antiferromagnets.

Several energy gaps in the spin excitation spectrum, which do not fit either into the conventional theory of the spin-Peierls transition with one triplet gap [12] or into the theory of the two-magnon excitation spectrum [13], have also been found in the singlet magnets $CuGeO_3$, $Na_2V_2O_5$, and $(VO)_2P_2O_7$. Of these compounds, $CuGeO_3$ has been studied in greatest detail. In this compound three temperature ranges have been found, $T_{c1} \sim (4–7)$ K, $T_{c2} \simeq 14$ K, and $T_{c3} \sim (20–25)$ K, where the EPR linewidth and intensity exhibit anomalous behavior [14, 15], and the magnetic thermal conductivity [16] and magnetostriction [17] possess maxima below and above the spin-Peierls transition temperature $T_N = 14$ K.

The present paper is devoted to an investigation of the region of stability of long-range antiferromagnetic order in an isotropic 3D antiferromagnet with a quite strong anisotropic distribution of exchange interactions in the lattice relative to the magnitude of the alternating exchange, and the determination of the site magnetic moment, the Néél temperature, and the critical temperatures at which the correlation radii assume their maximum values. According to the dynamic scaling hypothesis, the relaxation time τ is proportional to the correlation radius, $\tau \propto \xi^z$, and the temperatures indicated above can be found from the temperature dependence of the EPR linewidth, antiferromagnetic resonance, and diffuse neutron scattering. Additional spin excitations, spinons [18], will be proposed on the basis of the four-spin correlation function. These excitations have several excitation bands, which make it possible to explain previously reported experimental results and, using the computed values of the critical temperatures, predict the existence of additional spin modes and a number of new effects.

2. MODEL AND METHODS

Let us consider a Heisenberg model with negative interactions between nearest neighbors with spin $S = 1/2$ in an external magnetic field oriented in the Z direction. The alternating interaction is taken in the strong-coupling direction $I = J_1 + \delta$ and $K = J_1 - \delta$. The Hamiltonian has the form

$$H = -J_1\sum_{i,j}\mathbf{S}_{i,j}\mathbf{S}_{i+1,j} - \sum_{\substack{i,j,\gamma = 1\\ \alpha = x,y,z}} J_2^{\alpha}(\gamma)S^{\alpha}_{i,j}S^{\alpha}_{i,j+\gamma} - \sum_i H_i S_i^z, \tag{1}$$

where $J_1 < 0$ and $J_2 < 0$ are the intra- and interchain interactions, H is the external magnetic field, and γ signifies summation over the nearest neighbors between chains ($z = 4$). We transform the Hamiltonian of the $3D$ system to a one-dimensional chain of spins, which interact with the effective field, by means of the self-consistent molecular field approximation [19, 20]:

$$H = -\sum_{i=1}^{L/2} I_{2i,2i-1}\mathbf{S}_{2i}\mathbf{S}_{2i-1} - \sum_{i=1}^{L/2} K_{2i,2i+1}\mathbf{S}_{2i}\mathbf{S}_{2i+1} - \sum_{i=1}^{L} h_i^{\alpha}S_i^{\alpha} - \sum_{i=1}^{L} H_i S_i^z - 2NJ_2m_0^2, \tag{2}$$

where m_0 and h are the sublattice magnetization and field $\mathbf{h}_i(h^z, h^+, h^-)$, determined in [19, 20] as $m_0 = (1/L)\sum_{i=1}^{L}(-1)^i\langle S_i^z\rangle$ and $h = -4J_2m_0$. To take account of the quantum and temperature fluctuations, we shall determine these quantities from the spin–spin correlation function, which is a power-law function of on distance in a $1D$ antiferromagnet at $T = 0$. We shall assume that this dependence also holds also for the transverse spin components in the magnetically ordered region of a quasi-one-dimensional antiferromagnet, while in the paramagnetic and singlet states the instantaneous values of the sublattice field are proportional to the magnitude of the short-range order, i.e., the spin-spin correlation function of the nearest neighbors. The average value is $\langle h_i\rangle \approx 0$, and $\langle h_i^2\rangle \neq 0$ in the singlet state. Taking account of the fluctuations of the sublattice field in the singlet state leads to new effects, which will be described below. We shall study in the singlet and paramagnetic states two types of the sublattice field with respect to transverse spin components: isotropic $h^x = h^y = h^z$ and anisotropic, characteristic for $CuGeO_3$, $h^x = h^y = 1.4h^z$. In an antiferromagnet the sublattice fields have the form

$$\begin{aligned} m_0 &= \frac{2}{L}\sum_{i=L}^{L/2}\sqrt{\mathrm{abs}(S_0^zS_i^z)},\\ h_i^z &= 4J_2\,\mathrm{sign}(\langle S_0^zS_i^z\rangle)\sqrt{\mathrm{abs}(S_0^zS_i^z)},\\ b - h_i^{+,-} &= 4J_2(-1)^i\sqrt{\mathrm{abs}(S_0^+S_1^-)/i}. \end{aligned} \tag{3}$$

In this work, the quantum Monte Carlo method, which employs a trajectory algorithm of world lines, based on a transformation of a D-dimensional quantum system into a $(D + 1)$-dimensional classical system by discretization of the path integral in the space (imaginary time $0 < \tau < 1/T$, coordinate) [21, 22], is used. In the Monte Carlo calculations, Trotter's formula with the parameter $m = 32$, 64, 124, and 200 and periodic boundary conditions on a chain of length $L = 100$, 200, and 400 is used. One Monte Carlo step was determined by rotating all spins on a $L \times 2m$ lattice. From 4000 to 7000 Monte Carlo steps per spin were used to reach equilibrium, and 2000–5000 Monte Carlo steps per spin were used for averaging. The autocorrelation time τ required to establish thermodynamic equilibrium was estimated from the relation $\ln(\tau) = amT/J$ (T is the temperature) [23]. The systematic error due to quantum fluctuations is proportional to $\sim 1/(mT/J)^2$ and is of the order of 4% for the minimum temperature $T/J = 0.025$, used in the calculations. The rms errors of the computed quantities lie in the range (0.1–0.6)% for energy, (6–11)% for susceptibility, and ~10% for the correlation radius. The errors due to the finite dimensions of the lattice can be neglected, since $\xi < L/2$.

Let us consider the possible spin excitations in this model. If the wave function of the ground state is schematically represented as a sum of the Néél configuration and a set of singlet states of spins with different weight ratios, then, besides ordinary excitations of the spin-wave type, there can exist excitations in singlet regions that can be divided into two groups: the longitudinal component of the spin vector does not change, i.e., $\Delta S^z = 0$ (we call such excitations singlet excitations), and the longitudinal component of the spin vector changes by one unit, i.e., $\Delta S^z = 1$, which correspond

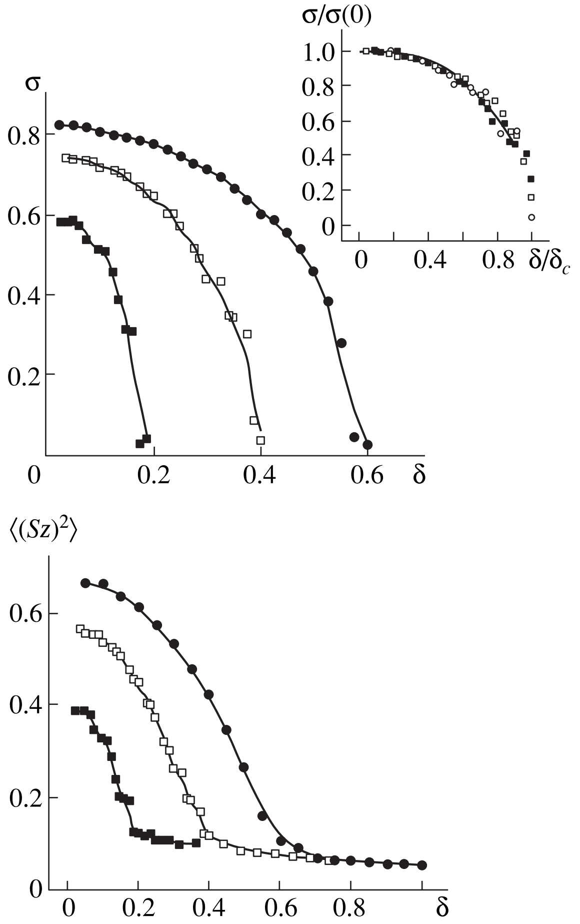


Fig. 1. Site magnetic moment (σ) and squared longitudinal component of the total spin ($\langle (S^z)^2\rangle$) of an antiferromagnet with $\lambda = 0.1$ (■), 0.2 (□), and 0.3 (●) versus the exchange alternation. Inset: Normalized values of the magnetization for the same parameters.

to triplet excitations. According to Anderson's theory [24], the singlet state is described well by a generalized resonance valence bond (RVB) model whose wave function is represented in the form of singlet pairs of spins over all possible configurations. In the presence of exchange alternation the generalized model reduces to a simple RVB model in which pairing of nearest spins is taken into account. Since two types of exchange interactions, differing in magnitude, exist here, the energies of the singlet pairs and the corresponding excitations on *K* bonds differ from the energy on *I* interactions. For this reason, the characteristic features of the temperature behavior of the magnetic characteristics of an antiferromagnet with alternating exchange can be calculated and understood on the basis of the four-spin correlation function of spin pairs $\langle \mathbf{S}_0\mathbf{S}_1\mathbf{S}_r\mathbf{S}_{r+1}\rangle$ and the dimmer ordering parameter *q*:

$$q^\alpha = \frac{4}{L}\sum_{i=2}^{L/4}\left(\langle S_0^\alpha S_1^\alpha S_{2i-2}^\alpha S_{2i-1}^\alpha\rangle - \langle S_1^\alpha S_2^\alpha S_{2i-1}^\alpha S_{2i}^\alpha\rangle\right), \tag{4}$$
$$\alpha = x, y, z.$$

A classical excitation of the kink type, which in what follows we shall call a spinon excitation, corresponds to the excited state in the simple RVB model with one dangling bond. We shall determine the correlation radius ξ_4 of the spinons and the parameter η_4 from the four-spin correlation function

$$\left|\langle S_0^z S_1^z S_{r-1}^z S_r^z\rangle - \langle S_0^z S_1^z S_{L/2-1}^z S_{L/2}^z\rangle\right| = \frac{A}{r^{\eta_4}}\exp(-r/\xi_4), \tag{5}$$

where $r = 2i + 1$ and $i = 1, 2, 3 \ldots$.

The following quantities will be calculated below: the energy, the specific heat $C = dE/dT$, the magnetization, the susceptibility $\chi = M/H$ in an external field, the spin-spin correlation function between longitudinal $\langle S^z(0)S^z(r)\rangle$ and transverse $\langle S^+(0)S^-(1)\rangle$ spin components, the Fourier spectrum $S(q) = (2/L)\sum_{r=1}^{L/2}\exp(-iqr)(S_0^z S_r^z)$, and the magnetic structure factor. We shall determine the correlation radius ξ_2 and the parameter η_2 from the spin-spin correlation function

$$\left|\langle S^z(0)S^z(r)\rangle - \langle S^z(0)S^z(L/2)\rangle\right| = \frac{B}{r^{\eta_2}}\exp(-r/\xi_2). \tag{6}$$

The squared total spin will be calculated from the longitudinal component $\langle (S^z)^2\rangle$. This parameter makes it possible to distinguish a singlet state from a paramagnetic state and is sensitive to a change in the spin excitation spectrum.

3. DISCUSSION

We shall use a number of criteria to determine the antiferromagnet–singlet state phase boundary in the interchain interactions–alternating exchange plane: the sublattice magnetization is zero, $\sigma \longrightarrow 0$, and the correlation radii ξ_2 and ξ_4 for δ_c have their maximum value. The singlet state is distinguished from the paramagnetic state or the spin-glass state according to the following indicators. In the singlet state, in a model with alternating exchange, the total spin is zero, $S = 0$, and the eigenvalue of the operator $\hat{S}_z^2$ is also zero on the basis of the equality $\langle (S^z)^2\rangle = S(S + 1)/3$. The dimmer ordering parameter is different from zero, $q \neq 0$, and the relation $\langle S_0^+ S_1^-\rangle \simeq 2\langle S_0^z S_1^z\rangle$ holds between the longitudinal and transverse components of the spins. We shall calculate the characteristics indicated above, some of which are displayed in Fig. 1, at low temperatures, $(0.1\text{–}0.2)T_N$, for a number of values of the interchain exchange parameters $\lambda = J_2/J_1 = 0.05, 0.075, 0.1, 0.125, 0.15, 0.2, 0.25$, and 0.3 as a function of the magnitude of alternation. The normalized values of the sublattice magnetization and the energy can be approximated well by power-law dependences $\sigma/\sigma(0) - 1 = 0.71(6)\delta^{2.5(2)}$ and

$E/E(0) - 1 = 0.02\delta^{3.6(3)}$, where $\sigma(0) = 1.9(1)\sqrt{\lambda}$, and are shown in the inset in Fig. 1. In the singlet state, the absolute value of the internal energy increases with the exchange alternation, $(E - 0.85) \approx 0.63\delta^{1.2(1)}$, which agrees well with the results for a one-dimensional chain, $\sim\delta^{4/3}$ [25]. The difference in the exponent could be due to the correlation effects of the interaction of the chains, which are taken into account in the form of the self-consistent sublattice fields **h** (3). The magnetic state for $\delta > \delta_c$ is a singlet state, and the finite quantity $\langle (S^z)^2 \rangle$ is due to the singlet excitations with $\Delta S^z = 0$, since Monte Carlo calculations are performed at finite temperatures. The phase boundary of the transition can be approximated well by the linear function $\lambda = 0.52(3)\delta$.

For an antiferromagnet with alternating exchange in the form of two sublattices with strong *I* and weak *K* exchange interactions, three types of spinon (pair) excitations can be distinguished: *I–I*, *K–K*, and *I–K*. In the sublattice field $(h^{+,-} \neq h^z)$, each of these spinon bands can split into subbands with transverse and longitudinal spin excitations. The wave functions of these excitations can be represented as

$$\psi^{s1} \propto c_1(|\ldots\uparrow\uparrow\downarrow\downarrow\ldots\rangle - |\ldots\downarrow\downarrow\uparrow\uparrow\ldots\rangle) + c_2(|\ldots\downarrow\uparrow\downarrow\uparrow\ldots\rangle - |\ldots\uparrow\downarrow\uparrow\downarrow\ldots\rangle)$$

on the *K–K* bonds and in the form

$$\begin{aligned}\psi_l^{s2} &\propto (|\ldots\uparrow\uparrow\ldots\downarrow\downarrow\ldots\rangle + |\ldots\downarrow\downarrow\ldots\uparrow\uparrow\ldots\rangle,\\ \psi_t^{s2} &\propto (|\ldots\uparrow\downarrow\ldots\uparrow\downarrow\ldots\rangle + |\ldots\downarrow\uparrow\ldots\uparrow\downarrow\ldots\rangle\end{aligned} \quad (7)$$

on the *I* bonds. Excitations of this type do not lead to a change in the *z* component of the total spin $(\Delta S^z = 0)$ and do not contribute to the longitudinal susceptibility, so that the minimum in the temperature dependence $\chi(T)$ for some temperature T_{si} corresponds to filling of the band of singlet spinon excitations. Excitations on the *I–K* bonds change the *z* component of the spin $\Delta S^z = 1$ and are spinons, or spin waves; this gives rise to a maximum in the temperature dependence of the susceptibility at T_{ti}. The filling of the singlet band of excitations in the *I–I* sublattice will lead to an increase in the dimmer ordering parameter *q* (4), and in the *K–K* sublattice it will lead to a sharp decrease of the parameter *q*. We shall determine the splitting into subbands according to the magnitude of the temperature variation of the dimerization parameter $q^{x,y,z}$ and the near-range correlation functions $\langle S_0^\alpha S_1^\beta \rangle$ with respect to the longitudinal and transverse spin components.

The temperature dependences of the above-indicated characteristics are calculated for three interchain exchange parameters, $\lambda = 0.05$, 0.1, and 0.25, and the corresponding values of the exchange alternation, $\delta = 0.05$, 0.075, 0.12, 0.14, and 0.2; $\delta = 0.1$, 0.15, 0.2, 0.3, 0.45, 0.6; and, $\delta = 0.15$, 0.3, 0.45, 0.6, 0.75. The critical temperatures are determined according to the maxima of the correlation radii $\xi_2(T)$ and $\xi_4(T)$, which are shown in Fig. 2 for antiferromagnetic and singlet states, according to the maximum change in the longitudinal component of the squared total spin $\langle (S^z)^2 \rangle$, i.e., according to the maxima of $d\langle (S^z)^2 \rangle/dT$ and the extremal points of the temperature dependence $q(t)$, presented in Fig. 3. On the basis of an analysis of the temperature behavior of the susceptibility $\chi(T)$ (Fig. 4), the critical temperatures were associated to the filling energy of triplet (χ = max) and singlet (χ = min) spin excitation bands. The filling of these bands forms three maxima in the temperature dependence of the specific heat (Fig. 4).

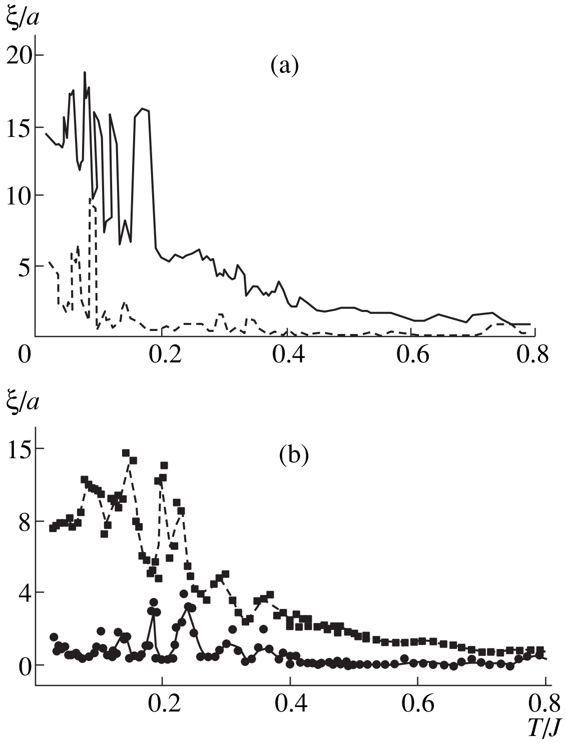


Fig. 2. Temperature dependence of the correlation radii of two-spin (solid line in Fig. a and ■ in Fig. b) and four-spin (broken line in Fig. a and ● in Fig. b) correlation functions for the antiferromagnetic state with $\lambda = 0.1$ and $\delta = 0.15$ (a) and for the singlet state with $\lambda = 0.05$, $\delta = 0.14$, and $h^{+,-} = 1.4h^z$ (b).

A qualitative estimate of the relations between these temperatures, $T_{ti} - T_{si} \propto \sqrt{\lambda^2 \pm 2\delta\lambda + \delta^2}$, where the minus sign corresponds to T_{s1} and the plus sign to T_{s2}, apparently, will also be valid for the energy gaps between these excitation bands. An even weaker effect is the splitting of the proposed spin bands by the sublattice field, which occurs for $\delta > 0.1$. The temperature at which the change in the correlation function between nearest neighbors with respect to the longitudinal spin components is much greater than this change with respect to transverse components (this appears most

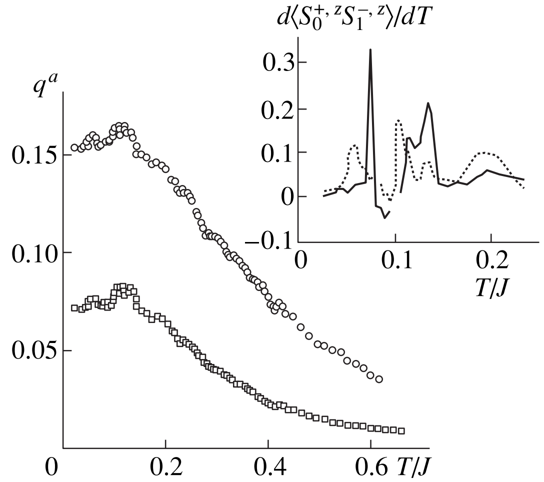


Fig. 3. Temperature dependence of the dimmer ordering parameter q^{α} ($\alpha = z$ (□), x, y = (○)) in an antiferromagnet with $\lambda = 0.1$ and $\delta = 0.15$. Inset: Temperature dependence of the derivative $d\langle S_0^{\alpha} S_1^{\beta}\rangle/dT$ with $\alpha, \beta = z$ (dotted line) and with $\alpha = +$, $\beta = -$ (solid line).

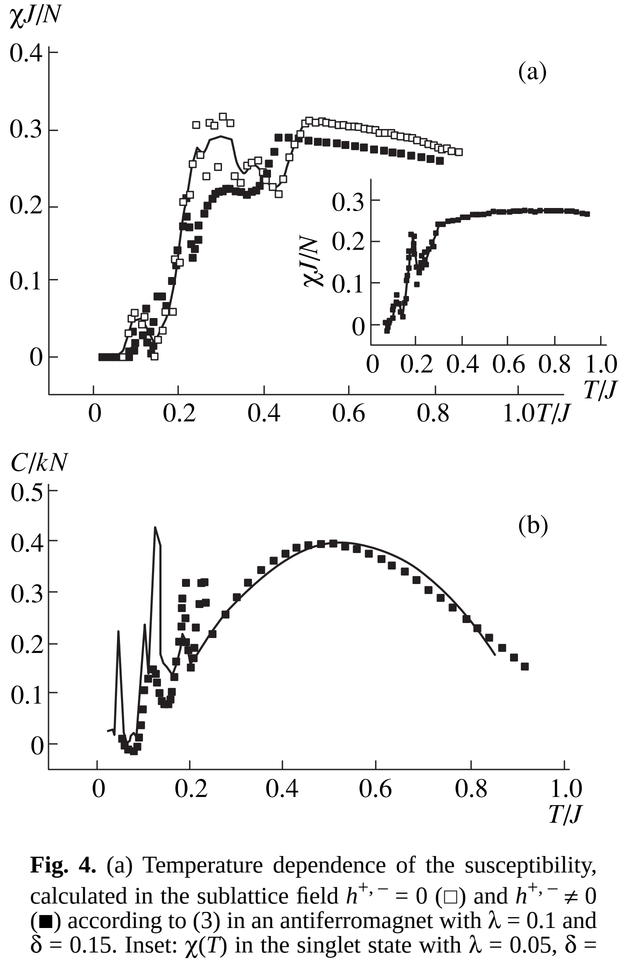


Fig. 4. (a) Temperature dependence of the susceptibility, calculated in the sublattice field $h^{+,-} = 0$ (□) and $h^{+,-} \neq 0$ (■) according to (3) in an antiferromagnet with $\lambda = 0.1$ and $\delta = 0.15$. Inset: $\chi(T)$ in the singlet state with $\lambda = 0.05$, $\delta = 0.14$, and $h^{+,-} = 1.4h^z$. (b) Temperature dependence of the specific heat in an antiferromagnet with alternating exchange and the parameters $\lambda = 0.1$ and $\delta = 0.15$ (solid line) and $\lambda = 0.05$ and $\delta = 0.14$ (■).

strikingly in the calculation of $d\langle S_0^{s,+} S_1^{z,-}\rangle/dT$ (inset in Fig. 3)) refers to the excitation energy of the longitudinal spin mode. In the singlet state, the sublattice field at a distance of the order of the correlation radius influences the spin excitations with wavelength $\sim\pi/\xi$, and for $\delta \gg \delta_c$ the interaction of chains has no effect on the redistribution of the spin excitation density (as compared with the one-dimensional chain) for $\lambda \leq 0.01$. It is possible that each spin subband is characterized by a definite wave vector of the structure $Q_i < \pi/a$, which can found from the Fourier spectrum of the spin correlation function $S(q)$, determined in the singlet state at a distance $k \sim 1/\xi_2$. Thus, $S(q)$ in the singlet state contains weak additional maxima at Q_i, whose number increases with temperature.

The Néél temperature was determined from the sublattice magnetization $\sigma \longrightarrow 0$. In the range of bond alternations close to the critical value, two sharp dropoffs appear clearly in the temperature behavior of the magnetization, for example, for $\delta = 0.15$ and $\lambda = 0.1$ for $T_{s1}/J_1 = 0.06$ and $T_t/J_1 = 0.11$; they are associated with the filling of the triplet spin excitation band in the temperature range $T_{s1} < T < T_t$ and in the spin-wave band at temperatures $T > T_t$. The computed critical temperatures for $\lambda = 0.1$ are presented in Fig. 5. The vanishing of the long-range antiferromagnetic order can be understood from this diagram. As exchange alternation increases, the spin-wave excitation density decreases and vanishes at $T_N \sim T_t$. When the sublattice fields are the same, $h^{+,-} = h^z$, the splitting into subbands vanishes, and for the singlet and paramagnetic states only two critical temperatures exist, above and below the spin-Peierls transition temperature, which are shown in Fig. 5 by the dashed and dotted lines. From the standpoint of symmetry, three phases can be distinguished in the temperature–exchange alternation plane: a region with long-range antiferromagnetic order, a region where the thermodynamic value of the spin is zero, i.e., a singlet state, and a region where $\langle S^z\rangle \sim H/k_BT$, i.e., the paramagnetic state. The phase diagrams (see Fig. 5b), calculated for the three parameters λ in normalized units, are the same to within the computational error.

A variety of incomprehensible experimental data for the antiferromagnet $CuWO_4$ with alternating exchange [2] can be explained on the basis of the results obtained: the existence of an energy gap at the center of the band of the spin excitation spectrum at $\omega = 1.4$ meV and the different temperature dependences of the intensity of spin modes, one of which vanishes at $T = 24$ K and the other (gapless) remains even at $T = 36$ K [2]. The temperature dependence of the susceptibility in the range $40 < T < 70$ K has a concave form [9]. Our computed intrachain exchange parameters $J = 11.6$ meV and $K = 8.9$ meV agree well with neutron diffraction measurements $J = 11.56$ and $K = 9.25$ meV [26]. The interchain exchange $J_2 \sim 1$ meV agrees fairly well with the aver-

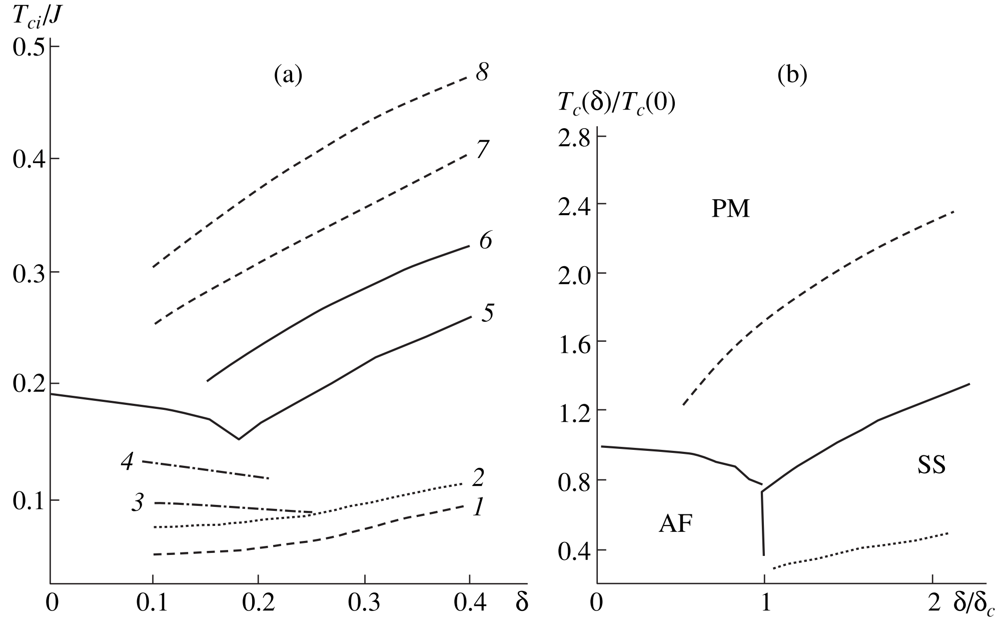


Fig. 5. Critical temperatures, associated with the filling energy of singlet (*1*, *2*, *7*, *8*) and triplet (*3*, *4*, *5*, *6*) spinon bands as a function of the exchange alternation for $h^{+,-} \neq h^z$ (a). Phase diagram of an antiferromagnet (AF), singlet state (SS), and paramagnet (PM) in the plane normalized temperature – normalized exchange alternation for $h^{+,-} = h^z$ (b).

age value $J_2 \approx 1.7$ meV [2]. According to the Monte Carlo calculations, the spin triplet band is filled in the temperature range $17 < T < 24$ K. At $T > 24$ K the gapless mode corresponds to spin-wave excitations, which should vanish at $T_c \approx 40$ K. The band of singlet excitations, not contributing to the magnetic susceptibility, is filled in the temperature range $52 < T < 86$ K, which gives rise to the inflection in the temperature behavior of $\chi(T)$. It is possible that another singlet mode with gap energy $\omega \sim 0.7$ meV exists at temperatures $T < 12$ K.

A series of energy gaps at the center of the band in Bi_2CuO_4 [4, 7] with tetragonal symmetry *P*4/*ncc* and alternating exchange in the [111] direction can be explained by the existence of singlet and triplet spin excitations. The temperature dependences of the resonance field H_0 and the anisotropy field H_a are nonmonotonic, and their derivatives dH_0/dT and dH_a/dT possess several maxima, with different magnitudes, at the corresponding temperatures, T_{ci} = 8, 12, 18, 26 K and T_{ci} = 7, 11, 17, 26, 38 K [10]. For the intra- and interchain exchange parameters chosen, $J_1 = 107$ K and $J_2 = 28$ K, the exchange alternation is $\delta = 0.2$, and our computed critical temperatures, related with the filling of the split singlet and triplet spinon bands at $T_{ci}^{MC} \simeq 7$, 11, 15, 25, 30, 35, agree well with the experimental results. It is possible that the energy gaps at $\omega_{s1} = 0.7$ meV and $\omega_{s2} = 3.4$ and 4 meV at the center of the Brillouin zone are due to singlet excitations of longitudinal and transverse spinon modes, and at ω_t = 1, 7, and 2.1 meV by triplet excitations. Several spin-excitation modes also exist in singlet magnets with alternating exchange. For example, in $(VO)_2P_2O_7$ two gaps were found at ω = 3.12 and 5.75 MeV [5]. It is possible that one other singlet mode with gap at band center at $\omega \simeq 0.6$ meV and with weak intensity exists in this compound. For $\lambda \simeq 0.02$ [5] our computed exchange alternation $\delta \approx 0.15(2)$ lies in the range $\delta_n = 0.12$ and $\delta_\chi = 0.18$, which were determined, respectively, in neutron resonance (*n*) experiments [5] and from the temperature dependence of the susceptibility [27].

The compound $CuGeO_3$ with a spin-Peierls transition was investigated in greatest detail. According to the neutron diffraction data, the ratio of the exchange interactions is $J_c : J_b : J_a = 100 : 10 : 1$ [28] and, according to our estimates, for $\delta > \delta_c \simeq 0.1$ a magnet with such an exchange ratio can be in a singlet state. Inelastic neutron scattering data revealed three gaps at band center at $\omega_i = 0.8$ meV (1.9–2.1) meV [29] and 4(±1) meV [30] and a wide maximum in neutron scattering $\omega \simeq 6$ meV [29], two values for which are close to the results $\omega = 1.86$ and 4.74 meV obtained from EPR measurements [14]. Light scattering at the boundary of the Brillouin zone also leads to three energy gaps ω_i = 2.2, 3.6, and 5.8 meV [31, 32]. The magnetic thermal conductivity in a magnetic field up to $H = 14$ T has two maxima at $T = 5.5$ and 22 K [16], and the width of the EPR line diverges at $T \simeq 4$ and 14 K [15]. In a magnetic field applied in the direction of the *c* axis, the temperature dependence of the magnetostriction constants has three maxima at $T \simeq 6$, 13, and 26 K and in a field $H \parallel b$ at $T \simeq 4$, 11, and 20 K [17].

These results can all be explained well by choosing the two intrachain exchange parameters $I = 145$ K and

$K = 109$ K, which agree fairly well with the data obtained based on the 1D Heisenberg model, in which the triplet gap $\omega \simeq 2$ meV has been calculated [33] and the spin-Peierls transition temperature is $T_{sp} = 14$ K. The observation of thermodynamic anomalies above and below T_{sp}, additional energy gaps, anisotropy of the critical magnetic fields, which are encountered in a ~10% range, where the field $H_c = 13.9$ T is applied in the direction of the exchange alternation axis and the field $H_b = 12.6$ T is applied perpendicular to this axis [34], remains unexplained. As indicated above, at a distance of the order of the correlation radius $\xi_2 \simeq 10c$, the interchain interaction leads to splitting of the low-energy band of singlet excitations with respect to its center into longitudinal and transverse excitation modes with gap energies $\omega_b \simeq 0.5$ meV and $\omega_c \simeq 0.8$ meV and corresponding critical temperatures $T_b \simeq 3.7$ K and $T_c \simeq 6$ K, splitting of the triplet modes with gap energies $\omega_b \simeq 1.7$ meV and $\omega_c \simeq 2$ meV and corresponding critical temperatures $T_b \simeq 14$ K and $T_c \simeq 18$ K, and splitting of the high-energy longitudinal and transverse singlet modes with $\omega_b \simeq 4.5$ meV and $\omega_c \simeq 5.8$ meV and corresponding temperatures $T_b \simeq 30$ K and $T_c \simeq 39$ K. However, at temperatures $T > T_{sp}$ small structural distortions are observed in $CuGeO_3$, which lead to a change in the magnitude of the exchange and its alternation. For this reason, in the temperature range $20 < T < 26$ K one can talk about a qualitative agreement between the temperature interval $\Delta T_{ex} = 6$ K and the Monte Carlo computational results $\Delta T_{MC} = 9$ K. It is possible that in $CuGeO_3$ in the singlet state the quantization axis is directed along the b axis, and then the anisotropy of the critical magnetic field is explained well. Thus, our calculations predict a polarization dependence of light scattering and inelastic neutron scattering along the dimerization axis of a chain. At low temperatures the system is nonlinear, so that the definition of the nonlinear susceptibility $M_\gamma^3 = \chi_{\gamma,\alpha,\beta,\delta}H_\beta H_\alpha H_\delta$ must be used to calculate the resonance absorption frequencies. It is possible that as a result of the nonlinear interaction of the field with the spin subsystem, a transition occurs from the singlet ground state into a singlet excited state, as is observed in EPR measurements at frequency $\omega = 294$ GHz [14], and transitions occur between the subbands $\psi_l^s \longrightarrow \psi_t^s$ at the frequency $f = 34$ GHz in a field $H \approx 12$ kOe [15]. The intensity of both resonances has a maximum at $T = 6$ K and vanishes at $T < 2$ K; this agrees well with our estimates for the energy of the singlet gap and the critical temperature.

Thus, the long-range antiferromagnetic order in the quasi-low-dimensional antiferromagnet with alternating exchange remains for $\lambda \le 0.52(3)\delta$. Alternation gives rise to quantum spin reduction at a site $\sigma/\sigma(0) - 1 = 0.71(6)\delta^{2.5(2)}$, where $\sigma(0) = 1.9\sqrt{\lambda}$. Several temperatures, at which the correlation radii are maximum and the thermodynamic characteristics exhibit features, which can be interpreted on the basis of the model of additional spinon singlet and triplet excitations, were found in an alternating antiferromagnet and in the singlet state. The sublattice self-consistent field ($h^{+,-} \neq h^z$) splits the spinon bands into longitudinal and transverse excitation modes. The temperatures corresponding to the maxima of the derivatives of the resonance field and the linewidths as a function of temperature in Bi_2CuO_4 were calculated using dynamic scaling between the relaxation time and the correlation radius. Possible spinon excitation modes and energy gaps at the center of the band in antiferromagnetic states of the compounds $CuWO_4$ and Bi_2CuO_4 and in the singlet states of the compounds $(VO)_2P_2O_7$ and $CuGeO_3$ were predicted. The temperatures corresponding to the maxima of the magnetic thermal conductivity, the magnetostriction constants, and the divergence of the EPR linewidth in $CuGeO_3$ were calculated. A polarization dependence of light and neutron scattering along the dimerization axis of a chain was predicted.

ACKNOWLEDGMENTS

This work was supported by INTAS (Grant no. 97-12124).

REFERENCES

1. D. Khomskii, W. Geertsma and M. Mostovoy, Czech. J. Phys. **46**, suppl., pt. 56, 3239 (1996).
2. B. Lake, D. A. Tennant, R. A. Cowley, *et al.*, J. Phys. Cond. Mat. **8**, 8613 (1996).
3. K. Yamada, K. Takada, S. Hosoya, *et al.*, J. Phys. Soc. Jpn. **60**, 2406 (1991).
4. B. Roessli, P. Fischer, A. Furrer, *et al.*, J. Appl. Phys. **73**, 6448 (1993).
5. A. W. Garrett, S. E. Nagler, D. A. Tennant, *et al.*, Phys. Rev. Lett. **79**, 745 (1997).
6. M. Hirotaka and Y. Isao, J. Phys. Soc. Jpn. **66**, 1908 (1997).
7. G. A. Petrakovskii, K. A. Sablina, A. I. Pankrats, *et al.*, J. Mag. Magnet. Mat. **140**, 1991 (1995).
8. M. Ain, G. Dhalenne, O. Guiselin, *et al.*, Phys. Rev. B **47**, 8167 (1993).
9. A. G. Anders, A. I. Zvyagin, M. I. Kobets, *et al.*, Zh. Éksp. Teor. Fiz. **62**, 1798 (1972) [Sov. Phys. JETP **35**, 934 (1972)].
10. A. I. Pankrats, G. A. Petrakovskii, and K. A. Sablina, Solid State Commun. **91**, 121 (1994).
11. Yu. P. Gaidukov, V. N. Nikiforov, and N. N. Samarin, Fiz. Nizk. Temp. **22**, 920 (1996) [Low Temp. Phys. **22**, 705 (1996)].
12. L. N. Bulaevskii, A. I. Buzdin, and D. I. Khomskii, Solid State Comm. **27**, 5 (1978).
13. G. S. Uhrig, Phys. Rev. Lett. **79**, 163 (1997).
14. T. M. Brill, J. P. Boucher, J. Voiron, *et al.*, Phys. Rev. Lett. **73**, 1545 (1994).
15. A. I. Smirnov, V. N. Glaznov, L. I. Leonyuk, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 1876 (1998) [JETP **87**, 1019 (1998)].

16. Y. Ando, J. Takeya, D. L. Sisson, *et al.*, Phys. Rev. B **58**, R2913 (1998).
17. G. A. Petrakovskiĭ, A. M. Vorotynov, G. Shimchak, *et al.*, Fiz. Tverd. Tela **40**, 1671 (1998) [Phys. Solid State **40**, 1520 (1998)].
18. L. D. Fadeev and L. A. Takhtajan, Phys. Lett. A **85**, 375 (1981).
19. D. J. Scalapino, Y. Imry, and I. Pincus, Phys. Rev. B **11**, 2042 (1975).
20. H. J. Schulz, Phys. Rev. Lett. **77**, 2790 (1996).
21. H. Raedt and A. Lagendijk, Phys. Rep. **127**, 233 (1985).
22. S. S. Aplesnin, Fiz. Tverd. Tela **38**, 1868 (1996) [Phys. Solid State **38**, 1031 (1996)].
23. N. Kawashima and J. E. Gubernatis, Phys. Rev. Lett. **73**, 1295 (1994).
24. P. W. Anderson, Mater. Res. Bull. **8**, 153 (1973).
25. M. C. Gross and D. S. Fisher, Phys. Rev. B **19**, 402 (1979).
26. J. P. Doumerc, J. M. Dance, J. P. Chaminade, *et al.*, Mater. Res. Bull. **16**, 985 (1981).
27. G. Barnes, Phys. Rev. B **35**, 219 (1987).
28. M. Nishi, O. Fugita, and J. Akimitsu, Technical Report of ISSR, Ser. A **2759**, 1 (1993).
29. M. Ain, J. E. Lorenzo, L. P. Regnault, *et al.*, Phys. Rev. Lett. **78**, 1560 (1997).
30. B. Roessli, P. Fischer, J. Schefer, *et al.*, J. Phys.: Cond. Matt. **6**, 8469 (1994).
31. A. Damascelli, van der Marel, F. Parmigiani, *et al.*, Phys. Rev. B **56**, R11373 (1997).
32. G. Els, M. van Loosdrecht, and P. Lemmens, Phys. Rev. Lett. **79**, 5138 (1997).
33. N. Nishi, O. Fujita, J. Akimitsu, *et al.*, Phys. Rev. B **52**, R6959 (1995).
34. H. Hori, M. Furusawa, S. Sugai, *et al.*, Physica B **211**, 180 (1995).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 202–216.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 227–242.
Original Russian Text Copyright © 2000 by Ginzburg, Savitskaya.

Self-Organization of the Critical State in Granular Superconductors

S. L. Ginzburg and N. E. Savitskaya*
St. Petersburg Institute of Nuclear Physics,
Gatchina, Leningrad oblast, 188350 Russia
**e-mail: savitska@thd.pnpi.spb.ru*

Received June 29, 1999

Abstract—The critical state in granular superconductors is studied using two mathematical models: systems of differential equations for the gauge-invariant phase difference and a simplified model that is described by a system of coupled mappings and in many cases is equivalent to the standard models used for studying self-organized criticality. It is shown that the critical state of granular superconductors is self-organized in all cases studied. In addition, it is shown that the models employed are essentially equivalent, i.e., they demonstrate not only the same critical behavior, but they also lead to the same noncritical phenomena. The first demonstration of the existence of self-organized criticality in a system of nonlinear differential equations and its equivalence to self-organized criticality in standard models is given in this paper. © *2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Interest in investigations of granular superconductors has increased substantially in the last few years primarily in connection with the fact that most HTSC materials can be obtained precisely in the form of a granular system. Specifically, there have appeared a number of theoretical works [1–3] on the magnetic properties of such systems where the granular superconductor was modeled by a system of superconducting granules joined by an insulator, i.e., an ordered lattice of Josephson junctions. It has been shown that the magnetic properties of such a system are similar to the properties of hard type-II superconductors; specifically, a critical state, which is described well by Bean's model [4], can also arise in them. However, the properties of a granular system (specifically, the magnetic-field profile in the sample in the critical state) depends strongly on its main parameter $V \sim j_c a^3/\Phi_0$ (j_c is the critical current density at the junction, a is the lattice constant, and Φ_0 is the quantum of magnetic flux). It has been shown in [1–3] that when this parameter is large ($V \gg 1$), each cell of such a system becomes a pinning center and is capable of confining a large number of magnetic flux quanta. It has also been shown [3] that for $V \gg 1$ the system as a whole possesses a large number of metastable states.

The capability of the system to arrive in the critical state without precise adjustment of the external parameters and the presence of a large number of metastable states are characteristic features of systems in which self-organized criticality (SOC) is observed [5]. It consists in the fact that the dynamical system naturally evolves to the critical state, which subsequently is self-maintaining, not requiring accurate adjustment of the external parameters for its existence. Structurally, the critical state that arises is a set of metastable states, transforming into one another by means of “avalanches,” which arise as a result of a local external perturbation. Such a critical state is said to be self-organized, and the criterion for the existence of self-organization in the system is a power-law dependence of the probability density of the sizes of the avalanches.

The concept of self-organized criticality can be used to describe the behavior of a large number of dissipative dynamical systems arising in various fields of modern science. The properties of granular superconductors are such that self-organized criticality can also be observed in them. This not only gives a new representation of the nature and structure of the critical state in these systems, but it also opens up the possibility of studying SOC experimentally. This is especially important, since thus far, despite the generality of the phenomenon, SOC has been studied only on model systems which are difficult to realize in practice, and experimental investigations have been performed only for a sandpile [6].

The possibility of self-organization of the critical state in granular superconductors for $V \gg 1$ has been considered in [7], where a simplified model of such a system was introduced and studied. In constructing this model, taking account of the physical characteristics of the behavior of granular superconductors for $V \gg 1$, the system of differential equations for the gauge-invariant phase difference on junctions that describes a superconductor was replaced by a system of mappings (equations with discrete time) for the currents through the junctions. It was found that for a two-dimensional multijunction SQUID, the system of mappings obtained is equivalent to the classical model for studying SOC: an Abelian model of a sandpile [8]. The magnitude of the current in the junction played the role of the pile height at a given node, and current injection was the analog of adding sand grains. In this manner, it was shown that the critical state in the simplified model of a multijunction SQUID described by a system of mappings is self-organized.

Subsequent study of the simplified model of granular superconductors showed that, on account of the physical properties incorporated in it, SOC in this system possesses a number of interesting properties, which have not been observed in the Abelian sandpile model.

Thus, in [9] the simplified model of a multijunction SQUID was studied with various methods of current injection into the system. Specifically, the case of current injected simultaneously into all nodes in small increments, which is not realizable in the sandpile model, since a sand grain in the sandpile model possesses a finite size, was studied. It was found that the probability density of the avalanche sizes for all injection methods is a universal function, while the finer characteristics (interavalanche correlators) differ.

In addition, in [10] the conjecture advanced earlier in [7] that in the simplified model of a multijunction SQUID placed in an increasing magnetic field SOC is realized even under closed boundary conditions was verified. This is impossible, in principle, to do in the sandpile model, since in this case the possibility for sand to leave the system is one of the main requirements for the existence of SOC. In the case of a multicontact SQUID, however, the process of current leaving from the system is replaced by a fundamentally different process, the annihilation of currents with opposite signs, and the SOC can also exist in a closed system.

Thus, the study of the simplified model makes it possible to infer that the critical state in granular superconductors is self-organized and possesses a number of interesting properties, which have not been observed in previously investigated models with SOC.

However, no conclusion can be drawn about the existence and properties of SOC in real superconductors, whose behavior is described by differential equations, on the basis of only the simplified model, since the simplifications introduced could have led to a loss of some properties of the initial system and to the appearance of new qualities which are uncharacteristic of the initial system. For this reason, in the first place, the problem of investigating the properties of the critical state, proceeding directly from a system of differential equations for the phase difference, remains open; in the second place, it is interesting to investigate the question of the degree to which the simplified model reflects the real behavior of the initial system.

Thus, the main objective of the present work was to study theoretically and using a computer the critical state of one- and two-dimensional granular superconductors under various conditions, proceeding directly from the system of differential equations for the gauge-invariant phase differences at junctions. The corresponding simplified models described by systems of mappings were studied in parallel. As a result, it was shown that the critical state in granular superconductors is self-organized, and the behavior of the simplified model is completely equivalent to the behavior of the initial system described by a system of differential equations.

In Sec. 2, the one- and two-dimensional granular superconductors (multijunction SQUIDs) are studied, the systems of differential equations describing them and the corresponding simplified models are presented, and the basic characteristics of the system are also introduced: the average current and voltage, whose statistical properties in the critical state will be studied.

In Sec. 3, the results of computer simulation of the critical state of multicomponent SQUIDs under various conditions are presented: for open boundary conditions with various methods of current injection and for closed boundary conditions and an increasing external magnetic field. For each case simulation was performed for the system of differential equations and for simplified models, and the results were compared.

The basic results of this work are formulated briefly in Sec. 4.

2. ONE- AND TWO-DIMENSIONAL MULTIJUNCTION SQUIDS

2.1. Basic Equations

A two-dimensional multicontact SQUID consists of two superconducting plates connected by Josephson junctions located at the nodes of a $N \times M$ lattice. The junction size l is much less than the lattice period a (Fig. 1a). In the geometry described, Josephson current flows along the z axis. Then the system of equations for the gauge-invariant phase difference $\varphi_{n,m}$ on the junctions with coordinates (na, ma) can be written on the basis of the equation for a large Josephson junction [11]. Using the resistive model of a Josephson junction and neglecting thermal fluctuations [12], we obtain in dimensionless form the following system of equations:

$$V\sin\varphi_{n,m} + \tau\frac{d\varphi_{n,m}}{dt} = \Delta_{n,m}(\varphi) + 2\pi F_{n,m},$$

$$V = \frac{2\pi j_c}{j_\varphi}, \quad j_\varphi = \frac{\Phi_0}{8\pi l^2\lambda_L},$$

$$F_{n,m} = \frac{j_{en,m}}{j_\varphi}, \quad \tau = \frac{\Phi_0}{\rho_0 j_\varphi}, \qquad (1)$$

$$\Delta_{n,m}(\varphi) = \varphi_{n+1,m} + \varphi_{n-1,m} + \varphi_{n,m-1} + \varphi_{n,m+1} - 4\varphi_{n,m},$$

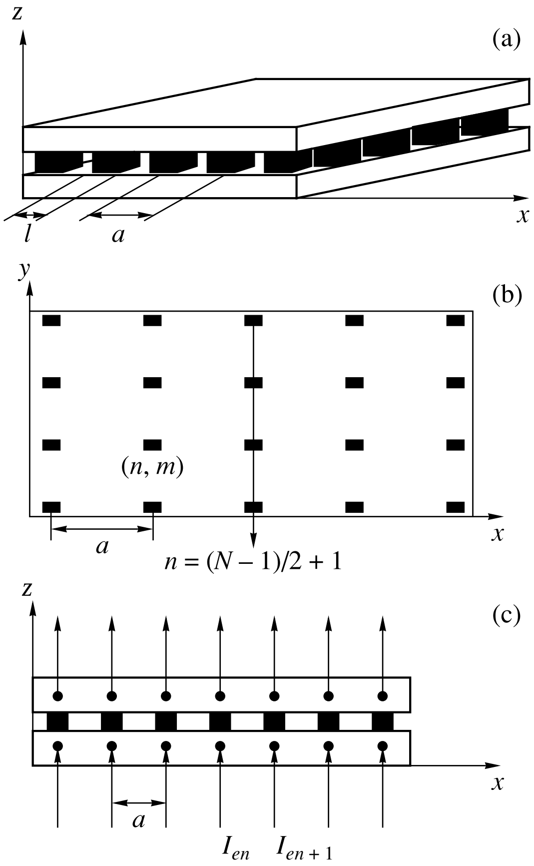


Fig. 1. (a) Two-dimensional multijunction SQUID, (b) projection of the two-dimensional multijunction SQUID on the (x, y) plane, and (c) one-dimensional multijunction SQUID.

where j_c is the critical current density, λ_L is the London penetration depth, Φ_0 is the flux quantum, ρ_0 is the surface resistivity of the junction, $j_{en,\,m}$ is the injection current density, and $\Delta_{n,\,m}$ is the two-dimensional discrete Laplacian.

For a one-dimensional multijunction SQUID (Fig. 1c) [12], which consists of two superconducting plates, which are infinite along the y axis and are connected with one another by Josephson junctions at the points with coordinates $x = an$, $n = 1, \ldots, N$, the system of equations (1) can be written as follows [2]:

$$V\sin\varphi_n + \tau\frac{\partial\varphi_n}{\partial t} = -2\varphi_n + \varphi_{n+1} + \varphi_{n-1} + 2\pi F_n,$$
$$V = 2\pi\frac{4\pi aSj_c}{\Phi_0}, \quad F_n = \frac{4\pi j_{en}aS}{\Phi_0}, \quad \tau = \frac{4\pi aS}{\rho}, \tag{2}$$

where S is the area of a cell between two junctions.

The parameter V appearing in the equations is the main characteristic of the SQUID. Thus, for $V \gg 1$ the energy of a single-junction SQUID has a large number (of the order of V) metastable states, and this fact leads to the existence of phenomena such as quantization of the magnetic flux and hysteresis [12]. These properties remain for a multijunction SQUID and $V \gg 1$. In what follows, we shall study multijunction SQUIDs precisely for $V \gg 1$.

We shall study, in parallel with the systems of differential equations (1) and (2), simplified models of multijunction SQUIDs, which have analogs among previous models used to study SOC, such as, the Abelian sandpile model and the one-dimensional “sandpile” model studied in [13, 14]. The transition to simplified models described by systems of coupled mappings has been examined in detail in [7]. In what follows, we shall present this in a brief form.

2.2. Simplified Model of Two- and One-Dimensional Multijunction SQUIDs

The simplified models of multijunction SQUIDs are formulated in terms of the dimensionless current $z_{n,\,m}$ through a junction:

$$z_{n,\,m} = z_c\sin(\varphi_{n,\,m}) + \frac{\tau}{2\pi}\frac{d\varphi_{n,\,m}}{dt}, \tag{3}$$

where $z_c = V/2\pi$. For the dimensionless current, equations (1) are rewritten in the form

$$z_{n,\,m}(\varphi) = \frac{1}{2\pi}\Delta_{n,\,m}(\varphi) + F_{n,\,m}. \tag{4}$$

The system (3) and (4) obtained for $V \gg 1$ can be simplified using the properties of the solution of equation (3), which is an equation for the phase for a single Josephson junction. It is obvious from this equation that if $z_{n,\,m} \le z_c$, then $d\varphi_{n,\,m}/dt = 0$, i.e., the phase does not change. However, if $z_{n,\,m} > z_c$, then for $V \gg 1$ and $|z_{n,\,m} - z_c| \ll z_c$ this equation has a solution $\varphi_{n,\,m}(t)$, which at first varies very little in a time interval T and then changes by 2π in a time τ_0, where $\tau_0 \ll T$.

Next, we neglect in the term $\Delta_{n,\,m}(\varphi)$ this weak (of the order of $1/V$) change and we approximate the phase by a step function

$$\varphi_{n,\,m} \approx 2\pi q_{n,\,m} + \pi/2, \tag{5}$$

where $q_{n,\,m}$ is an integer. In this approximation the system (3) and (4) can be rewritten as

$$z_{n,\,m} = z_c\sin(\varphi_{n,\,m}) + \frac{\tau}{2\pi}\frac{d\varphi_{n,\,m}}{dt},$$
$$z_{n,\,m} = \Delta_{n,\,m}(q) + F_{n,\,m}, \tag{6}$$
$$q_{n,\,m} = \mathrm{Int}\left(\frac{\varphi_{n,\,m}}{2\pi} + \frac{1}{4}\right),$$

where Int(x) is the integer part of x.

The equations (6) form a closed system in which the phases on different junctions do not interact with one another.

We shall describe the solution of such a simplified system in terms of the dimensionless current.

First, we examine the change in the phase at a site (n_0, m_0). We choose as initial conditions the situation where a current $F_{n,m} = K = \mathrm{Int}(z_c)$ less than the critical current is injected into all junctions, and the phases on all junctions are the same and are equal to

$$\varphi_{n,m}(0) = \arcsin\left(\frac{2\pi K}{V}\right) = \frac{\pi}{2} - \sqrt{2\left(1 - \frac{2\pi K}{V}\right)}. \qquad (7)$$

The $q_{n,m} = 0$ and $\Delta_{n,m}(q) = 0$.

We shall assume that the current is injected into the system in integral portions, i.e., $F_{n,m}$ is an integer, so that $z_{n,m}$ is also an integer. We add at the site n_0, m_0 one unit of current, i.e., $F_{n_0, m_0} = K + 1$. Then

$$z_{n_0, m_0} \longrightarrow z_{n_0, m_0} + 1. \qquad (8)$$

Now, the current z_{n_0, m_0} exceeds the critical current, and then $d\varphi_{n_0, m_0}/dt > 0$. From the first equation of our system (6) we find the solution for φ_{n_0, m_0}. It is such that the phase is essentially unchanged in a time $T = 2\pi\tau_0/\sqrt{2((K+1)/z_c - 1)}$, after which it rapidly changes by 2π in time τ_0. At the moment the phase changes by 2π the following changes occur:

$$\begin{aligned} &q_{n,m} \longrightarrow q_{n,m} + 1, \quad \Delta_{n,m} \longrightarrow \Delta_{n,m} - 4, \\ &\Delta_{n\pm1,m} \longrightarrow \Delta_{n\pm1,m} + 1, \quad \Delta_{n,m\pm1} \longrightarrow \Delta_{n,m\pm1} + 1. \end{aligned} \qquad (9)$$

Then z_{n_0, m_0} changes according to the law

$$\begin{aligned} &z_{n_0, m_0} \longrightarrow z_{n_0, m_0} - 4, \quad z_{n_0\pm1, m_0} \longrightarrow z_{n_0\pm1, m_0} + 1, \\ &z_{n_0, m_0\pm1} \longrightarrow z_{n_0, m_0\pm1} + 1. \end{aligned} \qquad (10)$$

We note that the rules (8) and (10) for z_{n_0, m_0} to change are identical to the rules for adding sand grains and for falling of sand grains in the Abelian sandpile model.

The following difficulty arises when switching to the entire lattice. The problem is that the time T for a weak change in phase to occur is different at different sites and constantly changes because during this weak change jumps in phase can occur at a neighboring site. The equations (6) take this change in T into account, but they are difficult to analyze. We shall make one other simplification. We shall assume that T is the same for all sites. Then, we can introduce into the system the discrete time $t_k = kT$. In this case, the advance of the phases can be ignored, and only the change in $q_{n,m}$ and $z_{n,m}$ need be considered. The rules for a change in $z_{n,m}$ in this case can be written in the form

$$\begin{aligned} q_{n,m}(k+1) &= q_{n,m}(k) + \theta[z_{n,m}(k) - z_c], \\ z_{n,m}(k) &= \Delta_{n,m}(q(k)) + F_{n,m}(k). \end{aligned} \qquad (11)$$

Then we have for $z_{n,m}$

$$\begin{aligned} z_{n,m}(k+1) &= z_{n,m}(k) + \Delta_{n,m}\{\theta[z(k) - z_c]\} + \xi_{n,m}(k) \\ &= z_{n,m}(k) + \theta[z_{n+1,m}(k) - z_c] + \theta[z_{n-1,m}(k) - z_c] \\ &\quad + \theta[z_{n,m+1}(k) - z_c] + \theta[z_{n,m-1}(k) - z_c] \\ &\quad - 4\theta[z_{n,m}(k) - z_c] + \xi_{n,m}(k), \\ \xi_{n,m}(k) &= F_{n,m}(k+1) - F_{n,m}(k). \end{aligned} \qquad (12)$$

The mappings obtained are completely equivalent to the algorithm describing the Abelian sandpile model. The quantity $z_{n,m}$ plays the role of the pile height at a given site, and current injection, described by the function $\xi_{n,m}(k)$, plays here the role of the addition of a sand grain.

For a one-dimensional multicomponent SQUID the system of mappings can be written in the form

$$\begin{aligned} q_n(k+1) &= q_n(k) + \theta[z_n(k) - z_c], \\ z_n(k) &= \Delta_n(q(k)) + F_n(k), \\ \Delta_n(q) &= -2q_n + q_{n+1} + q_{n-1}. \end{aligned} \qquad (13)$$

Then, for z_n we have

$$\begin{aligned} z_n(k+1) &= z_n(k) + \Delta_n\{\theta[z(k) - z_c]\} + \xi_n(k) \\ &= z_n(k) + \theta[z_{n+1}(k) - z_c] + \theta[z_{n-1}(k) - z_c] \\ &\quad - 2\theta[z_n(k) - z_c] + \xi_n(k), \\ \xi_n(k) &= F_n(k+1) - F_n(k). \end{aligned} \qquad (14)$$

The system of mappings (14) is equivalent to the rules for the change in the slopes in the one-dimensional sandpile model studied in [13].

We note that equations (12) and (14) describing the current dynamics can be briefly written in the form

$$z_{n,m}(k+1) = z_{n,m}(k) + \Delta_{n,m}\{\Psi[z(k)]\} + \xi_{n,m}(k), \qquad (15)$$

$$z_n(k+1) = z_n(k) + \Delta_n\{\Psi[z(k)]\} + \xi_n(k), \qquad (16)$$

$$\Psi(z(k)) = \theta(z - z_c). \qquad (17)$$

Let us consider the relation between the function $\Psi(z(k))$ and the electrodynamic system and its possible modifications. To understand the physical meaning of $\Psi(z(k))$, we shall examine the relation between the voltage $U_{n,m}(t)$ on a junction and the phase. This relation is expressed as follows:

$$U_{n,m} = \frac{\Phi_0}{2\pi}\frac{\partial\varphi_{n,m}}{\partial t}. \qquad (18)$$

If the current through the junction is greater than the critical current, then in a time T the phase $\varphi_{n,m}$ changes by 2π. Then the average voltage on the junction over this time is

$$U_{n,m}\left(k+\frac{1}{2}\right) = \frac{1}{T}\int_{t_k}^{t_{k+1}} U_{n,m}(t)dt = \frac{\Phi_0}{T}, \tag{19}$$

or, in a more general form,

$$\begin{aligned} U_{n,m}\left(k+\frac{1}{2}\right) &= \frac{\Phi_0}{T}\theta[z_{n,m}(k)-z_c] \\ &= \frac{\Phi_0}{T}[q_{n,m}(k+1)-q_{n,m}(k)]. \end{aligned} \tag{20}$$

Then we can write $z_{n,m}$ in the form

$$\begin{aligned} & z_{n,m}(k+1)-z_{n,m}(k) \\ &= \frac{T}{\Phi_0}\Delta_{n,m}\left[U\left(k+\frac{1}{2}\right)\right]+\xi_{n,m}(k), \end{aligned} \tag{21}$$

i.e.,

$$U_{n,m}\left(k+\frac{1}{2}\right) = \frac{\Phi_0}{T}\Psi[z_{n,m}(k)]. \tag{22}$$

Since $z_{n,m}$ is the dimensionless current through a junction, $U_{n,m}(k+1/2)$ and therefore $\Psi(z)$ have a simple physical interpretation: They are the current-voltage characteristic (IVC) of a junction.

We note that only one threshold is taken into account in the expression (15) for the IVC. The complete IVC, taking account of both thresholds, is given by the expression

$$\Psi(z) = \theta(z-z_c)-\theta(-z-z_c). \tag{23}$$

2.3. Boundary Conditions

The behavior of our system is not completely determined by equations (1) and (2) or the systems of mappings (12) and (14). In problems of studying the SOC, the boundary conditions and the method of perturbation of the system play an important role. Usually, two types of boundary conditions are considered in problems with SOC: open and closed.

In this section we shall examine what different boundary conditions signify for our system.

In our case, open boundary conditions can be realized, if it is assumed that our SQUID is shunted by a normal superconductor, whose critical current is several orders of magnitude greater than the critical current of Josephson junctions. This means that to a lattice of junctions with $1 \le n \le N$ and $1 \le m \le M$ are added junctions with $n = 0$ and $N+1$, $1 \le m \le M$ and $m = 0$ and $M+1$, and $1 \le n \le N$. We shall assume that the critical current density on them is infinite:

$$\begin{aligned} j_c &= \infty, \\ n &= 0, N+1; \quad 1 \le m \le M; \\ m &= 0, M+1; \quad 1 \le n \le N. \end{aligned} \tag{24}$$

Then the current on junctions with $n = 0$ and $N+1$, $1 \le m \le M$, $m = 0$ and $M+1$, and $1 \le n \le N$ can only increase, i.e., the current can leave the sublattice where $1 \le n \le N$ and $1 \le m \le M$, which corresponds to open boundary conditions. The condition (24) is equivalent to the condition

$$\begin{aligned} \varphi_{n,m} &= 0, \\ n &= 0, N+1; \quad 1 \le m \le M; \\ m &= 0, M+1; \quad 1 \le n \le N. \end{aligned} \tag{25}$$

The equations (1) for nodes at the boundary, for example, for $n, m = 1$ and $n = 1, m \ne 1, M$, and taking account of the condition (24), can be written as follows:

$$V\sin\varphi_{1,1}+\tau\frac{d\varphi_{1,1}}{dt} = \varphi_{2,1}+\varphi_{1,2}-4\varphi_{1,1}+2\pi F_{1,1},$$

$$\begin{aligned} & V\sin\varphi_{1,m}+\tau\frac{d\varphi_{1,m}}{dt} \\ &= \varphi_{2,m}+\varphi_{1,m-1}+\varphi_{1,m+1}-4\varphi_{n,m}+2\pi F_{1,m}. \end{aligned}$$

For the simplified models (12) and (14), open boundary conditions mean

$$\begin{aligned} z_c &= \infty, \\ n &= 0, N+1; \quad 1 \le m \le M; \\ m &= 0, M+1; \quad 1 \le n \le N, \end{aligned} \tag{26}$$

which is equivalent to the condition

$$\begin{aligned} z_{n,m} &= 0, \\ n &= 0, N+1; \quad 1 \le m \le M; \\ m &= 0, M+1; \quad 1 \le n \le N, \end{aligned} \tag{27}$$

which is usually studied in problems concerning systems with SOC.

With these boundary conditions the mappings (12) have the form

$$\begin{aligned} z_{1,1}(k+1) &= z_{1,1}(k)+\theta[z_{2,1}(k)-z_c] \\ &+\theta[z_{1,2}(k)-z_c]-4\theta[z_{1,1}(k)-z_c]+\xi_{1,1}(k), \\ z_{1,m}(k+1) &= z_{1,m}(k)+\theta[z_{2,m}(k)-z_c] \\ &+\theta[z_{1,m+1}(k)-z_c]+\theta[z_{1,m-1}(k)-z_c] \\ &-4\theta[z_{1,m}(k)-z_c]+\xi_{1,m}(k). \end{aligned} \tag{28}$$

Close or reflective boundary conditions mean that current is conserved and cannot leave the system. In this case, equations (1) for nodes on the boundary, for example, for $n, m = 1$, have the form

$$V\sin\varphi_{1,1} + \tau\frac{d\varphi_{1,1}}{dt} = \varphi_{2,1} + \varphi_{1,2} - 2\varphi_{1,1} + 2\pi F_{1,1}. \tag{29}$$

For the simplified model

$$\begin{aligned} z_{1,1}(k+1) &= z_{1,1}(k) + \theta[z_{2,1}(k) - z_c] \\ &+ \theta[z_{1,2}(k) - z_c] - 2\theta[z_{1,1}(k) - z_c] + \xi_{1,1}(k). \end{aligned} \tag{30}$$

2.4. Perturbation of the System

When studying SOC in the sandpile model, the system is usually considered in the following regime. Starting in a metastable state, the system is perturbed by a prechosen method, for example, a single sand grain is added at a randomly chosen node. The perturbation is followed by a relaxation process ("avalanche"), which puts the system into the next metastable state, the perturbation is repeated, and so on.

Two methods of perturbation are ordinarily used in problems studying SOC in the sandpile model.

1. Addition of a sand grain at a randomly chosen node. In our case, this is equivalent to injection of a unit current in a randomly chosen junction, i.e., when the SQUID is in a metastable state (all $(\partial\varphi_{n,m}/\partial t)(t_0) = 0$), the injection current on a randomly chosen junction increases by one unit:

$$F_{n_0, m_0}(t_0 + \delta t) = F_{n_0, m_0}(t_0) + 1, \quad \delta t \longrightarrow 0. \tag{31}$$

Subsequently, the magnitude of the injection current remains unchanged in the course of the relaxation process.

For the simplified model, this method of perturbation means that after the system arrives in the next metastable state (all $z_{n,m}(k_0) < z_c$) the function at the randomly chosen site (n_0, m_0)

$$\xi_{n_0, m_0}(k_0) = 1, \tag{32}$$

and $\xi_{n,m} = 0$ during relaxation at all sites.

2. Addition of a sand grain at the central node of the lattice. For the system (1), this is equivalent to injection of a unit current in the central junction, i.e., when the SQUID is in a metastable state, the injection current on the central junction increases by one unit:

$$F_{n_c, m_c}(t_0 + \delta t) = F_{n_c, m_c}(t_0) + 1, \quad \delta t \longrightarrow 0. \tag{33}$$

For the simplified model, this method of perturbation means that after the system arrives in the next metastable state the function in the central node (n_c, m_c)

$$\xi_{n_c, m_c}(k_0) = 1. \tag{34}$$

To study multijunction SQUIDs, besides the perturbation methods described above, other methods, which are physically meaningless for the Abelian sandpile model but are natural for SQUIDs can also be considered.

1. Current injection in all junctions simultaneously in small portions. For the system (1), this means that after arrival in the next metastable state the injection current on all junctions changes as follows:

$$\begin{gathered} F_{n,m}(t_0 + \delta t) = F_{n,m}(t_0) + \delta F, \\ \delta t \longrightarrow 0, \quad \delta F \ll 1. \end{gathered} \tag{35}$$

For the simplified model this method of perturbation means that after the system arrives in the next metastable state for all n, m the function

$$\xi_{n,m}(k_0) = p, \quad p \ll 1. \tag{36}$$

The next method of perturbation is closest to being realized in practice in an experiment.

2. To study the critical state in superconductors a sample is placed in an external ac magnetic field and current in the junctions is excited not by direct injection but rather by induction by an external field. In what follows we shall assume that the magnetic field is directed along the y axis (Fig. 1b). Then it can be shown (see [7] for a more detailed discussion) that the effect of an external increasing magnetic field effectively reduces to injection of a positive current in all junctions on the right-hand boundary of a SQUID and the same current in absolute magnitude but negative in sign in all junctions on the left-hand boundary. Then the total current injected into the system is zero. This means that when the system is in a metastable state the injection current changes as follows:

$$\begin{gathered} F_{1,m}(t_0 + \delta t) = F_{1,m}(t_0) - F_{ex}, \\ F_{N,m}(t_0 + \delta t) = F_{N,m}(t_0) + F_{ex}, \\ 1 \le m \le M. \end{gathered} \tag{37}$$

For the simplified system this will be

$$\xi_{1,m} = -F_{ex}, \quad \xi_{N,m} = F_{ex}, \quad 1 \le m \le M. \tag{38}$$

We note that in contrast to previous methods for perturbing the system, here not only a positive but also a negative current is injected into the system. Therefore, a negative threshold must also be taken into account in the IVCs of junctions in the simplified model of a granular superconductor. It was immaterial in previous cases, since there the system was always near the positive threshold. Thus, in the method described for exciting the system the simplified model will be described by the system of mappings (16) with the IVC (23).

The method for perturbing the system in a one-dimensional state also has its own specific features.

3. In the sandpile model the SOC does not occur in the one-dimensional case [5]. In [13, 14] it was shown that the SOC in the one-dimensional case can occur if the method for adding a sand grain is different from the

method ordinarily used in the Abelian sandpile model. For our case, it can be expressed as follows:

$$\begin{aligned} F_n(t_0+\delta t) &= F_n(t_0)+1/2, \\ F_{n+1}(t_0+\delta t) &= F_{n+1}(t_0)-1/2, \end{aligned} \tag{39}$$

where $n \neq N$ is a randomly chosen node.

For the simplified system this means

$$\xi_n = 1/2, \quad \xi_{n+1} = -1/2, \quad n \neq N. \tag{40}$$

We note that in [13, 14] a one-dimensional system was studied with one open and one closed boundary, and even though negative slopes appeared in the system, for them only one threshold was studied. We shall study in this case the simplified model, taking account of both thresholds (23), in order to be able to compare with the case of differential equations.

2.5. Average Current and Integral of the Voltage

The mathematical criterion for the existence of self-organization in the system is power-law behavior of the probability density of avalanche sizes. The size of an avalanche can be defined as the total number of topplings over the avalanche time [15]. This quantity can be expressed as follows:

$$S_i = \frac{1}{L}\sum_{k=t_{bi}}^{k=t_{ei}}\sum_{n,m}\theta(s_{n,m}(k)-s_c), \tag{41}$$

where L is the number of nodes in the lattice, t_{bi} and t_{ei} are the moments when the ith avalanche starts and ends, $s_{n,m}$ is the pile height at a given site in the Abelian sandpile model or the slope of the pile in the model of [13], and s_c is the critical value for $s_{n,m}$. In our case the analog of $s_{n,m}$ is the value of the current in a junction. When the current in a junction exceeds the critical value, a jump in phase by 2π occurs at the junction, and according to Eq. (18) this leads to a voltage pulse. Thus, in our case, the SQUID-averaged integral of the voltage over the avalanche time plays the role of the avalanche size:

$$U_i = \frac{1}{NM}\sum_{n,m}\int_{t_{ib}}^{t_{ie}} U_{n,m}(t)dt. \tag{42}$$

Taking account of Eq. (18), we have for U_i the following expression:

$$U_i = \frac{\Phi_0}{2\pi NM}\sum_{n,m}[\varphi_{n,m}(t_{ei})-\varphi_{n,m}(t_{bi})]. \tag{43}$$

For convenience in making comparisons below, we shall consider the dimensionless quantity U_i/Φ_0.

Besides the integral of the voltage, we shall also consider the properties of a different characteristic of the system, i.e., the system-averaged current at the moment the ith avalanche ends:

$$\begin{aligned} z_i &= \frac{1}{NM}\sum_{n,m} z^{st}_{n,m}(t_{ei}), \\ z^{st}_{n,m} &= \frac{V}{2\pi}\sin\varphi_{n,m}(t_{ei}). \end{aligned} \tag{44}$$

In the Abelian sandpile model the total mass of the system was the analog of this quantity [15].

For the simplified model of a multijunction SQUID, analogous quantities will have the index R. By analogy to the expression (41) we obtain for the simplified model the following expression for the integral of the voltage:

$$W^R_i = \frac{1}{NM}\sum_{k=t_{bi}}^{k=t_{ei}}\sum_{n,m}\theta(z_{n,m}(k)-z_c) \tag{45}$$

and for the average current

$$z^R_i = \frac{1}{NM}\sum_{n,m} z_{n,m}(k=t_{ei}). \tag{46}$$

3. RESULTS OF COMPUTER SIMULATION

In the present section the results obtained by computer simulation of the critical state of one- and two-dimensional multijunction SQUIDs directly from the system of differential equations with $V = 40$ are presented. The equations were solved using Euler's scheme with a time step $dt = 0.01$. Decreasing this quantity further and using other integration schemes did not change the physical results. Systems with various boundary conditions and with various methods of perturbation were studied.

Unless stipulated specially, the system was studied in the standard regime used for studying systems with SOC. Starting with "smooth" initial conditions, for which all $\varphi_{n,m} = 0$ and $F_{n,m} = 0$, we perturbed the system, changing the magnitude of the injection current by one of the methods described above. Then, the system relaxed to the next metastable state. The magnitude of the injection current remained unchanged during relaxation. When the system reached a metastable state, we once again perturbed the system by the same method, and so on. In the simulation it was assumed that the system reached a metastable state if $\partial\varphi_{n,m}/\partial t < 10^{-8}$ for all junctions.

After a transient process, the system reached a stationary critical state consisting of a set of metastable critical states transforming into one another after the next perturbation. We shall term the process whereby the system arrives in the next metastable state an "avalanche." For each avalanche the system-averaged current, z_i (Eq. (44)), at the moment the avalanche terminates and the system-averaged integral of the voltage, U_i/Φ_0 (Eq. (43)), were determined.

A corresponding simplified model described by the system of mappings with $z_c = V/2\pi$ was studied each time in parallel with the system of differential equations. The simplified model was studied in the same regime and with the same perturbation method as the initial system. For the simplified model, "smooth" boundary conditions means that all $z_{n,m} = 0$, and the metastable state is determined as the state in which all $|z_{n,m}| < z_c$. For the simplified model, for each avalanche the average current z_i^R (Eq. (46)) and the integral of voltage W_i^R (Eq. (45)) were also determined for each avalanche.

Next, in the simplified model and for the initial system, the probability density $\rho(x)$ was calculated for each quantity, and in some cases the interavalanche correlation functions $D_x(j)$ and the power spectra $S_x(f)$ of these quantities were also studied [16]:

$$\begin{aligned}\rho(x) &= \langle\delta(x-x_i)\rangle,\\ D_x(j) &= \langle x_i x_{i+j}\rangle - \langle x\rangle^2,\\ S_x(f) &= 2\sum_{j=-\infty}^{\infty} D_x(j)e^{-i2\pi fj}.\end{aligned}\tag{47}$$

Here $\langle\ldots\rangle$ denotes averaging over the number of the avalanche i, and x can be z, z^R, U/Φ_0, or W^R.

The results obtained for the simplified models and initial systems of differential equations were compared.

3.1. One-Dimensional SQUID

The one-dimensional multijunction SQUID has been studied in detail in [17]. We shall briefly describe the basic results obtained.

A one-dimensional multijunction SQUID described by the system of differential equations (2) was studied with open boundary conditions

$$\begin{aligned}V\sin\varphi_1 + \tau\frac{\partial\varphi_1}{\partial t} &= -\varphi_1 + \varphi_2 + 2\pi F_{e1},\\ V\sin\varphi_N + \tau\frac{\partial\varphi_N}{\partial t} &= -\varphi_N + \varphi_{N-1} + 2\pi F_N\end{aligned}\tag{48}$$

with the perturbation method (39).

In parallel with the system of differential equations, a simplified one-dimensional model (16) with the IVC (23) was also studied in the same regime with closed boundary conditions and with the perturbation method (40).

In both cases, after a transient process, the systems reached a critical state, which consisted of a set of metastable states passing into one another in which the values of the dimensionless currents on the left-hand side of the lattice are positive and close to the critical value z_c while on the right-hand side they are negative and close to $-z_c$ (Fig. 2a). This picture resembles the critical state in Bean's model [4], but Bean's model there is only one metastable state, into which the system returns each time after the next successive perturbation. In our case, however, after a perturbation the system relaxes to a different metastable state, in which the separation into "positive" and "negative" parts remains but the values of the currents at the junctions change somewhat. This is the behavior characteristic for systems with self-organization.

We note that in the case of a one-dimensional "sandpile" the separation of the system into positive and negative parts does not arise [13, 14], even though both positive and negative slopes do occur. In our case the separation is due to the fact that two threshold values of the currents are present in the system. An interesting feature of our system is that the separation occurs even though the positive and negative currents are injected in randomly chosen junctions.

In the numerical simulation the quantities U_i/Φ_0 for system (2) and W_i^R for the simplified model were calculated for each avalanche.

It should be noted that in calculating them, only the positive parts of the systems were considered. This is because our systems are closed, i.e., one of the most important conditions for the existence of self-organized criticality, the possibility of a sink, is not satisfied for them. However, a process replacing a sink exists separately for positive and negative parts of the system. It consists in the fact that the excess positive current, which cannot leave the system through the boundary, flows into the negative part of the system where it annihilates with the negative current, and similarly the excess negative current is shed into the positive part. Thus, it can be assumed that junctions on the boundary between the positive and negative subsystems act for each of them effectively as a reservoir for the current shed from the subsystems, i.e., the positive parts are systems with one open boundary and self-organization of the critical state is possible in them.

The probability densities $\rho(U_i/\Phi_0)$ and $\rho(W_i^R)$, displayed in Fig. 2b, were also calculated. It is evident in the figure that $\rho(U_i/\Phi_0)$ and $\rho(W_i^R)$ demonstrate power-law behavior, whence it follows that SOC is realized in the positive part of the system described by the differential equations (2) and in the simplified model.

It is clear from Fig. 2b that the results obtained for the simplified model and in the description of the system with differential equations are essentially identical. The agreement obtained shows that the simplified model accurately reflects the basic feature of the behavior of the initial system and belongs the same universality class and the simplying assumptions made do not

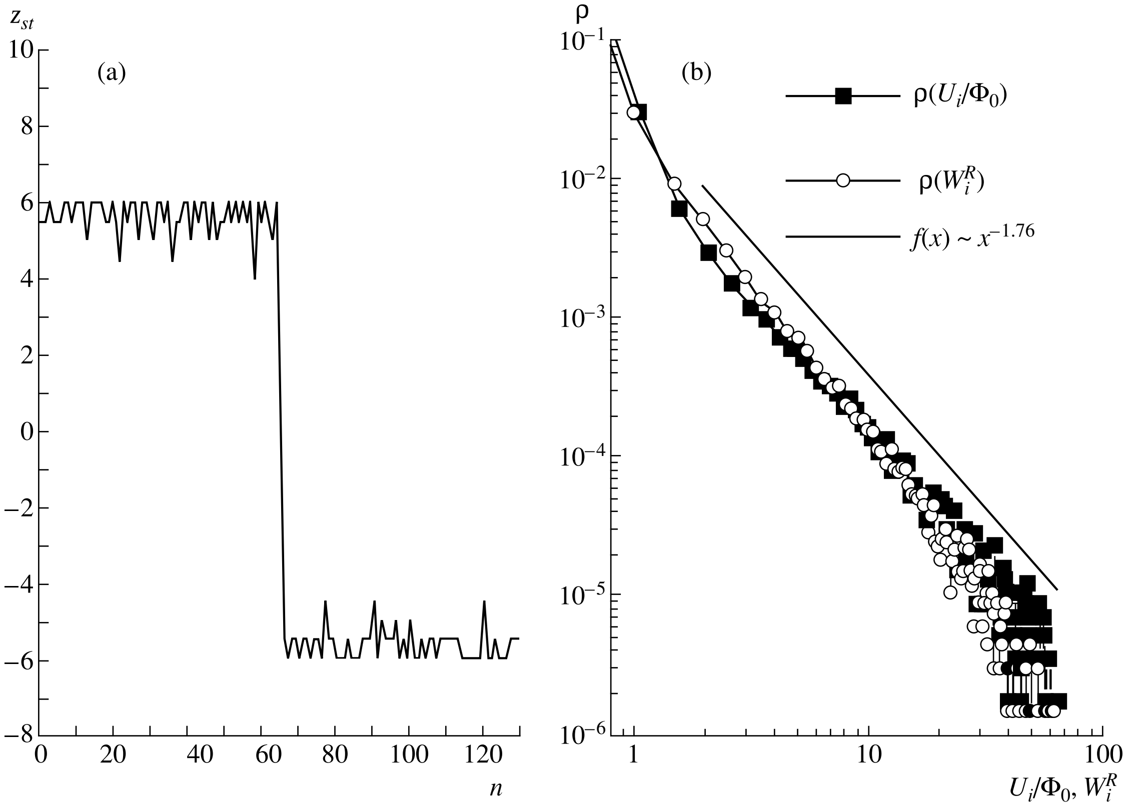


Fig. 2. Results of a computer simulation of the critical state of a one-dimensional multijunction SQUID for the system of differential equations (2) and for the simplified model (16) with the IVC (23). (a) One of the metastable states that comprise the self-organized critical state of a one-dimensional multijunction SQUID, obtained from a system of differential equations; (b) voltage probability density $\rho(U_i/\Phi_0)$ for the system (2) and the probability density $\rho(W_i^R)$ for the simplified model.

lead to a loss of the basic properties of the initial system by the model.

3.2. Two-Dimensional Multijunction SQUID in an External Magnetic Field

In [10] the critical state in the simplified model of a two-dimensional multijunction SQUID placed in an ac magnetic field was studied by computer simulation on an $N \times M$ lattice. In this case closed boundary conditions are realized in the system and the current in the junctions is excited not by the injection current but rather by an ac magnetic field, according to the rules (38). However, as noted in [10], for this method of perturbation SOC is not realized in the system. This is because the system degenerates into a set of one-dimensional strips in which the appearance of SOC is impossible for the perturbation method (38). For this reason, a different perturbation method was used in [10]. When the system reached a metastable state, a unit of current was added in a randomly chosen junction on the right-hand boundary and a current –1 was added at a randomly chosen contact on the left-hand boundary, i.e.,

$$\xi_{1,l} = -1, \quad \xi_{N,j} = 1, \tag{49}$$

where l and j are random and independent of one another. This perturbation rule does not violate the main physical requirement that the total current injected into the system in a single step be zero.

For such a perturbation method, after a certain transient period, the system reached a critical state, which consisted of a set of metastable states that transform into one another and in which the currents were positive in the right-hand side of the system and negative in the left-hand side. Just as for a one-dimensional SQUID, SOC was realized in the positive and negative parts of the system. This was indicated by the power-law probability density $\rho(W_i^R)$. The realization of SOC in a closed system is possible in this case also because of annihilation of the currents at the boundary of the positive and negative subsystems for odd N at contacts with the coordinates $n = (N-1)/2 + 1$ (Fig. 1b). It was shown in [10] that for odd N the properties of the critical state for the positive subsystem in a system with annihilation and in a system of corresponding size with one open boundary are completely equivalent.

In the present work the behavior of a two-dimensional multijunction SQUID was simulated by the system of differential equations (1) on a 31 × 15 lattice

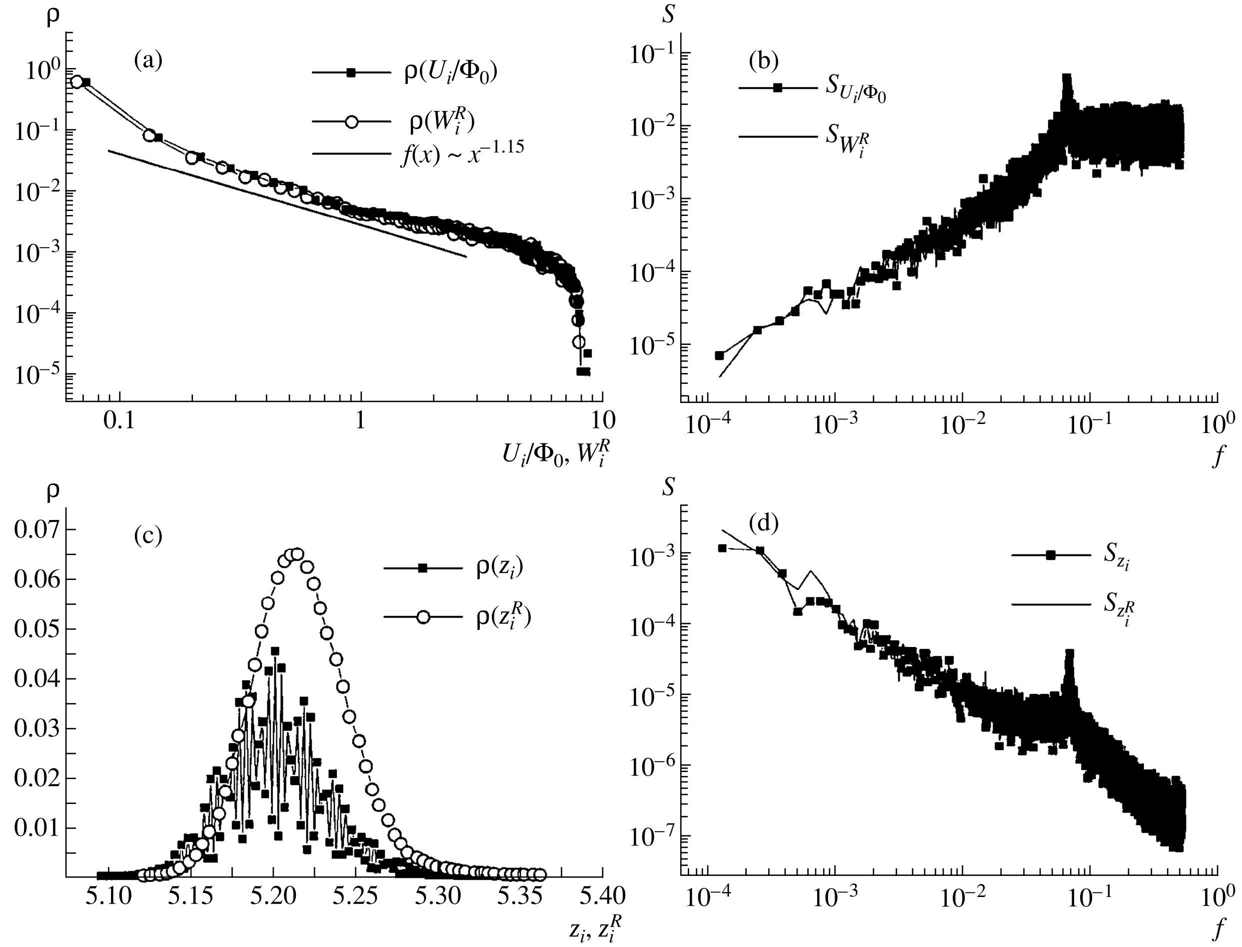


Fig. 3. Results of a computer simulation of the critical state of a two-dimensional multijunction SQUID, placed in an external magnetic field, for the system of differential equations (1) and the simplified model (16) with the IVC (23). (a) Voltage probability density $\rho(U_i/\Phi_0)$ for the system (1) and the probability density $\rho(W_i^R)$ for the simplified model, (b) voltage power spectra $S_{U_i/\Phi_0}(f)$ for the system (1) and $S_{W_i^R}(f)$ for the simplified model, (c) probability density $\rho(z_i)$ for the average current for the system (1) and $\rho(z_i^R)$ for the simplified model, and (d) power spectra of the average current $S_{z_i}(f)$ for the system (1) and $S_{z_i^R}(f)$ for the simplified model.

with closed boundary conditions. The system was perturbed by the rule (49), which for the system of differential equations means

$$\begin{aligned} F_{1,l}(t_0+\delta t) &= F_{1,l}(t_0)-1, \\ F_{N,j}(t_0+\delta t) &= F_{N,j}(t_0)+1, \end{aligned} \tag{50}$$

where l and j are random and independent of one another.

The structure of the critical state arising in the system is completely analogous to that of the critical state in the simplified model. Just as in the simplified model, a positive subsystem in which for each avalanche in the stationary critical state the subsystem-averaged current and integral of the voltage and the probability density and power spectra for these quantities were calculated for each avalanche. The results are displayed in Fig. 3. Figure 3 also displays similar characteristics for the positive subsystem in the simplified model.

It is evident in Fig. 3a that the probability density $\rho(U_i/\Phi_0)$ of the voltage for the positive subsystem, just as in the simplified system, possesses a power-law distribution, which indicates that SOC is realized in the positive subsystem of the closed system. In addition, just as in the one-dimensional case, the results for the simplified model and the initial system of differential equations are essentially identical. Therefore, even in this case both systems are in the same universality class.

We also compared the positive subsystem for the case where the SQUID is described by differential equations with a 15 × 15 system with one open boundary and injection of positive current on the boundary

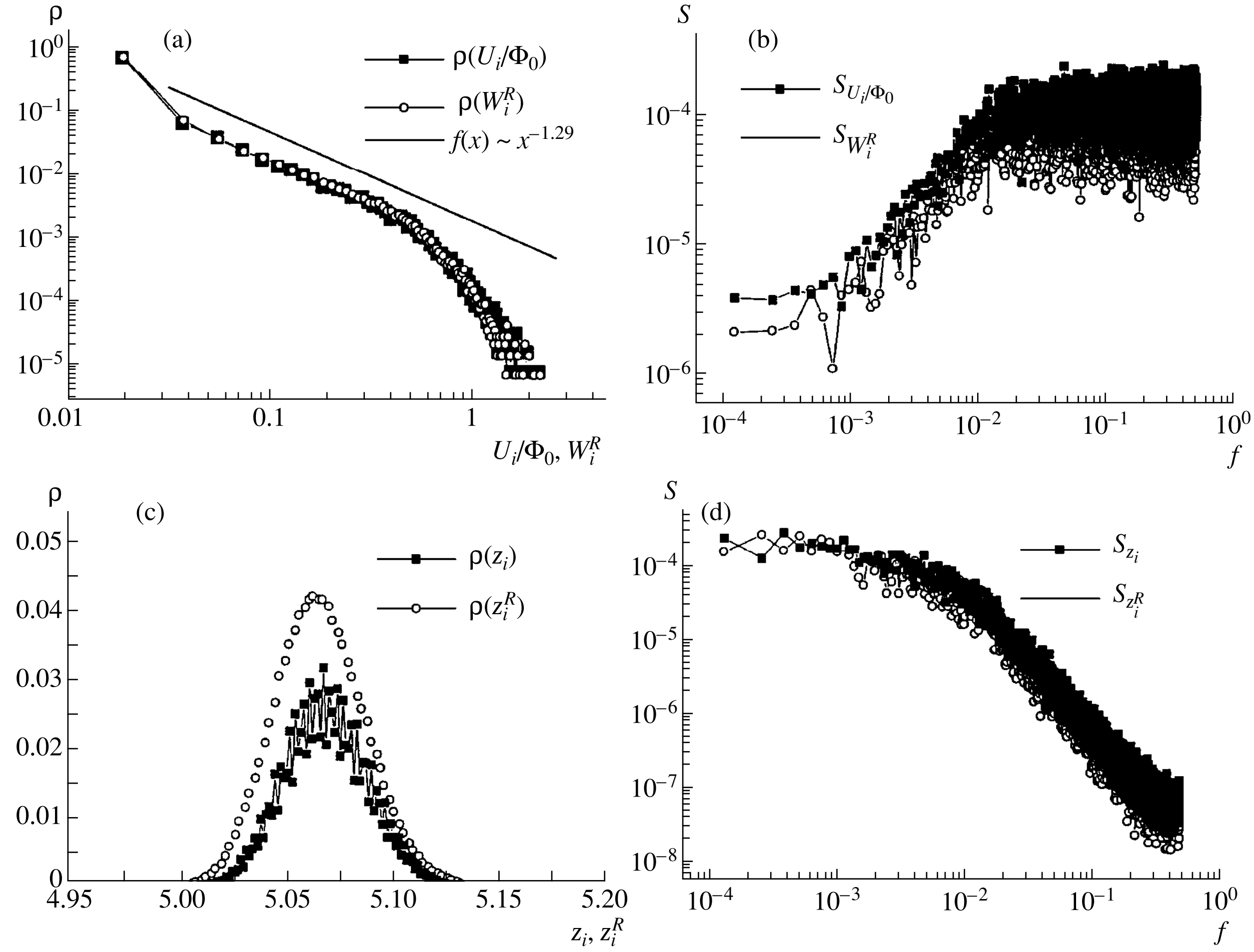


Fig. 4. Results of a computer simulation of the critical state of a two-dimensional multijunction SQUID with current injection into a randomly chosen junction for the system of differential equations (1) and the simplified model (12). (a) Voltage probability density $\rho(U_i/\Phi_0)$ for the system (1) and the probability density $\rho(W_i^R)$ for the simplified model; (b) voltage power spectra $S_{U_i/\Phi_0}(f)$ for the system (1) and $S_{W_i^R}(f)$ for the simplified model; (c) probability density $\rho(z_i)$ of the average current for the system (1) and $\rho(z_i^R)$ for the simplified model; and, (d) power spectra of the average current $S_{z_i}(f)$ for the system (1) and $S_{z_i^R}(f)$ for the simplified model.

opposite to the open boundary. Just as in the simplified model [10], the basic characteristics of the system in the critical state were essentially the same.

3.3. Two-Dimensional Multijunction SQUID with Open Boundary Conditions

The critical state in the simplified model of a two-dimensional multijunction SQUID with open boundary conditions has been studied in detail in [9] with various methods of current injection into the system. Here we present the results of a study of the critical state of a two-dimensional multijunction $N \times N$ ($N = 21$) SQUID described by the system of differential equations (1) with open boundary conditions (25) and the same perturbation methods as in [9].

The critical state arising in this case in the system, just as in the simplified model, is a set of metastable states can transform into one another. Just as in [9], we studied not only the probability density for the average current z_i and the voltage U_i/Φ_0, but also the interavalanche correlators and the power spectra of these quantities.

We obtained the following results for various perturbation methods.

1. Current injection into a randomly chosen junction (method (31) for (1) or (32) for the simplified model (12)).

Figure 4a shows the voltage probability densities in both cases. We can see that they demonstrate power-law behavior. This indicates that a self-organized critical state is realized in the system.

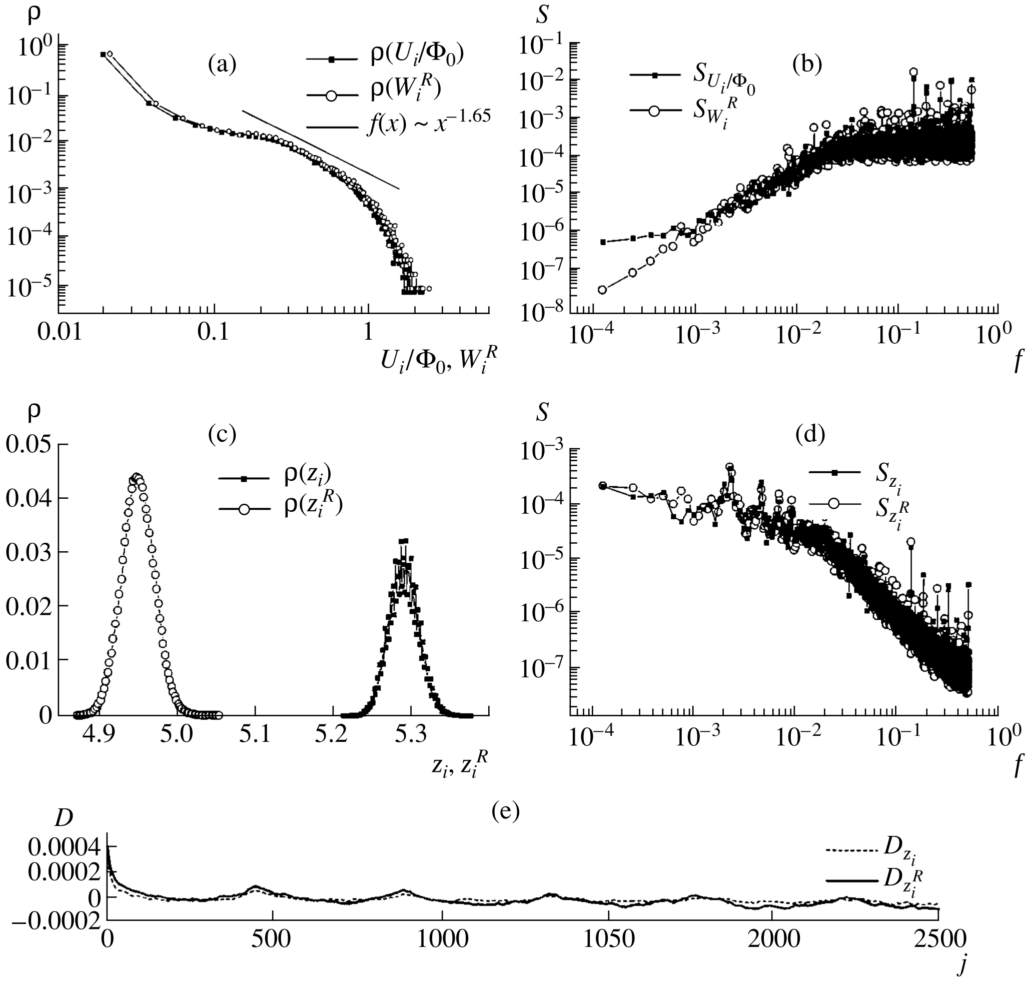


Fig. 5. Results of a computer simulation of the critical state of a two-dimensional multijunction SQUID with current injection into the central junction for the system of differential equations (1) and the simplified model (12) in the case of random initial conditions. (a) Voltage probability density $\rho(U_i/\Phi_0)$ for the system (1) and the probability density $\rho(W_i^R)$ for the simplified model; (b) voltage probability spectra $S_{U_i/\Phi_0}(f)$ for the system (1) and $S_{W_i^R}(f)$ for the simplified model; (c) probability density of the average current $\rho(z_i)$ for the system (1) and $\rho(z_i^R)$ for the simplified model; (d) power spectra of the average current $S_{z_i}(f)$ for the system (1) and $S_{z_i^R}(f)$ for the simplified model; (e) correlation functions of the average current $D_{z_i}(j)$ for the system (1) and $D_{z_i^R}(j)$ for the simplified model.

It is also clearly seen in Figs. 4a–4d that the characteristics considered are also essentially identical for the simplified model and the initial system in the present case. This shows that both of these systems in this situation fall into the same universality class.

2. Current injection in the central junction (method (33) for (1) and (34) for the simplified model (12)).

In this case, for the simplified model (12) and the initial system (1) both “smooth” and “random” initial conditions, for which $F_{n,m}(0)$ are random numbers and $0 \leq F_{n,m}(0) \leq V/2\pi$, were used.

For smooth initial conditions, the results for both systems were essentially the same. For random initial conditions, a difference is observed only in the probability density of the average current (Fig. 5). The behavior of the probability density remains the same but the magnitude of the average current decreases in the simplified model and increases in the initial system (1). This is because the mappings and differential equations respond differently to the introduction of nonintegral initial conditions.

Irrespective of the initial conditions, the voltage probability density for this method of perturbation, just as in the previous cases, shows power-law behavior,

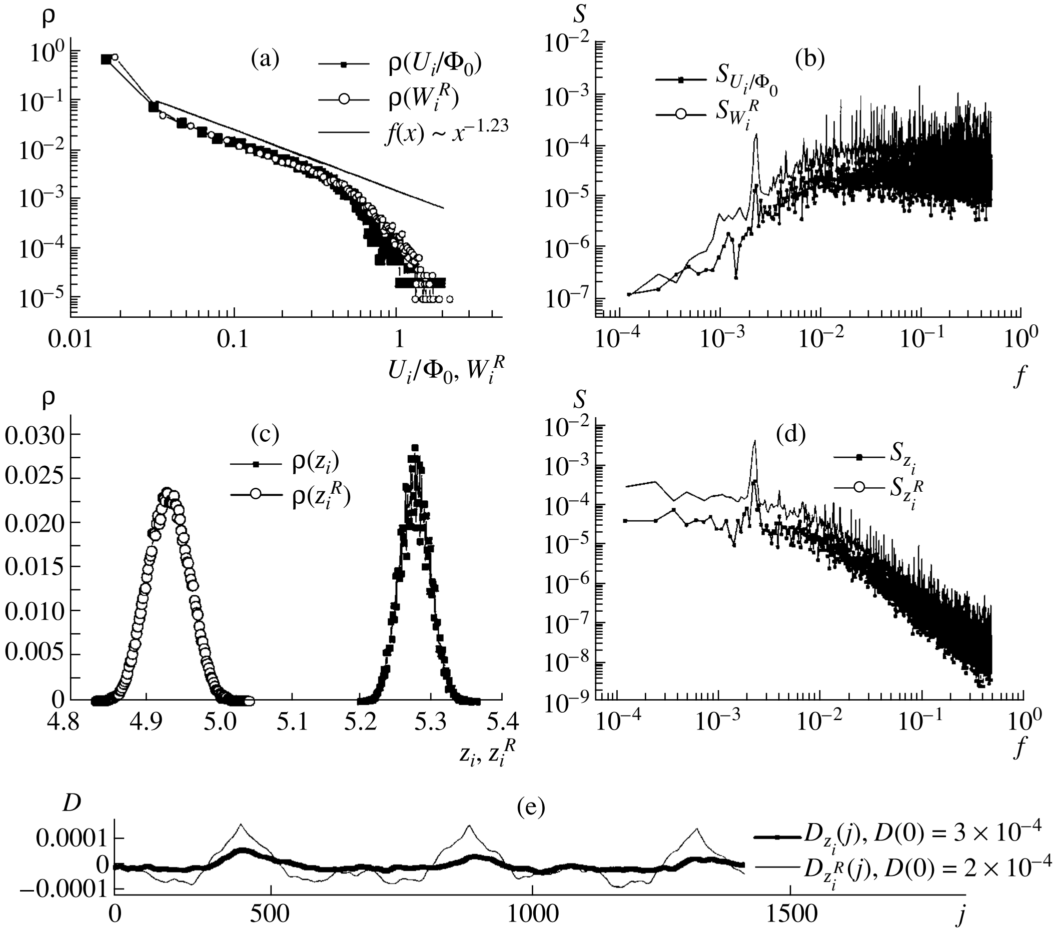


Fig. 6. Results of a computer simulation of the critical state of a two-dimensional multijunction SQUID with current injection simultaneously into all junctions for the system (1) with $\delta F = 1/N^2$ and the simplified model (12) with $p = 1/N^2$. (a) Voltage probability density $\rho(U_i/\Phi_0)$ for the system (1) and $\rho(W_i^R)$ for the simplified model; (b) voltage power spectra $S_{U_i/\Phi_0}(f)$ for the system (1) and $S_{W_i^R}(f)$ for the simplified model; (c) probability density of the average current $\rho(z_i)$ for the system (1) and $\rho(z_i^R)$ for the simplified model; (d) power spectra of the average current $S_{z_i}(f)$ for the system (1) and $S_{z_i^R}(f)$ for the simplified model; (e) correlation function of the average current $D_{z_i}(j)$ for the system (1) and $D_{z_i^R}(j)$ for the simplified model.

i.e., the system remains self-organized. However, the exponents for the cases of smooth and random initial conditions differ somewhat. This phenomenon was already noted in [18], where the sandpile model with sand grains added at the center and different initial conditions was studied. The difference in the exponents could be due to breaking of the internal symmetry of the system.

At the same time, as shown previously in [9], in contrast to the case where current is injected into a randomly chosen junction, a peak at frequency $f \sim 1/N^2$ is noticeable in the power spectrum of the current. This indicates the presence of a quasiperiodic process in the system. Figure 5e shows the interavalanche correlation function for the current. Quasiperiodicity is clearly seen in its behavior.

3. Current injection simultaneously into the entire lattice in small portions (method (35) for (1) or method (36) for the simplified model).

In this case, “random” initial conditions with a variance of the random quantities $F_{n,m}$, such that $0 \le F_{n,m} \le V/2\pi$, were used to study the critical state in the simplified model and for the initial system (1). This is because for smooth initial conditions the behavior of the system became purely periodic and SOC did not arise in it.

We chose the quantity δF for the system of differential equations or p for the simplified model in a manner so that in one step a unit of current was injected into the system, i.e., $N^2\delta F = N^2 p \approx 1$. Figure 6 displays the probability density for the average current and the voltage, the power spectra of these quantities, and the interavalanche correlation function of the average current for the simplified system (12) and the system (1). It is clear from the figure that, just as in the preceding cases, the results for both systems are essentially identical. A difference occurs only for the probability density of the currents.

As has already been shown in [9], the voltage probability density possesses a power-law distribution. This indicates the presence of SOC in the system. However, while the probability density of the voltage shows power-law behavior for all methods considered for perturbing the system, the power spectra for a given method differ from the spectra obtained for the central and random injections. The spectra contain a distinct peak attesting to the presence of a quasiperiodic process with period $T \sim p^{-1}$ or $T \sim \delta F^{-1}$ in the systems.

As noted in [9], the relative magnitude of the peak at frequency $f \sim p$ or $f \sim \delta F$ increases when the magnitude of the current injected into the system in one step increases. In general, however, the periodicity in the behavior of the system is observed not only with increasing current injected into the system in one step but also with “smoothing” of initial conditions, i.e., a decrease in the initial variance of $F_{n,m}$, as well as when the current was injected not into all junctions but into a definite group of junctions.

4. CONCLUSIONS

In the present paper, the critical state of two- and one-dimensional multijunction SQUIDs with a large parameter $V \gg 1$ was studied by computer simulation directly from a system of differential equations for the phase differences on the junctions. It was shown that in this case the critical state in the system is self-organized, i.e., it consists of a set of metastable states, which pass into one another by means of avalanches arising after a local external perturbation of the system. Thus, a physical system with SOC that can be studied not only theoretically but also experimentally was described.

In parallel with the initial systems described by differential equations, simplified models of one- and two-junction SQUIDs, described by systems of mappings and having analogs among standard models for studying SOC, such as the belian sandpile model, were studied. It was shown that for each of the boundary conditions and methods for perturbing the system which were studied in this paper, the behavior of the basic quantities is the same for both systems, i.e., they fall into the same universality class. Therefore the simplifications introduced do not change the basic physical properties of the system. This result is also important because the simplified model is much more convenient for study and computer simulation than the initial system of differential equations. Thus, SOC in granular superconductors can be further studied using the simplified models introduced in [7] for various cases.

In conclusion, we note once again that the physical properties of granular superconductors make it possible to study a system under conditions which are unphysical in previously proposed models for studying SOC. As a result of our study of the characteristic features of SOC in SQUIDs, it was shown that for all boundary conditions and methods for perturbing the system which were studied, the system remains self-organized in the generally accepted sense, i.e., the voltage probability densities have a power-law distribution. At the same time, interavalanche correlators, as characteristics which are more sensitive to the characteristic features of the system, behave differently. A study of these correlators showed that quasiperiodic processes can exist in a system with SOC. Moreover, it was shown that in granular superconductors SOC can exist even with closed boundary conditions. In this case the existence of SOC is ensured not by the possibility of a sink, as in all previously studied open systems, but rather by a fundamentally different mechanism – annihilation of the currents with opposite signs. Thus, the discovery of SOC in granular superconductors gives new information about this phenomenon.

We are grateful to O.V. Gerashchenko and M.A. Pustovoit for an interesting discussion of the problem and for valuable remarks.

This work was sponsored by the Russian Foundation for Basic Research (Grant no. 99-02-17545) and is supported by the Scientific Council on the direction “Superconductivity” of the program “Current Directions in Condensed-Media Physics” as part of project no. 96021 “Profile,” and also the subprogram “Statistical Physics” of the State Scientific and Technical Program “Physics of Quantum and Wave Processes” as part of project VIII-3 and the State Program “Neutron Investigations of Matter.” One of us (N. S.) is grateful to the Russian Foundation for Basic Research for partial financial support (projects 96-15-96775 and 96-15-96764).

REFERENCES

1. D.-X. Chen, J. J. Moreno, and A. Hernando, Phys. Rev. B **53**, 6579 (1996).
2. D.-X. Chen, A. Sánchez, and A. Hernando, Phys. Rev. B **50**, 10342 (1994).
3. A. Majhofer and T. Wolf, Phys. Rev. B **47**, 5383 (1993).
4. C. P. Bean, Rev. Mod. Phys. **36**, 31 (1964).
5. P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
6. G. A. Held, *et al.*, Phys. Rev. Lett. **65**, 1120 (1990).
7. S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **106**, 607 (1994) [JETP **79**, 334 (1994)].

8. D. Dhar, Phys. Rev. Lett. **64**, 1613 (1990).
9. S. L. Ginzburg, M. A. Pustovoit, and N. E. Savitskaya, Phys. Rev. E **57**, 1319 (1998).
10. S .L. Ginzburg and N. E. Savitskaya, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 119 (1999) [JETP Lett. **69**, 133 (1999)].
11. O. I. Kulik and I. K. Yanson, *The Josephson Effect in Superconducting Structures* (Nauka, Moscow, 1970).
12. K. K. Likharev, *Introduction to the Dynamics of Josephson Junctions and Circuits* (Nauka, Moscow, 1985; Gordon and Breach, New York, 1986).
13. L. Kadanoff, S. R. Nagel, L. Wu, *et al.*, Phys. Rev. A **39**, 6524 (1989).
14. S. T. R. Pinho, C. P. C. Prado, and S. R. Salinas, Phys. Rev. E **55**, 2159 (1997).
15. S. S. Manna and J. Kertesz, Physica A **173**, 49 (1991).
16. J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (Mir, Moscow, 1989; Wiley, New York, 1986).
17. S. L. Ginzburg and N. E. Savitskaya, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 688 (1998) [JETP Lett. **68**, 719 (1998)].
18. K. Wiesenfeld, J. Theiler, and B. McNamara, Phys. Rev. Lett. **65**, 949 (1990).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 217–221.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 243–247.
Original Russian Text Copyright © 2000 by Moskalenko, Perel', Yassievich.

SOLIDS

Effect of a Magnetic Field on Thermally Stimulated Ionization of Impurity Centers in Semiconductors by Submillimeter Radiation

A. S. Moskalenko[1], V. I. Perel'*[1], and I. N. Yassievich[1, 2]**
[1] *Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia*
[2] *Department of Theoretical Physics, Lund University, Lund, S-223 62 Sweden*
**e-mail: perel.theory@pop.ioffe.rssi.ru*
***e-mail: irina.yassievich@pop.ioffe.rssi.ru*

Received July 7, 1999

Abstract—The probability of electron tunneling from a bound state into a free state in crossed ac electric and dc magnetic fields is calculated in the quasiclassical approximation. It is shown that a magnetic field decreases the electron tunneling probability. This decreases the probability of thermally activated ionization of deep impurity centers by submillimeter radiation. The logarithm of the ionization probability is a linear function of the squared amplitude of the electric field and increases rapidly with the frequency of the electric field. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the last few years it has been shown experimentally and theoretically that the main mechanism of ionization of impurity centers in semiconductors by an electric field and by submillimeter radiation is thermally activated tunneling [1–4]. This process includes electron tunneling and tunneling of a vibrational system of the center. Recently, the effect of a magnetic field on this process in a dc electric field [5] was studied theoretically using the results of [6] for the electron tunneling probability.

The present paper is devoted to the theory of thermally activated tunneling ionization of centers by an ac electric field in the presence of a dc magnetic field perpendicular to the electric field. The paper consists of two parts. In part one the method used in [7] is applied to derive an expression for the probability of electron tunneling through a time-varying potential barrier. The calculations are performed for the short-range potential of a center. In the case of a charged center, there is hope that an attractive Coulomb potential will increase the tunneling probability but will not greatly change the linear dependence of the logarithm of the probability of thermally activated tunneling ionization on the squared amplitude of the electric field (equation (25); see discussion of a similar question in the case of a dc electric field in the book [3, §10.5]. In part two thermally stimulated tunneling is studied. We confine our attention to exponential accuracy. In this approximation, the logarithm of the ionization probability is a linear function of the squared amplitude of the electric field. A magnetic field decreases the ionization probability.

2. QUASICLASSICAL WAVE FUNCTION OF A TUNNELING ELECTRON

For exponential accuracy the wave function in the quasiclassical approximation is given by the expression

$$\Psi(r, t) \propto \exp\left[\frac{i}{\hbar}\tilde{S}(\mathbf{r}, t)\right], \tag{1}$$

where the action $\tilde{S}(r, t)$ satisfies the Hamilton–Jacobi equations

$$\frac{\partial \tilde{S}}{\partial t} = -\mathcal{H}(\mathbf{P}, \mathbf{r}, t), \quad \nabla\tilde{S} = \mathbf{P} \tag{2}$$

with the boundary condition $\tilde{S}(\mathbf{r}, t) = -\varepsilon t$ as $\mathbf{r} \longrightarrow 0$, where ε is the electron energy on a center ($\varepsilon < 0$). In equations (2) $\mathcal{H}$ is the Hamiltonian and $\mathbf{P}$ is the generalized momentum. In the present case

$$\begin{gathered} \mathcal{H}(\mathbf{P}, \mathbf{r}, t) = \frac{1}{2m}\left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2 - zF(t), \\ \mathbf{P} = m\dot{\mathbf{r}} + \frac{e}{c}\mathbf{A}, \end{gathered} \tag{3}$$

where $F(t) = F\cos\Omega t$ is the force exerted on an electron by the electric field of the wave and $F = e\mathcal{E}$. It is assumed that the field $\mathcal{E}$ is directed along the z axis, and the dc magnetic field $\mathbf{H}$ is directed along the x axis. In what follows, we use the gauge $A_x = A_z = 0$ and $A_y = -Hz$.

1063-7761/00/9001-0217$20.00 © 2000 MAIK "Nauka/Interperiodica"

The action $\tilde{S}$ satisfying the required conditions can be written in the form [7, 8]

$$\tilde{S} = S_0 - \varepsilon t_0, \quad S_0 = \int_{t_0}^{t} \mathcal{L}(r', \dot{r}', t')dt', \tag{4}$$

where $\mathcal{L}$ is the Lagrangian,

$$\mathcal{L}(\mathbf{r}', \dot{\mathbf{r}}', t') = \frac{m\dot{\mathbf{r}}'^2}{2} - z'\left[\frac{eH}{c}\dot{y}' - F(t')\right]. \tag{5}$$

The radius vector $\mathbf{r}'(t')$ satisfies the equations of motion

$$\ddot{x}' = 0, \quad \ddot{y}' = \omega_c \dot{z}', \quad \ddot{z}' = \frac{F}{m}\cos\Omega t' - \omega_c \dot{y}' \tag{6}$$

(where ω_c is the cyclotron frequency) under the conditions

$$\mathbf{r}'(t_0) = 0, \quad \mathbf{r}'(t) = \mathbf{r}. \tag{7}$$

In equations (4) t_0 should be taken to be a function of $\mathbf{r}$ and t, determined by the condition

$$\left(\frac{d\tilde{S}}{dt_0}\right)_{\mathbf{r},\, t} = 0. \tag{8}$$

We underscore that, according to the meaning of the wave function, the quantities $\mathbf{r}$ and t are real, while $\mathbf{r}'$, t', and t_0 can also be complex. Using the equality [2]

$$\left(\frac{dS_0}{dt_0}\right)_{\mathbf{r},\, t} = \mathcal{H}|_{t = t_0},$$

equation (8) can be written in the form

$$\frac{m}{2}(v_{0x}^2 + v_{0y}^2 + v_{0z}^2) = \varepsilon, \tag{9}$$

where $\mathbf{v}_0 = \dot{\mathbf{r}}'(t_0)$ is the velocity at the initial time t_0. According to equation (9), this velocity is purely imaginary, since $\varepsilon < 0$. This is natural, since the electron is located below the barrier.

3. PROBABILITY OF DIRECT TUNNELING OF AN ELECTRON

Having solved the equation of motion (6) under the conditions (7), we can find $\mathbf{v}_0$ as a function of $\mathbf{r}$, t, and t_0. Then equation (9) determines t_0 as a function of $\mathbf{r}$ and t. As a result, the action $\tilde{S}$ can be found as a function of $\mathbf{r}$ and t and therefore the wave function $\Psi(\mathbf{r}, t)$ can also be found.

However, to find the flux from the center (proportional to $|\Psi|^2$) it is sufficient to find $\mathrm{Im}\tilde{S}$ in the region of space where this quantity is maximum, i.e., for values of r such that

$$\nabla \mathrm{Im}\tilde{S} = 0.$$

Using the second equation of equations (2), we obtain the condition

$$\mathrm{Im}\mathbf{P} = 0. \tag{10}$$

From the first two equations of motion (6) and the second equation (3) it follows that P_x and P_y are integrals of motion: $P_x = mv_{0x}$ and $P_y = mv_{0y}$ and, hence, v_{0x} and v_{0y} are real. Then, it follows from equation (9) that v_{0z} is a purely imaginary quantity. It follows from equations (2) that for real values of t the imaginary part of the action does not depend on t in the region of space determining the result of tunneling. Therefore, in what follows, we can set $t = 0$.

We shall write out the solution of the equations of motion (6) that satisfies the first of the conditions (7):

$$\begin{gathered} x' = v_{0x}(t' - t_0), \\ y' = a_1(\cos\omega_c t' - \cos\omega_c t_0) + a_2(\sin\omega_c t' - \sin\omega_c t_0) \\ - \frac{\omega_c}{\Omega}b(\sin\Omega t' - \sin\Omega t_0), \\ z' = -a_1(\sin\omega_c t' - \sin\omega_c t_0) + a_2(\cos\omega_c t' - \cos\omega_c t_0) \\ - b(\cos\Omega t' - \cos\Omega t_0), \\ b = \frac{F}{m(\Omega^2 - \omega_c^2)}. \end{gathered} \tag{11}$$

Here a_1, a_2, and v_{0x} are, for the time being, arbitrary constants, but they are expressed in terms of x, y, z, and t_0 by the second of equations (7). However, the conditions (10) are sufficient to determine these constants in the region that is important for tunneling. These conditions can be rewritten (taking into account the fact that z is real) in the form

$$\mathrm{Im}\left(\frac{d\mathbf{r}'}{dt'}\right)_{t' = t} = 0.$$

Hence it follows that a_1, a_2, and v_{0x} are real.

We set $t_0 = \tau_0 + i\tau_e$ and we write out the condition for z to be real and v_{0z} imaginary, using the last of equations (11) with $t' = t$ and $z' = z$, and differentiating this equation with respect to t' and setting $t' = t_0$. Then, we obtain

$$\begin{aligned} a_1\cos\omega_c\tau_0 + a_2\sin\omega_c\tau_0 &= b\frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}\sin\Omega\tau_0 \\ &= b\frac{\Omega}{\omega_c}\frac{\cosh\Omega\tau_e}{\cosh\omega_c\tau_e}\sin\Omega\tau_0. \end{aligned} \tag{12}$$

It follows from the second equality in equation (12) that $\sin\Omega\tau_0 = 0$, i.e.,

$$\tau_0 = 0, \quad t_0 = i\tau_e, \quad a_1 = 0 \tag{13}$$

(strictly speaking, the condition $\sin\Omega\tau_0 = 0$ admits nonzero solutions for t_0, but it can be shown that they lead to the same results). The fact taht x and y are real (the first two of equations (11) with $t' = t$) leads to the equalities

$$v_{0x} = 0; \quad a_2 = \frac{F\omega_c}{m\Omega(\Omega^2-\omega_c^2)}\frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}. \tag{14}$$

Thus, the solution of the equations of motion that satisfies the first of the conditions (7), the condition (10), and the condition for x', y', and z' to be real at $t' = t$ and z' to be imaginary at $t' = t_0$, we have

$$x' = 0$$

$$y' = \frac{F\omega_c}{m\Omega(\Omega^2-\omega_c^2)}\left[-\sin\Omega t' + \frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}\sin\omega_c t'\right],$$

$$z' = \frac{F}{m(\Omega^2-\omega_c^2)}\left[\frac{\omega_c}{\Omega}\frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}(\cos\omega_c t' - \cosh\omega_c\tau_e) - (\cos\Omega t' - \cosh\Omega\tau_e)\right]. \tag{15}$$

Hence it is easy to obtain $\mathbf{v}_0$, differentiating these expressions with respect to t' and setting $t' = i\tau_e$. Then the relation (9) can be used to determine τ_e. This relation assumes the form

$$\sinh^2\Omega\tau_e\left[1-\left(\frac{\Omega\omega_c}{\Omega^2-\omega_c^2}\right)^2\left(\coth\Omega\tau_e - \frac{\omega_c}{\Omega}\coth\omega_c\tau_e\right)^2\right] = \frac{\Omega^2}{F^2}2m|\varepsilon|. \tag{16}$$

The ionization probability P_e can be written in the form

$$P_e \propto \exp[-2S_e(\varepsilon)], \tag{17}$$

where $S_e(\varepsilon) = \operatorname{Im}\tilde{S}/\hbar$, and $\tilde{S}$ is determined by the formulas (4) and (5) where $\mathbf{r}'$ is given by the formulas (15), $t_0 = i\tau_e$ and $t = 0$. Thus,

$$S_e(\varepsilon) = \frac{m}{2\hbar}\int_0^{\tau_e}(\dot{z}'^2 - \dot{y}'^2)d\tau - \frac{\varepsilon\tau_e}{\hbar}. \tag{18}$$

In deriving this expression, the last two terms in equation (5) were integrated by parts. The variable of integration $\tau = -it'$ was introduced in equation (18), so that

$$\dot{y}' = \frac{F\omega_c}{m(\Omega^2-\omega_c^2)}\left[-\cosh\Omega\tau + \frac{\omega_c}{\Omega}\frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}\cosh\omega_c\tau\right],$$

$$\dot{z}' = \frac{F\omega_c}{m(\Omega^2-\omega_c^2)} \times\left[\frac{\Omega}{\omega_c}\sinh\Omega\tau - \frac{\omega_c}{\Omega}\frac{\sinh\Omega\tau_e}{\sinh\omega_c\tau_e}\sinh\omega_c\tau\right]i. \tag{19}$$

Thus, the ionization probability is given to exponential accuracy by equation (17), where $S_e(\varepsilon)$ and $\tau_e(\varepsilon)$ are determined by equations (16) and (18).

We note the relation

$$\frac{d\tilde{S}}{d\varepsilon} = \left(\frac{\partial S_0}{\partial t_0} - \varepsilon\right)\frac{dt_0}{d\varepsilon_0} - t_0 = -t_0,$$

whence it follows that

$$\tau_e = -\hbar\frac{\partial S_e}{\partial\varepsilon}. \tag{20}$$

The quantity τ_e can be called the electron tunneling time. Without a magnetic field (for $\omega_c = 0$) equations (16)–(19) are identical to the result for the tunneling ionization in an ac electric field [7] and for $\Omega = 0$ they are identical with the result for tunneling ionization in a dc electric field in the presence of a magnetic field [6].

As the magnetic field increases, the ionization probability decreases and the tunneling time increases.

4. THERMALLY ACTIVATED TUNNELING IONIZATION OF A SHORT-RANGE CENTER

The probability of tunneling ionization with the participation of phonons can be regarded as being the result of three processes: thermal excitation of the system into a vibrational level E_1 in an adiabatic potential $U_1(x)$, corresponding to an electron bound on a center (x is the vibrational coordinate), tunneling transition of the vibrational system to the adiabatic potential $U_{2\varepsilon}(x)$, corresponding to a free electron with a negative energy ε, and a tunneling transition of an electron under the action of an electric field from a well into a free state with negative energy ε.

In the quasiclassical approximation, the ionization probability for fixed ε and E_1 is proportional to the expression

$$\exp\left(-\frac{E_1}{kT}\right)\exp[-2(S_{2\varepsilon}-S_{1\varepsilon})]\exp[-2S_e(\varepsilon)], \tag{21}$$

where the three cofactors give the probabilities of the three processes enumerated above,

$$S_{1\varepsilon} = \frac{\sqrt{2M}}{\hbar}\int_{a_1}^{x_{c\varepsilon}}\sqrt{U_1(x)-E}\,dx, \quad E = E_1 - \varepsilon_T,$$

$$S_{2\varepsilon} = \frac{\sqrt{2M}}{\hbar}\int_{a_{2\varepsilon}}^{x_{c\varepsilon}}\sqrt{U_{2\varepsilon}(x)-E}\,dx, \quad U_{2\varepsilon}(x) = U_2(x)+\varepsilon \tag{22}$$

(the notations are illustrated in Fig. 1), and M is the mass associated with the vibrational mode playing the main role in the process.

The total ionization probability is obtained by integrating the expression (21) with respect to E_1 and ε.

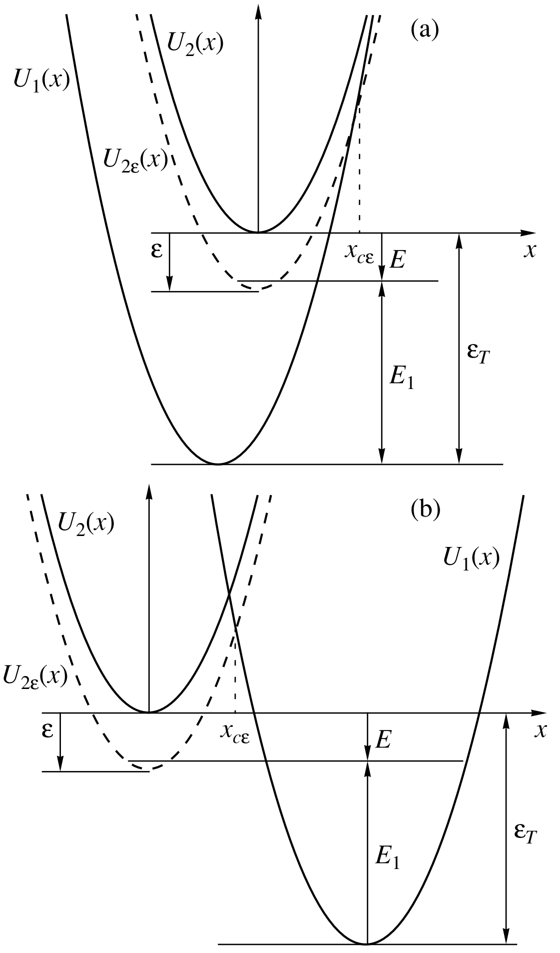


Fig. 1. Diagram of adiabatic potentials, illustrating the tunneling rearrangement of the vibrational system accompanying ionization for two possible arrangements of the potentials (a and b; autolocalization occurs in the case b). Explanations are presented in the text.

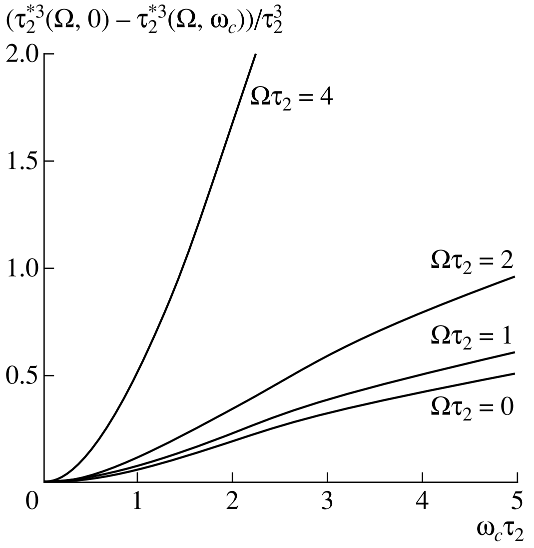


Fig. 2. $[\tau_2^{*3}(0, \Omega) - \tau_2^{*3}(\omega_c, \Omega)]/\tau_2^3$ versus $\omega_c\tau_2$ for various values of the parameter $\Omega\tau_2$.

The integrals can be calculated by the method of steepest descent. This gives two equations for the optimal quantities E_{1m} and ε_m:

$$\tau_{2\varepsilon} = \tau_{1\varepsilon} + \hbar/2kT, \quad \tau_{2\varepsilon} = \tau_e, \tag{23}$$

where the "tunneling times" $\tau_{n\varepsilon}$ of the vibrational system are determined by the expression

$$\tau_{n\varepsilon} = -\hbar\frac{\partial S_{n\varepsilon}}{\partial E}.$$

For not too strong fields, the quantity ε_m is small and we can set $\varepsilon = 0$ in the first equation in equations (23). It determines the energy E_{1m}, which does not depend on the electric field. In the expression (21) the quantity $S_{2\varepsilon} - S_{1\varepsilon}$ can be expanded in powers of ε, retaining only the first term in the expansion,

$$S_{2\varepsilon} - S_{1\varepsilon} = S_2 - S_1 + \frac{\varepsilon\tau_2}{\hbar}, \tag{24}$$

where S_2, S_1, and τ_2 are the values of $S_{2\varepsilon}$, $S_{1\varepsilon}$, and $\tau_{2\varepsilon}$ at $\varepsilon = 0$. Then, using equation (18) with $\tau_e = \tau_2$, the ionization probability $e(F)$ can be obtained as

$$e(F) = e(0)\exp\frac{F^2}{F_c^2}, \tag{25}$$

where it is convenient to write

$$F_c^2 = \frac{3m\hbar}{\tau_2^{*3}},$$

$$\begin{aligned}\tau_2^{*3} &= \frac{3\omega_c^2}{(\Omega^2 - \omega_c^2)^2} \\ &\times \int_0^{\tau_2}\left\{\left[-\cosh\Omega\tau + \frac{\omega_c}{\Omega}\frac{\sinh\Omega\tau_2}{\sinh\omega_c\tau_2}\cosh\omega_c\tau\right]^2\right. \\ &\left. + \left[\frac{\Omega}{\omega_c}\sinh\Omega\tau - \frac{\omega_c}{\Omega}\frac{\sinh\Omega\tau_2}{\sinh\omega_c\tau_2}\sinh\omega_c\tau\right]^2\right\}d\tau.\end{aligned} \tag{26}$$

The dependence of τ_2^* on the magnetic field is illustrated in Fig. 2.

Without a magnetic field (in the limit $\omega_c \longrightarrow 0$) the expressions (25) and (26) are identical to the expression for the probability of multiphonon thermally activated tunneling under the action of an ac electric field [2]. In the limit $\Omega \longrightarrow 0$, it follows from equation (26) that

$$\tau_2^{*3} = \frac{3\tau_2}{\omega_c^2}(\omega_c\tau_2\coth\omega_c\tau_2 - 1), \tag{27}$$

which gives the result for the probability of this process under the action of a dc electric field in the presence of a magnetic field [5].

We note that the applicability condition for equations (25) and (26) are the inequalities $F \gg F_c$ and $|\varepsilon_m| \ll \varepsilon_T$, where $|\varepsilon_m|$ is determined by equation (16) with $\tau_e = \tau_2$. According to the first of equations (23) with $\varepsilon = 0$

$$\tau_2 = \tau_1 + \hbar/2kT, \tag{28}$$

whence τ_1 is essentially temperature-independent. For adiabatic potentials of the type shown in Fig. 1b (autolocalization), we have $\tau_1 < 0$.

5. CONCLUSIONS

In summary, in the present paper an expression was obtained for the probability of a tunneling transition of an electron from a bound state into a free state under the action of an ac electric field in the presence of a magnetic field oriented perpendicular to the electric field (equations (17)–(19)).

The result was used to calculate the probability of thermally activated tunneling ionization of deep impurity centers in semiconductors by submillimeter radiation in a magnetic field (equations (25), (26), and (28)). All results were obtained to exponential accuracy.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (Grants 98-02-18268 and 96-15-96392). I. N. Yassievich is grateful to the Swedish NSRC (Grant O-AH/KG 03996-322).

REFERENCES

1. V. Karpus and V. I. Perel’, Pis’ma Zh. Éksp. Teor. Fiz. **42**, 403 (1985) [JETP Lett. **42**, 497 (1985)].
2. S. D. Ganichev, E. Ziemann, Th. Gleim, *et al.*, Phys. Rev. Lett. **80**, 2409 (1998).
3. V. N. Abakumov, V. I. Perel’, and I. N. Yassievich, *Nonradiative Recombination in Semiconductors* (Peterburgskiĭ Institut Yadernoĭ Fiziki im. B. P. Konstantinova RAN, St. Petersburg, 1988); *Modern Problems in Condensed Matter Sciences*, Ed. by V. M. Agranovich and A. A. Maradudin (North-Holland, Amsterdam, 1991), Vol. 33.
4. S. D. Ganichev, I. N. Yassievich, and W. Prettl, Fiz. Tverd. Tela **39**, 1905 (1997) [Phys. Solid State **39**, 1703 (1997)].
5. V. I. Perel’ and I. N. Yassievich, Pis’ma Zh. Éksp. Teor. Fiz. **68**, 763 (1998) [JETP Lett. **68**, 804 (1998)].
6. V. S. Popov and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **113**, 2047 (1998) [JETP **86**, 1122 (1998)].
7. L. D. Landau and E. M. Lifzhitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989; Pergamon, Oxford, 1977).
8. L. D. Landau and E. M. Lifzhitz, *Mechanics*, 3rd ed. (Nauka, Moscow, 1988; Pergamon, Oxford, 1976).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 222–226.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 248–252.
Original Russian Text Copyright © 2000 by Il’ichev.

MISCELLANEOUS

Recombination Accompanied by Teleportation of “Quantum Entanglement”

L. V. Il’ichev

Institute of Automatics and Electrometry, Siberian Division, Russian Academy of Sciences,
Novosibirsk, 630090 Russia;
e-mail: shalagin@iae.nsk.su

Received May 13, 1999

Abstract—It is remarked that the recently proposed [2] and partially realized [3] teleportation of quantum states of microparticles is (in a certain modification) a natural element of the process of cross recombination of radical spin pairs [4]. In the present paper the situation where singlet and triplet pairs are created in the medium and then diffuse and can recombine, if the particles encountering one another are in a singlet state, is examined in detail. As a result of cross recombination and teleportation, there arises a countable set of spin states of radical pairs. The distributions of the densities of radical pairs corresponding to this set are found under stationary and spatially uniform conditions. *© 2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The phenomenon of so-called quantum entanglement (QE) has been known for a long time, but it has aroused great interest in the last few years in connection with its role in the quantum theory of information and the theory of a quantum computer (see, for example, [1]). Quantum entanglement gives rise to specific correlations, which have no classical analog, between subsystems of a single quantum system. In the present paper the following general negative definition of QE is used: A state (the density matrix) $\hat{\rho}(1, 2)$ of the union of two subsystems S_1 and S_2 is said to be “nonentangled” if it can be represented as a convex linear combination of products of the states of the subsystems, specifically,

$$\hat{\rho}(1, 2) = \sum_i p_i \hat{\rho}_i(1) \otimes \hat{\rho}_i(2), \tag{1}$$

where $p_i > 0$ and $\sum_i p_i = 1$. We note that for a “pure” state $\hat{\rho}(1, 2) = |\psi(1, 2)\rangle\langle\psi(1, 2)|$ this definition of “nonentanglement” reduces to the well-known requirement that the wave function be factorizable: $|\psi(1, 2)\rangle = |\psi(1)\rangle \otimes |\psi(2)\rangle$.

The existence of QE is the main attribute in realizing the so-called quantum teleportation [2, 3]. We shall be interested in a modification of this phenomenon, specifically, quantum entanglement teleportation (QET). The crux of the QET process (in the form in which it appears below) is as follows. Two individuals, Alice and Bob, each possess a pair of spin-1/2 particles. We denote Alice’s particles by the numbers 1 and 3 and Bob’s particles by the numbers 2 and 4. We consider a situation where the state of all four particles has the form

$$\hat{\rho}(1, 2, 3, 4) = \hat{\rho}(1, 2) \otimes \hat{\rho}(3, 4). \tag{2}$$

In this case, as one can easily see, the pairs of Alice’s and Bob’s particles *a priori* are in the states

$$\begin{aligned} \hat{\rho}^{(pre)}(1, 3) &= \mathrm{Tr}_{2,4}\hat{\rho}(1, 2, 3, 4) = \hat{\rho}(1) \otimes \hat{\rho}(3), \\ \hat{\rho}^{(pre)}(2, 4) &= \mathrm{Tr}_{1,3}\hat{\rho}(1, 2, 3, 4) = \hat{\rho}(2) \otimes \hat{\rho}(4) \end{aligned} \tag{3}$$

and in accordance with equation (1) they do not possess the property of QE. Alice performs a measurement on her particles, checking to see whether or not her pair is in a singlet state,

$$\hat{P}_0(1, 3) \equiv |\Psi_0(1, 3)\rangle\langle\Psi_0(1, 3)|, \tag{4}$$

where

$$|\Psi_0(1, 3)\rangle = \frac{1}{\sqrt{2}}(|\psi_+(1)\rangle \otimes |\psi_-(3)\rangle - |\psi_-(1)\rangle \otimes |\psi_+(3)\rangle),$$

$|\psi_\pm(i)\rangle$ is the state of particle i with spin projection ±1/2 in a certain fixed direction. We note that the state $\hat{P}_0(1, 3)$ refers to the so-called “maximally entangled” state, in

1063-7761/00/9001-0222$20.00 © 2000 MAIK “Nauka/Interperiodica”

which the states of the particles 1 and 3 are maximally mixed:

$$\hat{\rho}_0(1) = \mathrm{Tr}_3\hat{P}_0(1,3) = \frac{1}{2}\hat{I}(1),$$

$$\hat{\rho}_0(3) = \mathrm{Tr}_1\hat{P}_0(1,3) = \frac{1}{2}\hat{I}(3).$$

All information carried by such states is contained in the interparticle correlations.

For a positive outcome of an experiment, Alice's pair at Bob's location are in the state

$$\hat{\rho}^{(post)}(2,4) = \frac{\mathrm{Tr}_{1,3}[\hat{P}_0(1,3)(\hat{\rho}(1,2)\otimes\hat{\rho}(3,4))\hat{P}_0(1,3)]}{\mathrm{Tr}_{1,2,3,4}[\hat{P}_0(1,3)(\hat{\rho}(1,2)\otimes\hat{\rho}(3,4))\hat{P}_0(1,3)]} \quad (5)$$

in accordance with von Neumann's projection postulate. It is convenient to introduce the QET mapping $\tau_{1,3}$, acting on the set of all states of particles 1, 2, 3, and 4, of the form (2). The result of this mapping is the state (5):

$$\tau_{1,3}: \hat{\rho}(1,2)\otimes\hat{\rho}(3,4) \longrightarrow \hat{\rho}^{(post)}(2,4). \quad (6)$$

In the next section it would be shown that in the presence of an initial "entanglement" within the pairs 1, 2 and 3, 4 (the so-called QE resource) the state $\hat{\rho}^{(post)}(2,4)$, in contrast to $\hat{\rho}^{(pre)}(2,4)$, is "entangled."

The fundamental condition, anticipating any use of quantum correlations of the QE type which exist in the state $\hat{\rho}^{(post)}(2,4)$, is Bob receiving a signal from Alice indicating a positive outcome for her experiment. In the literature, this circumstance is often interpreted as the necessity of the existence of two types of channels for transmitting information in realizing QET (and also standard quantum teleportation)—a classical channel, along which Alice reports about the outcome of her experiment, and the so-called Einstein–Podolsky–Rosen (EPR) quantum channels, formed by the world lines of "entangled" pairs of particles 1, 2 and 3, 4. It should be noted that in our exposition the states of the pairs 1, 2 and 3, 4 appear as two equivalent EPR channels. The pertinence of such a point of view is dictated by the specific nature of the problem considered below. Our symmetric interpretation differs from the conventional one [2], where the EPR channel is only one pair, for example, 1, 2, and this channel is used for teleportation of the "entanglement" existing between the spin states of particles 3 and 4 into a new pair 2, 4. A more accurate term, used in the literature for the process under study, *entanglement swapping*, is the transport or transfer of entanglement.

The view that QET is an exotic phenomenon is probably incorrect. In the present paper we study a simple model process of creation and recombination of spin radicals in a medium that is accompanied by QET. As far as I know, the aspect associated with teleportation was first brought up in [4]. The crux of the phenomenon is as follows. Let us assume that only singlet pairs can be created and recombined, and the spin states of the correlated pairs are not subjected to any perturbations and remain singlets in the intervals between the creation and recombination events. As a result of diffusion, particles from different pairs can approach one another.[1] Successful recombination is identical to a positive outcome of Alice's experiment. The photon emitted on recombination carries a classical message indicating success, and the remaining fragments of the two pairs are in a singlet state. The kinetics of this process, taking account of possible creation of triplet pairs also, is studied in Section 3, but first we need to develop a convenient language for quantitative description of QET in our model.

2. GENERAL RELATIONS FOR QET ACCOMPANYING RECOMBINATION

As already mentioned in the Introduction, QET in the presence of recombination appears to be especially simple if all correlated pairs are singlets. In this case, each recombination event annihilates two singlet pairs and creates one new pair—"teleportation of singletness" occurs. The problem of this section is to determine the quantitative characteristics of QET accompanying recombination of fragments of two pairs mixed in an arbitrary manner. These characteristics are required in the situation where triplet pairs appear in the medium for some reason. As will be shown below, this is equivalent to a decrease in the quality of EPR channels. We shall confine our attention to the spatially isotropic situation. In this case, the most general states of pairs of particles *i* and *j* have the form

$$\hat{\rho}^{(\alpha)}(i,j) = \frac{3\alpha+1}{4}\hat{P}_0(i,j) + \frac{1-\alpha}{4}(\hat{P}_{1,-1}(i,j) + \hat{P}_{1,0}(i,j) + \hat{P}_{1,+1}(i,j)), \quad (7)$$

where we have introduced the parameter $\alpha \in [-1/3, 1]$, and the operators $\hat{P}_{1,-1}(i,j)$, $\hat{P}_{1,0}(i,j)$, and $\hat{P}_{1,+1}(i,j)$, by analogy with equation (4), are projectors onto the states

$$|\Psi_{1,-1}(i,j)\rangle = |\psi_-(i)\rangle\otimes|\psi_-(j)\rangle,$$

---

[1] This fundamentally distinguishes the model under consideration from the process of geminal (cell) recombination [5]. The recombination of radicals from different pairs has been studied, for example, in [6].

$$|\Psi_{1,0}(i,j)\rangle = \frac{|\psi_+(i)\rangle \otimes |\psi_-(j)\rangle + |\psi_-(i)\rangle \otimes |\psi_+(j)\rangle}{\sqrt{2}},$$

$$|\Psi_{1,+1}(i,j)\rangle = |\psi_+(i)\rangle \otimes |\psi_+(j)\rangle,$$

respectively. The sum in parentheses in equation (7) is, to within a factor of 1/3, an isotropic triplet state. The parameter α, which gives the fraction of singletness and tripletness, is introduced in a manner so that $\alpha = 1$ corresponds to a purely singlet pair, $\alpha = 0$ corresponds to an uncorrelated maximally mixed state of the particles i and j, and $\alpha = -1/3$ corresponds to a purely triplet state.

It is not difficult to show that in the model under consideration the recombination of two particles from different pairs occurs with probability 1/4 (this immediately determines the normalization factor in equation (5)) and, which is especially important, it dictates the unique multiplicative character of the teleportation mapping from equation (6):

$$\tau_{1,3}\colon \hat{\rho}^{(\alpha)}(1,2) \otimes \hat{\rho}^{(\beta)}(3,4) \longrightarrow \hat{\rho}^{(\alpha\beta)}(2,4). \quad (8)$$

Since the parameters α and β of the initial pairs do not exceed 1 in magnitude, we have

$$0 \le |\alpha\beta| \le \min(|\alpha|, |\beta|).$$

This reflects the unavoidable degradation of the teleported state as compared with the states of the EPR channels 1–2, 3–4.

If it is assumed that triplet pairs, just like singlet pairs, arise as a result of a dissociation-type process, then it can be concluded from the multiplicative character of the mapping τ that only pairs whose parameter α is an integer power $n = 0, 1, 2, \ldots$ of the number $-1/3$, corresponding to a triplet state, are present in the medium:

$$\begin{aligned}\hat{\rho}^{(n)}(i,j) &= \frac{1}{4}\left[1 + 3\left(-\frac{1}{3}\right)^n\right]\hat{\rho}^{(0)}(i,j) \\ &+ \frac{3}{4}\left[1 - \left(-\frac{1}{3}\right)^n\right]\hat{\rho}^{(1)}(i,j).\end{aligned} \quad (9)$$

Here the state $\hat{\rho}$ is indexed by the power n and not by the parameter α itself, as in equation (7). In the chosen notation, $\hat{\rho}^{(0)}(i, j)$ corresponds to a singlet state and $\hat{\rho}^{(1)}(i, j)$ corresponds to a triplet state. In the new notation the teleportation mapping becomes additive:

$$\tau_{1,3}\colon \hat{\rho}^{(n)}(1,2) \otimes \hat{\rho}^{(m)}(3,4) \longrightarrow \hat{\rho}^{(n+m)}(2,4). \quad (10)$$

The maximally mixed state must now be designated as $\hat{\rho}^{(\infty)}(i, j)$.

3. KINETIC EQUATIONS AND THEIR SOLUTIONS

We shall now formulate the kinetic model. We shall consider an unbounded medium in which creation and recombination of pairs of particles with spin 1/2 occur.[2] For simplicity, we shall assume that the particles created are identical. We assume that only singlet pairs can recombine, while both singlet and triplet pairs can be created. In the intervals between creation and recombination events, the particles diffuse with diffusion coefficient D, which does not depend on the state of the particle. We introduce the set of functions $f^{(n)}(\mathbf{r}_1, \mathbf{r}_2)$, two-coordinate spatial densities of pairs of particles in the spin state $\hat{\rho}^{(n)}$ from equation (9). The densities satisfy the following system of kinetic equations:

$$\begin{aligned}&\partial_t f^{(n)}(\mathbf{r}_1, \mathbf{r}_2) \\ &= -\nu f(\mathbf{r}_1) f^{(n)}(\mathbf{r}_1, \mathbf{r}_2) - \nu f^{(n)}(\mathbf{r}_1, \mathbf{r}_2) f(\mathbf{r}_2) \\ &+ \nu \sum_{k=0}^{n} \int f^{(k)}(\mathbf{r}_1, \mathbf{r}') f^{(n-k)}(\mathbf{r}', \mathbf{r}_2) d^3r' \\ &+ D(\Delta_{\mathbf{r}_1} + \Delta_{\mathbf{r}_2}) f^{(n)}(\mathbf{r}_1, \mathbf{r}_2) + \gamma^{(n)}(\mathbf{r}_1)\delta(\mathbf{r}_1 - \mathbf{r}_2).\end{aligned} \quad (11)$$

Here the first two terms on the right-hand side correspond to annihilation of pairs accompanying recombination of one of its particles, located at the points $\mathbf{r}_1$ and $\mathbf{r}_2$;

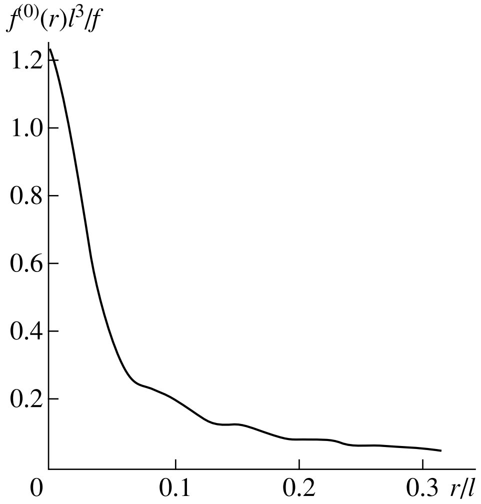
$$f(\mathbf{r}) = \sum_{n=0}^{\infty} \int f^{(n)}(\mathbf{r}, \mathbf{r}') d^3r'$$

is the total density of the recombining pairs. The third term corresponds to creation of a pair accompanying annihilation of two other pairs. The parameters of the created pair are related with the parameters of the annihilated pairs via the QET mapping (10). The last two terms are diffusion terms and the source of local pair creation. We assume that only the quantities $\gamma^{(0)}$ and $\gamma^{(1)}$ are different from zero. We note that in the case of singlet and triplet pairs we use singular sources in equation (11). At the same time, it is known that because of the statistics of spin-1/2 particles the coordinate density matrix of a created triplet pair and together with it the function $\gamma^{(1)}(\mathbf{r}_1, \mathbf{r}_2)$ must approach zero as $\mathbf{r}_1 \longrightarrow \mathbf{r}_2$. However, in our approach, where all spatial scales are much greater than the molecular scale (and this is the characteristic size of the radical pair just created), we are fully justified in replacing the true source of triplet pairs by its singular approximation.

We shall consider the spatially uniform and stationary situation where $f^{(n)}(\mathbf{r}_1, \mathbf{r}_2) = f^{(n)}(|\mathbf{r}_1 - \mathbf{r}_2|)$, $f(\mathbf{r}) = f$,

---

[2] Spin must be understood in the generalized sense. It can be the standard spin and any dichotomous observable, for example, chirality. In the latter case, the states "spin up" and "spin down" must be identified with right-hand and left-hand enantiomers.



Stationary density of singlet pairs (see explanation in the text).

$\gamma^{(n)}(\mathbf{r}) = \gamma^{(n)}$. Then the equations for the Fourier transforms of the functions $f^{(n)}(|\mathbf{r}_1 - \mathbf{r}_2|)$,

$$f_q^{(n)} = \int f^{(n)}(|\mathbf{r}|)\exp(-i\mathbf{q}\cdot\mathbf{r})d^3r,$$

acquire the especially simple form

$$2(\nu f + Dq^2)f_q^{(n)} = \nu\sum_{k=0}^{n} f_q^{(k)} f_q^{(n-k)} + \gamma^{(n)}. \tag{12}$$

The total stationary particle density $f = \sum_{n=0}^{\infty} f_0^{(n)}$ is determined by the same system of equations:

$$f = \sqrt{\frac{\gamma^{(0)}}{\nu} + \frac{\gamma^{(1)}}{\nu}}. \tag{13}$$

A quadratic equation is obtained for the density $f_q^{(0)}$ of singlet pairs (and only for $f_q^{(0)}$). The stable solution is

$$f_q^{(0)} = f + \frac{D}{\nu}q^2 - \sqrt{\left(f + \frac{D}{\nu}q^2\right)^2 - \frac{\gamma^{(0)}}{\nu}}. \tag{14}$$

The solution of the system (12) for the Fourier transforms $f_q^{(n)}$ of the densities for $n = 1, 2, \ldots$ has the form

$$f_q^{(n)} = C_n\left(\frac{\gamma^{(1)}}{2\nu}\right)^n\left[\left(f + \frac{D}{\nu}q^2\right)^2 - \frac{\gamma^{(0)}}{\nu}\right]^{1/2-n}, \tag{15}$$

where the constants C_n are determined from the system of recurrence relations

$$C_n = \frac{1}{2}\sum_{k=1}^{n-1} C_k C_{n-k},$$

$C_1 = 1$, $n = 2, 3, \ldots$. The characteristic form of the distribution $f^{(0)}(r)$ is displayed in the figure.

The plot was constructed for equal rates of creation of triple and singlet pairs ($\gamma^{(0)} = \gamma^{(1)}$). An expression for the characteristic length l is presented in the Conclusions (see equation (16)). We note the unique step character of the function $f^{(0)}(r)$.

4. CONCLUSIONS

We have described on the basis of a simple "mass action" type law the kinetics of the creation, diffusion, and recombination of pairs of particles in "entangled" quantum spin states. A nontrivial feature of the model is that the process of QET in the recombination of particles is taken into account. The stationary uniform regime admits, as shown above, a simple analytical description. We note that taking account of the evolution of the spin states between creation and recombination events should substantially change the results obtained.

As follows from equations (12) and the figure, as a result of diffusion and the QET process, a substantial fraction of "entangled" spin states with characteristic size $0.1l$ can appear, where

$$l = \left(\frac{D^2}{\nu(\gamma^{(0)} + \gamma^{(1)})}\right)^{1/4}. \tag{16}$$

Nonlocal, nonclassical spin correlations arise in the medium. This should have nontrivial thermodynamic consequences. We shall examine as an example the fluctuations of the total projection of the spin of the particles located in a certain region Ω. For simplicity we shall assume that only singlet pairs are present. We present without derivation the result of calculation of the distribution function $p(m)$ for the probability of finding the total spin projection m of all particles in a chosen direction:

$$p(m) = I_{|2m|}(\xi)e^{-\xi}. \tag{17}$$

Here $I_n(z)$ is a modified Bessel function, and the parameter

$$\xi = \int_{\Omega} d^3r_1 \int_{\bar{\Omega}} d^3r_2 f^{(0)}(|\mathbf{r}_1 - \mathbf{r}_2|) \tag{18}$$

is defined as the integral over the volume of the region Ω and its exterior part $\bar{\Omega}$. Physically, the parameter ξ is the average number of singlet pairs, one fragment of which lies inside Ω and the other outside Ω. It is obvious that only such pairs will contribute to the total spin of the region Ω. It is also obvious that ξ depends not only on the volume but also on the shape of the region Ω, especially if the characteristic size of Ω or its individual parts is of the order of l.

The situation where a source of triplet pairs is present is more complicated. This case, and the general problems of thermodynamics of a medium of spin-correlated pairs, require a separate analysis.

ACKNOWLEDGMENTS

I am grateful to S.V. Anishchik, N.N. Lukzen, and P.L. Chapovskiĭ for a fruitful discussion of the problem. This work was supported by the Russian Foundation for Basic Research (grant no. 98-03-33124a) and the Federal target program "Integration" (project no. 274).

REFERENCES

1. B. B. Kadomtsev, *Dynamics and Information* (Editorial office of the journal Usp. Fiz. Nauk, Moscow, 1997).
2. C. H. Bennett, G. Brassard, C. Crepeau, *et al.*, Phys. Rev. Lett. **70**, 1895 (1993).
3. D. Bouwmeester, J.-W. Pan, K. Mattle, *et al.*, Nature **390**, 575 (1997).
4. B. Brocklehurst, Int. Rev. Phys. Chem. **4**, 279 (1985).
5. A. L. Buchachenko, R. Z. Sagdeev, and K. M. Salikhov, *Magnetic and Spin Effects in Chemical Reactions* (Nauka, Novosibirsk, 1978).
6. S. V. Anischik, O. M. Usov, O. A. Anisimov, *et al.*, Radiat. Phys. Chem. **51**, 31 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 24–29.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 29–34.

Influence of a Polar Ambient Medium on Multiphoton Resonances in Dipole Molecules

V. A. Kovarskiĭ and O. B. Prepelitsa
Institute of Applied Physics, Academy of Sciences of Moldova,
Chisinau, 277028 Moldova;
e-mail: exciton@cc.acad.md

Received May 26, 1999

Abstract—The influence of fluctuations of a polar ambient medium on multiphoton resonances in dipole molecules interacting with an external electromagnetic field is studied. It is shown that the electric fields created by the fluctuating dipole moments of the medium strongly influence the fulfillment of the multiphoton resonance conditions. Cases of slow and fast fluctuations of the polar ambient medium are considered separately. It is established that the temperature dependence of the multiphoton resonance probability is non-Arrhenius because of the influence of the fluctuations on the profile of the potential barrier. For fast fluctuations it is predicted that the rate of the multiphoton transition will decrease as the intensity of the fluctuations of the polar ambient medium increases. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Multiphoton transitions in isolated molecules have formed the subject of numerous theoretical and experimental investigations (see, for example, the book by Kovarskiĭ *et al.* [1]). Recently, increasing interest has been directed toward a new class of nonlinear optical effects involving the generation of higher harmonics [2–4]. Investigating this problem revealed that the ambient medium may well play a role in high-order multiphoton processes. The simplest task in the theory of nonlinear interaction of electromagnetic radiation with atoms and molecules involves studying the multiphoton absorption process and thus, in the present paper we consider the influence of the ambient medium on this process and specifically analyze the influence of a polar ambient medium on multiphoton resonance in a dipole impurity molecule. The presence of fluctuating dipole moments of an ambient medium may bring about significant changes in the multiphoton absorption and scattering of light by an isolated molecule. This influence of a polar ambient medium on quantum nonradiative transitions involving intramolecular vibrations has already been examined in the literature [5–22]. For a molecule having a dipole-active vibration, the ambient medium strongly influences the absorption process as a result of dipole–dipole interaction. At the same time, the establishment of multiphoton resonances in a molecule in many respects depends on the contribution of electron–vibrational interactions. This particularly applies to the influence of the electric fields created by fluctuations of the dipole moments of the ambient medium on the dynamics of intramolecular vibrations and therefore on the establishment of multiphoton resonances. The dipole moment of a molecule is formed by electronic and vibrational states and determines its interaction with an external electromagnetic field. One model [20–22] frequently used for simplicity only allows for direct interaction of the electromagnetic field with the electron and neglects nonresonant interaction of the electromagnetic field with intramolecular vibrations and with a polar ambient medium. We also shall adopt this model.

If a molecule has a nonzero average electron dipole moment, this leads to strong interaction between this molecule and a high-intensity electromagnetic field. In the simplest case of a two-level molecule in an external electromagnetic field the system may be described by the Hamiltonian

$$H = H_e + H_{vib} + H_{int},$$

where

$$H_e = \frac{1}{2}(\varepsilon_2 - \varepsilon_1)\sigma_z + \frac{1}{2}\mathbf{E}_0(\mathbf{d}_{22} - \mathbf{d}_{11})$$
$$\times \sigma_z \cos(\Omega t) + \mathbf{E}_0 \mathbf{d}_{21} \sigma_x \cos(\Omega t)$$

is the Hamiltonian of a two-level system whose unperturbed states $|i\rangle$ have the energies ε_i and the intrinsic dipole moment $\mathbf{d}_{ii}$. Here $i = 1, 2$ denotes the ground and excited states of the system, $\mathbf{d}_{21}$ is the dipole moment of the transition between the ground and excited states, σ_z and σ_x are the Pauli matrices satisfying the familiar commutation relationships, and $\mathbf{E}_0$ and Ω are the amplitude and frequency of the applied linearly polarized electromagnetic field. In addition, H_{vib} is the Hamiltonian of the vibrational subsystem and H_{int} is the interaction Hamiltonian of the electronic and vibrational subsystems.

In the simple case of a multiphoton transition without any vibrations involved, the problem is in many respects similar to that of a multiphoton transition to a degenerate level of the hydrogen atom [23]. In this case, the transition rate W_{21} is described by the generating function $I_{21}(t, \tau)$ of the multiphoton process:

$$W_{21} = 2\int_0^{\infty} d\tau \exp\left(\frac{i}{\hbar}(\varepsilon_2 - \varepsilon_1)\tau\right) I_{21}(t, \tau), \tag{1}$$

where the generating function

$$\begin{gathered} I_{21}(t, \tau) = \left(\frac{\mathbf{E}_0\mathbf{d}_{21}}{\hbar}\right)^2 \\ \times \exp\{i(\rho_{22} - \rho_{11})[\sin(\Omega t) - \sin(\Omega(t - \tau))]\} \\ \times \cos(\Omega t)\cos(\Omega(t - \tau)), \end{gathered} \tag{2}$$

$$\rho_{22} = \frac{\mathbf{E}_0\mathbf{d}_{22}}{\hbar\Omega}, \quad \rho_{11} = \frac{\mathbf{E}_0\mathbf{d}_{11}}{\hbar\Omega},$$

is written in the lowest order of perturbation theory in terms of the interaction $(\mathbf{E}_0\mathbf{d}_{21})\sigma_x\cos(\Omega t)$, which mixes states $|1\rangle$ and $|2\rangle$, while the interaction of the intrinsic dipole moments $\mathbf{d}_{ii}$, $i = 1, 2$ is taken into account exactly [1].

The subsequent generalization of formula (1) corresponds to allowing for the contribution of the vibrational degrees of freedom to the multiphoton transition probability. We shall confine our analysis to the case of a single vibrational mode q of frequency ω active in the electronic transition. The adiabatic approximation is usually used to describe electron–vibrational interaction. In accordance with this, we take the electron–vibrational interaction in the electronic state $|2\rangle$ in the form

$$H_{int} = Vq.$$

Here V is the matrix element of the coefficient function of the electron–vibrational interaction. The assumed constraint of a linear term with respect to the coordinate q in the Hamiltonian H_{int} and the constraint of the harmonic approximation for the Hamiltonian H_{vib} are consistent with the “displaced parabola” model for the adiabatic terms of the ground and excited electronic states. In this case, calculations of the generating function using standard methods in the theory of multiphoton processes [1] yield the following formula:

$$W_{21}^{(a)} = 2\int_0^{\infty} d\tau I_{21}(t, \tau) S(\tau). \tag{3}$$

Here the generating function $I_{21}(t, \tau)$ is given by formula (2),

$$S(\tau) = \langle \exp(iH_2\tau)\exp(-iH_1\tau)\rangle_{av},$$

where H_i are the Hamiltonians of the vibrational subsystem in states $i = 1, 2$ and $\langle\ldots\rangle_{av}$ denotes averaging over the vibrational states in the initial electronic state with the density matrix

$$\rho_T = A^{-1}\exp(-\beta H_1),$$

$$A = \mathrm{Tr}(\exp(-\beta H_1)), \quad \beta = \frac{1}{k_0 T},$$

k_0 is the Boltzmann constant, and T is the absolute temperature. The value of $S(\tau)$ may be calculated exactly [1]:

$$S(\tau) = \exp\left\{\frac{i}{\hbar}(\varepsilon_1 - \varepsilon_2)\tau + z\cos(\omega\tau + \varphi)\right\},$$

$$z = a\sqrt{\bar{n}(\bar{n} + 1)}, \quad \bar{n} = \frac{1}{\exp(\hbar\omega/k_0 T) - 1}, \tag{4}$$

$$\varphi = \frac{i}{1 + 2\bar{n}}.$$

The value of a is the so-called heat release constant, defined as the square of the displacement of the adiabatic potentials relative to each other:

$$a = \left(\frac{V}{\hbar\omega}\right)^2 \frac{\hbar}{m\omega}. \tag{5}$$

It is easy to show that by using formulas (3)–(5) the rate of a multiphoton process involving vibrational quanta may be reduced to the form

$$\begin{gathered} W_{21}^{(a)} = 2\pi\frac{(\mathbf{E}_0\mathbf{d}_{21})^2}{\hbar}\sum_{n, p = -\infty}^{\infty} R_p(z) n^2 \frac{J_n^2(\rho)}{\rho^2} \\ \times \delta(\varepsilon_2 - \varepsilon_1 - n\hbar\Omega + p\hbar\omega), \\ R_p(z) = \exp\left(-a\left(\bar{n} + \frac{1}{2}\right)\right)\left(1 + \frac{1}{\bar{n}}\right)^{p/2} I_p(z), \\ \rho = \rho_{22} - \rho_{11}. \end{gathered} \tag{6}$$

Here $J_n(\rho)$ and $I_p(z)$ are Bessel functions of real and imaginary arguments, respectively.

We shall subsequently investigate the influence of a polar ambient medium on multiphoton processes in dipole molecules caused by their active interaction with fluctuations of the dipole moments of the ambient medium. We shall generalize formula (6) to the case of a polar ambient medium and identify the conditions under which this medium has a stabilizing or destabilizing influence on multiphoton resonances. For this we shall assume that under certain conditions (fast fluctuations) the optical absorption bands undergo a familiar narrowing effect with increasing fluctuation intensity [24] and the rates of the nonradiative processes decrease [21, 22].

For the real energy levels of an impurity molecule, the resonance condition involving electromagnetic field photons and vibrational quanta is not always satisfied. Any resonance detuning may be compensated by means of additional interactions introduced by the polar ambient medium, changes in the kinetic energy as a result of collisions, and other factors. However, we shall identify the particular role which may be played by fluctuations of a polar ambient medium since, as will be shown subsequently, the intensity of these fluctuations can not only increase the rate of the multiphoton transition but also, which is particularly important, can reduce the probability of multiphoton resonance. The reduction in the transition rate in this last case is occasioned by broadening of the energy levels caused by the interaction between an impurity molecule and the polar ambient medium.

2. INFLUENCE OF FLUCTUATIONS OF A POLAR AMBIENT MEDIUM ON THE MULTIPHOTON TRANSITION RATE

We shall consider an impurity dipole molecule in an external electromagnetic field interacting with a polar ambient medium. The proposed model is based on the fact that the electromagnetic field interacts actively with the impurity molecule and barely perturbs the polar ambient medium. If the ambient medium creating the fluctuations of the dipole moment is a plasma, this implies that the electromagnetic field frequency Ω is far from resonance with the plasma frequencies ω_p. Using the model for an impurity molecule described in the Introduction, i.e., intrinsic dipole moments exist in the ground and excited electronic states, where $|\mathbf{d}_{11}| \neq |\mathbf{d}_{22}|$, we shall use a Gauss–Markov autocorrelation function to describe the fluctuations of the polar ambient medium:

$$\langle\langle F(t)F(t')\rangle\rangle = F_0^2 \exp(-\gamma|t-t'|), \qquad (7)$$

where $F(t)$ is a random force, F_0^2 is the intensity of the fluctuations of the ambient medium which depends on temperature ($F_0^2 \propto T$) and γ^{-1} is the characteristic correlation decay time. It is assumed that the random forces act differently on the vibrational motion of the nuclei in different electronic states. Thus, for simplicity we can only assume $F(t) \neq 0$ for vibrations in the excited electronic state. In addition, we can neglect any direct change in the electronic state introduced by fluctuations of the ambient medium [20–22]. The action of the dipole ambient medium on the vibrational spectrum of the molecule in the excited electronic state will be described by the interaction Hamiltonian $F(t)q$. Hence, the vibrational Hamiltonians in electronic states 1 and 2 have the form

$$H_{1,2} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} + U_{1,2}(q),$$

$$U_1(q) = \frac{m\omega^2}{2}q^2,$$

$$U_2(q) = \frac{m\omega^2}{2}q^2 - Vq - F(t)q.$$

The rate of the transition between electronic states 1 and 2 of the impurity molecule under the action of an applied electromagnetic field in the presence of a dipole ambient medium is written in the form

$$W_{21} = 2\int_0^\infty d\tau \exp\left(\frac{i}{\hbar}(\varepsilon_2' - \varepsilon_1)\tau\right) I_{21}(t,\tau)$$

$$\times \int dq_1 dq_2 \left\langle\left\langle \varphi_0^*(q_2)\sum_p \Psi_p(q_2,t)\Psi_p^*(q_1,t-\tau)\varphi_0(q_1)\right\rangle\right\rangle.$$

Here we have $\varepsilon_2' = \varepsilon_2 + \hbar\omega a/2$, $\varphi_0(q) = \langle q|0\rangle$ is the initial (ground) state of the oscillator (here and subsequently it is assumed that $k_0T \ll \hbar\omega$ so that the probability of direct thermal excitation of the oscillator is negligible) relative to the electronic state $|1\rangle$ and in addition

$$\Psi_p(q,t) = \langle q|U(t,-\infty)|\varphi_p\rangle,$$

where $\varphi_p(q)$ are the oscillator wave functions of the electronic state $|2\rangle$, $U(t,-\infty)$ is the evolution operator determined by the vibrational Hamiltonian H_2, and the double angle brackets denote averaging over realizations of the random process $F(t)$ [see formula (7)].

Using the representation of the Green function

$$K(q_2, t, q_{12}, t-\tau)$$

$$= -\frac{i}{\hbar}\sum_p \Psi_p(q_2,t)\Psi_p^*(q_1,t-\tau), \quad \tau \geq 0,$$

we write the transition probability as follows:

$$W_{21} = 2i\hbar\int_0^\infty d\tau \exp\left(\frac{i}{\hbar}(\varepsilon_2' - \varepsilon_1)\tau\right)$$

$$\times I_{21}(t,\tau)\langle\langle K(q_2,t,q_1,t-\tau)\rangle\rangle,$$

and we write the Green function $K(q_2, t, q_1, t-\tau)$ in the form of the functional integral:

$$K(q_2,t,q_1,t-\tau) = \int Dq \exp\left(\frac{i}{\hbar}S(t,t-\tau)\right), \qquad (8)$$

where

$$S(t,t-\tau) = \int_{t-\tau}^{t} dt_1$$

$$\times\left[\frac{m}{2}\left(\frac{dq(t_1)}{dt_1}\right)^2 - \frac{m\omega^2}{2}q^2(t_1) + (V + F(t_1))q(t_1)\right]$$

is the classical action.

The continuous integral (8) is calculated over the trajectories $q(t)$ satisfying the boundary conditions $q(t-\tau) = q_1$, $q(t) = q_2$.

We shall use the well-known property of a Gauss–Markov process:

$$\left\langle\left\langle \exp\left(\frac{i}{\hbar}\int\limits_{t-\tau}^{t} dt_1 f(t_1)q(t_1)\right)\right\rangle\right\rangle = \exp\left(\frac{1}{2}\left(\frac{i}{\hbar}\right)^2\int\limits_{t-\tau}^{t} dt_1 \times \int\limits_{t-\tau}^{t} dt_2 q(t_1)\langle\langle F(t_2)F(t_1)\rangle\rangle q(t_2)\right),$$

and then the averaged Green function $\langle\langle K(q_2, t, q_1, t-\tau)\rangle\rangle$ can be written as

$$\langle\langle K(q_2, t, q_1, t-\tau)\rangle\rangle = \int Dq \exp\left(\frac{i}{\hbar}S_{eff}(t, t-\tau)\right).$$

The effective action has the form

$$S_{eff} = \int\limits_{t-\tau}^{t} dt_1\left[\frac{m}{2}\left(\frac{dq(t_1)}{dt_1}\right)^2 - \frac{m\omega^2}{2}q^2(t_1) + Vq(t_1) + \frac{iF_0^2}{2\hbar}\int\limits_{t-\tau}^{t} dt_2 \exp(-\gamma|t_2-t_1|)q(t_2)q(t_1)\right].$$

The extremal trajectory which minimizes the action satisfies the equation

$$q^{(2)}(t_1) + \omega^2 q(t_1) = \frac{iF_0^2}{m\hbar} \times \int\limits_{t-\tau}^{t} dt_2 \exp(-\gamma|t_2-t_1|)q(t_2) + \frac{V}{m}. \qquad (9)$$

The effective action on the extremal trajectory has the form

$$S_{eff}^{(cl)}(t, t-\tau) = \frac{m}{2}q(t_1)\frac{dq(t_1)}{dt_1}\Bigg|_{t-\tau}^{t} + \frac{V}{2}\int\limits_{t-\tau}^{t} dt_1 q(t_1). \qquad (10)$$

The continuous integral (8) may be written as

$$K(q_2, t, tq_1, t-\tau) = \sqrt{-\frac{1}{2\pi\hbar i}\frac{\partial^2 S_{eff}^{(cl)}}{\partial q_2 \partial q_1}} \times \exp\left(\frac{i}{\hbar}S_{eff}^{(cl)}(t, t-\tau)\right). \qquad (11)$$

Note that in general, expression (11) is exponentially exact but since the extremal trajectory $q(t)$ is a linear profile along q_2 and q_1, the pre-exponential factor in (11) does not depend on q_2 and q_1 so that in this case expression (11) is also exact.

After differentiating twice, the integrodifferential equation (9) is reduced to an ordinary fourth-order differential equation:

$$q^{(4)} + (\omega^2 - \gamma^2)q^{(2)} + \left(2i\gamma\frac{F_0^2}{m\hbar} - \gamma^2\omega^2\right)q + \gamma^2\frac{V}{m} = 0 \qquad (12)$$

with the boundary conditions $q(t-\tau) = q_1$ and $q(t) = q_2$.

The two additional integration constants are determined from equation (9) at the points $t_1 = t-\tau$, $t_1 = t$. In general, the solution of equation (12) is fairly cumbersome. Thus, at this point it is convenient to select the limiting cases of fast and slow fluctuations of the ambient medium.

In the case of slow fluctuations $b/\aleph \gg 1$, where $b = F_0^2/m\hbar\omega^3$, $\aleph = \gamma/\omega$, equation (12) gives an expression for the multiphoton transition similar to formula (6) where, however, the presence of the ambient medium is manifest in the appearance of an additional force in the heat release constant (5). This constant should be replaced by

$$\tilde{a} = \left(\frac{V+F}{\hbar\omega}\right)^2\frac{\hbar}{m\omega}.$$

In this case, the transition rate W_{12} should be averaged over slow fluctuations of F with the distribution function

$$\frac{1}{\sqrt{\pi}F_0}\exp\left(-\frac{F^2}{2F_0^2}\right).$$

Hence, the unknown transition rate is written as follows:

$$W_{21} = \frac{1}{2\sqrt{\pi}F_0}\int\limits_0^\infty dF \exp\left(-\frac{F^2}{2F_0^2}\right)W_{21}(\tilde{a}), \qquad (13)$$

where $W_{21}(\tilde{a})$ is determined by formula (6) in which the constant a should be replaced by $\tilde{a}$.

We shall confine ourselves to an analysis of processes involving photon emission. We calculate the integral (13) by the method of steepest descents, as a result of which we obtain with exponential accuracy

$$W_{21} \sim \exp\left(\frac{ab}{2}\left[\frac{2n_0}{a} - \ln\left(\frac{2n_0}{a} + 1\right)\right]^2\right)W_{21}^{(a)}, \qquad (14)$$

where n_0 is the photonicity of the process which makes the main contribution to the probability of a transition accompanied by phonon emission. It follows from this expression that the fluctuations of the ambient medium in the case $2n_0/a \ll 1$ and $bn_0/a > 1$ appreciably increase the probability of a multiphoton transition. For these relationships between the parameters expression (14) may be simplified and reduced to the form

$$W_{21} \sim \exp\left(\frac{bn_0^2}{a}\right) W_{21}^{(a)}.$$

Since $b \sim F_0^2 \sim T$, it follows from this last expression that for transitions accompanied by phonon emission, the transition probability has an "anomalous" temperature dependence $W_{21} \sim \exp(T/A)$, where A is a constant. (Here it is assumed that because of the condition $k_0 T \ll \hbar\omega$ the probability of transitions accompanied by phonon emission in the absence of an ambient medium does not depend on T.) The characteristic value of the parameter A for some biomolecules reaches 43 K [22].

We shall now consider the opposite case of fast fluctuations of the ambient medium. If the decay time for the correlations of a polar ambient medium is fairly short compared with the period of the molecular vibrations, the fluctuations of the ambient medium have little influence on the quantum transition processes. In the asymptotic limit $\aleph \longrightarrow \infty$ the influence of the ambient medium on the multiquantum transition process is "switched off", i.e., the transition probability ceases to depend on the fluctuations of the medium. The case $\aleph \geq 1$ for $b/\aleph \ll 1$ is more interesting. In this case, the extremal trajectory, being a solution of equation (12) with corresponding boundary conditions, has the form

$$q(t_1) = \frac{1}{\sin(\Omega_f \tau)}$$
$$\times [q_2 \sin(\Omega_f(t_1 - t + \tau)) - q_1 \sin(\Omega_f(t_1 - t))]$$
$$+ \frac{V}{m\Omega_f^2}\left[1 + \frac{\sin(\Omega_f(t_1 - t)) - \sin\Omega_f(t_1 - t + \tau)}{\sin(\Omega_f \tau)}\right],$$
$$\Omega_f = \omega(1 - iD), \quad D = b/\aleph.$$

Using this solution, we find after various transformations

$$W_{21} = 2\pi\left(\frac{e\mathbf{E}_0\mathbf{d}_{21}}{\rho}\right)^2 \sum_n n^2 J_n^2(\rho)$$
$$\times \frac{\Gamma + D}{(\varepsilon_2' - \varepsilon_1 - n\hbar\Omega + \hbar\omega)^2 + \hbar^2(\Gamma + D)^2}.$$

Here Γ is the width of the electronic state in the absence of a polar ambient medium.

In the particular case $\varepsilon_2' - \varepsilon_1 + \hbar\omega \sim n_0\hbar\Omega$ where the contribution of the resonant term (containing n_0) to the transition probability is decisive, the dependence of the multiphoton transition rate on the fluctuation intensity (in terms of the parameter D) is similar to that obtained in the case of dynamic narrowing of the optical transition rates [24]. In this case, the probability of a multiphoton transition decreases as the fluctuation intensity increases. In fact, for $\varepsilon_2' - \varepsilon_1 - n_0\hbar\Omega + \hbar\omega \ll \hbar(\Gamma + D)$ and $\Gamma \ll D$ expression (15) may be written in the form

$$W_{21} \approx 2\pi\left(\frac{e\mathbf{E}_0\mathbf{d}_{21}}{\hbar\rho}\right)^2 n_0^2 J_{n_0}^2(\rho)\frac{1}{D}.$$

From this it follows that as the fluctuation intensity increases, the multiphoton transition probability decreases. Moreover, since $D \sim b$ and $b \sim T$, the optical transition probability decreases with increasing temperature.

3. CONCLUSIONS

This analysis has shown that the fluctuations of a polar ambient medium lead to changes in the probability of multiphoton resonances. This influence is determined to a considerable extent by the relationship between the correlation decay time and the frequency of the molecular vibration active in the optical transition. The conclusion that the multiphoton transition probability does not have an Arrhenius temperature dependence for media with slow correlation decay is fundamentally new. This effect is entirely attributed to a change in the profile of the potential barrier caused by fluctuations of the ambient medium. In the other limiting case of fast fluctuations the theory predicts a decrease in the multiphoton transition probability with increasing fluctuation intensity. It is readily established that the characteristics attributable to the ambient medium are also found for the multiphoton scattering of coherent radiation by impurity dipole molecules.

REFERENCES

1. V. A. Kovarskiĭ, N. F. Perel'man, and I. Sh. Averbukh, *Multiquantum Processes* (Énergoatomizdat, Moscow, 1985).
2. M. Ferray, A. Huillier, X. Li, *et al.*, J. Phys. B **21**, L31 (1988).
3. J. H. Eberly, Q. Su, and J. Javanainen, Phys. Rev. Lett. **62**, 881, 1989 (1989).
4. T. P. Platonenko, Laser Phys. **6**, 1173 (1996).
5. R. A. Marcus, J. Phys. Chem. **90**, 3460 (1986).
6. A. Warshel and S.T. Russell, Quarterly Rev. Biophys. **17**, 283 (1984).
7. H. Sumi and R. A. Marcus, J. Chem. Phys. **84**, 4894 (1986).
8. A. I. Burshtein and A. G. Kofman, Chem. Phys. **40**, 289 (1979).
9. L. D. Zusman, Tekh. Éksp. Khim. **15**, 227 (1979).
10. L. D. Zusman, Chem. Phys. **49**, 295 (1980).

11. I. V. Aleksandrov, Tekh. Éksp. Khim. **16**, 435 (1980).
12. A. B. Gel’man, Tekh. Éksp. Khim. **19**, 281 (1983).
13. A. B. Helman, Chem. Phys. **65**, 271 (1982).
14. E. M. Kosower and D. Huppert, Chem. Phys. Lett. **96**, 433 (1983).
15. J. N. Onuchic and D. N. Beratan, J. Chem. Phys. **16**, 3707 (1986).
16. A. Gard, J. N. Onuchic, and V. J. Ambegankar, Chem. Phys. **83**, 4491 (1985).
17. I. Rips and J J. Jortner, Chem. Phys. **87**, 4491 (1987).
18. G. W. Lohson and D. W. Oxtby, Chem. Phys. **87**, 781 (1987).
19. A. Warshel and J.-R.Hwang, J. Chem. Phys. **84**, 4938 (1986).
20. I. Sh. Averbukh, V. A. Kovarskiĭ, A. A. Mosyak, *et al.*, Teor. Mat. Fiz. **81**, 271 (1989).
21. I. Sh. Averbukh, V. A. Kovarskiĭ, A. A. Mosyak, *et al.*, *Kinetic Effects in Molecules and Solid-State Systems in External Fields*, Ed. by V. A. Kovarskiĭ (Shtinitsa, Kishinev, 1990).
22. A. V. Belousov and V. A. Kovarskiĭ, Zh. Éksp. Teor. Fiz. **114**, 1944 (1998).
23. V. A. Kovarskiĭ, Zh. Éksp. Teor. Fiz. **57**, 1217 (1969).
24. S. A. Akhmanov, Yu. E. D’yakov, and A. S. Chirkin, *Introduction to Statistical Radiophsics and Optics* (Nauka, Moscow, 1989).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 30–44.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 35–50.
Original Russian Text Copyright © 2000 by Pozdneev.

ATOMS, SPECTRA, RADIATION

Collisions of Electrons with Molecules in Excited Vibrational–Rotational States

S. A. Pozdneev
Lebedev Physics Institute, Russian Academy of Sciences,
Moscow, 117924 Russia;
e-mail: pozdneev@sci.lebedev.ru

Received December 4, 1998

Abstract—Results are presented of calculations of cross sections for scattering of electrons by diatomic molecules in specific excited vibrational–rotational states. The calculations were made using an approximation based on a quantum theory of scattering in a system of several bodies which can be applied to calculations of direct reactions and reactions involving the formation of an intermediate transition complex. Results of calculations of cross sections for collisions of electrons with hydrogen, nitrogen, lithium, sodium, and hydrogen halide molecules are compared with existing experimental data and the results of calculations made by other authors. *© 2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Various electron–molecule collision processes

$$e + \mathrm{AB}(v_1, J_1) \longrightarrow \begin{cases} e + \mathrm{AB}(v_1, J_1) - \\ \text{elastic scattering processes} \\ e + \mathrm{AB}(v_2, J_2) - \\ \text{vibrational–rotational excitation} \\ \mathrm{A}^-(e\mathrm{A}) + \mathrm{B}, \quad \mathrm{A} + \mathrm{B}^-(e\mathrm{B}) - \\ \text{dissociative attachment of an} \\ \text{electron to a molecule} \\ e + \mathrm{A} + \mathrm{B} - \\ \text{molecular dissociation} \end{cases} \tag{1.1}$$

play an important role in negative ion production processes, fast generation of fluorine and chlorine atoms from halogen-containing molecules in excimer lasers, gas-discharge plasmas, and so on [1–4].

In addition, dissociative attachment is the simplest chemical reaction induced by electrons and has thus been studied on a considerable number of occasions [1–10]. However, studies dealing with collision processes between electrons and molecules in predefined excited vibrational–rotational states have only recently begun to appear [9–14]. This is because of the major difficulties involved in the experimental preparation of these excited states and the overall realization of the experiment.

Hence the present paper is devoted to a theoretical study of the dissociative attachment of electrons to hydrogen, nitrogen, sodium, and lithium molecules in specific vibrational–rotational states:

$$e + \mathrm{AB}(v = n, J = N) \longrightarrow \mathrm{A}^- + \mathrm{B}.$$

Various series of experiments[2, 6, 11–14] have shown that the cross sections of these processes depend on the degree of excitation of the vibrational–rotational state of the target molecule. A similar situation is observed for hydrogen [10], nitrogen [11], and other molecules [1, 2].

In addition, the dissociative attachment of an electron to a lithium molecule has the following characteristic features [12]:

(1) a fairly high rate of formation of negative lithium ions by dissociative attachment to molecules in highly excited vibrational states;

(2) a negative lithium ion will most probably play the same role in the future as a negative hydrogen ion because of the similarity between the electronic properties of these molecules;

(3) since the lithium molecule has the same valence as a hydrogen molecule, methods used to interpret the experimental data on collisions between electrons and hydrogen molecules can be used to study collisions between electrons and lithium molecules.

The well-known methods of investigating dissociative attachment [1–8] (the boomerang method, the *R*-matrix method, the time evolution of the wave function, the Feshbach operator method, and so on) are based on the treatment of this process as a multistage process:

(1) the first stage of the process involves the capture of an electron by the molecule with the consequent formation of a negative molecular ion;

(2) the second stage involves the decay (evolution) of this state into various states of the decay products: a

1063-7761/00/9001-0030$20.00 © 2000 MAIK “Nauka/Interperiodica”

negative ion and a neutral or excited atom (dissociative attachment); two neutral or excited atoms and electrons (molecular dissociation); an excited molecule and an electron (electron impact molecular excitation).

The basis of this formalism, i.e., the formation of an intermediate state of a molecular negative ion, is not always justified from the physical point of view. For instance, in the case of dissociative attachment of an electron to a hydrogen molecule the lifetime of this complex is comparable with the time taken for free drift of the electron over a distance equal to the diameter of the hydrogen molecule. A similar situation arises in the case of a molecular reaction [15] without the formation of an intermediate complex,

$$\mathrm{O}(^3P) + \mathrm{CS}(X^1\Sigma^+) \longrightarrow \mathrm{CO}(X^1\Sigma^+) + \mathrm{S}(^3P),$$

when an appreciable fraction of the translational energy (in accordance with the momentum limit $E_V/E_t \sim 0.88$ [14, 15]) is converted into vibrational energy of the CO molecule.

Finally there are many examples where a long-lived intermediate complex forms in the reaction process (for further details see [1–8, 10–14]), although the processes described above also exist, and these show that a preliminary analysis of the experimental data on the process is required for a theoretical analysis of various collisions. The absence of such an analysis fairly frequently leads to erroneous interpretations of the experimental data, as in the case of the dissociative attachment of an electron to a hydrogen molecule [4]. In this case, the dissociative attachment cross section calculated using the intermediate state model is ten times higher than the experimental data [16].

A similar situation is now encountered in quantum chemistry where "the potential energy surface plays a central role in the numerical modeling of chemical reactions" [17]. However, there are many examples, one of which is given above, where an intermediate complex does not appear during the reaction process and consequently, the existence of a potential energy surface presents problems.

Thus, in atomic and chemical physics there is a class of processes, described as direct by analogy with nuclear physics, whose basic characteristic is that no intermediate long-lived complex forms during the scattering process.

Consequently, in order to interpret these direct processes we need to develop suitable theoretical tools and methods, including the well-known methods to describe reactions involving the formation of an intermediate complex.

These methods were developed by L.D. Faddeev and O.A. Yakubovskiĭ (see [18]) who presented a quantum theory of scattering in systems of several particles, free from model assumptions on the formation of an intermediate complex in the collision process. This method can be used to describe both direct processes and those involving the formation of intermediate long-lived states.

In the present study, an approach based on the quantum theory of scattering in a system of several bodies is used to calculate cross sections for collision processes between electrons and diatomic hydrogen, nitrogen, lithium, sodium, and hydrogen halide molecules in predefined vibrational–rotational states.

In this approach the main approximation is that the interaction of an impinging electron with electrons and nuclei of the target molecule is replaced by interaction of an impinging electron with each atom as a whole, considering the atom to be a force center. Hence, the complex multiparticle problem needed to calculate cross sections for scattering of an electron by diatomic molecules is reduced to a collision problem in a three-body system which can be solved using the quantum scattering problem in a three-body system.

This approximation is reasonable at impinging electron energies lower than the electronic excitation energy of the molecule [9]. The initial data in this formulation of the problem are the interaction pair potentials, and the masses and energies of the colliding particles. For the pair potentials of the interaction of electrons with molecular atoms we used the zero-range potentials [8] and potentials having the form

$$V(r) = \lambda \exp(-\beta r)/r, \tag{1.2}$$

whose parameters were determined using the electron binding energy in the negative ion, the scattering lengths, and the effective radius where the spin (in the case of homonuclear molecules) is taken into account as follows. For the scattering length we used the quantity [19]

$$\frac{1}{a} = \frac{1}{a_1} = \frac{1}{a_2} = \frac{1}{4}\left(\frac{3}{a_t} + \frac{1}{a_s}\right),$$

where a_t and a_s are the triplet and singlet scattering lengths, respectively.

The pair potentials of the interaction between atoms in molecules were simulated by the Morse potentials

$$V(r) = D(1 - \exp[-\alpha(r - r_0)]), \tag{1.3}$$

whose parameters were determined using spectroscopic data [20].

2. FUNDAMENTAL APPROXIMATIONS AND METHODS OF CALCULATION

The quantum theory of scattering in a system of several particles (in our case reaction (1.1) and, in the approximation studied, three particles) is formulated for the three parts which make up the complete wave function of the three-body system

$$\Psi = \sum_{i=1}^{3} \Psi_i, \tag{2.1}$$

each part corresponding to all possible ways of dividing the three-particle system into noninteracting subgroups.

These equations in momentum space for the scattering of particle 1 at the bound pair (2,3) have the following form [18]:

$$\Psi_i = \Phi_i\delta_{i1} - G_0(Z)T_i(\Psi_j + \Psi_k),\qquad i, j, k = 1, 2, 3;\quad 3, 1, 2;\quad 2, 1, 3, \tag{2.2}$$

where Φ_1 describes the initial state of the three-body system: the free motion of particle 1 and the bound state of the pair (2, 3): $G_0(Z) = (H_0 - Z)^{-1}$, $Z = E + i0$, H_0 is the free motion operator of the three particles, E is the total energy of the three-body system which is equal to the sum of the kinetic energy of the impinging particle 1 and the binding energy of the pair (2, 3), and T_i is the pair T-matrix which is uniquely determined by the interaction pair potentials V_i using the Limpan–Schwinger equations [18]

$$T_i = V_i + V_iG_iT_i. \tag{2.3}$$

In order to describe the motion of three particles in a center-of-inertia system we use the usual Jacobi coordinates which are defined as follows:

$$\mathbf{k}_1 = \frac{m_3\mathbf{q}_2 - m_2\mathbf{q}_3}{m_2 + m_3},\qquad \mathbf{p}_1 = m_3(\mathbf{q}_1 + \mathbf{q}_2) - \frac{(m_1 + m_2)\mathbf{q}_3}{m_1 + m_2 + m_3}, \tag{2.4}$$

where m_i and $\mathbf{q}_i$ are the mass and momentum of each particle. The coordinates $\mathbf{k}_2$, $\mathbf{p}_2$, $\mathbf{k}_3$, and $\mathbf{p}_3$ are defined similarly.

In these variables

$$\Phi_1(\mathbf{k}_1, \mathbf{p}_1) = \varphi(\mathbf{k}_1)\delta(\mathbf{p}_1 - \mathbf{p}_1^0),$$

where φ is the wave function of the initial state of the system (2, 3) with the binding energy κ_1^2, $\mathbf{p}_1^0$ is the momentum of the impinging particle 1,

$$Z = \frac{p_1^{02}}{2n_1} + \frac{\kappa_1^2}{2m_{23}} + i0,$$

$$n_1 = \frac{m_1(m_2 + m_3)}{m_1 + m_2 + m_3},\qquad m_{23} = \frac{m_2m_3}{m_2 + m_3}.$$

The operators G_0 and T_i are integrals with kernels having the form:

$$T_i(\mathbf{k}_i, \mathbf{k}_i', \mathbf{p}_i, \mathbf{p}_i'; Z) = t_i\left(\mathbf{k}_i, \mathbf{k}_i'; Z - \frac{p_i^2}{2n_i}\right)\delta(\mathbf{p}_i - \mathbf{p}_i'),$$

$$G_0(\mathbf{k}_i, \mathbf{k}_i', \mathbf{p}_i, \mathbf{p}_i'; Z) = \frac{\delta(\mathbf{k}_i - \mathbf{k}_i')\delta(\mathbf{p}_i - \mathbf{p}_i')}{k_i^2/2m_{jk} + p_i^2/2n_i - Z}.$$

It should be borne in mind that any system of variables which is the most convenient may be taken as the variables of integration in (2.2). For example, in the integral corresponding to the expression $G_0T_1\Psi_2$ it is most convenient to take $\mathbf{k}_2$ and $\mathbf{p}_2$ as the variables of integration. In this case, the variables $\mathbf{k}_1$ and $\mathbf{p}_1$ on which the kernel of the operator T_1 depends, should be expressed in terms of the variables $\mathbf{k}_2$ and $\mathbf{p}_2$. In some cases, in this situation it is convenient to take $\mathbf{p}_1$ and $\mathbf{p}_2$ as the variables of integration.

The pair T-matrices $t_i(\mathbf{k}_i, \mathbf{k}_i'; Z)$ appearing in the kernels of the equations (2.2) have singularities in terms of the variable Z: poles corresponding to a discrete spectrum of pair subsystems and a branch cut across the positive part of the real axis generated by the spectrum of the two-body problem. The explicit form of these singularities gives the spectral representation of the T-matrix [18]:

$$t(\mathbf{k}, \mathbf{k}'; Z) = V(\mathbf{k} - \mathbf{k}') + \sum_i \frac{\psi_i(\mathbf{k})\psi_i(\mathbf{k}')}{\kappa_i^2 + Z} + \int t(\mathbf{k}, \mathbf{x}; x^2 \pm i0)t(\mathbf{k}', \mathbf{x}; x^2 \pm i0)\frac{dx}{Z - x^2/2m},$$

where $\psi_i(\mathbf{k}) = (\kappa_i^2 + k^2)\varphi_i(\mathbf{k})$, and κ_i^2 and $\varphi_i(\mathbf{k})$ are the energy and wave function of the bound state of the two particles in the momentum representation.

The poles of the T-matrix corresponding to the discrete spectrum generate singularities in the components of the wave function Ψ_i, and by separating these out we obtain the following representation:

$$\Psi_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0) = \varphi(\mathbf{k}_i)\delta(\mathbf{p}_I - \mathbf{p}_i^0) - \frac{B_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z)}{p_i^2/2n_i + k_i^2/2m_{jk} - Z}, \tag{2.5}$$

$$B_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z) = -\sum_{j=1}^{3}\left[Q_j(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z) - \frac{\varphi_j(\mathbf{p}_j)R_{ji}(\mathbf{k}_j; \mathbf{p}_i^0; Z)}{p_j^2/2n_j - \kappa_j - Z}\right],$$

Q_j, R_{ji} are smooth functions of their variables. This separation of the singularities occurs naturally when the integral equations (2.2) are solved numerically. To obtain a unique definition of the functions Q_j, R_{ji} we can proceed as follows: we substitute Ψ_i in the form (2.5) into the equations (2.2) and equate coefficients at the same singularities. Thus, we obtain equations for these functions in terms of which all the main characteristics of the three-body problem can be expressed explicitly, i.e., the wave function, the elements of the

S-matrix, and the amplitudes and cross sections of the various processes taking place in the three-body system:

$$\begin{array}{l} \qquad\qquad 1 + (2, 3) \\ \longrightarrow \left\{ \begin{array}{l} 1 + (2, 3)\text{—elastic scattering processes} \\ 1 + (2, 3)^*\text{—excitation processes,} \\ 3 + (1, 2)^*,\ \ 2 + (1, 3)^*\text{—rearrangement} \\ \text{processes with excitation,} \\ 1 + 2 + 3\text{—ionization processes.} \end{array} \right. \end{array} \tag{2.6}$$

The cross section of the elastic scattering process has the form

$$\frac{d\sigma_{11}}{d\theta} = (2\pi)^4 n_1 |R_{11}|^2, \tag{2.7}$$

the cross section of the rearrangement process is

$$\frac{d\sigma_{1i}}{d\theta} = (2\pi)^4 \frac{n_i p_f |R_{1i}|^2}{p_1^0}, \tag{2.8}$$

and the cross section of the breakup process

$$\frac{d\sigma_{1\to 3}}{d\theta dp} = (2\pi)^4 \frac{n_i p_f |B_{0i}|^2}{p_1^0}, \tag{2.9}$$

where $p_f = 2n_i(p_i^{02}/2n_i - \kappa_1^2 - \kappa_i^2)$.

It should be noted that the potentials do not participate in explicit form in equations (2.2), but contain the more general characteristic of the *T*-matrix and they are related to the potentials by equations (2.3). Hence, although this method formally uses the potentials, essentially the *T*-matrices are modeled, these being constructed using the Beteman method [9, 15, 16, 18–24].

According to this method [24], the partial harmonic of the local potential in momentum space,

$$V_l(k, k') = \frac{\pi}{2}\int r^2 j_l(kr) V(r) j_l(k'r) dr$$

is approximated by the expression

$$\tilde{V}_l(k, k') = \sum_{i,j}^{n} \frac{V_l(k, s_i) V_l(k', s_j)}{V_l(s_i, s_j)}, \tag{2.10}$$

and the parameters s_i, s_j are determined from the expression for the minimum of the functional

$$\chi(s_1, s_2, s_3, \ldots, s_n) = \frac{\iint |V_l(k, k') - V_l(k, k', s_1, s_2, s_3, \ldots, s_n)| dk dk'}{\iint V_l(k, k') dk dk'}.$$

The solution of the equations (2.3) with the potential (2.10) has the form

$$\tilde{t}(k, k', Z) = \sum_{i,j}^{n} C_{ij}(Z) V(k, s_i) V(k', s_j), \tag{2.11}$$

where

$$C_{ij}(Z) = V(s_i, s_j) + 8\pi m_{ij} \int_0^\infty\int_0^\infty \frac{V(k, s_i) V(k', s_j) k^2 dk dk'}{k^2/2m_{ij} - Z}.$$

It should be noted that this method of constructing the separable *T*-matrix (2.11) can be used for any local potential. In addition, this method appreciably simplifies the numerical solution of the system of integral equations (2.2) and in some cases an analytic solution can be obtained.

The integral equations (2.2) only possess good properties from the mathematical point of view (such as Fredholm properties, unique solubility, and so on) under certain conditions for two-particle data [18]:

(1) the pair potentials, which are generally nonlocal, $V_i(k, k')$ are smooth functions of k, k' and satisfy the condition

$$|V_i(k, k')| \le (1 - |k - k'|)^{1-\epsilon}, \quad \epsilon > 0;$$

(2) the point $Z = 0$ is not a singular point for the equations (2.3), i.e., all three scattering lengths in the pair channels are finite;

(3) the positive two-particle spectrum is continuous. This condition is important for nonlocal potentials since positive eigenvalues can only appear in this case, and it is satisfied for almost all real physical potentials.

The Coulomb potentials and the hard shell potentials do not satisfy the first condition, and the Coulomb potentials lead to singularities in *T*-matrices of the type $|k - k'|^{-2}$ while the hard shell potentials lead to a slow decay of the *T*-matrix at high momenta. When the second condition is violated, the equations (2.1) cease to be Fredholm equations for $Z = 0$, which leads to the Efimov effect [19] whereby at some critical value of the coupling constant, when the scattering length first goes to infinity, an infinite discrete spectrum may appear in the three-particle system under certain conditions.

The Faddeev integral equations (2.2) are exact equations to describe the dynamics of three pairwise interacting structureless particles, for which it is natural to assume pair interaction between the particles in the three-particle system since all the characteristics of the processes in this system will primarily be determined by the pair interaction. This is completely confirmed by the experimental results in direct reactions in nuclear and atomic physics where the pair potentials determine the entire dynamics of the reaction [8, 9, 21].

At this point, particular mention should be made of the fact that the classical trajectory method cannot generally be applied to calculate the collision between an atom and a diatomic molecule based on the surface potential energy, i.e., the potential energy surfaces can only appear when the chemical reaction involves an intermediate complex. The method of the quantum scattering problem can be applied to any chemical reactions for which the above conditions are satisfied.

In molecular reactions, the particles taking part in the reaction are intricate complexes (atoms with an electron shell) and thus their internal structure may play a significant role. As a result, the interaction potential in this system, which is generally nonlocal, will include many-particle potentials as well as pair potentials and in [9–17, 22] it was shown how these many-particles potentials appear and in what form, in a three-particle system consisting of three like atoms interacting using the potentials

$$V_i(r_i) = {}^1V_i^1(r_i)\,{}^1\chi_i + {}^3V_i^3(r_i)\,{}^3\chi_i,$$

where

$${}^1\chi_i = 1/4 + s_j s_k, \quad {}^3\chi_i = 3/4 - s_j s_k,$$

s_i is the spin of the ith atom, and the entire interaction has the form

$$V = \sum_i V_i. \tag{2.12}$$

Mention should be made of [22] in which similar reasoning was used to construct the potential energy surface of diatomic LiF and LiBr molecules.

By directly solving the equations (2.2) after separating the angular variables and expanding in terms of partial waves in momentum space, we also obtain a many-particle nonlocal potential [9, 11, 15, 16] for which the first terms of its Fourier transformation in the Born–Oppenheimer approximation are the same as (2.12). However, unlike (2.12), this potential has the asymptotic form $\sim C/\rho^2$ for $\rho \longrightarrow \infty$ where ρ is the distance between the impinging particle and the center of mass of the pair. This property of the interaction potential leads to threshold characteristics of the amplitudes and scattering cross sections which take the form of oscillations at the threshold for breakup into three particles [9, 16, 23]. Note that these characteristics do not depend on the specific type of pair forces and are determined by the presence of resonances in these and can thus exist in any system of three particles.

The equations of motion similar to (2.2) for three or more particles interacting using arbitrary pair potentials are generally nonintegrable and thus neither the bound states nor the scattering states in these systems can be expressed in the form of quadratures. This is because the number of existing integrals of motion is smaller than the number of dynamic variables required to describe the system [9, 18]. Hence, methods of constructing analytic solutions of the system of integral equations (2.2) are of particular importance. This can be achieved when systems of several bodies contain additional integrals of motion associated with specific features of the pair interaction or constraints are imposed on the system [9, 10].

The importance of constructing analytic solutions of the integral equations (2.2) describing various dynamic processes in systems of three or more particles arises because in the case of three particles, for example, new qualitative characteristics appear, completely uncharacteristic of the two-particle model: the Efimov effect [9, 18, 20, 24] described above, the use of T-matrices instead of the potentials in the equations, which include several new interaction characteristics not observed in the two-body problem [9, 18, 24].

The large ratio of the proton mass to the electron mass is a favorable circumstance which can be used to obtain analytic solutions of the system (2.2) and consequently cross sections of the processes (2.6)–(2.10) which can be expressed in terms of elementary and special functions [9–11, 15, 16]. Hence, we shall consider various particular cases when the systems of integral equations (2.2) can be solved analytically [9, 15, 16].

We consider the case of a single light particle $m_1 \equiv m$ and two like heavy particles $m_2 = m_3 \equiv M$. A similar situation is encountered in problems of proton scattering at hydrogen atoms or collision between an electron and a diatomic molecule consisting of like atoms where, in this last case, the condition $m/M \longrightarrow 0$ is the basis of the adiabatic approximation.

We set [21]

$$\Psi_2 = G_0F_2, \quad \Psi_3 = G_0F_3. \tag{2.13}$$

Then using (2.2) we obtain the following system of equations for F_2 and F_3

$$\begin{gathered}(1 - T_2G_0T_1G_0)F_2 \\ + (T_2G_0 - T_2G_0T_1G_0)F_3 = T_2\Phi, \\ (1 - T_3G_0T_1G_0)F_3 \\ + (T_3G_0 - T_3G_0T_1G_0)F_2 = T_3\Phi,\end{gathered} \tag{2.14}$$

and the following expression for the complete wave function:

$$\begin{gathered}\Psi = \Phi - G_0T_1G_0(F_2 + F_3) - G_0T_3(\Phi + G_0F_2) \\ - G_0T_2(\Phi + G_0F_3) - G_0(T_3 + T_2)G_0T_1(F_2 + F_3).\end{gathered} \tag{2.15}$$

The meaning of the various terms of the expansion becomes clear if we assume that the residues of expressions of the type $G_0T_iF_i$ at the poles T_i are (apart from a constant) the amplitudes of one of the possible particle excitation or redistribution processes as a result of collisions, and the residue at the pole G_0 gives the breakup amplitude.

Thus, the second term in (2.15) gives the amplitude of elastic scattering or excitation, and the third and fourth terms give the amplitudes of the rearrangement reactions in the three-body system. The last term makes contributions to all the amplitudes and thereby allows for the mutual influence of these processes.

It can be shown [9, 21] that the action of the operator $G_0T_iG_0$ in (2.15) on an arbitrary fairly smooth function F vanishes in the limit $m/M \longrightarrow 0$ provided that κ_n is fixed. Thus, the equations (2.14) in the limit $m/M \longrightarrow 0$ have the following form:

$$\begin{aligned} F_2 + T_2G_0F_3 &= T_2\Phi, \\ F_3 + T_3G_0F_2 &= T_3\Phi. \end{aligned} \qquad (2.16)$$

This approximation consists in determining F_2 and F_3 from the system (2.16) where $m/M \longrightarrow 0$ for finite, small values of m/M. Here it should be noted that if κ_n is fixed and $M \longrightarrow \infty$, all the energy levels of the coupled system (2, 3) tend to zero. However for finite M the energy of the coupled system (2, 3) is nonzero. Consequently, (2.16) will only be a reasonable approximation when the existence of coupling in the system (2, 3) plays no significant role in the elastic scattering of particle 1. However, this can only be achieved when the energy of the impinging particle is appreciably higher than the binding energy of the pair. The operator solution of the system (2.16) can be expressed in the form

$$\begin{aligned} F_2 &= \frac{T_2(1-G_0T_3)\Phi}{1-T_2G_0T_3G_0}, \\ F_3 &= \frac{T_3(1-G_0T_2)\Phi}{1-T_3G_0T_2G_0} \end{aligned} \qquad (2.17)$$

and agrees with the expressions obtained for multiple scattering by two intersecting systems of centers [9, 15, 16, 21], i.e., in the adiabatic approximation; all the characteristics of the amplitudes and the cross sections of the processes in the three-body system are determined in terms of multiple scattering.

Taking into account (2.15) and (2.17) for the complete wave function of the three-body system we have the following representation:

$$\begin{aligned} \Psi &\sim (1 - G_0T_3 - G_0T_2)\Phi \\ &- G_0T_3G_0T_2\Phi - G_0T_2G_0T_3\Phi - \ldots \end{aligned}$$

In this approximate expression, the second and third terms correspond to the single scattering approximation, while the fourth and fifth terms correspond to the double scattering approximation. Thus, in this approximation the terms $G_0T_3G_0F_2$ and $G_0T_2G_0F_3$ in (2.15) allow for multiple scattering at the potentials V_2 and V_3, i.e., all the characteristics of the amplitudes and the cross sections are determined by multiple scattering.

Hence, in this approximation the scattering in a three-body system (electron and atoms of a molecule) may be considered to be the result of successive two-particle collisions. Neglecting the operator $G_0T_1G_0$ is equivalent to replacing Ψ_1 by Φ in the equations (2.1). From this it follows that the system of equations (2.17) describes the scattering of particle 1 as if the momentum distribution of particles 2 and 3 conserved its initial value ψ_0 during the scattering process. This approximate description of the scattering process forms the basis of the well-known momentum approximation, which is based on the following assumptions.

(1) The effective radius of interaction of an impinging particle with the particles of a complex system is much smaller than the average distance between the particles in the system, i.e., the interaction of the impinging particle and a system particle is considerably stronger than the interaction between the particles in a bound system and consequently, the interaction potential in this system can be considered to be a pair potential.

(2) The pair interaction time is much shorter than the characteristic time for the system of scatterers.

Thus, it has been shown [9, 21] that in the three-body problem for the scattering of a light particle by a system of two heavy particles, the adiabatic and momentum approximations are related as the Fourier transform and the inverse transform.

A similar result is obtained in the quasiclassical description of a particle collision with a bound pair in which the scattering amplitude is also expressed in terms of the pair amplitudes [3, 9, 21]. It can be seen from these examples that the equations of quantum scattering theory in several-particle systems may serve as a universal mathematical basis for deriving these approximations, and the iterations of these equations form a representation for the scattering amplitudes in the form of a series (2.17) which contains the product of the pair amplitudes, and the number of cofactors in the terms determines the order of the multiple scattering. Multiple scattering approximations can be obtained by using the corresponding approximations for the T-matrices (for example, the separable approximation) in the iterations of the equations (2.2). In addition, pair single and double multiple scatterings of colliding particles are responsible for the appearance of nonintegrable δ-functions and pole singularities in the kernels of the scattering matrix for breakup processes which correspond to the free terms and the first iterations of the equations (2.2).

Similar expressions can also be obtained for scattering in systems of charged particles [18, 25]. In this case, the elements of the scattering matrix are determined by the resolvent

$$R_{00} = F_c + \sum_{\alpha} F_\alpha + \sum_{\alpha \neq \beta} F_{\alpha,\beta} + \tilde{F}_0,$$

$$F_c = A_c / |p - p'|^{5 + 2i\eta},$$

whose singularities are determined by the residual Coulomb interaction,

$$V_\alpha^0 = \frac{n_\alpha(1-\chi_\alpha)}{|x_\alpha|},$$

single scattering,

$$F_\alpha = A_\alpha \frac{f_\alpha(k_\alpha, k'_\alpha)}{|p-p'|^{3-ia_\alpha}},$$

double multiple scattering,

$$F_{\alpha\beta} = A_{\alpha\beta}\frac{f(k_\alpha, k'_{\alpha\beta})f_\beta(k'_\beta, k'_{\alpha\beta})}{(k'_{\alpha\beta}-k'_\beta-i0)^{1-ia_{\alpha\beta}}},$$

where f_α are the two-particle scattering amplitudes for the operators h_α and the form of the functions A_α, $A_{\alpha\beta}$ and the parameters a_α, $a_{\alpha\beta}$ is given in [5].

At this point it should be noted that the singularities of the S-matrix for the breakup processes most typical of a three-body system are concentrated in the same directions of configuration space as for neutral particles. However, unlike the case of breakup processes in systems of neutral particles, characterized by singularities of the type $\delta(p-p')\delta(p_\alpha - p'_\alpha)$, in the case of charged particles the generalized functions $|p_\alpha - p'_\alpha|^{-5-2i\eta}$ and $|p_\alpha - p'_\alpha|^{-3-ia_\alpha}$ appear and the pole $(k'^2_{\alpha\beta} - k'^2_\beta - i0)^{-1}$ is replaced by the distorted pole $(k'^2_{\alpha\beta} - k'^2_\beta - i0)^{-1-ia_{\alpha\beta}}$. The physical reason for these singularities is that the region where pair multiple scattering of free particles can take place is unbounded [18, 25]. In elastic scattering, rearrangement, and breakup processes, all the singularities are determined by two-particle collisions in a system of two Coulomb centers.

For the case of elastic scattering we can explicitly obtain the leading terms of the amplitudes by assuming $R_{\alpha\beta} \simeq R^\alpha$ or R^β and replacing the Green functions of the intermediate states by their eikonal approximation [25]. In this case, we have

$$f_{AA} = f_c + \tilde{f}_{AA},$$

$$\tilde{f}_{AA} \sim U_A^1\hat{V}_1(\hat{R}_2\hat{V}_2 + \hat{R}_3\hat{V}_3 + \hat{R}_2\hat{V}_2\hat{R}_3\hat{V}_3 + \hat{R}_3\hat{V}_3\hat{R}_2\hat{V}_2)U_A^1.$$

For breakup processes we have the following expression:

$$\tilde{f}_{0A} \sim U_0^1\hat{V}_1(\hat{R}_2\hat{V}_2 + \hat{R}_3\hat{V}_3 + \hat{R}_2\hat{V}_2\hat{R}_3\hat{V}_3 + \hat{R}_3\hat{V}_3\hat{R}_2\hat{V}_2)U_A^1,$$

where $U_A^\alpha(X, p_\alpha) = \psi_A(x_\alpha)\psi_c^\alpha(y_\alpha, p_\alpha)$ is the wave function of the discrete spectrum, ψ_α are the eigenfunctions of the Hamiltonian h_α; $A = (\alpha, i)$, $i = 1, 2, 3, \ldots, N$; ψ_c^α are the wave functions describing scattering in the Coulomb field $V_c^\alpha = n_{\alpha\alpha}/|y_\alpha|$; and $U_0^\alpha = \psi_\alpha(x_\alpha, k_\alpha)\psi_c^\alpha(y_\alpha, p_\alpha)$ is the wave function of the continuous spectrum.

It has therefore been shown that in general the adiabatic approximation in momentum space is equivalent to the momentum approximation and may be expressed in terms of multiple scattering. This approximation in configuration space corresponds to dividing the system into fast and slow subsystems in particular regions of the configuration space, where this division is not universal and has a dynamic nature. Thus, in various studies [18, 23, 25] attempts were made to use methods of differential geometry and algebraic topology to make such a division and in this approach the nontrivial covariant derivative and the differential topological invariants can be expressed in terms of the spectral geometry, i.e., in terms of the spectral properties of the fast Hamiltonian.

In momentum space the system of equations (2.16) has the form [21]

$$\begin{aligned} F_3(\mathbf{k}, \mathbf{p}) &= 2m_1\int\frac{t_3(\mathbf{p}, \mathbf{k}-\mathbf{k}'; Z)F_2(\mathbf{k}-\mathbf{k}', \mathbf{k}')}{(\mathbf{k}-\mathbf{k}')^2 - p_0^2 - i0}d\mathbf{k}' \\ &= t_3(\mathbf{p}, \mathbf{p}_0; Z)\varphi_0\left(\mathbf{k} - \frac{m_3}{m_2+m_3}\mathbf{p}_0\right), \\ F_2(\mathbf{k}, \mathbf{p}) &= 2m_1\int\frac{t_2(\mathbf{p}, \mathbf{k}-\mathbf{k}'; Z)F_3(\mathbf{k}-\mathbf{k}', \mathbf{k}')}{(\mathbf{k}-\mathbf{k}')^2 - p_0^2 - i0}d\mathbf{k}' \\ &= t_2(\mathbf{p}, \mathbf{p}_0; Z)\varphi_0\left(\mathbf{k} + \frac{m_2}{m_2+m_3}\mathbf{p}_0\right) \end{aligned} \tag{2.18}$$

and can be solved analytically when the T-matrices can be represented in the separable form (2.10), (2.11).

In the approximation of zero-range potentials, which are a particular case of separable potentials, the T-matrices do not depend on $\mathbf{k}$ and have the following form [8, 9, 18]:

$$t_i(Z) = [(2\pi)^2 m_{ij}(\alpha_i + i\sqrt{2m_{ij}Z})]^{-1},$$

where the values of α_i are the reciprocal scattering lengths $a_i^{-1} = \alpha_i$. Thus the functions F_2 and F_3 in the approximation of zero-range potentials satisfying the system of equations (2.18) will depend on one rather than two vector variables. The limiting expressions of t_i for $m/M \longrightarrow 0$ which should be substituted into (2.17) have the form

$$t_3 = [(2\pi)^2 m_1(\alpha_3 + ip_0)]^{-1},$$

$$t_2 = [(2\pi)^2 m_1(\alpha_2 + ip_0)]^{-1}.$$

For F_2 and F_3 we then obtain the following system of equations:

$$F_3(\mathbf{k}) = 2t_3 m_1 \int \frac{F_2(\mathbf{k}')}{(\mathbf{k}-\mathbf{k}')^2 - p_0^2 - i0} d\mathbf{k}' = t_3 \varphi_0\left(\mathbf{k} - \frac{m_3}{m_2+m_3}\mathbf{p}_0\right),$$

$$F_2(\mathbf{k}) = 2t_2 m_1 \int \frac{F_3(\mathbf{k}')}{(\mathbf{k}-\mathbf{k}')^2 - p_0^2 - i0} d\mathbf{k}' = t_2 \varphi_0\left(\mathbf{k} + \frac{m_2}{m_2+m_3}\mathbf{p}_0\right).$$

It should be noted that outside the limits of the adiabatic approximation it is impossible to use zero-range potentials [8, 9, 18, 21] because in a system of three particles 1, 2, 3 of which two, say 2 and 3, interact with 1 using zero-range potentials, there is an attraction between particles 2 and 3 proportional to $\sim C/r^2$, where the coefficient C depends on the mass ratio. This interaction leads to unique features in the behavior of a real physical system of three particles and, specifically, to coalescing to a central point. The nonuniqueness of the solution of the integral equation describing scattering in this system is related to this factor since it is found that the corresponding homogeneous equation has a nontrivial solution [9, 21].

However, the homogeneous equations obtained by discarding the right-hand sides in the system (2.18) have no nonzero solutions because t_2 and t_3 are complex numbers.

Hence, in the adiabatic approximation the features described above do not appear. This formally demonstrates that zero-range potentials can be used in this approximation.

Calculations of this effect for the dissociative attachment reaction of a hydrogen molecules are plotted in Fig. 1 where it can be seen that the dependence of the cross section on the impinging electron energy is nonmonotonic.

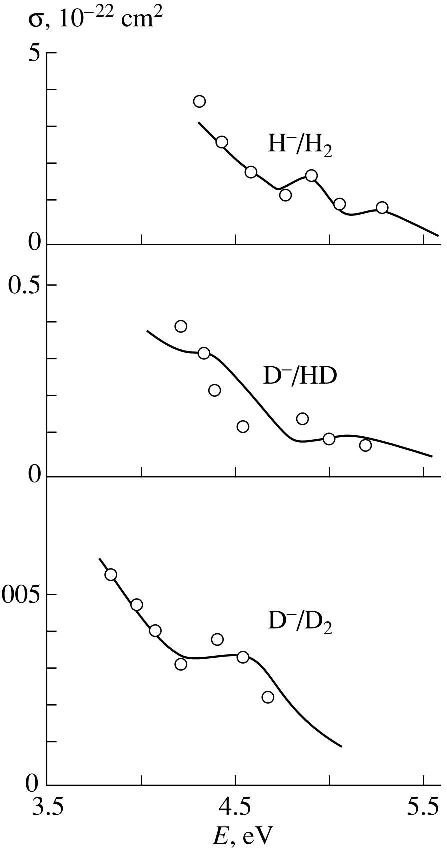


Fig. 1. Dissociative attachment of electrons to hydrogen molecules and their isotopically substituted analogs: circles—experimental data [28], curves—results of the present calculations.

The system of integral equations (2.18) in the approximation of zero-range potentials can be solved exactly by using a Fourier transformation and its solution has the following form:

$$F_3(\mathbf{r}) = \frac{(\alpha_2 + ip_0)\left[\exp(-\gamma\mathbf{p}_0\cdot\mathbf{r}) - \frac{\exp(i\mathbf{p}_0\cdot\mathbf{r} + i\kappa\mathbf{p}_0\cdot\mathbf{r})}{r}\right]\varphi(\mathbf{r})}{(2\pi)^2 m_1[(\alpha_2 + ip_0)(\alpha_3 + ip_0) - \exp(2i\mathbf{p}_0\cdot\mathbf{r})/r^2]},$$

$$F_2(\mathbf{r}) = \frac{(\alpha_3 + ip_0)\left[\exp(-\kappa\mathbf{p}_0\cdot\mathbf{r}) - \frac{\exp(i\mathbf{p}_0\cdot\mathbf{r} + i\gamma\mathbf{p}_0\cdot\mathbf{r})}{r}\right]\varphi(\mathbf{r})}{(2\pi)^2 m_1[(\alpha_2 + ip_0)(\alpha_3 + ip_0) - \exp(2i\mathbf{p}_0\cdot\mathbf{r})/r^2]},$$

$$\gamma = \frac{m_3}{m_2+m_3}, \quad \kappa = \frac{m_2}{m_2+m_3}.$$

Similarly, we can construct analytic solutions of the equations by using the approximation (2.10) with the form factors

$$V_i(k, \beta_i) = \frac{1}{k^2 + \beta_i^2},$$

which in coordinate space correspond to the local potential

$$V_i(r) = \frac{\lambda \exp(-\beta_i r)}{r}.$$

In this case the solution of the system has the following form [3, 9, 10]:

$$F_i(\mathbf{r}, \boldsymbol{\rho}) = \Lambda_i(\boldsymbol{\rho}, \beta_i) C_i(\mathbf{r}),$$

$$\Lambda_i = \frac{\lambda_i}{m_{jk}} \left[1 + 4\pi\lambda_i \int \frac{V_i(x) x^2 dx}{2m_{jk} Z_i - x^2} \right]^{-1},$$

$$C_2(\mathbf{r}) = \frac{V_2(\mathbf{p}_0) \exp(i\kappa \mathbf{p}_0 \cdot \mathbf{r}) - 2m_1 \Lambda_2 V_3(\mathbf{p}_0) \Phi(\mathbf{r}) \exp(-i\gamma \mathbf{p}_0 \cdot \mathbf{r}) \varphi(\mathbf{r})}{1 - \Lambda_2 \Lambda_3 [2m_1 \Phi(\mathbf{r})]^2},$$

$$C_3(\mathbf{r}) = \frac{V_3(\mathbf{p}_0) \exp(i\gamma \mathbf{p}_0 \cdot \mathbf{r}) - 2m_1 \Lambda_3 V_2(\mathbf{p}_0) \Phi(\mathbf{r}) \exp(-i\kappa \mathbf{p}_0 \cdot \mathbf{r}) \varphi(\mathbf{r})}{1 - \Lambda_2 \Lambda_3 [2m_1 \Phi(\mathbf{r})]^2},$$

$$\Phi(r) = \frac{2\pi}{r} \left\{ \frac{\exp(i p_0 r)}{(p_0^2 + \beta^2)^2} - \left[1 - (p_0^2 + \beta^2) \frac{r}{2\beta} \right] \frac{\exp(-i\gamma p_0 r)}{(p_0^2 + \beta^2)^2} \right\},$$

where ρ is the distance between the impinging particle and the center of gravity of the bound particle pair.

The amplitudes of the various processes (2.6) taking place in a three-particle system are determined using formulas (2.7)–(2.9) or from expression (2.18) for the complete wave function. In this last case, the amplitude is determined by the following scheme [9, 21] Near the pole corresponding to the bound state of the system we have

$$t_i(k, k'; Z - p^2) = \frac{\left(\frac{k^2}{2m_{jk}} - \varepsilon_n \right) \left(\frac{k'^2}{2m_{jk}} - \varepsilon_n \right)}{Z - p^2 - \varepsilon_n}$$

$$+ \varphi_n(k) \varphi_n(k') + \text{smooth part of function},$$

where φ_n is the wave function of the bound state of the particle pair (j, k) and ε_n is the binding energy of this pair.

Consequently, the expression $G_0 T_i G_0 F_i$ is factorized as the product of three factors:

$$G_0 T_i G_0 F_i = \varphi_n(k) \left[\frac{p^2}{2n} + \varepsilon_n - Z \right]^{-1} M.$$

The middle factor describes the free motion of particle i relative to the bound system (j, k) where in coordinate space this corresponds to a diverging wave. From this it is clear that the factor M, expressed in terms of the integral of the product of the functions φ_n and F_i, is the scattering amplitude.

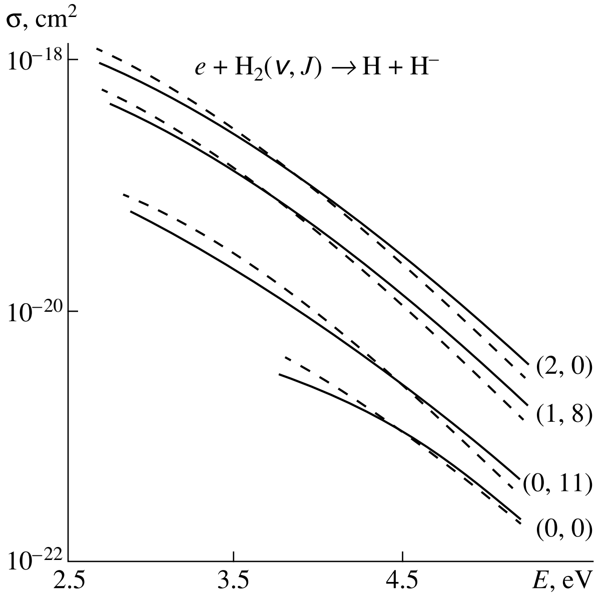


Fig. 2. Dissociative attachment of electrons to hydrogen molecules initially in excited vibrational–rotational states: dashed curves—results of calculations from [10], solid curves—results of the present calculations.

Thus, the amplitudes and cross sections of the elastic scattering and excitation processes are given by the expressions

$$M_{if} = (2\pi)^2 m_1 \int \varphi_i^*(r)[\exp(i\gamma \mathbf{p} \cdot \mathbf{r}) F_3(\mathbf{r}, \rho) + \exp(-i\kappa \mathbf{p}_0 \cdot \mathbf{r}) F_2(\mathbf{r}, \rho)] \varphi_f(r) d\mathbf{r}\boldsymbol{\rho},$$

$$\frac{d\sigma_{if}}{d\Omega} = |M_{if}|^2,$$

where φ_i and φ_f are the vibrational–rotational wave functions which have the following form in the adiabatic representation [8, 9, 21]

$$\varphi_i(r) = \psi_v(r) Y_{JM}(\theta, \xi), \quad \varphi_f(r) = \psi_v(r) Y_{J'M'}(\theta, \xi),$$

$Y_{JM}(\theta, \Omega)$ are spherical functions [26], ψ_v are the vibrational wave functions of the Morse potential,

$$\psi_v(r) = A_v F_1(-v, 1 + 2x_v, y) \exp\left(-\frac{y}{2}\right) \frac{y^{x_v}}{r},$$

$$y = L \exp[-a(r - r_0)], \quad x_v = \frac{L-1}{2} - v,$$

$$L = 2\sqrt{m_{23} D}/a, \quad A_v = \sqrt{\frac{a(2x_v + 1)\ldots(2x_v + v)}{\Gamma(2x_v) v!}},$$

$F_1(a, b, x)$ is a degenerate hypergeometric function [26].

The amplitude of the rearrangement process has the form

$$M_R = (2\pi)^2 m_{23} \int \varphi_{12, 13}^* \exp\left(i\sqrt{\frac{m_{23}}{m_1}}\right) F_{2, 3}(\mathbf{r}, \boldsymbol{\rho}) d\mathbf{r} d\boldsymbol{\rho},$$

where $\varphi_{12, 13}$ is the wave function of the bound system formed in the reaction process. The expression for the cross section of the rearrangement process is as follows:

$$\frac{d\sigma_R}{d\theta} = \sqrt{\frac{m_{12}}{m_{23}}} |M_R|^2.$$

For processes of molecular dissociation by electrons, the final state is the state of the continuous spectrum and thus functions having the following form [9, 27] should be used as the wave function ψ_v

$$\psi_\varepsilon(r) = A_\varepsilon \exp\left(-\frac{y}{2r}\right)$$

$$\times \left\{ \exp[i\delta(x)] F_1\left(\frac{1}{2} + x - \frac{L}{2}, 1 + 2x, y\right) y^x + \exp[i\delta(x)] F_1\left(\frac{1}{2} - x - \frac{L}{2}, 1 - 2x, y\right) \frac{1}{y^x} \right\},$$

$$x = i\sqrt{2\varepsilon m_{23}}/a, \quad \delta(x) = \frac{\mathrm{agr}\,\Gamma(-2x)}{\Gamma(1/2 - x - l/2)} = -\delta(-x),$$

$$A_\varepsilon = (m_{23}/8\pi\varepsilon)^{1/4},$$

ε is the kinetic energy of the expanding nuclei and in order to calculate the transitions to the continuous spectrum from high vibrational states of the molecule we can use their following asymptotic form [9, 27]:

$$\psi_\varepsilon(r) \sim 2A_\varepsilon \sin\left[\sqrt{2m_{23}(r - r_0) - \ln\left(\frac{L}{a}\right)\frac{1}{r}}\right].$$

For like particles $m_2 = m_3$, the amplitude of elastic scattering and excitation is given by

$$M_{if} = 2\int \varphi_f^*(r) \left[\frac{\cos(pr/2)\cos(p_0 r/2)}{\alpha + ip_0 + \exp(ip_0 r)/r} + \frac{\sin(pr/2)\sin(p_0 r/2)}{\alpha + ip_0 + \exp(ip_0 r)/r}\right] \varphi_i(r) dr. \tag{2.19}$$

The total cross sections of all the processes including dissociation and all possible reactions taking place in a three-particle system may be obtained using the optical theorem [9, 21, 24]. As a result, we have

$$\sigma_{tot} = \frac{4\pi}{p_0^2} \int |\varphi_0(r)|^2 \left[\frac{1}{1 + \left(\dfrac{\alpha r + \cos(pr)}{p_0 r + \sin(p_0 r)}\right)^2} + \frac{1}{1 + \left(\dfrac{\alpha r + \cos(pr)}{p_0 r - \sin(p_0 r)}\right)^2}\right] dr, \tag{2.20}$$

which is the same as the expressions obtained in [3–8].

The expressions (2.19) and (2.20) reproduce the main characteristics in the cross sections of the processes, especially processes involving collisions between electrons and diatomic molecules. For example, for collisions between electrons and homonuclear ($m_2 = m_3$) molecules, in the denominators of the functions f_i for fixed values of r and varying momentum p_0 it is possible to have cases when

$$\mathrm{Re}[J_\pm(p_0)] = \alpha - ip_0 \pm \exp(ip_0 r)/r = 0.$$

These points correspond to resonant maxima of the functions F_i which lead to the appearance of maxima in the cross sections of all the processes, and also in σ_{tot}. In addition, the presence of the ratio $\sqrt{m_{23}/m_1}$ in the argument of the exponential function for the reaction amplitudes results in an appreciable isotope effect, first predicted by Demkov [5] and clearly observed experimentally [28].

It has been noted that all the characteristics of the cross sections of the processes (2.6) are expressed in

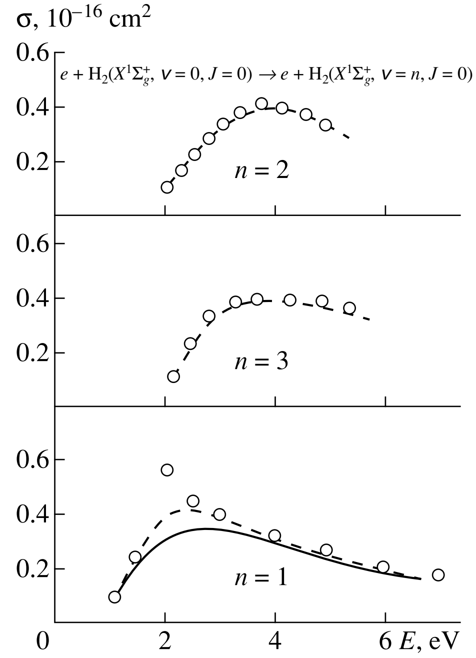


Fig. 3. Vibrational excitation of hydrogen molecules by electrons: circles—experimental data [28], solid curves—results of the present calculations, dashed curves—calculations using quasiclassical approximation [3].

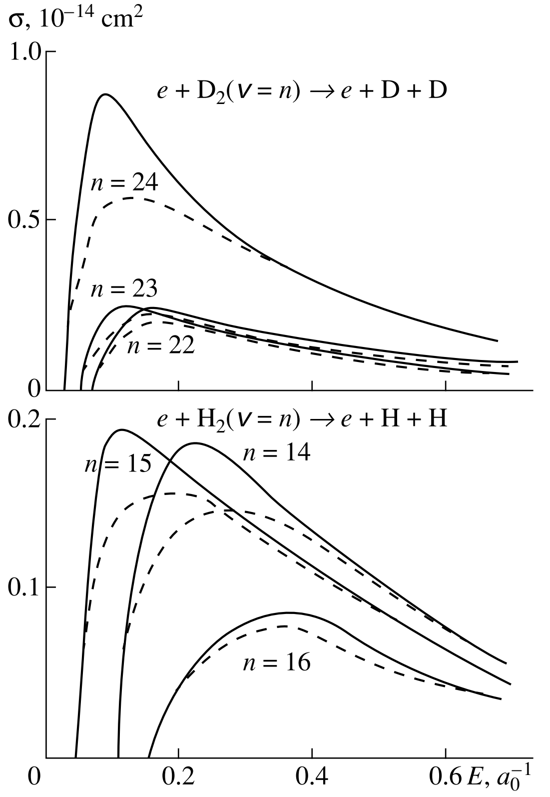


Fig. 4. Dissociation of vibrationally excited hydrogen and deuterium molecules by electrons: dashed curves—results of calculations using the approximation of zero-range potentials [27], solid curves—results of the present calculations.

terms of multiple scattering whose action is characterized by the term $\exp(ip_0r)/r$ in the functions F_i.

However, it has also been noted that the traditional interpretation of resonant phenomena accompanying the interaction of electrons or atoms with molecules is based on the assumption that transition complexes are formed during the collision process. The relationship between the multiple scattering approximation and the theory of a transition state can be seen immediately if we bear in mind that the imaginary zeros of the function $J_{\pm}(p_0)$ determine the quasi-steady-state terms of the transition state [9, 21] and the complex zeros correspond to the continuations of the terms into the region of quasi-steady states.

Another method [9] of obtaining an analytic solution of the system of integral equations (2.2) involves first going to the limit at $m/M \longrightarrow 0$ directly in the equations (2.2) and then using the approximation of separable variables. In this case it is convenient to take $\mathbf{p}_2$ and $\mathbf{p}_3$ as independent variables.

Similarly we can obtain an analytic solution of the equations (2.2) for the case of one heavy particle, $m_3 \equiv M$ and two identical light particles $m_1 = m_2 \equiv M$, for example, for the scattering of an electron at a hydrogen atom. In this case, the momenta of the light particles $\mathbf{k}_1$ and $\mathbf{k}_2$ can be conveniently assumed to be independent variables.

It should be noted that the smallness of the mass ratio and specific models for the T-matrices or Green functions were used to obtain analytic solutions of the systems of equations (2.2), (2.14). This is undoubtedly the main advantage of the formulation of these equations which allows us to explicitly isolate objects for which approximate expressions need to be constructed.

In addition, in analyses of various limiting cases such as a single light particle and two identical heavy ones or one heavy particle and two identical light ones, a preferred system of coordinates is obtained, which can be used to express the characteristics of all the dynamic processes taking place in these systems.

3. RESULTS AND DISCUSSION

Figures 1–10 give the results of calculations of the cross sections for scattering of electrons by hydrogen molecules and their isotopically substituted analogs (Figs. 1–4), hydrogen halides (Figs. 5, 6), nitrogen (Figs. 7, 8), lithium (Fig. 9), and sodium (Fig. 10), in both the ground and excited vibrational–rotational states. The calculations were made using the approximation of zero-range potentials and in the more general case of separable variables. It can be seen that the cross sections of these processes are accurately reproduced by formulas (2.7)–(2.9), which give a quantitative description of the processes (1.1), (2.6) and a qualitative picture of the effect whereby it is sufficient to allow only for the action of multiple scattering in this approximation.

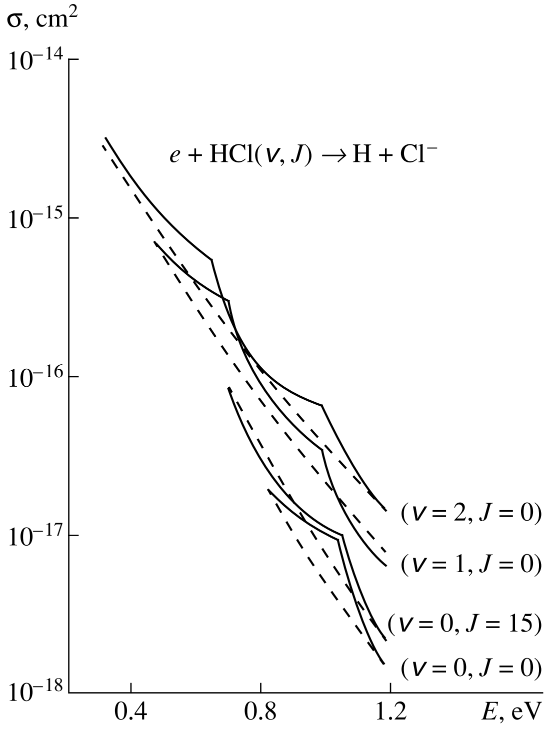


Fig. 5. Dissociative attachment of electrons to HCl molecules initially in excited vibrational–rotational states: solid curves—results of calculations from [10], dashed curves—results of the present calculations.

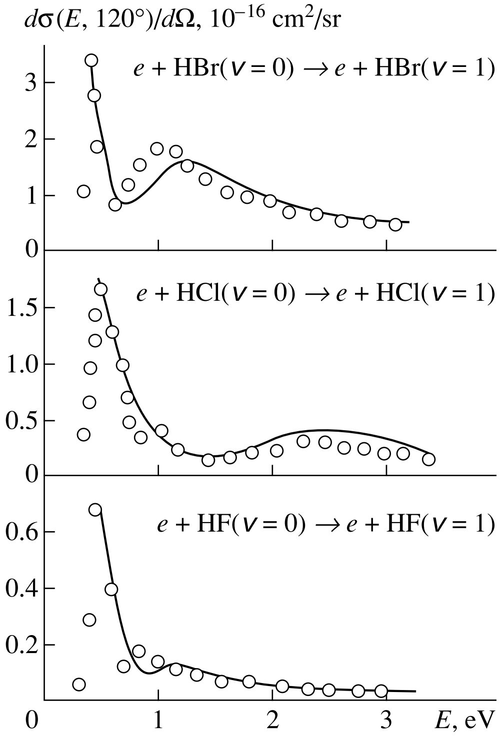


Fig. 6. Vibrational excitation of hydrogen halide molecules (HBr, HCl, HF) by electrons: circles—experimental data [28], curves—results of the present calculations.

In this approximation an isotope effect is observed, which was first predicted by Demkov [5] (Figs. 1, 2).

We note that the results of calculations of the cross sections of the processes (2.6) using the single scattering approximation [9] differ substantially from the experimental data since the single scattering approximation is inappropriate and multiple scattering must be taken into account.

Figure 1 gives the results of calculating the cross sections for dissociative attachment of an electron to hydrogen molecules and their isotopically substituted modifications made by solving the system of equations (2.2) numerically using a method described in [9, 29]. The energy dependences of the cross sections of the dissociative attachment processes are nonmonotonic. This indicates that the cross sections of the reactions taking place in a three-particle system have threshold characteristics which are observed most clearly in systems consisting of two like heavy particles and one light particle.

A comparison between these calculations [9, 23] and the experimental data [28] shows that modeling the interaction of an electron with each of the atoms in a molecule using the potentials given above in the multiple scattering approximation can give satisfactory agreement with the experiment (the orders of the cross sections agree, including the isotope effects and the threshold characteristics).

Note that various theories have now been put forward to calculate the cross sections of processes accompanying the interaction between electrons and molecules [1–8, 10–14] for which the terms of molecular negative ions and various other empirical parameters are required as initial data. However, in many cases it is impossible to make specific calculations according to these theories, even for diatomic molecules, because for the vast majority of molecules the data are lacking. In addition, there is a large discrepancy between the experimental results and the results of these calculations, as is demonstrated clearly by the calculations made using the Firsov–Smirnov model [4] which uses approximate values of the energy and the width of the H_2^- quasi-steady-state term obtained using the approximation of zero-range potentials. At energies below the dissociation threshold the cross sections were too high, whereas at energies above the threshold the agreement between the experiment and our calculations was considerably better.

Thus, this particular case clearly reveals the advantages of an approach based on quantum scattering theory in a system of several particles, in which the required initial data are the pair *T*-matrices, which can be constructed using the pair interaction potentials and wave functions corresponding to the initial state of the scattering process.

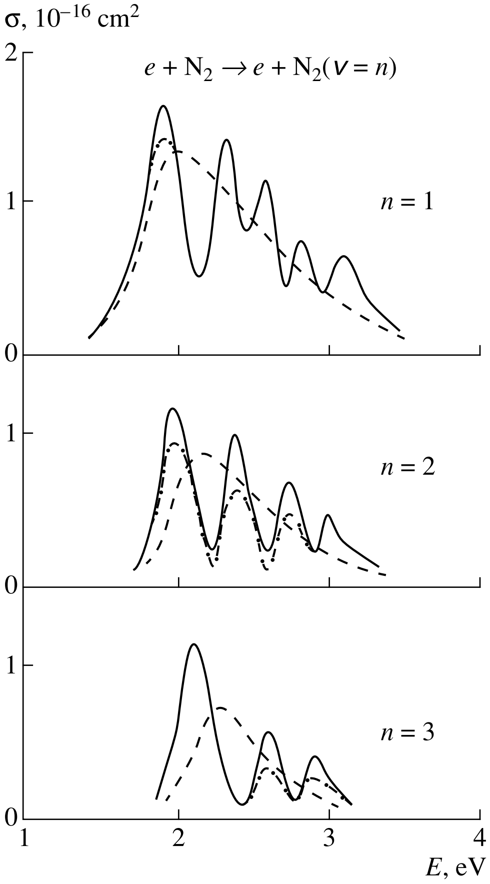


Fig. 7. Vibrational excitation of nitrogen molecules by electrons: solid curves—experimental data [28], dot-dash curves—results of calculations using quasiclassical approximation [3], dashed curves—results of the present calculations.

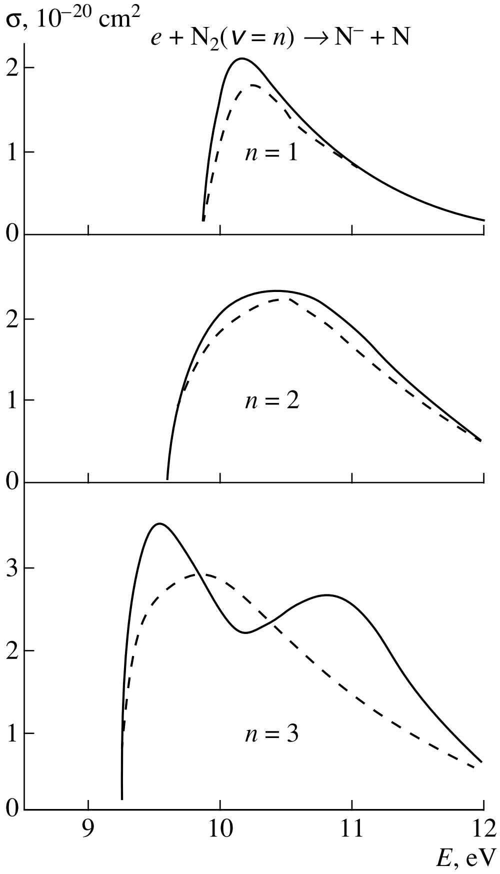


Fig. 8. Dissociative attachment of electrons to nitrogen molecules initially in excited vibrational states: solid curves—results of calculations from [11], dashed curves—results of the present calculations.

Figures 7 and 8 give the results of calculations of the cross sections for electron collisions with nitrogen molecules based on the quantum scattering theory in a three-body system, experimental data from [28], results of quasiclassical calculations [3], and calculations made using the strong coupling approximation [9, 11].

A distinguishing feature of the dissociative attachment of an electron to a nitrogen molecule is that as a result of the reaction a negative nitrogen ion forms in the *p*-state:

$$e + N_2 \longrightarrow N(^4S) + N^-(^3P).$$

Consequently, in order to describe this reaction in the separable approximation we need to use the potentials (2.10) with the form factors [11]

$$g(k) = [1 + \mu k Y_{10}(\rho)]/(k^2 + \beta^2).$$

Formulas linking the parameters λ, μ, β with the binding energy, scattering length, and effective radius are given in [8].

Unfortunately, only the binding energies have been determined experimentally so far [7, 28] and only theoretical estimates are available for the scattering length [30]. No experimental data or theoretical estimates are available for the effective radius. Thus, the absolute value of the cross section for dissociative attachment of an electron to a nitrogen molecule was used to determine the effective radius and the other parameters required [9, 11].

The interaction potential between the atoms of the molecule was simulated by the Morse potential, In order to construct the separable *T*-matrix from this potential we used the Beteman method described above and the method described in [9, 29] to obtain a numerical solution of the system of integral equations in this approximation.

A comparison between the results of the calculations and the experimental data [11, 28] shows that modeling the interaction of an electron with the atoms forming a molecule using the nonlocal separable potentials (2.11), (2.12) can give satisfactory agreement with the experimental data on average, which is a natural consequence of the properties of the nonlocal potentials. It should also be noted that by selecting the parameters of the separable potentials using the experimental data [9, 11, 28] it is possible to calculate all

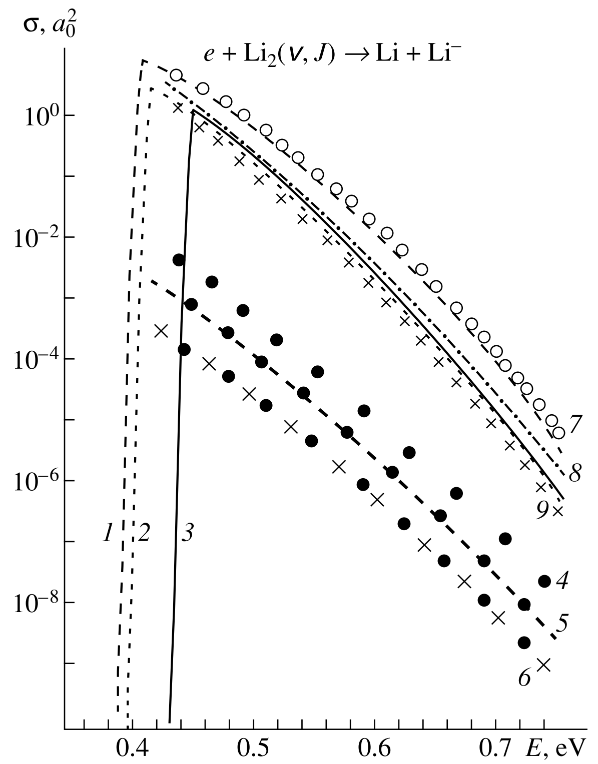


Fig. 9. Dissociative attachment of electrons to lithium molecules initially in excited vibrational–rotational states: *1*, *2*, *3*—results of calculations from [12], *4*, *5*, *6*—results of the present calculations using zero-range potential approximation, *7*, *8*, *9*—results of the present calculations using separable potential approximation (1.2).

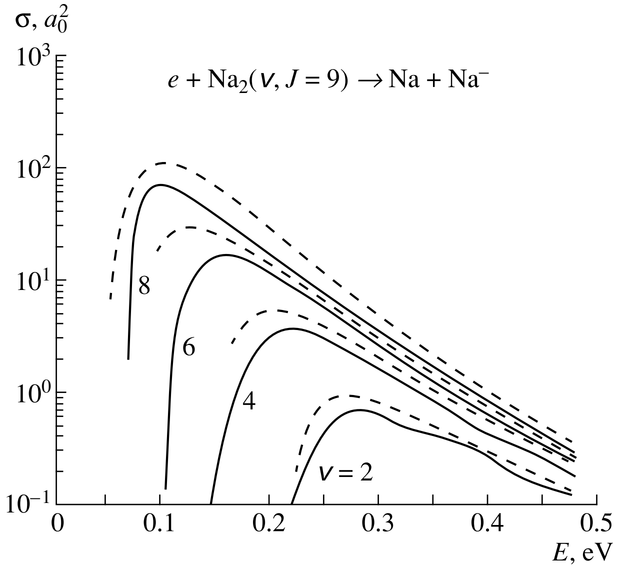


Fig. 10. Dissociative attachment of electrons to sodium molecules initially in excited vibrational–rotational states: solid curves—results of calculations from [13], dashed curves—results of the present calculations using separable potential approximation (1.2).

possible electron–molecule collision processes, which confirms the self-consistency of the proposed model.

In cases of dissociative attachment of electrons to vibrational–rotationally excited lithium, sodium, nitrogen, and hydrogen halide molecules (Figs. 5, 8–10) at specific impinging electron energies the calculations show that the vibrational excitation of the molecules predominates over their rotational excitation.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project 98-002-17266) and by the Academy of Sciences of Taiwan (project NCS-85-2112-M-007-009).

REFERENCES

1. L. G. Christophorou, *Electron Molecule Interaction and their Application* (Academic, New York, 1984).
2. A. Chutjian, A. Garscadden, and J. M. Wadehra, Phys. Rep. **264**, 393 (1995).
3. A. K. Kazanskiĭ and I. I. Fabrikant, Usp. Fiz. Nauk **143**, 602 (1984).
4. E. Illenberger and B. M. Smirnov, Usp. Fiz. Nauk **168**, 731 (1998).
5. Yu. N. Demkov, Phys. Lett. **15**, 235 (1965).
6. W. Domcke, Phys. Rep. **208**, 98 (1991).
7. A. Herzenberg, *Electron-Molecular Collision* (Plenum, New York, 1984, 191).
8. Yu. N. Demkov and V. N. Ostrovskiĭ, *Zero-Range Potentials and Their Application in Atomic Physics* (Plenum, New York, 1988).
9. S. Pozdneev, Nova Sci. Publ. **212**, 99 (1996).
10. J. M. Wadehra and J. N. Bardsley, Phys. Rev. Lett. **41**, 1791 (1978); Phys. Rev. Lett. **41**, 1795 (1978); S. Pozdneev and G. F. Drukharev, J. Phys. B: Atom Mol. Phys. **13**, 2611 (1980).
11. A. Huetz, F. Grestean, J. Mazeau, *et al.*, J. Chem. Phys. **72**, 5297 (1980); S. Pozdneev, J. Phys. B: Atom Mol. Phys. **16**, 867 (1983).
12. J. M. Wadehra, Phys. Rev. A **41**, 3607 (1990).
13. M. Kulz, M. Keil, B. Kortyna, *et al.*, Phys. Rev. A **53**, 3324 (1996).
14. E. E. Nikitin, *Theory of Elementary Atomic and Molecular Processes in Gases* (Khimiya, Moscow, 1970; Clarendon, Oxford, 1974).
15. S. A. Pozdneev, Khim. Vys. Énerg. **18**, 290 (1984).
16. S. A. Pozdneev, Zh. Éksp. Teor. Fiz. **77**, 38 (1979).
17. N. F. Stepanov and V. I. Pupyshev, Molecular Quantum Mechanics and Quantum Chemistry (Mosk. Gos. Univ., Moscow, 1991); N. F. Stepanov, Sorosov. Obraz. Zh., No. 10, 33 (1996); A. I. Nemukhin, Sorosov. Obraz. Zh., No. 6, 48 (1998).

18. S. P. Merkur’ev and L. D. Faddeev, *Quantum Scattering Theory for Systems of Several Particles* (Nauka, Moscow, 1985).
19. V. Efimov, Nucl. Phys. A **362**, 1024 (1981).
20. K. P. Hube and G. Gerzberg, *Constants of Diatomic Molecules* (Springer-Verlag, New York, 1979).
21. G. F. Drukarev, Zh. Éksp. Teor. Fiz. **67**, 38 (1974).
22. D. A. Micha, Adv. Chem. Phys. **30**, 7 (1975).
23. S. A. Pozdneev, Kratk. Soobshch. Fiz., No. 6, 61 (1987).
24. V. B. Belyaev, *Lectures on the Theory of Few-Particle Systems* (Énergoatomizdat, Moscow, 1986).
25. A. M. Veselova, S. P. Merkur’ev, and L. D. Faddeev, *Diffraction Interaction of Hadrons with Nuclei* (Naukova Dumka, Kiev, 1987), p. 107.
26. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Nauka, Moscow, 1979; Dover, New York, 1965).
27. R. I. Boĭkova, Vestn. Leningr. Gos. Univ., Ser. Fiz. Khim., No. 10, 31 (1977).
28. G. J. Schultz, Rev. Mod. Phys. **45**, 423 (1973).
29. N. M. Larson and J. H. Hetherington, Phys. Rev. C **9**, 699 (1974).
30. N. B. Berezina and Yu. N. Demkov, Zh. Éksp. Teor. Fiz. **68**, 848 (1975).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 45–49.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 51–56.
Original English Text Copyright © 2000 by Kravtsov, Mikhaĭlov.

ATOMS, SPECTRA, RADIATION

Temperature Dependence of the Formation Rates of Hydrogen-Helium Mesic Molecules in Collisions of Slow Hydrogen Atoms with Helium¶

A. V. Kravtsov* and A. I. Mikhaĭlov**
St. Petersburg Nuclear Physics Institute
Gatchina, Leningradskaya oblast, 188350 Russia
**e-mail: kravtsov@hep486.pnpi.spb.ru*
***e-mail: mikhailo@thd.pnpi.spb.ru*

Received February 8, 1999

Abstract—The rates of molecular muon transfer from the ground-state muonic hydrogen to helium isotopes are calculated in an improved adiabatic approximation. The results obtained by us at various temperatures are compared with the available experimental data. *© 2000 MAIK "Nauka/Interperiodica".*

Muon transfer from the ground state of the muonic hydrogen to nuclei with $Z > 1$ is one of the important problems of mesic-atom physics for more than 30 years. The transfer to helium is of special interest, since it is connected with the problem of muon-catalyzed fusion in the deuterium–tritium mixture. Helium nuclei are unavoidably accumulated in the mixture due to the nuclear fusion reaction and tritium decay. Muon transfer in collisions of muonic hydrogen atoms with helium stops the cycle of the catalysis; therefore, the transfer rate is an important characteristic of the muon catalyzed fusion.

The rate of the direct muon transfer to helium is rather low ($\sim 10^6$ s^{-1}) [1, 2], because the crossing point of the lowest terms of the system ($2p\sigma$ and $1s\sigma$), which corresponds to the initial and final states of the reaction, turns out to be deep under the barrier at energies ~1 eV.

Another possible mechanism for muon transfer with the formation of the intermediate molecular state via the conversion of the atomic electron was proposed in [3]. The resulting molecular ion is in an excited state (in the muon motion) and undergoes deexcitation to the lower term $1s\sigma$ in a time of $\sim 10^{-12}$ s. As a result, the muon turns out to be bound by a helium nucleus, forming a muonic helium atom in the ground state $(\mathrm{He}^{++}\mu)_{1s}$.

Since the rate of formation of the muonic molecule ($\sim 10^8$ s^{-1}) is much lower than that of its decay ($\sim 10^{12}$ s^{-1}), the muon-transfer rate nearly coincides with the formation rate. The latter was calculated in a number of papers [3–6] in various approximations.

In this paper, we pay special attention to the construction of the effective potential in accordance with the prescription given in [7] ("simple-approach approximation"). We calculated the muon-transfer rates in a low energy collision of hydrogen isotopes with helium isotopes and compared them with new experimental data.

The rate of the formation of the muonic molecule[1] is determined by a dipole transition with the conversion of the atomic electron and is given by [6]

$$d\lambda = \frac{16\pi^2}{3} N_0 a_e^3 \xi^5 q^{-1} I^2(q) \langle d \rangle^2 \nu_e, \tag{1}$$

where $\nu_e = m_e e^4/\hbar^3 = 4.134 \times 10^{16}$ s^{-1} is an atomic frequency unit, $a_e = \hbar^2/m_e e^2 = 0.529 \times 10^{-8}$ cm is the Bohr radius of the hydrogen atom, $\xi = m_e/m$ (here, m_e is the electron mass and m is the reduced mass of mesic hydrogen; $m^{-1} = m_\mu^{-1} + M_1^{-1}$, where m_μ is the muon mass, and M_1 is the mass of the nucleus of the hydrogen isotope), and $N_0 = 4.25 \times 10^{22}$ cm^{-3} is a liquid-hydrogen density. The quantity $q^{-1}I^2(q)$ is calculated in atomic units and determines the rate of the electron transition from the bound $1s$-state of the helium atom to the continuum. The quantity $\langle d \rangle^2$ is calculated in mesic-atom units (see below) and determines the rate of the dipole transition of the system of three bodies from the continuum to the bound state on the term $2p\sigma$.

The momentum q and the energy ε_e of the conversion electron are given by

$$q = \sqrt{2m_e\varepsilon_e}, \quad \varepsilon_e = \varepsilon + |\varepsilon_{J\nu}| - I_e, \tag{2}$$

where ε is the collision energy, $\varepsilon_{J\nu}$ is the energy of the mesic molecule in the rotational–vibrational state (J, ν) (here, J and ν are the rotational and the vibrational quantum number, respectively), and I_e is the binding energy of the $1s$-electron in the helium atom.

The integral $I(g)$ is the matrix element of the variable $1/\rho^2$ (ρ is the distance between the electron and the

---

¶This article was submitted by the authors in English.

[1] For brevity, we shall call a molecular ion $(\mathrm{H}\mu\mathrm{He})^{++}$, $\mathrm{H} \equiv p$, d or t, mesic molecule.

Table 1

ε_e, eV	q	$I(q)$	$q^{-1}I^2(q)$
4	0.542	0.491	0.445
8	0.767	0.607	0.481
12	0.939	0.712	0.540
16	1.084	0.805	0.598
20	1.212	0.887	0.649
24	1.328	0.964	0.700
28	1.435	1.030	0.739
32	1.534	1.089	0.774
36	1.627	1.144	0.805
40	1.715	1.196	0.834
44	1.798	1.248	0.866
48	1.878	1.300	0.899
52	1.955	1.352	0.934
56	2.029	1.403	0.971
60	2.100	1.455	1.008

center of mass of the system) calculated with the electron wave functions

$$I(q) = \int_0^\infty d\rho R_{q1}(\rho)R_{1s}(\rho), \tag{3}$$

where $R_{1s}(\rho)$ is the radial ground-state wave function of the electron in the helium atom and $R_{q1}(\rho)$ is the radial wave function of the p-wave electron ($l = 1$) in the continuum.

The functions $R(\rho)$ are normalized by the conditions

$$\begin{gathered}\int_0^\infty R_{1s}^2(\rho)\rho^2 d\rho = 1,\\ \int_0^\infty R_{q'1}(\rho)R_{q1}(\rho)\rho^2 d\rho = \delta(q'-q).\end{gathered} \tag{4}$$

The Hartree–Fock wave functions were used as $R(\rho)$. The wave function of the emitted electron was calculated in the frozen core model (in the model with the core reconstruction, the results for $I^2(q)$ are about 10% smaller [6]). The values of the integral $I(q)$ for several energies ε_e are given in Table 1.

The dipole matrix element calculated with the wave functions of the system of three bodies has the form [5]

$$\langle d\rangle = aI_1 + bI_2, \tag{5}$$

$$\begin{gathered}a = \frac{M_2 - M_1}{M_t} - \frac{1}{2}, \quad b = 1 + \frac{2m_\mu}{M_t},\\ M_t = M_1 + M_2 + M_\mu,\end{gathered} \tag{6}$$

$$\begin{gathered}I_1 = \frac{1}{k}\int_0^\infty \chi_f(R)\chi_i(R)RdR,\\ I_2 = \frac{1}{k}\int_0^\infty \chi_f(R)\chi_i(R)dR\int \phi_{2p\sigma}^2(R, \mathbf{r})\frac{\mathbf{R}\cdot\mathbf{r}}{R}d\mathbf{r}.\end{gathered} \tag{7}$$

M_2 is the mass of the helium isotope, $\phi_{2p\sigma}(R, \mathbf{r})$ is the wave function of the muon in the $2p\sigma$ state in the field of the two Coulomb centers, R is the internuclear distance, and $\mathbf{r}$ is the muon coordinate calculated from the middle of the internuclear distance. The radial wave functions $\chi_i(R)$ and $\chi_f(R)$ describe the relative motion of the nuclei in the initial and final states, respectively. In the one-channel approximation they are obtained from the Schrödinger equation

$$\left[\frac{d^2}{dR^2} + 2M(\varepsilon - W) - \frac{J(J+1)}{R^2}\right]\chi(R) = 0. \tag{8}$$

Here J is the total orbital angular momentum of the system of three particles, M is the reduced mass of the nuclei, $\varepsilon = k^2/2M$ is the collision energy, and k is the asymptotic momentum of the relative nuclear motion in the effective potential W [the same momentum enters formulas (7)]. All the quantities in Eq. (8) are used in the mesic atom units $e = \hbar = m = 1$, where $m = M_1 m_\mu/(M_1 + m_\mu)$. In these units,

$$m_\mu = 1 + \lambda_1, \quad \lambda_1 = m_\mu/M_1. \tag{9}$$

The effective potential W consists of the term E, the adiabatic corrections, and the energy of the Coulomb repulsion of the nuclei [8]:

$$W = E + \frac{2}{R} + \frac{1}{2M}[H^+ - H^* + \kappa(H^- - 2H^*)], \tag{10}$$

$$M = \frac{M_2 M_1}{M_2 + M_1}, \quad \kappa = \frac{M_2 - M_1}{M_2 + M_1}. \tag{11}$$

For $R \longrightarrow \infty$, we have $E_{2p\sigma} \longrightarrow -1/2$, $H_{2p\sigma}^+ \longrightarrow 1/4$, $H_{2p\sigma}^* \longrightarrow 1/4$, $H_{2p\sigma}^- \longrightarrow 1/2$, so that

$$W_{2p\sigma} \longrightarrow -\frac{1}{2} = E_{1s}(H\mu). \tag{12}$$

The asymptotic value of $W_{2p\sigma}$ coincides with the energy of the $H\mu$-atom in the ground state (which corresponds to the initial conditions of the collision). However, the reduced mass M in Eq. (11) (and hence the asymptotic momentum k), which enters Eq. (8) differs from the true reduced mass of the system $H\mu + He$, which is

$$\begin{gathered}\tilde{M} = \frac{M_2(M_1 + m_\mu)}{M_2 + M_1 + m_\mu} = M\frac{1 + \lambda_1}{1 + \lambda},\\ \lambda = \frac{m_\mu}{M_2 + M_1}.\end{gathered} \tag{13}$$

If one makes the replacement $M \longrightarrow \tilde{M}$ in Eq. (8), the calculated energy levels of the mesic molecule ε_{Jv} [9] turn out to be in better agreement with accurate calculations[2] [10–12] than those obtained earlier [3] with the mass M. Even better agreement can be obtained if one replaces κ in Eq. (10) by $\tilde{\kappa}$, which gives the correct asymptotic value of the effective potential in the $1s\sigma$ channel. We are solving a one-level problem; however, the $1s\sigma$ channel is the second open channel in the problem of slow collisions of mesic hydrogen with helium, and its influence can be taken into account indirectly[3] by the replacement $\kappa \longrightarrow \tilde{\kappa}$. In order to obtain $\tilde{\kappa}$, let us write the asymptotics of the potentials in the $1s\sigma$ channel: for $R \longrightarrow \infty$, we have $E_{1s\sigma} \longrightarrow -2$, $H^{+}_{1s\sigma} \longrightarrow 1$, $H^{*}_{1s\sigma} \longrightarrow 1$, $H^{-}_{1s\sigma} \longrightarrow -2$, so that

$$\begin{aligned} W_{1s\sigma} &\longrightarrow -2 + \frac{1}{2\tilde{M}}[1 - 1 + \tilde{\kappa}(-2-2)] \\ &= -2\left(1 + \frac{\tilde{\kappa}}{\tilde{M}}\right). \end{aligned} \tag{14}$$

Actual asymptotic value of the potential in the $1s\sigma$ channel should coincide with the ground-state energy of the $He\mu$-atom:

$$\begin{gathered} E_{1s\sigma}(He\mu) = -2m', \\ m' = \frac{m_\mu M_2}{M_2 + \mu} = \frac{1+\lambda_1}{1+\lambda_2}, \quad \lambda_2 = \frac{m_\mu}{M_2}. \end{gathered} \tag{15}$$

Let us choose such $\tilde{\kappa}$ that $W_{1s\sigma} \longrightarrow E_{1s\sigma}(He\mu)$. Then,

$$\tilde{\kappa} = \kappa\frac{(1+\lambda_1)^2}{(1+\lambda)(1+\lambda_2)}. \tag{16}$$

So, the wave functions $\chi_{i,f}(R)$ are solutions to the equation

$$\begin{gathered} \left[\frac{d^2}{dR^2} + 2\tilde{M}(\varepsilon - \tilde{W}_{2p\sigma}) - \frac{J_{i,f}(J_{i,f}+1)}{R^2}\right] \\ \times \chi_{i,f}(R) = 0, \end{gathered} \tag{17}$$

where

$$\begin{aligned} \tilde{W}_{2p\sigma} &= E_{2p\sigma} + \frac{2}{R} \\ &+ \frac{1}{2\tilde{M}}[H^{+}_{2p\sigma} - H^{*}_{2p\sigma} + \tilde{\kappa}(H^{-}_{2p\sigma} - 2H^{*}_{2p\sigma})] \end{aligned} \tag{18}$$

with the boundary conditions

$$\begin{gathered} \chi_i(0) = \chi_f(0) = 0, \\ \chi_i(R)_{R\to\infty} = \sin(kR - J\pi/2 + \delta_J), \quad k = \sqrt{2\tilde{M}\varepsilon}, \\ \chi_f(R)_{R\to\infty} = \exp[-(2\tilde{M}|\varepsilon_{Jv}|)^{1/2}R]. \end{gathered} \tag{19}$$

The energy of the rotational-vibrational state (J, v) of the molecular ion $(H\mu He)^{++}$ is obtained together with the wave function $\chi_f(R)$ when solving the Schrödinger equation. For slow collisions, it is enough to consider $J_i = 0$ for initial and $J_f = 1$ for final states (the molecule is formed via the dipole transition). There exist only three bound states in the $2p\sigma$ channel, which have $v = 0$ and $J = 0, 1, 2$. The level energies ε_{00} and ε_{10} obtained in this paper are given in Table 2 together with the results of other papers.

As can seen from the table, the level energies calculated in this paper are very close to the high-accuracy calculations with large number of basis functions [10–12]. For this reason, when calculating the rates of the formation of mesic molecules, we use Eq. (17), where $\tilde{M}$ and $\tilde{\kappa}$ are defined by Eqs. (13) and (16).

When calculating the wave function of the initial state $\chi_i(R)$, the electron screening was taken into account by an additional term in the potential of Eq. (17). The influence of the electron shell of the helium atom on the final (bound) state of the molecule is negligible because of the short length of the corresponding wave function (see Refs. [4–6] for more details).

Table 3 shows the rates of the molecule formation λ (10^8 s^{-1}) averaged over the Maxwellian energy distribution. The electron screening is taken into account.

Recently new experimental data of λ have been obtained. In the experiment, the total muon transfer rate from the ground-state muonic hydrogen to helium,

$$\lambda_{p\mathrm{He}} = \lambda_{p\mu\mathrm{He}} + \lambda^{\mathrm{dir}}_{p\mathrm{He}}, \tag{20}$$

is measured, where $\lambda^{\mathrm{dir}}_{p\mathrm{He}}$ is the rate of the direct muon transfer without the molecule formation. This rate was calculated in [1] for the systems $p\mu + {}^{3,4}\mathrm{He}$ and $d\mu + {}^{3,4}\mathrm{He}$. For collision energies $\varepsilon \le 0.1$ eV, the rate of the direct muon transfer does not depend on energy and amounts to

$$\lambda^{\mathrm{dir}}_{p\mathrm{He}} \approx 0.06 \times 10^8 \ \mathrm{s}^{-1} \ \text{ and } \ \lambda^{\mathrm{dir}}_{d\mathrm{He}} \approx 10^6 \ \mathrm{s}^{-1}. \tag{21}$$

As was mentioned, the rate of molecular transfer coincides with the rate of the molecule formation (see Table 3). So, when comparing experimental rates with theoretical ones (Table 4), the latter were enlarged according to Eqs. (20) and (21).

As a matter of fact, the genuine muon transfer to the helium nucleus occurs when the molecular ion decays into hydrogen and muonic helium. Such a decay may be radiative (or via the electron conversion), as well as via predissociation. The latter channel was not considered in the first papers on the molecular charge exchange. This led to large discrepancies when comparing the calculations

---

[2] These calculations make use of about 300 to 3000 basis functions.

[3] A similar procedure for solving the two-channel problem was proposed earlier in [7] ("simple approach" of "improved two-channel approximation").

Table 2. The binding energies (eV) the hydrogen-helium muonic molecules

Energy	Reference	$p\mu^3$He	$p\mu^4$He	$d\mu^3$He	$d\mu^4$He	$t\mu^3$He	$t\mu^4$He
$-\varepsilon_{00}$	[3][a]	67.2	73.9	69.5	77.6	71.6	80.5
	[9][b]	69.0	75.4	70.6	78.7	72.3	81.3
	This paper[c]	73.2	80.8	71.0	79.4	72.3	81.4
	[10]			70.7			
	[13]	67.7	74.4	70.0	78.0	71.9	80.8
	[14]	72.8	80.6	69.4	77.5		
$-\varepsilon_{10}$	[3][a]	34.9	41.6	46.5	55.9	52.3	62.9
	[9][b]	38.1	45.4	48.2	57.6	53.4	63.9
	This paper[c]	41.5	50.0	48.5	58.3	53.4	64.0
	[10]		50.0	47.9	57.8		
	[11]			48.4	58.2		
	[12]			48.4	58.2		
	[13]	33.8	41.2	46.8	56.1	52.7	63.1
	[14]	38.8	47.4	46.3	55.7		

Notes: [a] One-channel approximation with M and κ.
[d] One-channel approximation with $\tilde{M}$ and κ.
[c] One-channel approximation with $\tilde{M}$ and $\tilde{\kappa}$.

Table 3. The rates of the molecule formation λ (10^8 s^{-1}) averaged over the Maxwellian energy distribution. The electron screening is taken into account

T, K	$p\mu^3$He	$p\mu^4$He	$d\mu^3$He	$d\mu^4$He	$t\mu^3$He	$t\mu^4$He
15	0.52	0.33	2.40	12.7	51.2	1.89
20	0.52	0.33	2.34	11.8	45.5	1.86
25	0.51	0.32	2.28	11.0	41.1	1.84
30	0.51	0.32	2.24	10.4	37.6	1.82
35	0.51	0.32	2.20	9.8	34.7	1.79
40	0.50	0.31	2.16	9.3	32.3	1.77
50	0.50	0.31	2.09	8.5	28.6	1.74
100	0.47	0.29	1.85	6.2	18.8	1.60
150	0.46	0.28	1.70	5.1	14.4	1.51
200	0.45	0.27	1.60	4.4	11.9	1.43
250	0.44	0.27	1.52	3.8	10.2	1.37
300	0.43	0.26	1.45	3.5	9.0	1.31
350	0.42	0.25	1.39	3.2	8.1	1.27
400	0.42	0.25	1.34	2.9	7.3	1.22
450	0.41	0.25	1.30	2.7	6.7	1.19
500	0.40	0.24	1.26	2.6	6.2	1.15

Table 4. Muon transfer rates (10^8 s^{-1}) for various isotopes and temperatures

	Experiment		Theory
$\lambda(p^3\mathrm{He}, 30\ \mathrm{K})$	0.46 ± 0.15	[15]	0.57
$\lambda(p^4\mathrm{He}, 30\ \mathrm{K})$	0.42 ± 0.07	[15]	0.38
$\lambda(p^4\mathrm{He}, 300\ \mathrm{K})$	0.36 ± 0.10	[16]	0.32
$\lambda(p^4\mathrm{He}, 300\ \mathrm{K})$	0.44 ± 0.20	[17, 9]	0.32
$\lambda(p^4\mathrm{He}, 300\ \mathrm{K})$	0.5 ± 0.1	[18, 19]	0.32
$\lambda(d^3\mathrm{He}, 30\ \mathrm{K})$	1.86 ± 0.08	[20]	2.25
$\lambda(d^3\mathrm{He}, 40\ \mathrm{K})$	2.25 ± 0.15	[21]	2.17
$\lambda(d^3\mathrm{He}, 300\ \mathrm{K})$	1.24 ± 0.05	[22]	1.46
$\lambda(d^4\mathrm{He}, 30\ \mathrm{K})$	10.50 ± 0.21	[20]	10.4
$\lambda(d^4\mathrm{He}, 300\ \mathrm{K})$	3.68 ± 0.18	[23]	3.48
$\lambda(t^3\mathrm{He}, 15\ \mathrm{K})$	46 ± 4	[24]	51

with the experiments, in which the probability of the transfer was obtained by the measurement of the X-ray yield. These discrepancies were removed when the predissociation channel was pointed out [13, 9] and taken into account.

Comparing the experimental and theoretical values for muon transfer from ground-state muonic hydrogen to helium, one can see that they are in reasonable agreement. This means that the main features of the process are understood correctly.

The authors are grateful to S. Tresch and B. Gartner for sending the experimental results prior to publication.

REFERENCES

1. A. V. Matveenko and L. I. Ponomarev, Zh. Éksp. Teor. Fiz. **63**, 48 (1972) [Sov. Phys. JETP 36, **24** (1972)].
2. S. S. Gershteĭn, Zh. Éksp. Teor. Fiz. **43**, 706 (1962) [Sov. Phys. JETP **16**, 501 (1963)].
3. Yu. A. Aristov, A. V. Kravtsov, N. P. Popov, *et al.*, Yad. Fiz. **33**, 1066 (1981) [Sov. J. Nucl. Phys. **33**, 564 (1981)].
4. A. V. Kravtsov, A. I. Mikhaĭlov, and N. P. Popov, J. Phys. B: At. Mol. Phys. **19**, 1323 (1986).
5. A. V. Kravtsov, A. I. Mikhaĭlov, and N. P. Popov, J. Phys. B: At. Mol. Phys. **19**, 2579 (1986).
6. V. K. Ivanov, A. V. Kravtsov, A. I. Mikhaĭlov *et al.*, Zh. Éksp. Teor. Fiz. **91**, 358 (1986) [Sov. Phys. JETP **64**, 210 (1986)].
7. L. I. Ponomarev, L. N. Somov, and M. P. Faĭfman, Yad. Fiz. **29**, 133 (1979) [Sov. J. Nucl. Phys. **29**, 67 (1979)].
8. S. I. Vinitsky and L.I. Ponomarev, Fiz. Elem. Chastits At. Yadra **13**, 1336 (1982) [Sov. J. Part. Nucl. **13**, 557 (1982)].
9. A. V. Kravtsov, A. I. Mikhaĭlov, and V. I. Savichev, Hyperfine Interact. **82**, 1205 (1993).
10. S. Hara and T. Ishihara, Phys. Rev. A **39**, 5633 (1989).
11. Y. Kino and M. Kamimura, Hyperfine Interact. **82**, 195 (1993).
12. V. I. Korobov, Hyperfine Interact. **101/102**, 329 (1996).
13. S. S. Gershteĭn and V. V. Gusev, Hyperfine Interact. **82**, 185 (1993).
14. V. B. Belyaev, O. I. Kartavtsev, V. I. Kochkin, *et al.*, Phys. Rev. A **52**, 1765 (1995).
15. S. Tresch, F. Mulhauser, C. Piller, *et al.*, Phys. Rev. A **58**, 3528 (1998).
16. V. M. Bystritsky, V. P. Dzhelepov, V. I. Petrukhin, *et al.*, Zh. Éksp. Teor. Fiz. **84**, 1257 (1983) [Sov. Phys. JETP **57**, 728 (1983)].
17. H. P. von Arb, F. Dittus, H. Hofer, *et al.*, Muon Catal. Fusion **4**, 61 (1989).
18. C. Piller *et al.*, Helv. Phys. Acta **67**, 779 (1994).
19. S. Tresch, P. Ackerbauer, W. H. Breunlich, *et al.*, Hyperfine Interact. **101/102**, 221 (1996).
20. B. Gartner, P. Ackerbauer, M. Augsburger, *et al.*, in *Proceedings of Workshop on Exotic Atoms, Molecules and Muon Catalyzed Fusion, Ascona, 1998* (submitted to Hyperfine Interact.).
21. D. V. Balin, T. Case, K. M. Crowe, *et al.*, in *Proceedings of Workshop on Exotic Atoms, Molecules and Muon Catalyzed Fusion, Ascona, 1998* (submitted to Hyperfine Interact.).
22. D. V. Balin, V. N. Baturin, Yu. A. Ghestnov *et al.*, in *Muonic Atoms and Molecules*, Ed. by L. A. Schaller and C. Petitjean (Birkhäuser Verlag, Basel, 1993), p. 25.
23. D. V. Balin, A. A. Vorobyov, An. A. Vorobyov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 236 (1985) [JETP Lett. **42**, 293 (1985)].
24. T. Matsuzaki, K. Nagamine, K. Ishida, *et al.*, in *Proceedings Workshop on Exotic Atoms, Molecules and Muon Catalyzed Fusion, Ascona, 1998* (submitted to Hyperfine Interact).

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 50–58.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 57–66.
Original Russian Text Copyright © 2000 by Panteleev, Rerikh, Starostin.

ATOMS, SPECTRA, RADIATION

Influence of Interference Effects of Spontaneous Emission on the Behavior and Spectra of Resonance Fluorescence of a Three-Level Atom in a Strong Wave Field

A. A. Panteleev*, Vl. K. Rerikh, and A. N. Starostin**
Russian State Scientific Center, Troitsk Institute of Innovative and Thermonuclear Research (TRINITI),
Troitsk, Moscow oblast, 142092 Russia
**e-mail: mediana@mail.ru*
***e-mail: vroerieh@fly.triniti.troitsk.ru*

Received June 7, 1999

Abstract—An investigation is made of the influence of quantum interference processes accompanying radiative relaxation of excited states on the population dynamics, total intensity, and spectra of the resonance fluorescence of three-level *V*-type atoms. Analytic expressions are obtained for the total intensity and spectra of the resonance fluorescence taking into account the off-diagonal nature of the radiative relaxation operator. It is shown that quantum interference process can substantially alter the total spontaneous emission intensity of the atoms and the population dynamics of the atomic levels, as well as the resonance fluorescence spectra. *© 2000 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Interference effects accompanying the spontaneous emission of atoms have recently formed the subject of numerous studies in connection with the development of lasers with no population inversion, the generation of squeezed states, studies of laser-induced transparency, suppression of absorption and emission, and other effects. Interference effects accompanying the spontaneous emission of atoms have been known for some time: modification of the polarization structure of spontaneous emission in a magnetic field (Hanle effect), level crossing, quantum resonances, transient quantum beats, and so on [1]. These effects may be classified as linear. In addition, nonlinear effects may be observed as a result of nonlinear parametric processes (the Stark effect, the dynamic Stark effect, two- and multiphoton scattering, saturation effects, multiple harmonic generation, bleaching under multiphoton resonance, and so on [2–10]). Interference effects do not usually influence the rate of radiative relaxation of the atomic states. From the mathematical point of view their contribution becomes zero either when integrated over angles (for polarization effects) or when integrated over the spectrum (for nonlinear interference effects).

The influence of interference effects on the rate of radiative decay was taken into account in studies of the spontaneous emission spectra [7, 11], in analyses of the dynamics of molecular quasienergy states [12], and in studies of the spontaneous Raman scattering spectra of two adjacent states [13]. Note that in these studies the off-diagonal terms in the radiative relaxation operator corresponding to interference effects were introduced formally.

The simplest system in which interference effects are observed in one-photon processes contains three levels since a minimum of two correlated paths is required. Three-level systems satisfying this condition may have three configurations: stepped, Λ-type, and *V*-type. Interference effects only influence the spontaneous relaxation rate for transitions having the same final states. Hence, interference terms influencing the rate of radiative relaxation can only occur for the *V*-configuration. In the present paper we study the influence of interference effects accompanying the spontaneous emission of a three-level *V*-type atom in a laser field on its population dynamics and also on its absorption and emission spectra. Note that the resonance fluorescence properties of a doublet-type three-level atom were investigated experimentally and theoretically in [14] where it was shown that these spectra can contain up to seven peaks. This is because of the high-frequency Stark effect: the presence of a strong field binding the N levels leads to the formation of quasienergy states and the transitions between these are responsible for the multiplet structure of the spectra. In this case, the maximum number of peaks is determined by the relationship [15] $(N^3 + 1)/(N + 1)$. Note that we have already investigated the characteristics of quantum interference accompanying the spontaneous emission of a three-level atom in problems involving the emission of an atom in a cavity and also in analyses of the characteristics of scattered radiation [16]. We investigated the influence of interference effects on scattered radiation transport processes in [17].

The present paper is constructed as follows. In the second section we give the equations of motion for a three-level *V*-type atom in a monochromatic electromagnetic field and we analyze the influence of interference effects on the radiative relaxation operator. In the third section we investigate the dynamics of the atomic states and the total intensity of the resonance fluorescence. Finally, in the fourth section we study the influence of interference effects on the emission spectra and discuss the results.

2. EQUATIONS OF MOTION FOR A THREE-LEVEL ATOM IN AN ELECTROMAGNETIC FIELD

The behavior of the atomic system will be described using the equation for the density matrix

$$i\hbar\dot{\rho} = [\hat{H}_0 + \hat{V}, \rho] + i\hat{R}, \tag{1}$$

where $\hat{H}_0$ is the Hamiltonian of the free atom:

$$\hat{H}_0 = \hbar \sum_{j=2,3} \Delta_j \hat{S}_{jj}, \tag{2}$$

and $\hat{V}$ is the Hamiltonian of the interaction between the atom and a monochromatic electromagnetic wave having the field strength E_L and the frequency ω_L (in the rotating wave approximation):

$$\hat{V} = \hbar \sum_{j=2,3} (V_j \hat{S}_j^- + \hat{S}_j^+ V_j^*). \tag{3}$$

Here $V_j = -\mu_{j1}E_L/2\hbar$, μ_{j1} is the matrix element of the dipole moment, $\hat{S}_{jm} = |m\rangle\langle j|$, $\hat{S}_j^- = |1\rangle\langle j|$, $\hat{S}_j^+ = |j\rangle\langle 1|$, $|j\rangle$ are the projection operators of the *j*th state [18] ($j = 1$ corresponds to the ground state, and $j = 2, 3$ corresponds to the excited states), $\Delta_j = \omega_{j1} - \omega_L$ (ω_{j1} is the frequency corresponding to the $|j\rangle \longrightarrow |1\rangle$ transition). We express the radiative relaxation operator $\hat{R}$ in the form [18]

$$\begin{aligned}\hat{R} = \sum_{m,j=2,3} \frac{\hbar}{2}[(\Gamma_{jm} + \Gamma_{mj}^+)S_m^-\rho S_j^+ \\ - \Gamma_{jm}\rho S_m^+ S_j^- - S_j^+ S_m^- \rho \Gamma_{mj}^+],\end{aligned} \tag{4}$$

where for $j = m$ we have $\Gamma_{jj} = \gamma_j + i\Delta_j^L$, and γ_j and Δ_j^L determine the radiative relaxation rate and the Lamb shift corresponding to the transition $|j\rangle \longrightarrow |1\rangle$. The first term in expression (4) describes the radiative influx to the ground state, and the next two terms describe radiative damping and the level shift. A distinguishing feature of the last terms is that they are off-diagonal. We postulate that the excited states are adjacent, $\delta = \omega_2 - \omega_3 \ll \omega_2$ whereas for Γ_{jm} where $j \neq m$, we have

$$\Gamma_{jm} = p[\sqrt{\gamma_j\gamma_m} + i\alpha|\Delta_j^L\Delta_m^L|^{1/2}], \tag{5}$$

where $\alpha = +1$ or -1 depending on the specific properties of the atomic system. The factor *p* determines the correlation between the interference contributions of differently polarized photons. In [7–12] *p* was assumed to be unity, whereas in descriptions of the Raman scattering spectra in [13] *p* varied between 0 and 1 depending on the direction of observation of the scattered quanta. It should be stressed that this approach is incorrect: the rate of the relaxation processes of the atomic system does not depend on the direction of observation of the scattered radiation. In the present paper we investigate a model three-level system neglecting any degeneracy of the magnetic moment. Thus, *p* will be taken as a model parameter. Expressions for the elements of the radiative relaxation operator Γ_{js} for real systems are given in the Appendix. It should be stressed that for most real systems the parameter is $p = 0$ because of the phase quenching of the contributions from photons of different polarizations [see (A.2)].

However, in some cases, for example when the density of states of the electromagnetic field varies, the quenching is destroyed and the value of *p* will be nonzero. In this case, we need to bear in mind the change in the diagonal elements Γ_{jj}. Since the radiative relaxation constants will also be taken to be model parameters in the following investigation, we shall assume that the interference change in the diagonal elements is already taken into account. The proposed model may be used for a qualitative description of the spontaneous luminescence on the *D*-line of alkali metal vapor in the field of a linearly polarized laser wave near the metal surface. A detailed analysis of the resonance fluorescence of the sodium *D*-line in the field of a high-intensity electromagnetic wave in a formulation of the problem allowing for the complete angular momentum of the states requires a separate analysis.

From expressions (1)–(4) we obtain equations for the components of the atomic density matrix $\rho_{jm} = \langle j|\rho|m\rangle$:

$$\begin{aligned}i\dot{\rho}_{jm} &= (\omega_{j1} - \omega_{m1})\rho_{jm} \\ &- \frac{i}{2}\sum_{s=2,3}(\Gamma_{js}\rho_{sm} + \rho_{js}\Gamma_{sm}) + V_{j1}\rho_{1m} - \rho_{j1}V_{1m}, \\ i\dot{\rho}_{j1} &= (\omega_{j1} - \omega_L)\rho_{j1} \\ &- \frac{i}{2}\sum_{s=2,3}\Gamma_{js}\rho_{s1} + V_{j1}\rho_{11} - \sum_{r=2,3}\rho_{jr}V_{rj}, \\ \rho_{1j} &= \rho_{j1}^*, \quad j, m = 2, 3,\end{aligned} \tag{6}$$

subject to the conditions $\rho_{11} + \rho_{22} + \rho_{33} = 1$. It follows from the equations (6) that the total intensity of the scattered radiation is given by

$$I_{\text{rad}} \propto \hbar\omega(\gamma_2\rho_{22} + \gamma_3\rho_{33} + 2\Gamma'_{32}\text{Re}\,\rho_{23}). \tag{7}$$

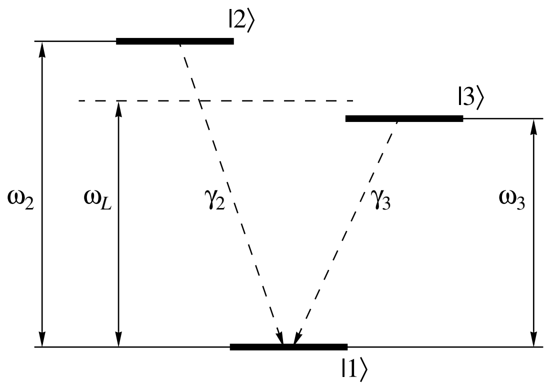


Fig. 1. Diagram of a *V*-type three-level atom.

Here we assume that $\Gamma_{jm} = \Gamma'_{jm} + i\Gamma''_{jm}$. Note that expression (7) derives from the off-diagonal structure of the relaxation operator (4). In the standard description [2, 18] the drift terms have a diagonal structure:

$$R^{(1)}_{jk} = \hbar(\gamma_j + \gamma_k)\rho_{jk}/2 \tag{8}$$

and correspond to the usual expression for the total scattered radiation intensity:

$$I_{\mathrm{rad}} \propto \hbar\omega(\gamma_2\rho_{22} + \gamma_3\rho_{33}) \tag{9}$$

[here and in (7) $\omega \approx \omega_{32} \approx \omega_{21}$]. We shall subsequently normalize the total intensity of the spontaneous emission I_{rad} to $I_0 \propto \hbar\omega\gamma_2$ which is proportional to the total luminous intensity of atoms excited only to the second state. We checked the accuracy of equations (4) and (6) using the Keldysh diagram technique for the nonequilibrium Green functions [19] and also the atom–photon density matrix [20–22] using an approach developed in [23]. Note that some aspects associated with describing the spectral relaxation characteristics of the excited states were taken into account in [11, 12] using the technique of probability amplitudes.

3. INVESTIGATION OF THE DYNAMICS OF ATOMIC STATES AND THE TOTAL RESONANCE FLUORESCENCE INTENSITY

In this section we consider the influence of quantum interference effects on the total spontaneous emission intensity of the atoms and the populations of the upper levels. As we noted in the previous section, the off-diagonal nature of the radiative relaxation operator leads to the appearance of an additional term $2\Gamma'_{32}\,\mathrm{Re}\rho_{23}$ in the standard expression for the total intensity and allowance for this term may be significant.

Figures 2 and 3 give the total intensity of the resonance fluorescence as a function of the laser field strength for various relationships between the radiative relaxation times and various intensities of the interference effects (various values of the parameter *p*). It can be seen from these graphs that interference effects may have a substantial influence on the emission properties of the atom, which for specific values of the parameters may lead to a sharp drop (as far as zero) in the intensity (see Figs. 2b and 3b).

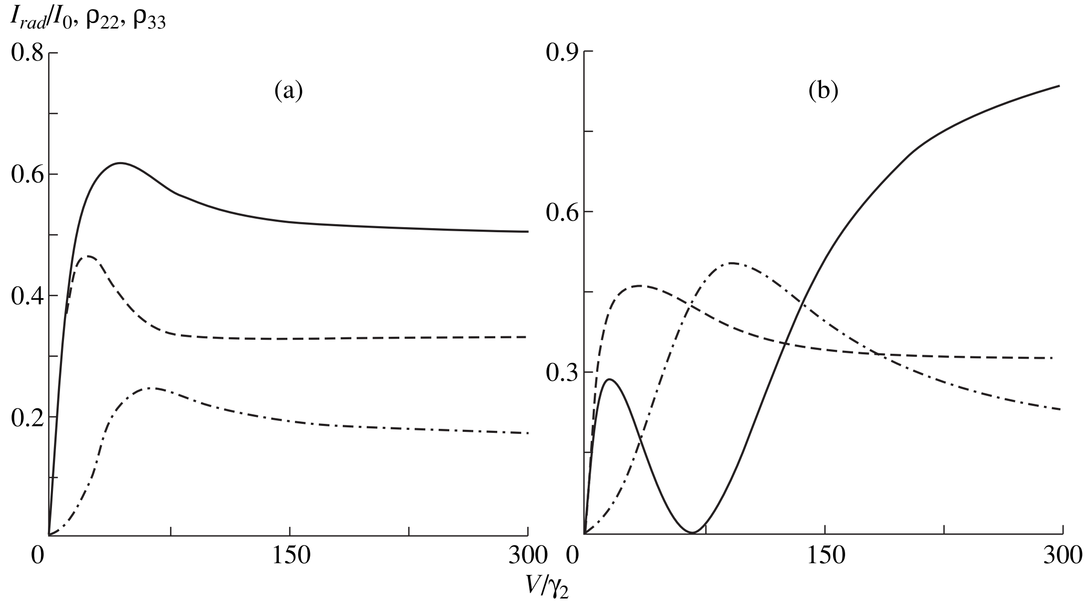


Fig. 2. Dependences of the upper level populations and the total intensity of the resonance fluorescence on the laser field strength for the parameters $\gamma_2 = \gamma_3$, $\Delta_2/\gamma_2 = 10$, $\Delta_3/\gamma_2 = -60$, $V = V_{21} = \sqrt{2}\,V_{31}$, $I_0 = \hbar\omega\gamma_2$, and the Lamb shift $\Delta^L_j = 0$, $p = 0$ (a), $p = 1$ (b): ρ_{22}—dashed curve, ρ_{33}—dot-dash curve, I_{rad}—solid curve.

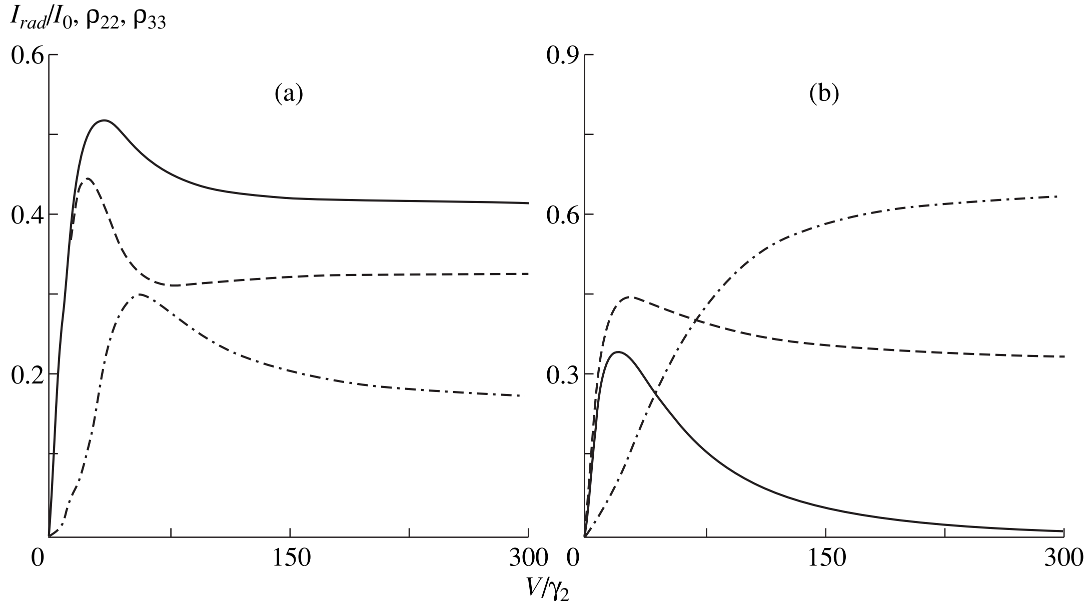


Fig. 3. Dependences of the upper level populations and the total intensity of the resonance fluorescence on the laser field strength for the parameters $\gamma_3/\gamma_2 = 0.5$ (the values of the other parameters and the notation are the same as in Fig. 2).

We can identify several main types of behavior of the total resonance fluorescence intensity as a function of the pump wave intensity.

When the relationship $\gamma_2/\gamma_3 > (\mu_{21}/\mu_{31})^2$ is satisfied, the dependence of the fluorescence intensity on the wave intensity does not change drastically as the contribution of the interference effects increases, because of the slight changes in the populations of the upper levels and the value of $\mathrm{Re}\rho_{23}$. As a result, we observe a linear dependence of the intensity on the parameter p (see expressions (7) and (5)) whose behavior for a given laser field strength is determined by the magnitude and sign of $\mathrm{Re}\rho_{23}$.

We may observe similar behavior for the case $\gamma_2/\gamma_3 < (\mu_{21}/\mu_{31})^2$ at high ($V_{j1} \gg \Delta_{j1}$) or low ($V_{j1} \ll \Delta_{j1}$) pump wave intensities. However, for $V_{j1} \sim \Delta_{j1}$ the influence of quantum interference effects may lead to substantial changes in the upper level populations and the coherence between them (see Fig. 4) which gives rise to a sharp drop in the luminescence of the atoms in this range of parameters. Formally for $p = 1$ the total intensity of the resonance fluorescence may even have zero values (see Fig. 2b). The reduction in the total intensity of the resonance fluorescence is observed particularly clearly under the condition $\gamma_2/\gamma_3 = (\mu_{21}/\mu_{31})^2$ and $p = 1$ (see Fig. 3b). This case should be an isolated one because, despite an increase in the laser field and total population inversion $\rho_{11} < \rho_{22} < \rho_{33}$, the intensity of the atomic spontaneous emission decreases.

We also investigated the dependences of the populations and the total intensity of the resonance fluorescence on the Lamb component of the off-diagonal relaxation constant Γ''_{jm}. We established that in quantum interference effects the Lamb shift has little influence on the populations and the total intensity. However, as will be shown later, the Lamb shift can substantially influence the form of the spectra in regions where the intensity of the atomic spontaneous emission is low.

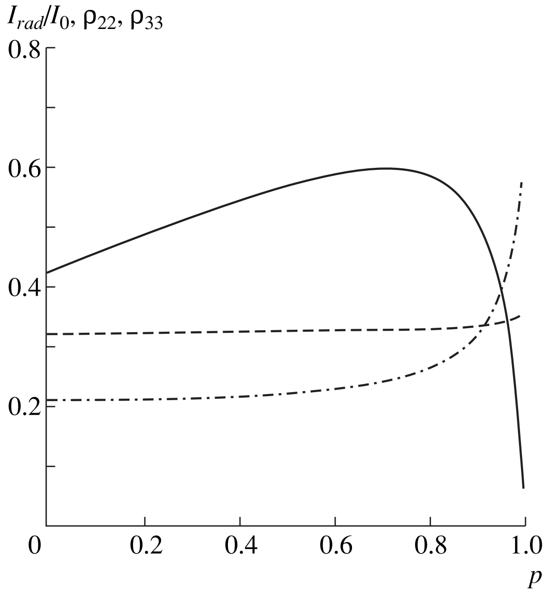


Fig. 4. Dependences of the upper level populations and the total intensity of the resonance fluorescence on the parameter p for $\gamma_2 = \gamma_3$, $V = V_{21} = 140$, $p = 1$ (the values of the other parameters and the notation as the same as in Fig. 2).

4. INVESTIGATION OF THE SPECTRA

In this section we investigate the resonance fluorescence spectra taking into account interference effects and the angular distribution of the spontaneous emission. In order to construct the spectra we use the technique of the atom–photon density matrix [20–23]. Following the general procedure, we add the equation for the lower state ρ_{11}

$$\frac{d\rho_{11}}{dt} = \frac{i}{2}\sum_{j,\,s\,=\,2,\,3}(\Gamma_{js}\rho_{sj} + \rho_{js}\Gamma_{js}) + \sum_{j\,=\,2,\,3}(V_{1j}\rho_{j1} - \rho_{1j}V_{j1}) \tag{10}$$

to the system (6) and rewrite it in the following form:

$$i\frac{d\bar{\rho}}{dt} = M\bar{\rho}, \tag{11}$$

where

$$\bar{\rho} = (\rho_{11}, \rho_{22}, \rho_{33}, \rho_{21}, \rho_{12}, \rho_{31}, \rho_{13}, \rho_{32}, \rho_{23})^T,$$

and the matrix M is determined by the system of equations (6) and (10).

The system of equations for the atom–photon density matrix $\bar{\rho}^\nu$ has the form

$$i\frac{d\bar{\rho}^\nu}{dt} = (M - \nu)\bar{\rho}^\nu + g_2^*\bar{F}_2 + g_3^*\bar{F}_3. \tag{12}$$

Here the vector

$$\bar{\rho}^\nu = (\rho_{11}^\nu, \rho_{22}^\nu, \rho_{33}^\nu, \rho_{21}^\nu, \rho_{12}^\nu, \rho_{31}^\nu, \rho_{13}^\nu, \rho_{32}^\nu, \rho_{23}^\nu)^T,$$

$$\rho_{js}^\nu = \langle j|\bar{\rho}^\nu|s\rangle$$

are the projections of the atom–photon density matrix on the atomic variables, $\nu = \omega - \omega_L$, ω is the frequency of a scattered quantum), g_j is the coupling constant of the scattered radiation with the transition $|j\rangle \longrightarrow |1\rangle$:

$$g_j = \mu_{1j}\left(\frac{2\pi\omega_L}{\hbar W}\right)^{1/2}\left(1 + O\left(\frac{\nu}{\omega_L}\right)\right)$$

(W is the quantization volume). The expression for the coupling constant g_j for real physical systems is given in the Appendix (see (A.1)). The atomic density matrix is given by

$$\rho = \mathrm{Sp}(\rho^\nu)_{ph},$$

where the subscript ph denotes the photon variables. The vectors $\bar{F}_2$ and $\bar{F}_3$ are expressed in terms of the elements of the atomic density matrix as follows:

$$\bar{F}_2 = (\rho_{12}, 0, 0, \rho_{22}, 0, \rho_{32}, 0, 0, 0)^T,$$

$$\bar{F}_3 = (\rho_{13}, 0, 0, \rho_{23}, 0, \rho_{33}, 0, 0, 0)^T.$$

Following [15], the coupling constants g_j can be related to the elements of the radiative relaxation operator and the angular distribution function of the spontaneous emission as follows:

$$\Gamma_{jk} = \int g_j g_k^* a_{jl} d\mathbf{k} = \int \bar{\Gamma}_{jk} dO = \int \Gamma_{jk}\Phi_{jk}(\phi, \theta) dO. \tag{13}$$

Here $\mathbf{k}$ is the wave vector of the vacuum mode, a_{jk} is the complex line profile, dO is an element of the solid angle, and the quantity $\Phi_{jk}(\phi, \theta)\sin(\theta)d\theta d\phi$ determines the resonance fluorescence intensity in the angle $d\phi d\theta$ in spherical coordinates and possesses the normalization

$$\int \Phi_{jk}(\phi, \theta) dO = 1.$$

Generally in this model the angular distribution function of the spontaneous emission $\Phi_{jk}(\phi, \theta)$ cannot be obtained from expression (13). In order to determine this, we need to make a complete analysis which takes into account the magnetic sublevels and the polarization composition of the spontaneous emission and the pump wave. The expression for the function $\Phi_{jk}(\phi, \theta)$ obtained in this analysis is given in the Appendix [see (A.3)]. In this model the function $\Phi_{jk}(\phi, \theta)$ is introduced phenomenologically.

The general expression for the fluorescence spectrum for observations in a certain direction has the form

$$A(\nu) d\nu dO = \mathrm{Im}\left(\sum_{j\,=\,2,\,3} g_j \rho_{j1}^\nu\right)\frac{\nu^2 d\nu}{c^3} dO. \tag{14}$$

Using (13), we can reduce this to the simpler form:

$$A(\nu) = \mathrm{Im}\left(\sum_{j,\,k\,=\,2,\,3} \bar{\Gamma}_{jk}(M^{-1}\bar{F}_j)_{k1}\right), \tag{15}$$

where $(M^{-1}\bar{F}_j)_{k1}$ are the elements of the vector $M^{-1}\bar{F}_j$ whose positions correspond to those of the elements ρ_{k1} of the vector $\bar{\rho}$. The explicit form of $\bar{\Gamma}_{jk}$ depends on the geometry of the problem. For most problems where we are not interested in the angular distribution of the scattered radiation, we assume $\Phi_{jk}(\phi, \theta) = 1$ and $\bar{\Gamma}_{jk} = \gamma_j\delta_{jk}$ (δ_{jk} is the Kronecker delta). In the present paper we shall assume that the system has a preferred axis and then, using expression (A.3), we can show that the values of $\bar{\Gamma}_{jk}$ have the form

$$\bar{\Gamma}_{jk} = \Gamma_{jk}(c_{jk} + b_{jk}\cos(2\theta)). \tag{16}$$

The first term in (16) describes the contribution of the standard spontaneous emission and the second describes the polarization effects whose contribution vanishes when integration is performed over angles. In this case, the expression for the spectrum has the form

$$A(\nu) = \operatorname{Im}\left(\sum_{j,k=2,3}\Gamma_{ik}(c_{jk}+b_{jk}\cos(2\theta))(M^{-1}F_j)_{k1}\right). \quad (17)$$

In order to interpret the spectra of a three-level atom, formula (17) can be transformed to give

$$A(\nu) = \sum_{n=1}^{9}\frac{f_n}{\nu-\Omega_n+i\gamma_n}+\text{c.c.}, \quad (18)$$

where $\Omega_n - i\gamma_n$ are the eigenvalues of the matrix M, and f_n is a complex quantity which determines the weight of the corresponding term. The values of Ω_n have threefold degeneracy near zero while the remaining six ($\pm\Omega_1$, $\pm\Omega_2$, $\pm\Omega_3$) are symmetric relative to the ν axis. These are the Rabi frequencies of the three-level system and describes the high-frequency Stark effect. In addition, one of the eigenvalues of the matrix M is zero, i.e., for one of the degenerate $\Omega_n = 0$ and also for $\gamma_n = 0$, this corresponds to elastic Rayleigh scattering taking place without any change in frequency. Bearing this in mind and using the Sokhotskiĭ formula, we can isolate the unshifted Rayleigh scattering component in (18):

$$A^{el} = \sum_{j,k=2,3}\bar{\Gamma}_{jk}\rho_{1j}\rho_{k1}\delta(\nu). \quad (19)$$

Here $\delta(\nu)$ is the Dirac delta function. Note that in strong fields, and also in the presence of collisions, the contribution of the elastic component is small so that we shall neglect this and rewrite formula (18) in the form

$$A(\nu) \approx \sum_{j=1}^{8}\frac{\operatorname{Re}(f_n)(\nu-\Omega_n)+\operatorname{Im}(f_n)\gamma_n}{(\nu-\Omega_n)^2+\gamma_n^2}+A^{el}\delta(\nu). \quad (20)$$

In expression (20) the central peak with $\Omega_9 = 0$ is doubly degenerate and is the sum of two Lorenzians of different widths. Consequently, in the resonance fluorescence spectrum we predict that up to seven components will be present.

The dispersion term

$$\sum_{n=1}^{8}\frac{\operatorname{Re}(f_n)(\nu-\Omega_n)}{(\nu-\Omega_n)^2+\gamma_n^2}$$

may lead to asymmetry of the peaks and reduce their intensity to zero for some frequencies. This possibility was investigated in [11] for the spontaneous emission spectra. It can be seen from the graphs that the dispersion component only makes a significant contribution for $p = 1$ in regions where the total intensity of the resonance fluorescence is low, which can be seen from the slight asymmetry of the peaks (see Fig. 5c). Thus, the main changes to the spectrum are associated with the term

$$\sum_{n=1}^{8}\frac{\operatorname{Im}(f_n)\gamma_n}{(\nu-\Omega_n)^2+\gamma_n^2},$$

and since the value of Ω_j remains unchanged for constant E_L and δ_j, as p increases the main changes take place for $\operatorname{Im}(f_n)$ and γ_n.

It can be seen from the spectra that the changes in $\operatorname{Im}(f_n)$ and γ_n may not be the same for all the peaks. Moreover, this is difficult to investigate because of the large number of parameters which are generally interrelated for real systems (for example, the constants c_{jk}, b_{jk}, and p). However, some general dependences can be identified.

For $p = 0$ and $\Delta_j^L = 0$ the system has a symmetric spectrum regardless of the relationships between γ_j, Δ_j, and V_{j1} (see Fig. 5a). This agrees with the well-known result obtained by Mollow for the spectra of a two-level atom [26]. It should be noted that the asymmetry of the spectra obtained in [14] using correct overall methodology is attributed to the incomplete averaging over $g_j g_k^*$ (see (13) and (14)).

Another interesting result concerns the influence of quantum interference on the form of the spectrum. In most cases, as p increases the intensity of the outer lines decreases relative to the central ones while the symmetry of the spectrum is conserved (see Figs. 5b and 5c) (an exception is the region where the total intensity of the resonance fluorescence is almost zero where the dispersion term (20) makes an important contribution).

An investigation of the influence of the Lamb shift on the spectra shows that this does not lead to any significant changes at high spontaneous emission intensities. However, in regions of low total intensity (see Section 3) allowance for the Lamb shift may lead to substantial modification of the spectrum (see Fig. 5d).

Dependences of the spectra on the angular distribution function [see (13)–(17)] are plotted in Fig. 6. It can be seen from the figures that these characteristics may have a substantial influence on the form of the spectra and in certain directions, they may lead to changes in the total intensity of the resonance fluorescence and to degeneracy or, conversely, to strong amplification of isolated lines in the spectrum.

It should be reemphasized that this investigation is to a certain extent illustrative. The explicit form of the angular distribution $\Phi_{jk}(\phi, \theta)$ depends strongly on the geometry of the problem which, like the parameter p, depends on the quantum interference effects. Thus, a comprehensive analysis of real systems involves allowing for the degeneracy of the angular momentum and the relationships between the constants c_{jk}, b_{jk}, and p.

We have therefore shown that quantum interference processes accompanying the radiative relaxation of

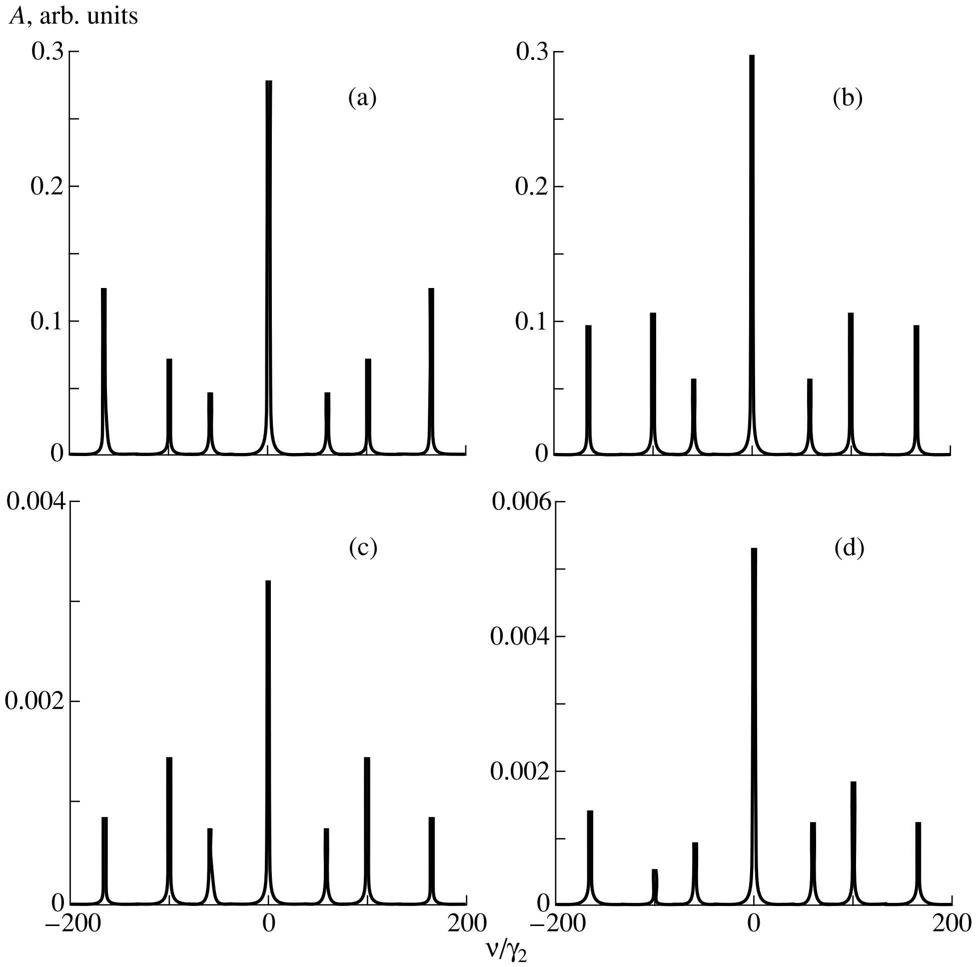


Fig. 5. Resonance fluorescence spectra for $\gamma_2 = \gamma_3$, $\Delta_2/\gamma_2 = 10$, $\Delta_3/\gamma_2 = -60$, $V = V_{21} = \sqrt{2}\,V_{31} = 60$, $b_{jk} = 0$, $a_{jk} = 1$: (a) $p = 0$, $\Delta_j^L/\gamma_2 = 0$; (b) $p = 0.5$, $\Delta_j^L/\gamma_2 = 0$; (c) $p = 1$, $\Delta_j^L/\gamma_2 = 0$; (d) $p = 1$, $\Delta_j^L/\gamma_2 = 3$.

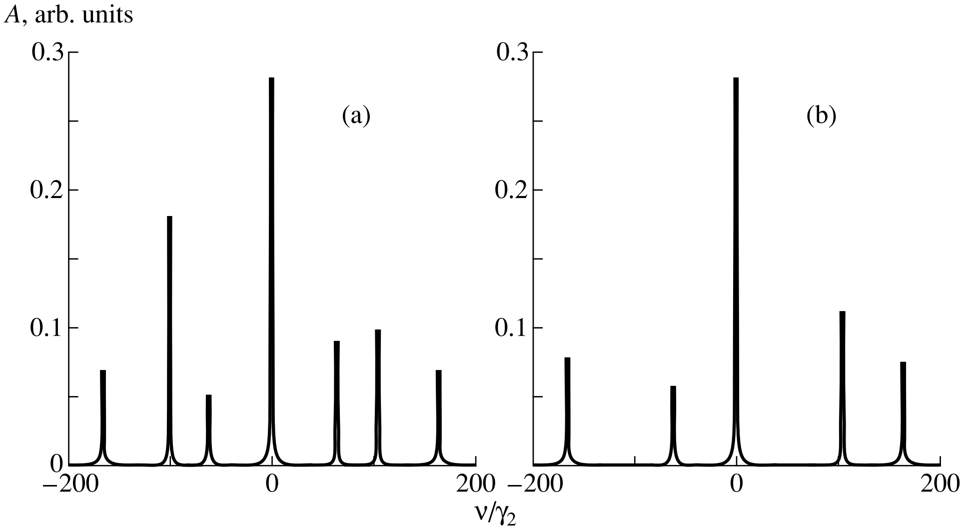


Fig. 6. Resonance fluorescence spectra for $p = 0.8$, $b_{jj} = 0.2$, $\theta = 0$, $\Delta_j^L = 0$, $a_{jk} = 1$, $b_{23} = b_{32} = 0$ (a), $b_{23} = b_{32} = 0.5$ (b) (the values of the other parameters are as in Fig. 5).

excited states may substantially alter the total intensity of the atomic spontaneous emission and its angular distribution, as well as the dynamics of the atomic level populations and the resonance fluorescence spectra. In addition, we have derived analytic expressions for the total intensity and the spectra of the resonance fluorescence taking into account the off-diagonal nature of the radiative relaxation operator and the angular distribution of the spontaneous emission. These results may be used for a qualitative description of the *D*-line spontaneous luminescence of alkali metal vapor in the field of a linearly polarized laser wave near a metal surface.

This work was partly financed by the Russian Foundation for Basic Research (Grants nos. 99-02-17063, 96-15-96446, and 99-02-18176).

APPENDIX

Structure of Interference Terms in the Radiative Relaxation Operator for Real Atomic Systems

Expression (5) for the radiative relaxation operator allows for the influence of interference effects in the spontaneous emission of atoms but it describes a model nondegenerate atomic subsystem. We shall consider the role of these effects for a real system allowing for angular momentum degeneracy. The radiative lifetime, or the rate of radiative relaxation, of the excited state of an atomic system is determined using the standard Wigner–Weisskopf procedure using second-order perturbation theory in terms of the coupling constant in the resonance approximation. We assume that adjacent excited states have the angular momentum J_2 and J_3 and then the explicit expression for the coupling constant may be given in the form [15]

$$g_{j1}^{\mathbf{k}\sigma}(M_j, M_1) = i(-1)^{\sigma} C_{J_j M_j J_1 - M_1}^{1-\sigma} \times \frac{\|\mu_{j1}\|}{\sqrt{3}} U_{k\sigma}(\mathbf{r}) \sqrt{\frac{2\pi\omega_k}{\hbar W}}. \tag{A.1}$$

Using the analog of the general expression (13) for the elements of the radiative relaxation operator, we have

$$\begin{aligned}\Gamma_{js}(M_j M_s) &= \sum_{M_1} \int g_{j1}^{\mathbf{k}\sigma}(M_j, M_1) g_{1s}^{\mathbf{k}\sigma'}(M_1, M_s) a_{js} \frac{d\mathbf{k}}{(2\pi)^3} \\ &= \frac{2\pi\|\mu_{j1}\|\|\mu_{s1}\|}{\hbar\sqrt{(2J_j+1)(2J_s+1)}} \sum_{M_1, \lambda, \sigma, \sigma'} C_{J_1 M_1 1\sigma}^{J_j M_j} C_{J_1 M_1 1\sigma'}^{J_s M_s} \\ &\quad \times \int U_{\mathbf{k}\sigma} U_{\mathbf{k}\sigma'} D_{\lambda\sigma'}^{1} (D_{\lambda\sigma}^{1})^* \omega a_{js} \frac{d\mathbf{k}}{(2\pi)^3}.\end{aligned} \tag{A.2}$$

In (A.1) and (A.2) $C_{J_1 M_1 1\sigma}^{J_j M_j}$, $C_{J_1 M_1 1\sigma}^{J_s M_s}$, and $C_{J_j M_j J_1 - M_1}^{1-\sigma}$ are the Clebsch–Gordan coefficients, $U_{\mathbf{k}\sigma}(\mathbf{r})$ is the spatial mode factor of σ-polarized photons having the wave vector $\mathbf{k}$ (λ is the polarization index in rotated Eulerian coordinates), $D_{\lambda\sigma}^{1}$ is the Wigner function, and $a_{js} = a_{js}(\omega, \sigma, \sigma')$ is the complex line profile.

Expression (A.2) can be used to isolate all the main spontaneous emission characteristics and is obtained in energy balance with the equation for the photon field dynamics (see, for example, [15]). This expression can be used to describe the angular structure of the scattered radiation [24] and polarization effects. Using (A.2) we can write the angular distribution function of the spontaneous emission

$$\begin{aligned}\Phi_{js}^{M_j\sigma M_s\sigma'}(\phi, \theta) &= \sum_{M_1, \lambda} C_{J_1 M_1 1\sigma}^{J_j M_j} C_{J_1 M_1 1\sigma'}^{J_s M_s} U_{\mathbf{k}\sigma} U_{\mathbf{k}\sigma'} \\ &\quad \times D_{\lambda\sigma'}^{1}(\phi, \theta, 0) (D_{\lambda\sigma}^{1}(\phi, \theta, 0))^*.\end{aligned} \tag{A.3}$$

When analyzing (A.2), we need to bear in mind the following properties of the sums [25]:

$$\sum_{M_1} C_{J_1 M_1 1\sigma}^{J_j M_j} C_{J_1 M_1 1\sigma'}^{J_s M_s} = \begin{cases} 1, & \sigma \neq \sigma', \\ \delta_{J_j, J_s}, & \sigma = \sigma', \end{cases} \tag{A.4}$$

and also the orthogonal property of the Wigner functions:

$$\int D_{\lambda\sigma'}^{1} (D_{\lambda\sigma}^{1})^* dO = \delta_{\sigma'\sigma}. \tag{A.5}$$

It can be seen from (A.4) and (A.5) that for most real atomic systems no interference damping occurs, i.e., $\Gamma_{js} = 0$, where $j \neq s$. The only exceptions are adjacent states having the same values of the angular momentum. Our analysis for atomic systems, including hyperfine splitting, confirms this result. However, we note that in many cases where $J_2 \neq J_3$ the coefficient p in (5) may be nonzero.

(1) If the atoms are located in a high-Q resonator which changes the density of states of the electromagnetic field and the resonator axis defines a specific direction (quantization axis). The density of states for photons of different polarization may differ substantially and in the expression (A.2) interference quenching disappears. Naturally, the Parcel effect must be taken into account here, i.e., variation of the ordinary rates of radiative relaxation γ_j in the resonator.

(2) If the atoms are in a highly dispersive medium having different values of $\varepsilon(\omega, \sigma)$. In this case, the value of Γ_{23} is a function of the projection M of the angular momenta. Then, for example, for sodium *D*-lines excited by linearly polarized radiation we have for $\Gamma_{23}(1/2, 1/2)$

$$p = \frac{\sqrt{2}}{3}(\phi(0) - \phi(+1)) \tag{A.6}$$

and for $\Gamma_{23}(-1/2, -1/2)$

$$p = \frac{\sqrt{2}}{3}(\phi(-1) - \phi(0)). \qquad (A.7)$$

The quantity $\phi(\sigma)$ is determined by the convolution

$$\phi(\sigma) = \int \varepsilon(\omega, \sigma) a_{23}(\omega, \sigma) \frac{d\omega}{2\pi}, \qquad (A.8)$$

where $\sigma = 0, \pm 1$. If this transition makes a significant contribution to the permittivity of the medium, its value should be calculated self-consistently. If this contribution is small and $\varepsilon(\omega, \sigma)$ is determined by a different type of particle in the medium and varies only slightly within the line profile of these transitions, $\varepsilon(\omega_0, \sigma)$ can be removed from the integral in (A.6).

3. If the atoms are in a strong electromagnetic field where nonlinear dynamic relaxation effects become substantial and there is an appreciable difference between the line profiles for photons of different polarization. In this case, we have the following estimate for the parameter p:

$$p \propto \frac{V}{\omega_0}. \qquad (A.9)$$

Here we have $V \sim V_{j1}$. We note that in all three cases discussed above the parameter p may be either positive or negative. The specific sign of p depends on the physical conditions.

REFERENCES

1. E. B. Aleksandrov, G. I. Khvostenko, and M. P. Chaĭka, *Interference of Atomic States* (Nauka, Moscow, 1991).
2. S. G. Rautian, G. I. Smirnov, and A. I. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectra* (Nauka, Novosbirsk, 1990).
3. S. Stenholm, *Foundations of Laser Spectroscopy* (Mir, Moscow, 1987; Wiley, New York, 1984).
4. S. G. Rautian, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 461 (1995).
5. F. A. Lomaya and A. A. Panteleev, Zh. Éksp. Teor. Fiz. **103**, 1970 (1993).
6. E. G. Pestov, Tr. Fiz. Inst. Akad. Nauk SSSR **187**, 60 (1988).
7. D. Agassi, Phys. Rev. A **30**, 2449 (1984).
8. E. A. Manykin and A. M. Afanas'ev, Zh. Éksp. Teor. Fiz. **48**, 931 (1965).
9. E. A. Manykin and A. M. Afanas'ev, Zh. Éksp. Teor. Fiz. **52**, 1246 (1967).
10. Ce Chen, Yi-Yian Yin, and D. S. Elliot, Phys. Rev. Lett. **64**, 507 (1990).
11. S. Y. Zhu, R. C. F. Chan, and C. P. Lee, Phys. Rev. A **52**, 710 (1995).
12. A. E. Devdariani, V. N. Ostrovskiĭ, and Yu. N. Sebyakin, Zh. Éksp. Teor. Fiz. **71**, 909 (1976).
13. S. Y. Zhu and M. O. Scully, Phys. Rev. Lett. **76**, 388 (1996).
14. A. G. Leonov, A. A. Panteleev, A. N. Starostin, *et al.*, Zh. Éksp. Teor. Fiz. **105**, 1536 (1994).
15. A. A. Panteleev, Zh. Éksp. Teor. Fiz. **111**, 440 (1997).
16. V. Savchenko, A. A. Panteleev, and A. N. Starostin, in *Proc. 13th Intern. Conf. of Spectral Line Change* (Firenze, 1996), p. 132.
17. V. I. Savchenko, N. J. Fisch, A. A. Panteleev, *et al.*, Phys. Rev. A **59**, 708 (1999).
18. G. S. Agarwal, in *Progress in Optics*, Ed. E. Wolf, (North-Holland, Amsterdam, 1973), Vol. 11, p. 2.
19. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964).
20. M. O. Scully and W. E. Lamb, Jr., Phys. Rev. **159**, 208 (1967).
21. S. Stenholm, D. A. Holm, and M. Sargent, III, Phys. Rev. A **31**, 3124 (1985).
22. M. Sargent, III, D. A. Holm, and M. S. Zubary, Phys. Rev. A **31**, 3112 (1985).
23. A. A. Panteleev and A. N. Starostin, Zh. Éksp. Teor. Fiz. **106**, 1606 (1994).
24. L. D. Landau and E. M. Lifshits, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1984; Nauka, Moscow, 1982).
25. D. A. Varshalovich, A. P. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988; Nauka, Leningrad, 1975).
26. B. R. Mollow, Phys. Rev. **188**, 1969 (1969).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 59–65.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 67–74.
Original Russian Text Copyright © 2000 by Karshenboĭm, Ivanov, Shabaev.

ATOMS, SPECTRA, RADIATION

Polarization of Vacuum in a Hydrogen-Like Relativistic Atom: Hyperfine Structure

S. G. Karshenboĭm*, V. G. Ivanov, and V. M. Shabaev*****
**State Science Center, Mendeleev All-Russia Research Institute of Metrology,*
St. Petersburg, 198005 Russia;
e-mail: ksg@hm.csa.ru
***Main Astronomical Observatory, Russian Academy of Sciences,*
St. Petersburg, 196140 Russia
****St. Petersburg State University,*
St. Petersburg, 198904 Russia

Received July 19, 1999

Abstract—An analysis is made of the contribution of the polarization of vacuum to the hyperfine splitting of the ground state of a hydrogen-like atom. An expression for the correction to the energy is obtained as an explicit function of the parameter $Z\alpha$. The final expression derived in terms of generalized hypergeometric functions and their derivatives is a function of the particle mass ratio in orbit and in a vacuum loop, and is therefore valid for both ordinary and muonic atoms. Various asymptotic forms are also given. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Vacuum polarization effects influence the energy levels in ordinary (electronic) and muonic atoms. The most important is the Lamb shift which occurs in muonic atoms in the principal approximation as a result of these effects (Fig. 1a). An analytic calculation of this correction was presented in [1]. The present paper is concerned with the contribution to the hyperfine splitting of the ground state of a hydrogen-like atom and is a direct continuation of [1].

Calculations of the leading polarization correction to the hyperfine splitting are more complex than those for the case of the Lamb shift. Contributions are made by the single-potential (Fig. 1b) and the double-potential diagrams (Fig. 1c):

$$\Delta E_{hfs} = E_{TU} + E_{UT}. \qquad (1)$$

The single-potential contribution is exactly the same as the contribution to the Lamb shift analyzed in [1, 2] whereas the contribution of the diagrams in Fig. 1c containing the Uehling potential and the magnetic interaction, with the Coulomb Green function connecting them is far more complex [3]. Nevertheless it can be calculated in closed analytic form.

We recall the main results for the hyperfine splitting of the ground state of hydrogen-like electronic and muonic atoms. In the electronic case, the first terms of the expansion in terms of the small parameter $Z\alpha$ are known [4–6] while numerical calculations are available for high values of the nuclear charge [7, 8].

A nonrelativistic result was recently published for a muonic atom [9] where the single-potential contribution was obtained in analytic form and the double-potential contribution was obtained as a single integral.

It should be noted that all these calculations, apart from the numerical ones [7, 8], were made for a point nucleus. Obviously charge distribution and magnetic moment effects become important as Z increases. However, it should be noted that similar calculations can also be made for more highly excited levels for which these effects are not so important. For electronic atoms the most interesting are the 1*s* and 2*s* levels for which we found new terms of the expansion for low values of the parameter $Z\alpha$ relevant to muonium and hydrogen. For muonic atoms the calculations can be made for higher levels since hyperfine structure effects in muonic atoms are more important than in ordinary atoms.

The present study is organized as follows: we first discuss a general expression for the single-potential contribution and then one for the double-potential contribution. We then analyze various asymptotic forms. Finally we discuss the results and compare them with the numerical calculations.

In the present study we use exactly the same notation as in [1] and, in particular

$$\varepsilon = 1 - \sqrt{1-(Z\alpha)^2}, \quad \kappa = Z\alpha m/m_l,$$

where m_l is the particle mass in the loop, m is the mass of a bound particle, and $\hbar = c = 1$.

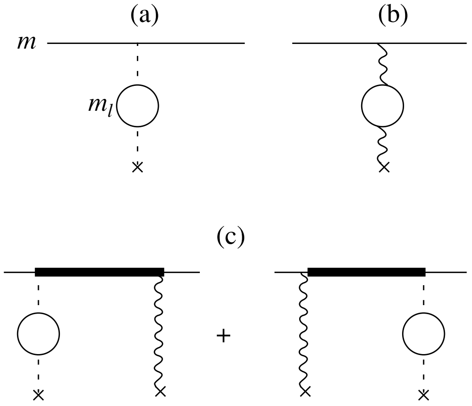


Fig. 1. Contribution of the vacuum polarization to the energy. The bound particle has the mass m and the particle in the polarization loop has the mass m_l. The thick line gives the reduced Green function of a particle in the Coulomb field of the nucleus.

2. GENERAL EXPRESSION

2.1. Single-Potential Contribution

It is convenient to begin our calculations of the single-potential contribution (Fig. 1b) with a substitution in the momentum representation, which is equivalent to inserting the free polarization of vacuum (cf. (4) in [1]):

$$\varepsilon_{abc}\frac{q_b\mu_c}{\mathbf{q}^2}\longrightarrow\frac{\alpha}{\pi}\int_0^1 dv\frac{v^2(1-v^2/3)}{1-v^2}\varepsilon_{abc}\frac{q_b\mu_c}{\mathbf{q}^2+\lambda^2}, \tag{2}$$

where μ is the magnetic moment of the nucleus,

$$\lambda(v) = \frac{2m_l}{\sqrt{1-v^2}}, \tag{3}$$

and a loop particle has the mass m_l. In the coordinate representation the substitution has the form

$$\varepsilon_{abc}\nabla_b\mu_c\frac{1}{r}\longrightarrow\frac{\alpha}{\pi}\int_0^1 dv\frac{v^2(1-v^2/3)}{1-v^2}\varepsilon_{abc}\nabla_b\mu_c\frac{e^{-\lambda r}}{r}. \tag{4}$$

The single-potential contribution is determined by

$$E_{TU} = \frac{\alpha}{\pi}E_T R_{TU}(\kappa). \tag{5}$$

Here E_T is the matrix element of the magnetic exchange without any radiation correction, which was obtained for the ground state in [10]

$$E_T(1s) = \frac{E_F}{2[1-(Z\alpha)^2]-\sqrt{1-(Z\alpha)^2}}, \tag{6}$$

where the Fermi energy E_F corresponds to the nonrelativistic interaction of two magnetic moments:

$$E_F = \frac{4}{3}\alpha(Z\alpha)^3\frac{\mu}{\mu_N}\frac{m}{m_p}\frac{2I+1}{2I}m, \tag{7}$$

μ_N is the nuclear magneton, I is the nuclear spin, and m_p is the proton mass.

In general, the dipole magnetic interaction contains spin–spin and spin–orbit terms. For s-states the second term is obviously zero which appreciably simplifies the expression for the correction

$$\mathcal{R}_{TU}(\lambda) = \frac{\int_0^\infty drr^2 f(r)g(r)(\partial/\partial r)(e^{-\lambda r}/r)}{\int_0^\infty drr^2 f(r)g(r)(\partial/\partial r)(1/r)}. \tag{8}$$

The radial integral is easily taken

$$\begin{aligned}\mathcal{R}_{TU}(\lambda) &= \frac{\int_0^\infty drr^2 e^{-2\gamma r}r^{-2\varepsilon}(\partial/\partial r)(e^{-\lambda r}/r)}{\int_0^\infty drr^2 e^{-2\gamma r}r^{-2\varepsilon}(\partial/\partial r)(1/r)} \\ &= \left(\frac{2\gamma}{2\gamma+\lambda}\right)^{1-2\varepsilon}\left[1+(1-2\varepsilon)\frac{\lambda}{2\gamma+\lambda}\right],\end{aligned} \tag{9}$$

and the ensuing single integral

$$\begin{aligned}R_{TU}(\kappa) &= \int_0^1 dv\frac{v^2(1-v^2/3)}{\sqrt{1-v^2}}\left(\frac{\kappa}{1+\kappa\sqrt{1-v^2}}\right) \\ &\times\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{-2\varepsilon}\left[1+(1-2\varepsilon)\frac{1}{1+\kappa\sqrt{1-v^2}}\right]\end{aligned} \tag{10}$$

is easily expressed in terms of the base integrals (see (16) in [1])

$$\begin{aligned}E_{TU}(1s) &= \frac{\alpha}{\pi}\frac{E_F}{2[1-(Z\alpha)^2]-\sqrt{1-(Z\alpha)^2}} \\ &\times\left\{\left(I_{121}-\frac{1}{3}I_{221}\right)+\frac{1-2\varepsilon}{\kappa}\left(I_{132}-\frac{1}{3}I_{232}\right)\right\}.\end{aligned} \tag{11}$$

The analyticity of the result is determined by the factor (6):

$$(Z\alpha) < \sqrt{3}/2, \text{ or } Z < 118.676\ldots. \tag{12}$$

2.2. Double-Potential Contribution

We shall now analyze the contribution of the diagrams in Fig. 1c. The difference between this contribution and the single-potential one consists in the presence of a Coulomb Green function connecting the potentials. However, the contribution may be rewritten in the form of some matrix element of the Uehling potential examined by us in [1] in terms of the wave function perturbed by the magnetic field. This function was obtained explicitly in [11]. Using the explicit expression for the perturbed wave function (see the Appendix) we obtain (cf. [1])

$$E_{UT} = \frac{\alpha}{\pi} E_T R_{UT}(\kappa), \tag{13}$$

where

$$R_{UT}(\kappa) = \int_0^1 dv \frac{v^2(1-v^2/3)}{1-v^2} \mathcal{R}_{UT}(\lambda(v)), \tag{14}$$

$$\mathcal{R}_{UT}(\lambda) = \frac{\int_0^\infty (f_{1s}X_{1s} + g_{1s}Y_{1s})[(Z\alpha)e^{-\lambda r}/r]dr}{\int_0^\infty f_{1s}g_{1s}dr}, \tag{15}$$

and explicit expressions for the radial components of the wave functions are given in the Appendix.

It is not difficult to calculate the radial integrals in (15) and thus obtain

$$\begin{aligned} R_{UT}(\kappa) &= \frac{3-8\varepsilon+8\varepsilon^2-2\varepsilon^3}{(1-\varepsilon)^2} J_{20}(\kappa) \\ &- \frac{2}{1-\varepsilon} J_{21}(\kappa) + \frac{2(\varepsilon+1)}{1-2\varepsilon} J_{20}(\kappa) - 2(1-\varepsilon) J_{30}(\kappa), \end{aligned} \tag{16}$$

where the J-integrals

$$\begin{aligned} J_{mn}(\kappa) &= \int_0^1 dv \frac{v^2(1-v^2/3)}{1-v^2} \\ &\times \left(\frac{\kappa y}{1+\kappa y}\right)^{m-2\varepsilon} \ln^n \frac{\kappa y}{1+\kappa y} \end{aligned} \tag{17}$$

are expressed in terms of the derivatives of the integrals I_{abc} (see (16) in [1]) [3]:

$$J_{mn} = \frac{\partial^n}{\partial \mu^n} \left[I_{12\mu} - \frac{1}{3} I_{22\mu} \right]_{\mu = m}. \tag{18}$$

As for the single-potential term, the analyticity of Z is determined by condition (12).

3. CORRECTION TO THE HYPERFINE SPLITTING FOR $Z\alpha \ll 1$

For large Z the finite dimensions of the nucleus play a significant role, so these formulas are most topical in the nonrelativistic limit ($Z\alpha \ll 1$). We shall begin our analysis with the single-potential contribution.

3.1. Single-Potential Contribution

In the Schrödinger approximation the result is known analytically [9]:

$$r_{TU}(\kappa) = \frac{2-\kappa^2+2\kappa^4}{4\kappa^3} \mathcal{A}(\kappa) - \frac{2}{3\kappa^3}\frac{\pi}{2} + \frac{6+\kappa^2}{9\kappa^2}, \tag{19}$$

where

$$\mathcal{A}(\kappa) = \frac{\arccos(\kappa)}{\sqrt{1-\kappa^2}} = \frac{\ln(\kappa+\sqrt{\kappa^2-1})}{\sqrt{\kappa^2-1}}, \tag{20}$$

and by analogy with expression (18) in [1] we determine

$$R_{TU}(\kappa) = r_{TU}(\kappa) - 2\varepsilon p_{TU}(\kappa) + \mathcal{O}(\varepsilon^2). \tag{21}$$

Further expansion in the nonrelativistic approximation yields the result for small κ:

$$\begin{aligned} R_{TU} &= \frac{3\pi}{8}\kappa - \frac{4}{5}\kappa^2 + \frac{5\pi}{24}\kappa^3 \\ &- \left(\frac{3\pi}{8}\ln\frac{\kappa}{2} + \frac{\pi}{32}\right)\kappa(Z\alpha)^2. \end{aligned} \tag{22}$$

Substituting the expanded normalization factor

$$E_T = E_F\left[1 + \frac{3}{2}(Z\alpha)^2\right], \tag{23}$$

for the electronic atom ($\kappa = Z\alpha$) we obtain the result

$$\begin{aligned} E_{TU} &= \frac{\alpha}{\pi} E_F \left\{ \frac{3\pi}{8}(Z\alpha) - \frac{4}{5}(Z\alpha)^2 \right. \\ &+ \left. \left[-\frac{3\pi}{8}\ln\frac{Z\alpha}{2} + \frac{71\pi}{96} \right](Z\alpha)^3 \right\}, \end{aligned} \tag{24}$$

which confirms all the previously known coefficients (see [4, 6]) and contains a single new one, denoted by C_{30}.

3.2. Double-Potential Contribution

Unlike the single-potential contribution, even in the leading nonrelativistic approximation it is impossible to obtain a fairly compact expression for the double-potential contribution. In particular, the integral J_{21} cannot be simplified substantially [9]:

$$\begin{aligned} R_{UT}(\kappa) &= \pi\frac{\kappa^2-2}{2\kappa^3} + \frac{5\kappa^4-8\kappa^2+6}{3\kappa^2(1-\kappa^2)} \\ &+ \frac{2\kappa^6-3\kappa^4+4\kappa^2-2}{\kappa^3(\kappa^2-1)} \mathcal{A}(\kappa) - 2J_{21}. \end{aligned} \tag{25}$$

In [9] asymptotic forms of this expression were also obtained for large and small values of the parameter κ.

A further nonrelativistic expansion for small κ yields the result

$$R_{UT}(\kappa) = \frac{3\pi}{8}\kappa + \left(\frac{214}{225} - \frac{8}{15}\ln(2\kappa)\right)\kappa^2 + \left(-\frac{43\pi}{288} + \frac{5\pi}{24}\ln\frac{\kappa}{2}\right)\kappa^3 + \left(\frac{23\pi}{32} - \frac{3\pi}{8}\ln\frac{\kappa}{2}\right)\kappa(Z\alpha)^2. \tag{26}$$

For electronic atoms, assuming κ = Zα we find [12]

$$R_{UT}(\kappa) = \frac{3\pi(Z\alpha)}{8} + \left(\frac{214}{225} - \frac{8}{15}\ln(2Z\alpha)\right)(Z\alpha)^2 + \left(\frac{41\pi}{72} - \frac{\pi}{6}\ln\frac{Z\alpha}{2}\right)(Z\alpha)^3 \tag{27}$$

or

$$E_{UT} = \frac{\alpha}{\pi}E_F\left\{\frac{3\pi Z\alpha}{8} + \left[\frac{214}{225} - \frac{8}{15}\ln(2Z\alpha)\right](Z\alpha)^2 + \left[\frac{163\pi}{144} - \frac{\pi}{6}\ln\frac{Z\alpha}{2}\right](Z\alpha)^3\right\}. \tag{28}$$

As in the single-potential case (24), this expression confirms all the expansion coefficients known so far [4, 6] and also gives the result for one new one (C_{30}).

4. HYPERFINE SPLITTING IN MUONIC ATOMS

4.1. Single-Potential Contribution

For nonrelativistic muonic atoms the asymptotic form for small Zα and large κ is more relevant:

$$R_{TU}(\kappa) = \left\{\frac{2}{3}\ln(2\kappa) + \frac{1}{9}\right\} + \frac{\pi^2}{9}(Z\alpha)^2 + \mathcal{O}((Z\alpha)^4) + \mathcal{O}\left(\frac{1}{\kappa^2}\right). \tag{29}$$

In the more general case we can obtain the expression

$$R_{TU}(\kappa) = I_{21} + (1-2\varepsilon)\frac{1}{\kappa}I_{32}, \tag{30}$$

where the integrals I_{bc} are determined in (32) in [1]. One of these (I_{21}) was obtained in [1] (see (37)) while the other (I_{32}) is known only partly:

$$I_{32} = I_{32}^{(1)} + I_{32}^{(2)}, \tag{31}$$

where (see (34) in [1])

$$I_{bc}^{(1)} = \frac{2}{3}\kappa^{b-2}B_{1-\delta}(2-b+c-2\varepsilon, b-2), \qquad \delta = \frac{1}{1+\kappa} \ll 1 \tag{32}$$

and

$$B_{1-\delta}(c-2\varepsilon-1, 1) = \frac{1}{c-2\varepsilon-1} - \frac{1}{\kappa} + \frac{c-2\varepsilon}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^2}\right). \tag{33}$$

The second term is given by

$$I_{3c}^{(2)} = -\left(\frac{\pi}{4} - \frac{2}{3}\right) + \frac{c-2\varepsilon}{6\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right), \tag{34}$$

and ultimately we obtain

$$\frac{1}{\kappa}I_{3c} = \frac{2}{3}\frac{1}{c-2\varepsilon-1} - \frac{\pi}{4}\frac{1}{\kappa} + \frac{c-2\varepsilon}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right). \tag{35}$$

Finally the single-potential correction for the hyperfine splitting in a muonic atom is given by

$$R_{TU}(\kappa) = \left[\frac{2}{3}\ln(2\kappa) + \frac{2}{3}[\psi(1) - \psi(1-2\varepsilon)] + \frac{1}{9}\right] + \left[\frac{2}{3} - \frac{5}{6}\varepsilon - \frac{5}{3}\varepsilon^2\right]\frac{1}{\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right). \tag{36}$$

The expression contains elementary functions and a ψ-function which is a logarithmic derivative of a gamma function. The result agrees with the double asymptotic form (Zα ≪ 1 and κ ≫ 1) given in (29).

4.2. Double-Potential Contribution

We first introduce the double limit Zα ≪ 1 and κ ≫ 1:

$$R_{UT}(\kappa) = 2\ln(2\kappa) - 3 + \frac{2\pi^2}{9} + \left(2\ln(2\kappa) - \frac{4}{3}\psi''(2) - 3 + \frac{4\pi^2}{9}\right)(Z\alpha)^2 + \mathcal{O}((Z\alpha)^4) + \mathcal{O}\left(\frac{1}{\kappa^2}\right). \tag{37}$$

In the more general case (κ ≫ 1 for arbitrary Zα), derivatives of the ψ-function appear in the expression:

$$R_{UT}(\kappa) = \frac{3-6\varepsilon+4\varepsilon^2-2\varepsilon^3}{(1-\varepsilon)^2(1-2\varepsilon)}$$

Analytic results of the present study and numerical results from [8] (denoted in [8] as E_{VP}^{ML-Ue} and E_{VP}^{EL-Ue}, respectively

	Single-potential contribution			Double-potential contribution		
Z	F_{TU} (11)	F_{TU}, $Z \longrightarrow 0$ (24)	Numerical value [8]	F_{TU} (16)	F_{TU}, $Z \longrightarrow 0$ (28)	Numerical value [8]
1	0.0085578	0.0085579	0.0085578	0.0087703	0.0087702	0.0087691
3	0.0254869	0.0254877	0.025487	0.0271133	0.0271086	0.027112
5	0.0422566	0.0422619	0.042257	0.0464153	0.0463843	0.046414
7	0.0589599	0.0589771	0.058960	0.0666723	0.0665652	0.066671
10	0.0840724	0.0841284	0.084072	0.0989572	0.0985560	0.098955
25	0.2192738	0.2195350	–	0.3068919	0.2936806	–

$$\times\left[\frac{2}{3}\ln(2\kappa)+\frac{2}{3}[\psi(1)-\psi(1-2\varepsilon)]+\frac{1}{9}\right]+\left[\frac{4}{3(1-\varepsilon)}\psi'(2-2\varepsilon)-\frac{2}{3}\frac{3-\varepsilon}{3(1-\varepsilon)(1-2\varepsilon)}\right]+\frac{3-2\varepsilon+\varepsilon^2}{\kappa^2}+\mathbb{O}\left(\frac{1}{\kappa^3}\right). \tag{38}$$

5. DISCUSSION OF RESULTS

We shall briefly discuss these results. For the case of ordinary (electronic) atoms we obtained an expansion for small $Z\alpha$

$$\Delta E_{hfs} = \frac{\alpha}{\pi}E_F\left\{\frac{3\pi}{4}Z\alpha+\left[\frac{34}{225}-\frac{8}{15}\ln(2Z\alpha)\right](Z\alpha)^2+\left[\frac{539\pi}{288}-\frac{13\pi}{24}\ln\frac{Z\alpha}{2}\right](Z\alpha)^3\right\}, \tag{39}$$

which contains one new coefficient in addition to the known ones. The possibility of expanding as far as arbitrary order of smallness is important for the polarization correction and for studying the structure of the series in terms of the parameter $Z\alpha$ in general. The appearance of the square of the logarithm of this parameter in various corrections [13, 6] makes accurate calculations preferable. This result together with calculations of the Uehling correction was first obtained in a closed analytic form as an explicit function of the parameter $Z\alpha$.

It is easy to obtain expansions of higher orders as high as necessary. For example, in the fourth order of the parameter $Z\alpha$ we obtain

$$\Delta E_{TU}^{(4)} = \frac{\alpha}{\pi}E_F(Z\alpha)^4\left(\frac{4}{5}\ln(2Z\alpha)-\frac{303}{175}\right)$$

and

$$\Delta E_{UT}^{(4)} = \frac{\alpha}{\pi}E_F(Z\alpha)^4\times\left(\frac{8}{15}\ln^2(2Z\alpha)-\frac{5336}{1575}\ln(2Z\alpha)+\frac{767881}{165375}-\frac{2\pi^2}{45}\right).$$

It has been noted that numerical results are also available for calculations of this correction [7, 8]. For comparison we first note that a slightly different parametrization of the result is frequently used for numerical and analytic calculations in electronic atoms ($\kappa = Z\alpha$)

$$E_{hfs}(1s) = \frac{\alpha}{\pi}E_F F(Z\alpha). \tag{40}$$

Results for the single-potential and double-potential contributions to the total expression F (see (11) and (16)), the expansions (see (24) and (28)), and the numerical results are compared in table. We only make a comparison with the results [8] for $Z \le 10$ since for larger Z we used a nonpoint nucleus in this study, as in the calculations made in [7].

We also give the asymptotic forms obtained above for muonic atoms

$$\Delta E_{hfs} = \frac{\alpha}{\pi}E_T\left\{\frac{4-10\varepsilon+9\varepsilon^2-4\varepsilon^3}{(1-\varepsilon)^2(1-2\varepsilon)}\times\left[\frac{2}{3}\ln(2\kappa)+\frac{2}{3}[\psi(1)-\psi(1-2\varepsilon)]+\frac{1}{9}\right]+\left[\frac{4}{3(1-\varepsilon)}\psi'(2-2\varepsilon)-\frac{2}{3}\frac{3-\varepsilon}{(1-\varepsilon)(1-2\varepsilon)}\right]+\left(\frac{11}{3}-\frac{17}{3}\varepsilon-\frac{2}{3}\varepsilon^2\right)\frac{1}{\kappa^2}+\mathbb{O}\left(\frac{1}{\kappa^3}\right)\right\},$$

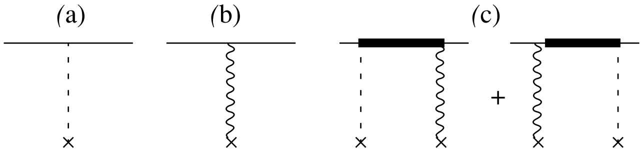


Fig. 2. Skeletal diagrams.

which for low nuclear charge have the form

$$\Delta E_{hfs} = \frac{\alpha}{\pi}E_F\left\{\frac{8}{3}\ln(2\kappa) - \frac{26}{9} + \frac{2\pi^2}{9} + \frac{11}{3\kappa^2}\right.$$

$$+\left(6\ln(2\kappa) - \frac{4}{3}\psi''(2) - \frac{22}{3} + \frac{8\pi^2}{9} + \frac{103}{24\kappa^2}\right)(Z\alpha)^2$$

$$\left. + \mathcal{O}((Z\alpha)^4) + \mathcal{O}\left(\frac{1}{\kappa^3}\right)\right\}.$$

Both these asymptotic forms contain a large logarithm (ln(2κ)) which can be determined using the running coupling constant and the known matrix elements for the skeletal diagrams (Fig. 2) [12], in particular

$$E^{\mathrm{Sc}}_{TU}(1s) = E_T(1s) \quad \text{and}$$

$$E^{\mathrm{Sc}}_{UT}(1s) = Z\alpha\frac{\partial E_T(1s)}{\partial(Z\alpha)} = \frac{3 - 6\varepsilon + 4\varepsilon^2 - 2\varepsilon^3}{(1-\varepsilon)^2(1-2\varepsilon)}E_T(1s).$$

The formula for calculating the logarithmic terms involves multiplying the skeletal matrix elements by

$$\frac{2}{3}\frac{\alpha}{\pi}\ln\frac{\bar{q}_{\mathrm{at}}}{m_l},$$

which is obtained from the known expression for the running coupling constant (see [1]). The characteristic atomic momentum is $\bar{q}_{\mathrm{at}} = Z\alpha m = \kappa m_l$. Finally we obtain the logarithmic contributions contained in (36) and (38) which serves as further confirmation of these results.

To conclude, we note that similar calculations may be made for the hyperfine splitting of any level. The expression for the relativistic energy E_T is known [14] as is the method of calculating the correction to the wave function perturbed by the magnetic field of the nucleus [11, 15].

This work was supported by the program “Fundamental Metrology” and also by the Russian Foundation for Basic Research (project no. 98-02-18350) and the program “Universities of Russia. Fundamental Research” (project no. 3930).

APPENDIX

Perturbation of the Dirac Coulomb Wave Function by the Magnetic Field of the Nucleus

The first-order correction for the hyperfine interaction to the Dirac Coulomb wave function may be obtained analytically using the method of generalized virial relations developed in [15]. An explicit solution for the ground state in a convenient form for practical calculations was given in [11] (in the second line of equation (18) in this article the sign before $\varepsilon_{n\kappa}$ should be changed). Below we follow the notation used in [16]. The Dirac wave function is defined by us as follows:

$$\Psi_{njlm} = \begin{pmatrix} f_{njl}(r)\Omega_{jlm}(\mathbf{n}) \\ -i(\boldsymbol{\sigma}\cdot\mathbf{n})g_{njl}(r)\Omega_{jlm}(\mathbf{n})\end{pmatrix}, \qquad (A.1)$$

where

$$f_{1s} = 2\sqrt{\frac{(Z\alpha)^3m^3(2-\varepsilon)}{\Gamma(3-2\varepsilon)}}(2Z\alpha mr)^{-\varepsilon}e^{-Z\alpha mr}, \qquad (A.2)$$

$$g_{1s} = -2\sqrt{\frac{(Z\alpha)^5m^3}{(2-\varepsilon)\Gamma(3-2\varepsilon)}}(2Z\alpha mr)^{-\varepsilon}e^{-Z\alpha mr} \qquad (A.3)$$

are the components of the pure Coulomb function (see [16]).

The linear correction to the ground state for the magnetic moment of the nucleus is a complex expression containing contributions of *s*- and *d*-waves. We are interested in the matrix element of the central potential (see (15)), so we merely give the first of these two contributions here

$$(\delta\Psi_{1,\,1/2,\,0,\,m})^s = -\frac{2}{3}\alpha\frac{\mu}{\mu_N}\frac{1}{m_p}$$

$$\times\frac{F(F+1) - I(I+1) - 3/4}{2I} \qquad (A.4)$$

$$\times\begin{pmatrix} X_{1s}(r)\Omega_{1/2,\,0,\,m}(\mathbf{n}) \\ -i(\boldsymbol{\sigma}\cdot\mathbf{n})Y_{1s}(r)\Omega_{1/2,\,0,\,m}(\mathbf{n})\end{pmatrix},$$

where the total moment of the system F may have values of $I \pm 1/2$ and the radial components have the form [11]

$$X_{1s} = \frac{1}{1-4(1-\varepsilon)^2}\left[\left(\frac{2(Z\alpha)^3m}{1-\varepsilon} - \frac{3}{r}\right)f_{1s}\right.$$

$$\left. + \left(3(2-\varepsilon)m - \frac{2Z\alpha}{r}\right)g_{1s} - \frac{2Z\alpha(3-2\varepsilon)m}{r}F_{1s}\right], \qquad (A.5)$$

$$Y_{1s} = \frac{1}{1-4(1-\varepsilon)^2}\left[\left(\frac{2(Z\alpha)^3m}{1-\varepsilon} + \frac{3}{r}\right)g_{1s}\right.$$

$$\left. + \left(-3\varepsilon m + \frac{6Z\alpha}{r}\right)f_{1s} - \frac{2Z\alpha(3-2\varepsilon)m}{r}G_{1s}\right], \qquad (A.6)$$

$$F_{1s}(r) = \sqrt{\frac{Z\alpha m(2-\varepsilon)}{\Gamma(3-2\varepsilon)}}e^{-Z\alpha mr}(2Z\alpha mr)^{1-\varepsilon}$$

$$\times\left[\frac{\psi(3-2\varepsilon)}{1-\varepsilon}+2-\varepsilon-\frac{1}{2(1-\varepsilon)}\right. \tag{A.7}$$

$$\left.-Z\alpha mr-\frac{1}{1-\varepsilon}\ln(2Z\alpha mr)\right],$$

$$G_{1s}(r) = -\sqrt{\frac{(Z\alpha)^3 m}{(2-\varepsilon)\Gamma(3-2\varepsilon)}}e^{-Z\alpha mr}(2Z\alpha mr)^{1-\varepsilon}$$

$$\times\left[\frac{\psi(3-2\varepsilon)}{1-\varepsilon}+2-\varepsilon+\frac{1}{2(1-\varepsilon)}\right. \tag{A.8}$$

$$\left.-Z\alpha mr-\frac{1}{1-\varepsilon}\ln(2Z\alpha mr)\right].$$

REFERENCES

1. S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **116** (11) (1999).
2. S. G. Karshenboĭm, Can. J. Phys. **76**, 168 (1998).
3. S. G. Karshenboĭm, V. G. Ivanov, and V. M. Shabaev, Can. J. Phys. **76**, 503 (1998).
4. S. L. Brodsky and G. W. Erickson, Phys. Rev. **148**, 26 (1966).
5. S. M. Schneider, W. Greiner, and G. Soff, Phys. Rev. A **50**, 118 (1994).
6. S. G. Karshenboĭm, Z. Phys. D **36**, 11 (1996).
7. V. M. Shabaev, M. Tomaselli, T. Kühl, *et al.*, Phys. Rev. A **56**, 252 (1997).
8. P. Sunnergren, H. Persson, S. Salomonson, *et al.*, Phys. Rev. A **58**, 1055 (1998).
9. S. G. Karshenboĭm, U. Jentschura, G. Soff, *et al.*, Eur. J. Phys. D **2**, 209 (1998).
10. G. Breit, Phys. Rev. **35**, 1477 (1930).
11. M. B. Shabaeva and V. M. Shabaev, Phys. Rev. A **52**, 2811 (1995).
12. S. G. Karshenboĭm, V. G. Ivanov, and V. M. Shabaev, Physica Scripta **80**, 491 (1999).
13. S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **103**, 1105 (1993).
14. P. Pyykko, E. Pajanne, and M. Inokuti, Int. J. Quantum Chem. **7**, 785 (1973).
15. V. M. Shabaev, J. Phys. B **24**, 4479 (1991).
16. V. B. Berestetskiĭ, E. M. Lifshits, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Pergamon, Oxford, 1982; Nauka, Moscow, 1980).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 66–78.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 75–89.

Quasiclassical Green Function in an External Field and Small-Angle Scattering Processes

R. N. Lee, A. I. Milsteĭn*, and V. M. Strakhovenko
Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
Novosibirsk, 630090 Russia
**e-mail: A.I.Milstein@inp.nsk.su*

Received July 22, 1999

Abstract—A representation is obtained for the quasiclassical Green functions of the Dirac and Klein–Gordon equations allowing for the first nonvanishing correction in an arbitrary localized potential which generally possesses no spherical symmetry. This is used to obtain a solution of these equations in an approximation similar to the Furry–Sommerfeld–Maue approximation. It is shown that the quasiclassical Green function does not reduce to the Green function obtained in the eikonal approximation and has a wider range of validity. This is illustrated by calculating the amplitude of small-angle scattering of a charged particle and the amplitude of Delbrück forward scattering. A correction proportional to the scattering angle was obtained for the amplitude of charged particle scattering in a potential possessing no spherical symmetry. The real part of the Delbrück forward scattering amplitude was calculated in a screened Coulomb potential. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

As we are well aware, quantum electrodynamics processes in external fields can be conveniently analyzed using wave functions which are a solution of the Dirac equation in a given field, and the corresponding Green functions. In cases where the particle energy is high compared with the mass and the scattering angles are small when the characteristic angular momenta in the process are large, the quasiclassical approximation can be applied. Moreover, using the quasiclassical Green functions considerably simplifies the calculations.

The quasiclassical electron Green function $G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ was first obtained for a Coulomb potential in [1, 2] using the exact Green function of the Dirac equation [3]. A more convenient representation of the quasiclassical Coulomb Green function for almost collinear vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ was obtained in [4, 5]. The same geometry was used to obtain the quasiclassical Green function for an arbitrary spherically symmetric decreasing potential in [6, 7].

In Section 2 of the present study, expressions for the quasiclassical Green functions of the Dirac and Klein–Gordon equations are derived for the case of a localized potential which does not generally possess spherical symmetry. By localized we understand a potential which has a maximum at a certain point and decreases fairly rapidly with increasing distance. The Green functions are obtained allowing for the first correction. These are then used to find the quasiclassical wave functions of the Dirac and Klein–Gordon equations and the corrections to them (Section 3). These wave functions generalize the results obtained in the Furry–Sommerfeld–Maue approximation [8–10] to the case of an arbitrary localized potential.

The eikonal approximation is frequently used to calculate the amplitudes of processes at high energies. The corresponding wave functions and Green functions generally differ from the quasiclassical ones and have a narrower range of validity. Thus, using the eikonal approximation without any special justification may yield incorrect results. In Section 4 we demonstrate this for the small-angle scattering of a charged particle in an external field. In this Section we put forward a systematic derivation of an expression for the amplitude of the process using quasiclassical wave functions. In particular, we find that for the case of a screened Coulomb potential the quasiclassical wave function rather than the eikonal one must be used to obtain a correct description in cases of momentum transfer larger than the reciprocal screening radius (when screening can be neglected).

In Section 5 the expressions obtained for the quasiclassical Green function are used to calculate the amplitude of elastic forward scattering of a photon in the electric field of an atom (Delbrück scattering). It has been shown [2] that corrections to the quasiclassical Green function must be taken into account to calculate this amplitude. These corrections are significant in the range of validity of the eikonal approximation whereas the contribution of higher orders of perturbation theory in terms of the external field (Coulomb corrections) is determined by the range of variables where the eikonal Green function cannot be used. The real part of the Delbrück scattering amplitude is obtained for the first time

1063-7761/00/9001-0066$20.00 © 2000 MAIK "Nauka/Interperiodica"

for a screened Coulomb potential. For scattering at comparatively low photon energies the real parts of the Delbrück amplitude and the amplitude of Compton scattering at atomic electrons are of the same order. This factor may prove important in descriptions of photon propagation in matter.

2. GREEN FUNCTION

It has been shown [4, 6] that in order to calculate the amplitudes of various quantum electrodynamic processes it is convenient to use the Green function of the "squared" Dirac equation $D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$, which is related to the ordinary Green function $G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ by

$$G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = [\gamma^0(\varepsilon - V(\mathbf{r}_2)) - \boldsymbol{\gamma}\mathbf{p}_2 + m]D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon), \tag{1}$$

where γ^μ are the Dirac matrices, $\mathbf{p} = -i\nabla$ is the momentum operator, and $V(\mathbf{r})$ is the potential energy. In the quasiclassical approximation the function $D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ may be expressed in the form

$$\begin{aligned} &D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) \\ &= \left[1 - \frac{i}{2\varepsilon}(\boldsymbol{\alpha}\cdot(\nabla_1 + \nabla_2))\right]D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon), \end{aligned} \tag{2}$$

where

$$\begin{gathered} D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \left\langle \mathbf{r}_2 \left| \frac{1}{\kappa^2 - H + i0} \right| \mathbf{r}_1 \right\rangle, \\ H = \mathbf{p}^2 + 2\varepsilon V(\mathbf{r}) - V^2(\mathbf{r}), \end{gathered} \tag{3}$$

$\kappa^2 = \varepsilon^2 - m^2$. Thus the problem is reduced to calculating the quasiclassical Green function $D^{(0)}$ of the Klein–Gordon equation with the potential $V(\mathbf{r})$ (the Schrödinger equation with the Hamiltonian H).

In the function $D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ we now convert from the variables $\mathbf{r}_1$ and $\mathbf{r}_2$ to the variables $\mathbf{r}_1$ and $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. In these variables the function D_0 satisfies the equation

$$[\kappa^2 - 2\kappa\phi(\mathbf{r}_1 + \mathbf{r}) - \mathbf{p}^2]D^{(0)}(\mathbf{r} + \mathbf{r}_1, \mathbf{r}_1|\varepsilon) = \delta(\mathbf{r}), \tag{4}$$

where $\phi = \lambda V - V^2/2\kappa$, $\lambda = \varepsilon/\kappa$, and $\mathbf{p} = -i\nabla_r$. In the ultrarelativistic case we have $\lambda = +1$ for $\varepsilon > 0$ and $\lambda = -1$ for $\varepsilon < 0$.

We shall seek a solution of this equation in the form

$$D^{(0)}(\mathbf{r} + \mathbf{r}_1, \mathbf{r}_1|\varepsilon) = -\frac{\exp(i\kappa r)}{4\pi r}F(\mathbf{r}, \mathbf{r}_1). \tag{5}$$

Note that the factor at F in (5) is the Green function of equation (4) for $\phi = 0$. The function F satisfies the equation

$$\left[i\frac{\partial}{\partial r} - \phi(\mathbf{r}_1 + \mathbf{r})\right]F = -\frac{1}{2\kappa}r\Delta(F/r) \tag{6}$$

with the boundary condition $F(\mathbf{r} = 0, \mathbf{r}_1) = 1$. For the following reasoning we define the effective impact parameter ρ_* of the rectilinear trajectory Γ linking the points $\mathbf{r}_1$ and $\mathbf{r}_2$ as

$$\rho_* = \min_{\mathbf{x}\in\Gamma}\frac{|\phi(\mathbf{x})|}{|\nabla_\perp\phi(\mathbf{x})|}. \tag{7}$$

Let us assume that this minimum is achieved at a certain point $\mathbf{x} = \mathbf{r}_*$. A necessary condition for the validity of the quasiclassical approximation is that $\kappa\rho_* \gg 1$, which we assume to be satisfied. We introduce the notation $a_{1,2} = |\mathbf{r}_{1,2} - \mathbf{r}_*|$. We consider the two overlapping regions:

$$1)\ \min(a_1, a_2) \ll \kappa\rho_*^2, \quad 2)\ \min(a_1, a_2) \gg \rho_*. \tag{8}$$

In the first region the right-hand side of equation (6) is small. The solution of this equation with zero right-hand side has the form

$$F_0 = \exp\left[-ir\int_0^1\phi(\mathbf{r}_1 + x\mathbf{r})dx\right], \tag{9}$$

which corresponds to the eikonal approximation. In order to calculate the first correction to F_0, we shall seek a solution of equation (6) in the form $F = F_0(1 + g)$ and we shall neglect g on the right-hand side of this equation. As a result, we obtain the equation for g:

$$\begin{aligned} 2i\kappa\frac{\partial}{\partial r}g &= \phi^2(\mathbf{r}_1 + \mathbf{r}) + ir\int_0^1 dx x^2\Delta_1\phi(\mathbf{r}_1 + x\mathbf{r}) \\ &+ r^2\left[\int_0^1 dx x\nabla_{1\perp}\phi(\mathbf{r}_1 + x\mathbf{r})\right]^2, \end{aligned} \tag{10}$$

where the index 1 in the derivatives denotes differentiation with respect to $\mathbf{r}_1$, $\nabla_{1\perp}$ is the component of the gradient perpendicular to $\mathbf{r}$, and $\Delta_1 = \nabla_1^2$. After integrating over r, we obtain the expression for g:

$$\begin{aligned} g = \frac{1}{2\kappa}\Bigg[&r^2\int_0^1 dx x(1-x)\Delta_1\phi(\mathbf{r}_1 + x\mathbf{r}) \\ &- ir\int_0^1 dx\phi^2(\mathbf{r}_1 + x\mathbf{r}) \\ &- 2ir^3\int_0^1 dx(1-x)\nabla_{1\perp}\phi(\mathbf{r}_1 + x\mathbf{r})\int_0^x dy y\nabla_{1\perp}\phi(\mathbf{r}_1 + y\mathbf{r})\Bigg]. \end{aligned} \tag{11}$$

Finally, taking into account the correction, the Green function $D^{(0)}$ in the first region has the form

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = -\frac{1}{4\pi r}\exp\left[i\kappa r - i\lambda r\int_0^1 V(\mathbf{r}_1 + x\mathbf{r})dx\right]$$

$$\times\left\{1 + \frac{1}{2\kappa}\left[\lambda r^2\int_0^1 dx x(1-x)\Delta_1 V(\mathbf{r}_1 + x\mathbf{r})\right.\right. \qquad (12)$$

$$\left.\left. - 2ir^3\int_0^1 dx(1-x)\nabla_{1\perp}V(\mathbf{r}_1 + x\mathbf{r})\int_0^x dy y\nabla_{1\perp}V(\mathbf{r}_1 + y\mathbf{r})\right]\right\}.$$

In this expression we used the relationship between ϕ and V and assumed that $\varepsilon \gg m$ (we set $\lambda^2 = 1$). Substituting (12) into (2), we obtain an expression for the function D in the first region:

$$D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = -\frac{1}{4\pi r}\exp\left[i\kappa r - i\lambda r\int_0^1 V(\mathbf{r}_1 + x\mathbf{r})dx\right]$$

$$\times\left\{1 - \frac{r}{2\kappa}\int_0^1 dx\boldsymbol{\alpha}\nabla_1 V(\mathbf{r}_1 + x\mathbf{r})\right.$$

$$+ \frac{\lambda r^2}{2\kappa}\int_0^1 dx x(1-x)\Delta_1 V(\mathbf{r}_1 + x\mathbf{r}) \qquad (13)$$

$$\left. - \frac{ir^3}{\kappa}\int_0^1 dx(1-x)\nabla_{1\perp}V(\mathbf{r}_1 + x\mathbf{r})\int_0^x dy y\nabla_{1\perp}V(\mathbf{r}_1 + y\mathbf{r})\right\}.$$

We shall now analyze the second region. In this case the right-hand side of equation (6) is not small. Using spherical coordinates we can rewrite this in the form:

$$\left[i\frac{\partial}{\partial r} - \phi(\mathbf{r}_1 + \mathbf{r}) - \frac{\mathbf{L}^2}{2\kappa r^2}\right]F = -\frac{1}{2\kappa}\frac{\partial^2}{\partial r^2}F, \qquad (14)$$

where $\mathbf{L}^2$ is the square of the angular momentum operator. In this equation $\mathbf{r}$ is an independent variable. We are interested in the value of the function $F(\mathbf{r}, \mathbf{r}_1)$ for $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. The polar axis of the spherical coordinates is for convenience directed along $\mathbf{r}_2 - \mathbf{r}_1$. The term on the left-hand side of the equation containing $\mathbf{L}^2$ is significant for $r \geq a_1$. This becomes clear if the operator $\mathbf{L}^2$ is made to act on the eikonal function (9). Then in order to calculate the function $F(\mathbf{r}_2 - \mathbf{r}_1, \mathbf{r}_1)$ it is sufficient to analyze a narrow region of polar angles of the vector $\mathbf{r}$ of the order of $\theta \sim \rho_*/a_1 \ll 1$. The right-hand side of equation (14) is a small correction. We shall seek a solution of this equation in the form

$$F = e^{iA}\mathcal{F}, \quad A = \left(\frac{1}{r} - \frac{1}{a_1}\right)\frac{\mathbf{L}^2}{2\kappa}. \qquad (15)$$

Substituting (15) into (14) we obtain an equation for $\mathcal{F}$

$$\left[i\frac{\partial}{\partial r} - \tilde{\phi}\right]\mathcal{F}$$

$$= -\frac{1}{2\kappa}\left[\frac{\partial^2}{\partial r^2} - \frac{i\mathbf{L}^2}{\kappa r^2}\left(\frac{\partial}{\partial r} - \frac{1}{r} - \frac{i\mathbf{L}^2}{4\kappa r^2}\right)\right]\mathcal{F}, \qquad (16)$$

where

$$\tilde{\phi} = e^{-iA}\phi(\mathbf{r}_1 + \mathbf{r})e^{iA}. \qquad (17)$$

We shall seek a solution of this equation taking into account the first correction in terms of the parameter $1/\kappa\rho_*$. For this purpose in the expansion of $\tilde{\phi}$ in terms of commutators of the operator A with ϕ it is sufficient to retain only the first two terms:

$$\tilde{\phi} \approx \phi - i[A, \phi].$$

In the principal approximation in terms of the parameter $1/\kappa\rho_*$ we can neglect the right-hand side of (16) and replace $\tilde{\phi}$ by ϕ. In this approximation the function $\mathcal{F}$ is the same as the eikonal function F_0 [see (9)]. In order to find the first correction, we express $\mathcal{F}$ in the form

$$\mathcal{F} = F_0(1 + g_1).$$

With the necessary accuracy we obtain the following equation for g_1:

$$2i\kappa\frac{\partial}{\partial r}g_1 = \phi^2(\mathbf{r}_1 + \mathbf{r}) - \left(\frac{1}{r} - \frac{1}{a_1}\right)$$

$$\times\left[i\mathbf{L}^2\phi(\mathbf{r}_1 + \mathbf{r}) + 2r(\mathbf{L}\phi(\mathbf{r}_1 + \mathbf{r}))\int_0^1 dx\mathbf{L}\phi(\mathbf{r}_1 + x\mathbf{r})\right]. \qquad (18)$$

Integrating this over r, we find

$$g_1 = -\frac{ir}{2\kappa}\left[\int_0^1 dx\phi^2(\mathbf{r}_1 + x\mathbf{r})\right.$$

$$+ ir\int_0^1 dx x\left(1 - \frac{r}{a_1}x\right)\Delta_{1\perp}\phi(\mathbf{r}_1 + x\mathbf{r})$$

$$+ 2r^2\int_0^1 dx\left(1 - \frac{r}{a_1}x\right)(\nabla_{1\perp}\phi(\mathbf{r}_1 + x\mathbf{r})) \qquad (19)$$

$$\left.\times\int_0^x dy y\nabla_{1\perp}\phi(\mathbf{r}_1 + y\mathbf{r})\right].$$

Here we have $\Delta_{1\perp} = \nabla_{1\perp}^2$. Using an expansion in terms of spherical functions we can show that when $\beta \ll 1$ for an arbitrary function $g(\mathbf{r})$, the following relationship is satisfied with the required accuracy

$$\exp[-i\beta^2\mathbf{L}^2]g(\mathbf{r}) \approx \int\frac{d\mathbf{q}}{i\pi}e^{iq^2}g(\mathbf{r}+2\beta r\mathbf{q}), \tag{20}$$

where $\mathbf{q}$ is a two-dimensional vector perpendicular to $\mathbf{r}$. Using (15), (19), and (20) we obtain the following expression for the Green function $D^{(0)}$ in the second region allowing for the correction:

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa r}}{4\pi^2 r}\int d\mathbf{q}\exp\left[iq^2 - i\lambda r\int_0^1 dx V(\mathbf{r}_x)\right]$$
$$\times\left\{1 - \frac{ir}{2\kappa}\left[i\lambda r\int_0^1 dx x\left(1-\frac{r}{a_1}x\right)\Delta_{1\perp}V(\mathbf{r}_x)\right.\right. \tag{21}$$
$$\left.\left. + 2r^2\int_0^1 dx\left(1-\frac{r}{a_1}x\right)(\nabla_{1\perp}V(\mathbf{r}_x))\int_0^x dy y\nabla_{1\perp}V(\mathbf{r}_y)\right]\right\},$$

where $\mathbf{r}_x = \mathbf{r}_1 + x\mathbf{r} + \mathbf{q}x\sqrt{2r(r-a_1)/(\kappa a_1)}$ and $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Note that the gradients with respect to $\mathbf{r}_1$ in this formula are taken for fixed $\mathbf{r}$.

We note that in $\mathbf{r}_x$ the term proportional to $\mathbf{q}$ need only be taken into account in the narrow region $|x - a_1/r| \sim \rho_*/r \ll 1$. Utilizing this fact, we rewrite formula (21) in a form which does not depend explicitly on a_1. For this we write $\mathbf{r}_x = \mathbf{R}_x + \delta\mathbf{r}_x$, where

$$\begin{aligned}\mathbf{R}_x &= \mathbf{r}_1 + x\mathbf{r} + \mathbf{q}\sqrt{2x(1-x)r/\kappa},\\ \delta\mathbf{r}_x &= \sqrt{2r/\kappa}\left(x\sqrt{r/(a_1-1)} - \sqrt{x(1-x)}\right)\mathbf{q}.\end{aligned} \tag{22}$$

We expand $V(\mathbf{r}_x)$ in phase in terms of $\delta\mathbf{r}_x$ as far as the first term, we replace $\mathbf{r}_x$ with $\mathbf{R}_x$ in the preexponential factor, and integrate by parts over $\mathbf{q}$. As a result we obtain

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa r}}{4\pi^2 r}\int d\mathbf{q}\exp\left[iq^2 - i\lambda r\int_0^1 dx V(\mathbf{R}_x)\right]$$
$$\times\left\{1 + \frac{ir^3}{2\kappa}\int_0^1 dx\int_0^x dy(x-y)(\nabla_{1\perp}V(\mathbf{R}_x))(\nabla_{1\perp}V(\mathbf{R}_y))\right\}. \tag{23}$$

As we have already discussed, the ranges of validity of formulas (12) and (23) overlap. In fact for $\rho_* \ll \min(a_1, a_2) \ll \kappa\rho_*^2$ we can expand $V(\mathbf{R}_x)$in terms of $\mathbf{q}$ as far as quadratic terms and integrate over $d\mathbf{q}$. Then formula (23) yields (12) to within terms $O(1/\kappa\rho_*)$ inclusive. Thus, we can write a formula which is valid in both regions subject to the condition $\kappa\rho_* \gg 1$:

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa r}}{4\pi^2 r}\int d\mathbf{q}\exp\left[iq^2 - i\lambda r\int_0^1 dx V(\mathbf{R}_x)\right]$$
$$\times\left\{1 - \frac{\lambda}{2\kappa}\left[2\int_0^1 dx V(\mathbf{R}_x) - V(\mathbf{r}_1) - V(\mathbf{r}_2)\right]\right. \tag{24}$$
$$+ \frac{ir^3}{\kappa}\int_0^1 dx\int_0^x dy[\sqrt{x(1-x)y(1-y)} - (1-x)y]$$
$$\left.\times(\nabla_{1\perp}V(\mathbf{R}_x))(\nabla_{1\perp}V(\mathbf{R}_y))\right\},$$

where the value of $\mathbf{R}_x$ is determined in (22). In fact, in the first region we can expand $V(\mathbf{R}_x)$ in terms of $\mathbf{q}$ as far as quadratic terms and integrate over $d\mathbf{q}$ which directly yields formula (12). In the second region we need to bear in mind that, as we have already noted, the main contribution to the integrals is made by a narrow region of x and y near a_1/r of width $\delta x \sim \delta y \sim \rho_*/r$. Thus we obtain

$$\begin{aligned}2[\sqrt{x(1-x)y(1-y)} - (1-x)y] &= x - y - (x-y)^2\\ &- (\sqrt{x(1-x)} - \sqrt{y(1-y)})^2 \approx x - y\end{aligned}$$

to within quadratic terms with respect to ρ_*/r. In addition, the linear terms with respect to V in the preexponential factor also make a negligible contribution in terms of the parameter ρ_*/r. Hence, in the second region (24) yields (21).

Using (2) we obtain a final expression for the Green function $D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$

$$D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa r}}{4\pi^2 r}\int d\mathbf{q}\exp\left[iq^2 - i\lambda r\int_0^1 dx V(\mathbf{R}_x)\right]$$
$$\times\left\{1 - \frac{r}{2\kappa}\int_0^1 dx\boldsymbol{\alpha}\nabla_1 V(\mathbf{R}_x)\right.$$
$$- \frac{\lambda}{2\kappa}\left[2\int_0^1 dx V(\mathbf{R}_x) - V(\mathbf{r}_1) - V(\mathbf{r}_2)\right] \tag{25}$$
$$+ \frac{ir^3}{\kappa}\int_0^1 dx\int_0^x dy[\sqrt{x(1-x)y(1-y)} - (1-x)y]$$

$$\times(\nabla_{1\perp}V(\mathbf{R}_x))(\nabla_{1\perp}V(\mathbf{R}_y))\Bigg\}.$$

An advantage of the representations (24) and (25) is that they retain their form in any coordinate system.

Integration over the variable $\mathbf{q}$ in formulas (24) and (25) may be interpreted as allowing for quantum fluctuations near the rectilinear trajectory between the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. The main contribution to the integral over $\mathbf{q}$ is made by the region $q \sim 1$. Quantum fluctuations may be neglected if

$$\int_0^1 dx\sqrt{x(1-x)r/\kappa}\,|\nabla_{1\perp}V(\mathbf{r}_1+x\mathbf{r})| \ll \int_0^1 dx|V(\mathbf{r}_1+x\mathbf{r})|,$$

which is in fact equivalent to the first condition in formula (8) which ensures the validity of the eikonal approximation.

In formulas (24) and (25), the terms in the preexponential factor containing the potential determine the correction to the quasiclassical Green function. The condition for the validity of these expressions is that these corrections are small. Our results were obtained for the case of a localized potential. In fact, when these corrections are small, they can also be used for potentials which are not localized, for example, for a superposition of localized potentials.

For a spherically symmetric potential the quasiclassical Green function without corrections can also be obtained from the results of [6, 7] in which this function was calculated for the case of almost collinear vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. In these studies quasiclassical radial wave functions were used to obtain the following expression for the function $D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ for a small angle θ between the vectors $\mathbf{r}_2$ and $-\mathbf{r}_1$:

$$D^{(0)}(\mathbf{r}_2,\mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa(r_1+r_2)}}{4\pi\kappa r_1 r_2}\int_0^\infty dl\,lJ_0(l\theta)\exp\left\{i\left[\frac{l^2(r_1+r_2)}{2\kappa r_1 r_2} + 2\lambda\delta\left(\frac{l}{\kappa}\right) + \lambda(\Phi(r_1)+\Phi(r_2))\right]\right\}. \tag{26}$$

In this formula we have

$$\Phi(r) = \int_r^\infty V(\zeta)d\zeta,$$

$$\delta(\rho) = -\int_0^\infty V(\sqrt{\zeta^2+\rho^2})d\zeta, \quad \lambda = \varepsilon/\kappa.$$

If the angle $\theta' = \pi - \theta$ between the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is small, we find

$$D^{(0)}(\mathbf{r}_1,\mathbf{r}_2|\varepsilon) = -\frac{1}{4\pi|\mathbf{r}_1-\mathbf{r}_2|}\exp\{i\kappa|\mathbf{r}_1-\mathbf{r}_2| + i\lambda\,\mathrm{sign}(r_1-r_2)(\Phi(r_1)-\Phi(r_2))\}. \tag{27}$$

For the following transformations it is convenient to rewrite formula (26) in a different form, using the identity

$$\int dl\,lJ_0(l\theta)g(l^2) = \frac{1}{2\pi}\int d\mathbf{q}\exp(i\mathbf{q}\cdot\boldsymbol{\theta})g(q^2),$$

where $g(x)$ is an arbitrary function and $\mathbf{q}$ is a two-dimensional vector. We substitute this relation into (26) and change the variables

$$\mathbf{q} \longrightarrow \sqrt{\frac{2\kappa r_1 r_2}{r_1+r_2}}\mathbf{q} - \frac{\kappa r_1 r_2}{r_1+r_2}\boldsymbol{\theta}.$$

We define the impact parameter $\boldsymbol{\rho}$ by

$$\boldsymbol{\rho} = \frac{\mathbf{r}\times[\mathbf{r}_1\times\mathbf{r}_2]}{r^3},$$

where $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Bearing in mind that $\theta \ll 1$, the impact parameter is $\boldsymbol{\rho} \approx \boldsymbol{\theta} r_1 r_2/(r_1+r_2)$ and $\rho \ll r$ we can reduce formula (26) to the form

$$D^{(0)}(\mathbf{r}_2,\mathbf{r}_1|\varepsilon) = \frac{ie^{i\kappa r}}{4\pi^2 r} \times\int d\mathbf{q}\exp\left[iq^2 - i\lambda r\int_0^1 dxV(\mathbf{r}_1 + x\mathbf{r} + \mathbf{q}\sqrt{2r_1r_2/\kappa r})\right]. \tag{28}$$

Here $\mathbf{q}$ is a two-dimensional vector lying in the plane perpendicular to the vector $\mathbf{r}$. This formula was obtained for small angles θ. However, it is also valid for $\theta \sim 1$ since in this case we can neglect the term proportional to $\mathbf{q}$ in the argument of the potential and the Green function (28) becomes the eikonal function. In particular, for $\theta' \ll 1$ it is the same as (27). If the correction is omitted in formula (24), it agrees with (28) since in situations where the term containing $\mathbf{q}$ needs to be retained in the argument of the potential, we can replace $\sqrt{x(1-x)}$ with $\sqrt{r_1r_2/r^2}$.

3. WAVE FUNCTIONS IN THE QUASICLASSICAL APPROXIMATION

The expressions obtained for the quasiclassical Green function can be used to find the quasiclassical wave functions taking into account the corrections. In earlier studies [8, 9] the quasiclassical wave functions were obtained for a Coulomb field (Furry–Sommerfeld–Maue functions) and in [10] they were obtained for an arbitrary decreasing central field. These wave functions were calculated in the principal approximation. In [11] the wave functions and corrections to them were determined for an arbitrary potential in the eikonal approximation.

In order to calculate the wave functions we use the well-known relations (see, for example [12])

$$\begin{aligned}\lim_{r_2\to\infty} D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) &= -\frac{\exp[i\kappa r_2]}{4\pi r_2}\psi^{(-)*}_{\mathbf{p}_2}(\mathbf{r}_1),\\ \lim_{r_1\to\infty} D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) &= -\frac{\exp[i\kappa r_1]}{4\pi r_1}\psi^{(+)}_{\mathbf{p}_1}(\mathbf{r}_2).\end{aligned} \tag{29}$$

Here $\mathbf{p}_1 = -\kappa\mathbf{r}_1/r_1$, $\mathbf{p}_2 = -\kappa\mathbf{r}_2/r_2$, and $\psi^{(+)}_{\mathbf{p}}(\mathbf{r})$ $(\psi^{(-)}_{\mathbf{p}}(\mathbf{r}))$ denotes the solution of the Klein–Gordon equation containing at infinity a plane wave having the momentum $\mathbf{p}$ and a diverging (converging) spherical wave. From (24) and (29) we obtain the quasiclassical wave functions for the Klein–Gordon equation allowing for the correction:

$$\begin{aligned}\psi^{(\pm)}_{\mathbf{p}}(\mathbf{r}) &= \pm\int\frac{d\mathbf{q}}{i\pi}\exp\left[i\mathbf{p}\cdot\mathbf{r} \pm iq^2 \mp i\lambda\int_0^\infty dx V(\mathbf{r}_x)\right]\\ &\times\left\{1 + \frac{\lambda}{2\kappa}V(r)\right.\\ &\left.\pm\frac{i}{\kappa}\int_0^\infty dx\int_0^x dy[\sqrt{xy} - y](\nabla_\perp V(\mathbf{r}_x))(\nabla_\perp V(\mathbf{r}_y))\right\},\end{aligned} \tag{30}$$

$$\mathbf{r}_x = \mathbf{r} \mp \mathbf{p}x/\kappa + \mathbf{q}\sqrt{2x/\kappa}, \quad \kappa = |\mathbf{p}|.$$

Here $\mathbf{q}$ is the two-dimensional vector perpendicular to $\mathbf{p}$ and $\nabla_\perp$ is the component of the gradient perpendicular to $\mathbf{p}$. Similarly for the Dirac equation the quasiclassical wave functions with the correction are obtained from (25):

$$\begin{aligned}\Psi^{(\pm)}_{\mathbf{p}}(\mathbf{r}) &= \pm\int\frac{d\mathbf{q}}{i\pi}\exp\left[i\mathbf{p}\cdot\mathbf{r} \pm iq^2 \mp i\lambda\int_0^\infty dx V(\mathbf{r}_x)\right]\\ &\times\left\{1 \mp \frac{1}{2\kappa}\int_0^\infty dx\,\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}_x) + \frac{\lambda}{2\kappa}V(\mathbf{r})\right.\\ &\left.\pm\frac{i}{\kappa}\int_0^\infty dx\int_0^x dy(\sqrt{xy} - y)(\nabla_\perp V(\mathbf{r}_x))(\nabla_\perp V(\mathbf{r}_y))\right\}u^\lambda_{\mathbf{p}},\end{aligned} \tag{31}$$

where $u^\lambda_{\mathbf{p}}$ are free positive- ($\lambda = 1$) and negative- ($\lambda = -1$) frequency Dirac spinors having the momentum $\mathbf{p}$. When $V(\mathbf{r}_x)$ can be expanded in terms of $\mathbf{q}$, formula (31) yields the wave function in the eikonal approximation with corrections and agrees with the result obtained in [11].

4. SCATTERING OF CHARGED PARTICLES

In this section we use the quasiclassical Green function to calculate the amplitude of small-angle scattering of high-energy charged particles in a localized potential. For a spin zero particle the scattering amplitude has the form

$$f(\mathbf{p}_2, \mathbf{p}_1) = -\frac{\kappa}{2\pi}\int d\mathbf{r}\,\psi^{(-)*}_{\mathbf{p}_2}(\mathbf{r})\phi(\mathbf{r})e^{i\mathbf{p}_1\cdot\mathbf{r}}. \tag{32}$$

We note that $\phi = \lambda V - V^2/2\kappa$.

For convenience, we position the origin at the point of maximum potential and convert to cylindrical coordinates with the z axis parallel to the vector $\mathbf{p}_2$. Substituting (30) into (32) we obtain, allowing for the first correction,

$$\begin{aligned}f &= \frac{i\kappa}{2\pi^2}\int_{-\infty}^{\infty} dz\int d\boldsymbol{\rho}\int d\mathbf{q}\\ &\times\exp\left[-iQ_z z - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} + i\mathbf{q}^2 - i\lambda\int_0^\infty dx V(x+z, \boldsymbol{\rho}_x)\right]\\ &\times\lambda V(z, \boldsymbol{\rho})\left[1 - \frac{i}{\kappa}\int_0^\infty dx\int_0^x dy[\sqrt{xy} - y]\right.\\ &\left.\times(\nabla_\rho V(x+z, \boldsymbol{\rho}_x))(\nabla_\rho V(y+z, \boldsymbol{\rho}_y))\right],\end{aligned} \tag{33}$$

where $\boldsymbol{\rho}_x = \boldsymbol{\rho} + \sqrt{2x/\kappa}\,\mathbf{q}$, $\mathbf{Q} = \mathbf{p}_2 - \mathbf{p}_1$.

We first demonstrate the need to use quasiclassical rather than eikonal wave functions in the scattering problem. To do this, we calculate the amplitude in the principal approximation which corresponds to replacing the brackets in the pre-exponential factor in formula (33) by one. We divide the region of integration over z into two: $(-\infty, 0)$ and $(0, \infty)$. In the region between zero and infinity we can replace $\boldsymbol{\rho}_x$ by $\boldsymbol{\rho}$ (neglecting quantum fluctuations) and integrate over $\mathbf{q}$. The contribution of this region then has the form

$$\begin{aligned}f_+ &= -\frac{\kappa}{2\pi}\int_{-\infty}^{\infty} dz\int d\boldsymbol{\rho}\\ &\times\exp\left[-iQ_z z - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} - i\lambda\int_0^\infty dx V(x+z, \boldsymbol{\rho})\right]\lambda V(z, \boldsymbol{\rho}).\end{aligned} \tag{34}$$

Now integrating by parts over z using the relation

$$\lambda V(z,\boldsymbol{\rho})\exp\left[-i\lambda\int_0^\infty dx V(x+z,\boldsymbol{\rho})\right]$$
$$= -i\frac{\partial}{\partial z}\left\{\exp\left[-i\lambda\int_0^\infty dx V(x+z,\boldsymbol{\rho})\right]-1\right\}, \tag{35}$$

we obtain

$$f_+ = -\frac{i\kappa}{2\pi}\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}]\left(\exp\left[-i\lambda\int_0^\infty dx V(x,\boldsymbol{\rho})\right]\right.$$
$$\left.-1-iQ_z\int_0^\infty dz\exp[-iQ_z z]\left(\exp\left(\left[-i\lambda\int_0^\infty dx V(x,\boldsymbol{\rho})\right]-1\right)\right)\right). \tag{36}$$

The main contribution to the integral over z in this formula is made by the region $z \sim \rho$. The relative value of the contribution proportional to this integral compared with the term outside the integral is $Q_z\rho \sim Q_z/Q_\perp \ll 1$ so that in the principal approximation this contribution can be neglected. In the region between $-\infty$ and zero in the principal approximation x may be replaced by $|z|$ in $\boldsymbol{\rho}_x$. We then make the change $\boldsymbol{\rho} \longrightarrow \boldsymbol{\rho} - \mathbf{q}\sqrt{2|z|/\kappa}$ and obtain

$$f_- = \frac{i\kappa}{2\pi^2}\int_{-\infty}^0 dz\int d\boldsymbol{\rho}\int d\mathbf{q}$$
$$\times\exp\left[-i\left(Q_z-\frac{\mathbf{Q}_\perp^2}{2\kappa}\right)z - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} + i\left(\mathbf{q}+\mathbf{Q}_\perp\sqrt{\frac{|z|}{2\kappa}}\right)^2\right. \tag{37}$$
$$\left.-i\lambda\int_0^\infty dx V(x+z,\boldsymbol{\rho})\right]\lambda V\left(z,\boldsymbol{\rho}-\mathbf{q}\sqrt{\frac{2|z|}{\kappa}}\right).$$

We note that for $\kappa\rho \gg 1$ the condition $\sqrt{2|z|/\kappa} \ll \max(|z|, \rho)$is satisfied for any z which means that the term proportional to $\mathbf{q}$ can be neglected in the argument of the potential. In the small-angle approximation we have $Q_z = \mathbf{Q}_\perp^2/(2\kappa)$, i.e., the coefficient of z in the argument of the exponential function becomes zero. Calculating the integral over $\mathbf{q}$ and then integrating over z using the identity (35), we find

$$f_- = -\frac{i\kappa}{2\pi}\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}]$$
$$\times\left(\exp\left[-i\lambda\int_{-\infty}^\infty dx V(x,\boldsymbol{\rho})\right]-\exp\left[-i\lambda\int_0^\infty dx V(x,\boldsymbol{\rho})\right]\right). \tag{38}$$

Adding f_+ and f_- we obtain the expression familiar to us as the scattering amplitude in the eikonal approximation:

$$f = -\frac{i\kappa}{2\pi}\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}]$$
$$\times\left(\exp\left[-i\lambda\int_{-\infty}^\infty dx V(x,\boldsymbol{\rho})\right]-1\right). \tag{39}$$

This result is usually obtained using the eikonal wave function over the entire range of integration over z and neglecting the term $Q_z z$ in the argument of the exponential function. It follows from our calculations that, for arbitrary transfers of momentum, both these approximations are generally inaccurate. To illustrate this we consider the screened Coulomb potential with the screening radius r_c. The main contribution to the amplitude is made by the region of integration $z \sim r_c$, $\rho \sim 1/Q_\perp$. If the value of $Q_z z \sim Q_\perp^2 r_c/\kappa$ in this region is small compared with unity, the term containing $\mathbf{q}$ must be neglected in $\boldsymbol{\rho}_x$ since $\rho \sim 1/Q_\perp \le \sqrt{r_c/\kappa}$. Thus, it is not permissible to use the eikonal wave function. At the same time, the quantity $Q_z z$ in the argument of the exponential function cannot be neglected. If this term is retained but the eikonal wave function is used, under the condition $Q_\perp^2 r_c/\kappa \ge 1$ the result for the scattering amplitude is incorrect. In particular, if this procedure is adopted, the well-known result (Rutherford formula) is not obtained in the limit $r_c \longrightarrow \infty$ (unscreened Coulomb potential). Hence, formula (39) is valid for any $Q_\perp \ll \kappa$, but it cannot be derived correctly without using the quasiclassical wave function.

We shall now calculate the scattering amplitude allowing for the correction. For convenience integration over z will be performed between $-L$ and $+\infty$ and we then make L tend to infinity. Using the identity

$$\left[\lambda V(z,\boldsymbol{\rho}) + i\frac{\partial}{\partial z} + \lambda\int_0^\infty\frac{dx}{\sqrt{2\kappa x}}\mathbf{q}\cdot\nabla_\rho V(z+x,\boldsymbol{\rho}_x)\right]$$
$$\times\exp\left[-i\lambda\int_0^\infty dx V(x+z,\boldsymbol{\rho}_x)\right] = 0, \tag{40}$$

we can integrate by parts over z in the first term in brackets in formula (33). As a result, the expression for f has the form

$$f = f_0 + f_1,$$

where

$$f_0 = -\frac{\kappa}{2\pi^2}\lim_{L\to\infty}\int d\boldsymbol{\rho}\int d\mathbf{q}\exp\left[i\frac{\mathbf{Q}_\perp^2}{2\kappa}L - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}\right.$$
$$\left.+i\mathbf{q}^2 - i\lambda\int_{-L}^\infty dx V(x,\boldsymbol{\rho}_{x+L})\right], \tag{41}$$

$$\begin{aligned}
f_1 &= \frac{i\kappa}{2\pi^2}\int\limits_{-\infty}^{\infty} dz\int d\boldsymbol{\rho}\int d\mathbf{q} \\
&\times \exp\left[-i\frac{\mathbf{Q}_\perp^2}{2\kappa}z - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} + i\mathbf{q}^2 - i\lambda\int\limits_0^\infty dx V(x+z, \boldsymbol{\rho}_x)\right] \\
&\times\Bigg[\frac{i\lambda}{\kappa}V(z,\boldsymbol{\rho})\int\limits_0^\infty dx\int\limits_0^x dy(\sqrt{xy}-y) \\
&\times(\nabla_\rho V(x+z,\boldsymbol{\rho}_x))(\nabla_\rho V(y+z,\boldsymbol{\rho}_y)) \\
&+\frac{\mathbf{Q}_\perp^2}{2\kappa} - \lambda\int\limits_0^\infty\frac{dx}{\sqrt{2\kappa x}}(\mathbf{q}\cdot\nabla_\rho V(z+x,\rho_x))\Bigg].
\end{aligned}\tag{42}$$

In the expression for f_0, the term which does not depend on the potential and vanishes for $\mathbf{Q}_\perp \neq 0$ is omitted. We taken this into account explicitly in the final expression for the amplitude. In order to find the limit in the function f_0, we make the transposition $\boldsymbol{\rho} \longrightarrow \boldsymbol{\rho} - \mathbf{q}\sqrt{2L/\kappa}$ and $\mathbf{q} \longrightarrow \mathbf{q} - \mathbf{Q}_\perp\sqrt{L/(2\kappa)}$. Thereafter, calculation of the limit $L \longrightarrow \infty$ and integration over $\mathbf{q}$ become elementary. Taking into account the first correction with respect to $Q_\perp/\kappa$ we obtain

$$\begin{aligned}
f_0 &= -\frac{i\kappa}{2\pi}\int d\boldsymbol{\rho}\exp\left[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} - i\lambda\int\limits_{-\infty}^{\infty}dx V(x,\boldsymbol{\rho})\right] \\
&\times\left[1 + \frac{i\lambda}{2\kappa}\int\limits_{-\infty}^{\infty}dx x\mathbf{Q}_\perp\cdot\nabla_\rho V(x,\boldsymbol{\rho})\right].
\end{aligned}\tag{43}$$

We now calculate f_1. The integral over $\boldsymbol{\rho}$ in (42) converges in the region $\rho \sim 1/Q_\perp$. Quantum fluctuations are only important for $z < 0$ and x near $-z$. Since f_1 is a correction, in formula (42) we can replace $\boldsymbol{\rho}_x$ with $\boldsymbol{\rho}_{|z|}$. In addition $V(z, \boldsymbol{\rho})$ in the pre-exponential factor may be replaced by $V(z, \boldsymbol{\rho}_{|z|})$ since the difference between them is small for any z. The integral over z between zero and infinity converges in the range $z \leq \rho$. We then have

$$z\mathbf{Q}_\perp^2/2\kappa \sim \rho\mathbf{Q}_\perp^2/\kappa \sim Q_\perp/\kappa \ll 1.$$

Hence, in the exponential function we can replace $z\mathbf{Q}_\perp^2/2\kappa$ with $-|z|\mathbf{Q}_\perp^2/2\kappa$. We integrate the term proportional to $\mathbf{q}$ by parts over $\mathbf{q}$ so that in the preexponential factor this vector only appears in the argument of the potential. Then the expression for f_1 is transformed to give

$$\begin{aligned}
f_1 &= \int\limits_{-\infty}^{\infty}dz\int d\boldsymbol{\rho}\int\frac{d\mathbf{q}}{i\pi} \\
&\times\exp\left[i\frac{\mathbf{Q}_\perp^2}{2\kappa}|z| - i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} + i\mathbf{q}^2\right]g(z, \boldsymbol{\rho} + \sqrt{2|z|/\kappa}\mathbf{q})
\end{aligned}\tag{44}$$

with a certain function q. After making the transposition $\boldsymbol{\rho} \longrightarrow \boldsymbol{\rho} - \mathbf{q}\sqrt{2|z|/\kappa}$, the integral over $\mathbf{q}$ becomes elementary. As a result we obtain

$$f_1 = \int\limits_{-\infty}^{\infty}dz\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}]g(z,\boldsymbol{\rho}).\tag{45}$$

Following this procedure we obtain the following expression for f_1:

$$\begin{aligned}
f_1 &= -\frac{1}{2\pi}\int d\boldsymbol{\rho}\int\limits_{-\infty}^{\infty}dz\exp\left[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} - i\lambda\int\limits_0^\infty dx V(x+z,\boldsymbol{\rho})\right] \\
&\times\Bigg[i\lambda V(z,\boldsymbol{\rho})\int\limits_0^\infty dx\int\limits_0^x dy(\sqrt{xy}-y) \\
&\times(\nabla_\rho V(x+z,\boldsymbol{\rho}))(\nabla_\rho V(y+z,\boldsymbol{\rho})) \\
&+\frac{1}{2}\int\limits_0^\infty dx\int\limits_0^\infty dy(1-\sqrt{x/y})(\nabla_\rho V(x+z,\boldsymbol{\rho}))(\nabla_\rho V(y+z,\boldsymbol{\rho}))\Bigg].
\end{aligned}\tag{46}$$

When deriving this formula we integrated by parts over $\boldsymbol{\rho}$ the term proportional to $\mathbf{Q}_\perp^2$ in the preexponential factor. We can check that the expression in the integrand of (46) is a total derivative with respect to z so that integration over z is trivial. Combining these expressions with (43) and integrating by parts over $\boldsymbol{\rho}$ the term proportional to $\mathbf{Q}_\perp$, we finally obtain the scattering amplitude allowing for the first correction:

$$\begin{aligned}
f &= -\frac{i\kappa}{2\pi}\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}]\Bigg\{\exp\left[-i\lambda\int\limits_{-\infty}^{\infty}dx V(x,\boldsymbol{\rho})\right] \\
&-1 + \exp\left[-i\lambda\int\limits_{-\infty}^{\infty}dx V(x,\boldsymbol{\rho})\right]\Bigg[\frac{\lambda}{2\kappa}\int\limits_{-\infty}^{\infty}dx x\Delta_\rho V(x,\boldsymbol{\rho}) \\
&-\frac{i}{\kappa}\int\limits_{-\infty}^{\infty}dx\int\limits_{-\infty}^{x}dy y(\nabla_\rho V(x,\boldsymbol{\rho}))(\nabla_\rho V(y,\boldsymbol{\rho}))\Bigg]\Bigg\}.
\end{aligned}\tag{47}$$

Here we have $\Delta_\rho = \nabla_\rho^2$. Using the wave function (31), we can show (see, for example [13]) that for small-angle scattering the amplitude for spin 1/2 particles is the same as the amplitude (47) for spin zero particles to within terms $Q_\perp/\kappa$ inclusive.

We now show that this amplitude has regular properties relative to shifts of the origin. It follows from formula (33) that when the potential is transposed $V(\mathbf{r}) \longrightarrow V(\mathbf{r} + \mathbf{r}_0)$, the amplitude $f(\mathbf{p}_2, \mathbf{p}_1)$ becomes $f(\mathbf{p}_2, \mathbf{p}_1)\exp[i\mathbf{Q} \cdot r_0]$. The amplitude (47) clearly has this property for a shift perpendicular to $\mathbf{p}_2$ along which the z axis is directed. We now analyze the shift along the z axis: $V(z, \boldsymbol{\rho}) \longrightarrow V(z + z_0, \boldsymbol{\rho})$. Transposing the variables $x \longrightarrow x - z_0$ and $y \longrightarrow y - z_0$ in the integrals over x and y, we obtain a correction to the amplitude proportional to z_0:

$$\begin{aligned}\delta f &= \frac{iz_0}{4\pi}\int d\boldsymbol{\rho}\exp\left[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho} - i\lambda\int\limits_{-\infty}^{\infty} dx V(x, \boldsymbol{\rho})\right] \\ &\times\left[\lambda\int\limits_{-\infty}^{\infty} dx\Delta_\rho V(x, \boldsymbol{\rho}) - i\left(\int\limits_{-\infty}^{\infty} dx\nabla_\rho V(x, \boldsymbol{\rho})\right)^2\right] \\ &= -\frac{z_0}{4\pi}\int d\boldsymbol{\rho}\exp[-i\mathbf{Q}_\perp\cdot\boldsymbol{\rho}] \\ &\times\Delta_\rho\left(\exp\left[-i\lambda\int\limits_{-\infty}^{\infty} dx V(x, \boldsymbol{\rho})\right] - 1\right).\end{aligned} \tag{48}$$

Integrating by parts over $\boldsymbol{\rho}$ and replacing $\mathbf{Q}_\perp^2/2\kappa$ with Q_z, we find that $f + \delta f \approx f\exp(iQ_z z_0)$ provided that $Q_z z_0 \ll 1$. Thus, the regular law for transformation of the scattering amplitude in translation is satisfied with the same accuracy as formula (47).

For a potential satisfying the condition $V(z, \boldsymbol{\rho}) = V(-z, \boldsymbol{\rho})$, expression (47) agrees with that obtained in [11] in the eikonal approximation allowing for the correction. However, as we have already discussed, formula (47) may also be valid for a relationship between momentum transfers and potential parameters for which the eikonal approximation cannot be applied.

5. DELBRÜCK SCATTERING

The coherent scattering of photons in the electric field of atoms via virtual electron–positron pairs (Delbrück scattering) has been studied intensively both theoretically and experimentally (see, for example, the review [14]). In this section we analyze the amplitude of Delbrück scattering in a screened Coulomb field as another example of using this quasiclassical Green function. The photon energy ω is assumed to be large compared with the electron mass m. In accordance with the optical theorem, the imaginary part of this amplitude is expressed in terms of the cross section for creation of electron–positron pairs by a photon in an atomic field. In order to describe the propagation of light in the medium we also need to know the real part of the scattering amplitude.

In order to calculate the amplitudes of Delbrück scattering using Green functions, we need to use various approximations for these functions for various transfers of momentum $\Delta = \mathbf{k}_2 - \mathbf{k}_1$ ($\mathbf{k}_1$ and $\mathbf{k}_2$ are the momenta of the initial and final photons). For $\Delta \sim \omega$ the quasiclassical approximation cannot be used since the main contribution to the amplitude is made by the orbital moments $l \sim \omega/\Delta \sim 1$. The amplitude for these transfers of momentum was calculated in [15]. In the region $\omega \gg \Delta \gg m^2/\omega$ we can use the quasiclassical Green function without any correction [1, 16]. Calculation of the Delbrück scattering amplitude in a Coulomb field for $\Delta \le m^2/\omega$ necessitated a special analysis (see [2, 17]). In this case a contribution to the amplitude is made by the impact parameters ρ as far as $\rho \sim \omega/m^2$. For these impact parameters we need to allow for a correction to the quasiclassical Green function [2]. For a screened Coulomb potential where $\omega/m^2 \gg r_c$ (r_c is the screening radius, in the Thomas–Fermi model $r_c \sim (m\alpha)^{-1}Z^{-1/3} \gg 1/m$) a contribution to the amplitude is made by the impact parameters $\rho \le r_c \ll \omega/m^2$. In this case we can use the quasiclassical Green function without any correction for arbitrary $\Delta \ll \omega$ [6, 7]. If Δ, $r_c^{-1} \le m^2/\omega$, the correction must be taken into account. Moreover the expression for the forward scattering amplitude ($\Delta = 0$) obtained using the Green quasiclassical function without any correction is indeterminate. Subsequently we obtain the Delbrück scattering amplitude for an arbitrary relation between r_c and ω/m^2.

It was shown in [6] that the Delbrück scattering amplitude at high photon energies may be expressed in the form

$$\begin{aligned}M &= i\alpha\int\limits_0^\omega d\varepsilon\int d\mathbf{r}_1 d\mathbf{r}_2\exp[i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)] \\ &\times \mathrm{Sp}[(2\mathbf{e}^*\cdot\mathbf{p}_2 - \hat{e}^*\hat{k})D(\mathbf{r}_2, \mathbf{r}_1|\omega - \varepsilon)] \\ &\times[(2\mathbf{e}\cdot\mathbf{p}_1 + \hat{e}\hat{k})D(\mathbf{r}_1, \mathbf{r}_2|-\varepsilon)],\end{aligned} \tag{49}$$

where e and k are the photon polarization and four-momentum, $\mathbf{p}_{1,2} = -i\nabla_{1,2}$. In this formula it is implied that its value in a zero external field is subtracted from the expression in the integrand. Since in a central field the scattering amplitude does not depend on the photon polarization, in formula (49) it is convenient to make the substitution $e_i^* e_j \longrightarrow (\delta_{ij} - k_i k_j/\omega^2)/2$.

In (49) we convert from the variables $\mathbf{r}_{1,2}$ to

$$\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1, \quad \boldsymbol{\rho} = \frac{\mathbf{r}\times[\mathbf{r}_1\times\mathbf{r}_2]}{r^3}, \quad z = -\frac{\mathbf{r}\cdot\mathbf{r}_1}{r^2}.$$

Since $\boldsymbol{\rho}\cdot\mathbf{r} = 0$ integration over $\boldsymbol{\rho}$ is performed in the plane perpendicular to the vector $\mathbf{r}$. The main contribution to the amplitude is made by the region of integration in which $r \sim \omega/m^2$, $|z| \sim 1$ and the angles between the vector $\mathbf{r}$ and $\mathbf{k}$ are of the order $\theta_r \sim m/\omega \ll 1$. As a result of the smallness of the angles θ_r we can assume that the vector $\boldsymbol{\rho}$ is perpendicular to $\mathbf{k}$. Quite clearly the main contribution is made by $\rho \le 1$, $m^2 r_c/\omega$.

We divide the region of integration over ρ into two: between 0 and ρ_0 and between ρ_0 and ∞, where $m/\omega \ll \rho_0 \ll 1$, $m^2 r_c/\omega$. In the first region (where $\rho < \rho_0$) we can use the following representation for the Green function

$$\begin{aligned} D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) &= \frac{ie^{i\kappa r}}{4\pi^2 r}\int d\mathbf{q}\left[1 + \frac{\boldsymbol{\alpha}\cdot\mathbf{q}}{\varepsilon}\sqrt{\frac{\kappa r}{2r_1 r_2}}\right] \\ &\times \exp\left[iq^2 - i\lambda r\int_0^1 dx V(\mathbf{r}_1 + x\mathbf{r} + \mathbf{q}\sqrt{2r_1 r_2/\kappa r})\right], \end{aligned} \tag{50}$$

which is obtained by substituting (28) into (2). In this case in the term containing $\boldsymbol{\alpha}$ we omitted the longitudinal components of the gradient which have additional ρ smallness compared with the transverse ones, and we integrated by parts over $\mathbf{q}$. Moreover, because of the definition of ρ_0 we can neglect screening in this region and replace $V(r)$ by the Coulomb potential $V_c(r) = -Z\alpha/r$ after which the integral over x is easily taken in phase.

Screening is only significant in the second region where the representation (13) can be used, i.e., the eikonal Green function with a correction. By substituting (13) into (49), the phases of the Green functions which depend on the potential are reduced. Hence the contribution of the second region is quadratic in terms of the potential, i.e., it corresponds to the first Born approximation. In addition, the reduction of the phases means that it is necessary to allow for the correction to the Green function. In this region the contributions to the amplitude from the correction and the principal term in the Green function are of the same order.

We shall now calculate the contribution M_1 of the first region to the amplitude of the process. We substitute into (49) the Green function (50) for the case of a Coulomb potential, differentiate, and calculate the trace of the γ-matrices. Using the smallness of the angle between the vectors $\mathbf{r}$ and $\mathbf{k}$ and the relationship $\sqrt{\varepsilon^2 - m^2} \approx |\varepsilon| - m^2/2|\varepsilon|$ we obtain

$$\begin{aligned} M_1 &= -\frac{i\alpha}{(2\pi)^4}\int_0^\omega d\varepsilon\varepsilon\kappa\int dr r^5 d\boldsymbol{\theta}_r \int_{\rho<\rho_0} d\boldsymbol{\rho}\int_0^1 \frac{dz}{(z(1-z))^3} \\ &\times \int d\mathbf{q}_1 d\mathbf{q}_2\left[\mathrm{Re}\left(\frac{|\boldsymbol{\rho}-\mathbf{q}_1|}{|\boldsymbol{\rho}-\mathbf{q}_2|}\right)^{2iZ\alpha} - 1\right] \\ &\times \exp\left\{i\frac{r}{2}\left[\omega\boldsymbol{\theta}_r^2 - \frac{m^2\omega}{\varepsilon\kappa} + \frac{\varepsilon q_1^2 + \kappa q_2^2}{z(1-z)}\right]\right\} \\ &\times\left[2\varepsilon\kappa[\mathbf{q}_1\cdot\mathbf{q}_2 - z(1-z)\boldsymbol{\theta}_r^2]\right. \\ &\left. + \frac{\omega}{4(z(1-z))}(\varepsilon\mathbf{q}_1 - \kappa\mathbf{q}_2)\cdot(\mathbf{q}_1 - \mathbf{q}_2) - i\frac{\omega}{r}\right], \end{aligned} \tag{51}$$

where $\kappa = \omega - \varepsilon$, and $\boldsymbol{\theta}_r$ is a two-dimensional vector in the plane perpendicular to κ (like the vectors $\mathbf{q}_{1,2}$). We draw attention to the fact that the integral over z in this formula is taken between zero and one. This is because outside this interval the Green functions are eikonal and the potential-dependent phases are reduced. For $\rho < \rho_0 \ll 1$ this leads to a smaller contribution of the region of integration outside the interval $0 \le z \le 1$ compared with that made by this interval.

We now integrate over $\boldsymbol{\theta}_r$, convert from the variables $\mathbf{q}_{1,2}$ to $\mathbf{Q} = (\mathbf{q}_1 + \mathbf{q}_2)/2$ and $\mathbf{q} = (\mathbf{q}_1 - \mathbf{q}_2)/2$, and make the transposition $\boldsymbol{\rho} \longrightarrow \boldsymbol{\rho} + \mathbf{Q}$. Then the integral over $\boldsymbol{\rho}$ has the form

$$J = \int_{|\boldsymbol{\rho}+\mathbf{Q}|<\rho_0} d\boldsymbol{\rho}\left[\mathrm{Re}\left(\frac{|\boldsymbol{\rho}-\mathbf{q}|}{|\boldsymbol{\rho}+\mathbf{q}|}\right)^{2iZ\alpha} - 1\right]. \tag{52}$$

The main contribution to the amplitude (51) is made by $Q, q \sim m/\omega \ll \rho_0$ so that in (52) we can omit Q in the limits of integration. We use the following procedure to calculate the integral (52). We subtract and add to the expression in the integrand the function $-2(Z\alpha)^2[2\boldsymbol{\rho}\cdot\mathbf{q}/(\rho^2 + q^2)]^2$ which can easily be integrated:

$$\begin{aligned} J_1 &= \int_{\rho<\rho_0} d\boldsymbol{\rho}\left[-2(Z\alpha)^2\left(\frac{2\boldsymbol{\rho}\cdot\mathbf{q}}{\rho^2 + q^2}\right)^2\right] \\ &= -4\pi(Z\alpha)^2 q^2\left(\ln\frac{\rho_0^2}{q^2} - 1\right). \end{aligned} \tag{53}$$

Then in the integral of the difference, we can replace ρ_0 by infinity. To calculate this integral it is convenient to multiply the expression in the integrand by

$$\begin{aligned} 1 &\equiv \int_{-1}^1 dy\,\delta\left(y - \frac{2\boldsymbol{\rho}\cdot\mathbf{q}}{\rho^2 + q^2}\right) \\ &= (\rho^2 + q^2)\int_{-1}^1 \frac{dy}{|y|}\delta((\boldsymbol{\rho} - \mathbf{q}/y)^2 - q^2(1/y^2 - 1)) \end{aligned} \tag{54}$$

and change the order of integration over $\boldsymbol{\rho}$ and y. As a result, the integral over $\boldsymbol{\rho}$ becomes elementary and we obtain

$$J_2 = 4\pi q^2\int_0^1 \frac{dy}{y^3}\left[\mathrm{Re}\left(\frac{1-y}{1+y}\right)^{iZ\alpha} - 1 + 2(Z\alpha)^2 y^2\right]. \tag{55}$$

After replacing $y = \tanh\tau$, the integral over τ becomes tabular and for $J = J_1 + J_2$ we obtain

$$J = 8\pi q^2(Z\alpha)^2 \times\left[\ln\frac{2q}{\rho_0} - 1 + \mathrm{Re}\,\psi(1 + iZ\alpha) + C\right], \tag{56}$$

where $C = 0.577\ldots$ is the Euler constant, $\psi(x) = d\ln\Gamma(x)/dx$.

After calculating the integral over $\boldsymbol{\rho}$ in (51), the remaining integrals can be conveniently taken in the order: over $\mathbf{Q}$, $\mathbf{q}$, r, z, and ε. Finally the contribution of the first region is given by

$$M_1 = i\frac{28\alpha(Z\alpha)^2\omega}{9m^2} \times\left[\ln\frac{\omega\rho_0}{m} - i\frac{\pi}{2} - \mathrm{Re}\,\psi(1 + iZ\alpha) - C - \frac{47}{42}\right]. \tag{57}$$

The contribution of higher orders of perturbation theory with respect to the external field (Coulomb corrections) corresponds to the term $\mathrm{Re}\,\psi(1 + iZ\alpha) - C$ in (57) and agrees with the well-known result [17]. Thus, the Coulomb corrections are completely determined by the first region in which the quasiclassical Green function is not reduced to the eikonal one.

We shall now calculate the contribution M_2 of the second region. After differentiating over $\mathbf{r}_{1,2}$, calculating the trace of the γ-matrices, and taking the integral over $\boldsymbol{\theta}_r$, we arrive at the following representation for M_2:

$$M_2 = \frac{\alpha}{2\pi\omega}\int_0^{\omega} d\varepsilon\int dr\,r^2 \exp\left[-i\frac{\omega r m^2}{2\varepsilon\kappa}\right]\int_{\rho>\rho_0} d\boldsymbol{\rho}\int_{-\infty}^{\infty} dz \times\int_0^1\!\!\int_0^1 dx\,dy\left[2y(1-x)[2\vartheta(x-y)+1] - \frac{\omega^2}{2\varepsilon\kappa}\right] \times[\nabla_\rho V(R_{z-x})]\cdot[\nabla_\rho V(R_{z-y})], \tag{58}$$

where $R_s = r\sqrt{s^2 + \rho^2}$. In this formula we omit terms which are antisymmetric relative to the substitution $\varepsilon \longrightarrow \omega - \varepsilon$, $z \longrightarrow 1 - z$ whose integral is zero. Note that unlike the first region, in the second integration over z is performed within infinite limits. Expression (58) is simplified if in the third integral over x, y, z we transpose the variables $z \longrightarrow z + x$, $y \longrightarrow y + x$ and integrate over x. We obtain

$$M_2 = \frac{\alpha}{2\pi\omega}\int_0^{\omega} d\varepsilon\int dr\,r^2 \exp\left[-i\frac{\omega r m^2}{2\varepsilon\kappa}\right]\int_{\rho>\rho_0} d\boldsymbol{\rho}\int_{-\infty}^{\infty} dz$$

$$\times\int_0^1 dy(1-y)\left[\frac{4}{3}(1-y)^2 + 2y - \frac{\omega^2}{\varepsilon\kappa}\right] \times[\nabla_\rho V(R_z)]\cdot[\nabla_\rho V(R_{z-y})]. \tag{59}$$

For $r \ll r_c$ the potential is $V(r) \approx -Z\alpha/r$. It is therefore convenient to express the contribution (59) in the form $M_2 = M_2^c + \delta M$, where M_2^c is the value of M_2 for $V(r) = V_c(r) = -Z\alpha/r$.

For $V = V_c$ the integrals over r and ε are easily obtained and we find

$$M_2^c = -\frac{2i\alpha(Z\alpha)^2}{m^2}\int_{\rho_0}^{\infty} d\rho\,\rho^3\int_{-\infty}^{\infty} dz \times\int_0^1 dy(1-y)\frac{2(1-y)^2/9 + y/3 - 1}{[z^2+\rho^2]^{3/2}[(z-y)^2+\rho^2]^{3/2}}. \tag{60}$$

Using the Feynman parametrization of the denominators

$$\frac{1}{(AB)^{3/2}} = \frac{8}{\pi}\int_0^1 dv\frac{\sqrt{v(1-v)}}{[Av + B(1-v)]^3},$$

we take the integrals over ρ and z, and then over y and v bearing in mind that $\rho_0 \ll 1$. Finally for this (Coulomb) contribution we obtain

$$M_2^c = -i\frac{28\alpha(Z\alpha)^2\omega}{9m^2}\left(\ln\frac{\rho_0}{2} + \frac{31}{21}\right). \tag{61}$$

The sum of the contributions (57) and (61) gives the well-known expression for a Coulomb field [17]:

$$M_c = i\frac{28\alpha(Z\alpha)^2\omega}{9m^2} \times\left[\ln\frac{2\omega}{m} - i\frac{\pi}{2} - \mathrm{Re}\,\psi(1 + iZ\alpha) - C - \frac{109}{42}\right]. \tag{62}$$

In order to calculate the correction δM caused by the presence of screening, by subtracting from the expression in the integrand in (59) its value for $V = V_c$ we can replace ρ_0 in the lower limit of integration over ρ by zero. For the following calculations it is convenient to use the momentum representation for the potentials:

$$V(r) = \int\frac{d\mathbf{p}}{(2\pi)^3}e^{i\mathbf{p}\cdot\mathbf{r}}\tilde{V}(p).$$

Using this representation we take the integrals over $\boldsymbol{\rho}$ and z. The result of this integration is proportional to

$\delta(\mathbf{p}_1 - \mathbf{p}_2)$. Integrating over $\mathbf{p}_2$ and the angles of the vector $\mathbf{p}_1$, we have

$$\delta M = \frac{\alpha}{2\pi^3\omega}\int_0^\omega d\varepsilon\int dr\, r\exp\left[-i\frac{\omega r m^2}{2\varepsilon\kappa}\right]$$
$$\times\int_0^1 dy(1-y)\left[\frac{4}{3}(1-y)^2 + 2y - \frac{\omega^2}{\varepsilon\kappa}\right] \tag{63}$$
$$\times\int_0^\infty dp[p^4\tilde{V}^2(p) - (4\pi Z\alpha)^2]\left(\frac{\sin\zeta}{\zeta^3} - \frac{\cos\zeta}{\zeta^2}\right),$$

where $\zeta = rpy$. Changing from the variable r to ζ, we integrate by parts over p and take the integral over ζ. Making the transposition $\varepsilon \longrightarrow \omega\varepsilon$, we obtain

$$\delta M = \frac{\alpha\omega}{2\pi^3 m^2}\int_0^1 d\varepsilon\varepsilon(1-\varepsilon)\int_0^1 dy\left(\frac{1}{y} - 1\right)$$
$$\times\left[\frac{4}{3}(1-y)^2 + 2y - \frac{1}{\varepsilon(1-\varepsilon)}\right]\int_0^\infty dp(\partial_p p^4\tilde{V}^2(p)) \tag{64}$$
$$\times\left[\frac{1}{\eta} + \frac{1}{2}\left(1 - \frac{1}{\eta^2}\right)\ln\left(\frac{1+\eta}{1-\eta-i0}\right)\right],$$

where $\eta = 2\omega\varepsilon(1-\varepsilon)py/m^2$. We then transpose the variables $y \longrightarrow y/(2\varepsilon(1-\varepsilon))$, change the order of integration over ε and y, and take the integral over ε. Finally the contribution to the amplitude of Delbrück scattering associated with the screening is given by

$$\delta M = -\frac{\alpha\omega}{18\pi^3 m^2}\int_0^\infty dp(\partial_p p^4\tilde{V}^2(p))$$
$$\times\int_0^{1/2}\frac{dy}{y}\left[\frac{1}{\eta} + \frac{1}{2}\left(1 - \frac{1}{\eta^2}\right)\ln\left(\frac{1+\eta}{1-\eta-i0}\right)\right]$$
$$\times\left[(6y^2 + 7y + 7)\sqrt{1-2y}\right. \tag{65}$$
$$\left. + 3y(2y^2 - 3y - 3)\ln\left(\frac{1+\sqrt{1-2y}}{1-\sqrt{1-2y}}\right)\right],$$

where $\eta = \omega py/m^2$. This formula holds for an arbitrary type of screened Coulomb potential. In particular cases, this can be simplified substantially. For instance, for the potential

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

which corresponds to

$$\tilde{V}(p) = -4\pi Z\alpha/(p^2 + \beta^2),$$

all the integrals in (65) are taken:

$$\delta M = \frac{4i\alpha(Z\alpha)^2\omega}{9m^2}\left[33 - 13\tau^2 + \frac{3}{2}\tau^4\right.$$
$$\left. + \frac{1}{2}\tau(24 - 13\tau^2 + 3\tau^4)L + \frac{3}{8}(8 - 9\tau^2 + \tau^6)L^2\right], \tag{66}$$
$$\tau = \sqrt{1 + \frac{2im^2}{\omega\beta}}, \quad L = \ln\frac{\tau - 1}{\tau + 1}.$$

More realistic is the Moliére potential [18] for which

$$\tilde{V}(p) = -4\pi Z\alpha\sum_{n=1}^{3}\frac{\alpha_n}{p^2 + \beta_n^2}, \tag{67}$$

$$\alpha_1 = 0.1, \quad \alpha_2 = 0.55, \quad \alpha_3 = 0.35, \quad \beta_n = \beta_0 b_n,$$
$$b_1 = 6, \quad b_2 = 1.2, \quad b_3 = 0.3, \quad \beta_0 = \frac{mZ^{1/3}}{121}.$$

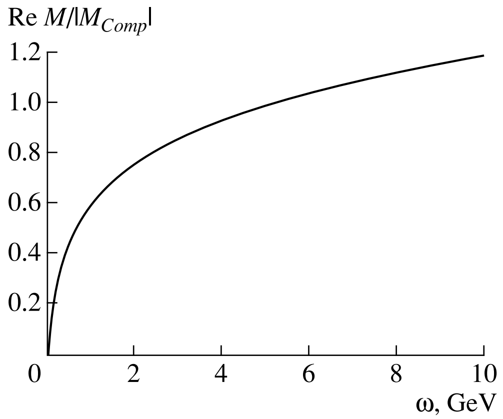
In this case we have

$$\delta M = -\frac{4i\alpha(Z\alpha)^2\omega}{9m^2}\int_0^{1/2}\frac{dy}{y}\left[(6y^2 + 7y + 7)\sqrt{1-2y}\right.$$
$$\left. + 3y(2y^2 - 3y - 3)\ln\left(\frac{1+\sqrt{1-2y}}{1-\sqrt{1-2y}}\right)\right]$$
$$\times\left\{\sum_{n\neq k}\alpha_n\alpha_k\left[1 - \frac{2i}{\gamma_n + \gamma_k} - \frac{2}{\gamma_n^2 - \gamma_k^2}\ln\left(\frac{\gamma_n + i}{\gamma_k + i}\right)\right]\right. \tag{68}$$
$$\left. + \sum_n\alpha_n^2\frac{\gamma_n}{\gamma_n + i}\right\},$$

where $\gamma_n = \omega\beta_n y/m^2$.

We know that the imaginary part of the Delbrück scattering amplitude is related to the total cross section for photon creation of electron–positron pairs in an external field by $\sigma = \mathrm{Im}M/\omega$. Using (62) and (68), we obtain the cross section for creation of electron–positron pairs for the Moliére potential:

$$\sigma = \frac{28\alpha(Z\alpha)^2}{9m^2}\left\{\ln\frac{2\omega}{m} - \mathrm{Re}\psi(1 + iZ\alpha) - C - \frac{109}{42}\right.$$
$$-\int_0^{1/2}\frac{dy}{y}\left[\left(\frac{6}{7}y^2 + y + 1\right)\sqrt{1-2y}\right. \tag{69}$$
$$\left. + \frac{3}{7}y(2y^2 - 3y - 3)\ln\left(\frac{1+\sqrt{1-2y}}{1-\sqrt{1-2y}}\right)\right]$$



Real part of the Delbrück scattering amplitude as a function of ω for the potential $V(r) = -Z\alpha\exp(-m\alpha Z^{1/3}r)/r$ in units of $4\pi Z\alpha/m$, $Z = 82$.

$$\times\left[\sum_{n\neq k}\alpha_n\alpha_k\left[1-\frac{1}{\gamma_n^2-\gamma_k^2}\ln\left(\frac{1+\gamma_n^2}{1+\gamma_k^2}\right)\right]+\sum_n\alpha_n^2\frac{\gamma_n^2}{1+\gamma_n^2}\right]\Bigg\},$$

which agrees with the well-known expressions in the literature (see, for example [19]).

We now discuss the dependence of the real part of the Delbrück amplitude on the photon energy ω. As long as $\omega/m^2 \ll r_c$, screening can be neglected and

$$\mathrm{Re}M = \mathrm{Re}M_c = 14\pi\alpha(Z\alpha)^2\omega/9m^2.$$

As ω increases, the real part increases more slowly and for $\omega/m^2 \gg r_c$ we obtain from (62) and (65)

$$\mathrm{Re}M \approx \frac{\alpha}{2\pi^3}\ln^2\left(\frac{\omega}{m^2r_c}\right)\int_0^\infty\frac{dp}{p}(\partial_p p^4\tilde{V}^2(p)). \qquad (70)$$

In this asymptotic form we only retained the term containing the leading power of the large logarithm. For illustration figure gives the real part of the Delbrück scattering amplitude as a function of ω for $Z = 82$ and the potential

$$V(r) = -Z\alpha\exp(-m\alpha Z^{1/3}r)/r.$$

One of the main mechanisms for elastic scattering of a photon is Compton scattering at atomic electrons. For forward scattering the amplitude of this process is a real quantity and does not depend on ω:

$$M_{Comp} = -4\pi Z\alpha/m.$$

It can be seen from figure that for photon scattering we need to allow for interference between the amplitudes of Compton and Delbrück scattering even at comparatively low energies.

For the case of Delbrück scattering we reaffirm that using the eikonal approximation to describe small-angle scattering processes at high energies without due substantiation may lead to incorrect results. For instance, if the Delbrück scattering amplitude is calculated using the eikonal Green function, the expression for the imaginary part of the amplitude proportional to the cross section for creation of electron–positron pairs will contain no Coulomb corrections.

REFERENCES

1. A. I. Milstein and V. M. Strakhovenko, Phys. Lett. A **95**, 135 (1983).
2. A. I. Milstein and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **85**, 14 (1983).
3. A. I. Milstein and V. M. Strakhovenko, Phys. Lett. A **90**, 447 (1982).
4. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **112**, 1921 (1997).
5. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Phys. Rev. A **57**, 2325 (1998).
6. R. N. Lee and A. I. Milstein, Phys. Lett. A **198**, 217 (1995).
7. R. N. Lee and A. I. Milsteĭn, Zh. Éksp. Teor. Fiz. **107**, 1393 (1995).
8. W. Furry, Phys. Rev. **46**, 391 (1934).
9. A. Sommerfeld and A. Maue, Ann. Phys. **22**, 629 (1935).
10. H. Olsen, L. C. Maximon, and H. Wergeland, Phys. Rev. **106**, 27 (1957).
11. A. I. Akhiezer, V. F. Boldyshev, and N. F. Shul'ga, Teor. Mat. Fiz. **23**, 11 (1975).
12. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay in Nonrelativistic Quantum Mechanics* (Israel Program for Scientific Translations, Jerusalem, 1966; Nauka, Moscow, 1971)
13. V. N. Baĭer and V. M. Katkov, Dokl. Akad. Nauk SSSR **227**, 325 (1976).
14. A. I. Milstein and M. Schumacher, Phys. Rep. **243**, 183 (1994).
15. A. I. Milstein and R. Zh. Shaisultanov, J. Phys. A **21**, 2941 (1988).
16. R. N. Lee, A. I. Milsteĭn, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **116**, 78 (1999).
17. M. Cheng and T. T. Wu, Phys. Rev. D **2**, 2444 (1970).
18. G. Z. Moliére, Z. Naturforsch. A **2**, 133 (1947).
19. H. Davies, H. A. Bethe, and L. C. Maximon, Phys. Rev. **93**, 788 (1954).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 79–92.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 90–104.
Original Russian Text Copyright © 2000 by Vagin, Silin, Uryupin.

PLASMA, GASES

Stimulated Brillouin Scattering in a Plasma with Ion-Acoustic Turbulence

K. Yu. Vagin*, V. P. Silin, and S. A. Uryupin**

Lebedev Physics Institute, Russian Academy of Sciences,
Moscow, 117924 Russia
**e-mail: vagin@sci.lebedev.ru*
***e-mail: silin@sci.lpi.ac.ru*

Received June 18, 1999

Abstract—A theory of stimulated Brillouin scattering (STBS) in a plasma with ion-acoustic turbulence is developed using concepts of parametric instability under conditions when equations of two-temperature hydrodynamics can be used to describe ion-acoustic perturbations of the electron density. The temporal growth rate of the absolute instability and the spatial gain of the scattered wave are determined. The dependence of the threshold density of the radiation flux on the angle between the scattering wave vector and the direction of anisotropy of the turbulent noise is described. A new effect of STBS forbiddenness caused by anomalous turbulent heating of the ions is predicted for a plasma with a high level of turbulent noise. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The phenomenon of stimulated Brillouin scattering has formed the subject of numerous experimental (see, for example, [1–5]) and theoretical [6–9] investigations. In general, the theoretical description of this phenomenon is based on a joint analysis of equations for the fields of the incident and scattered waves and an equation for the perturbation of the electron density created by the acoustic wave field. These equations have been studied in the greatest detail for a laminar plasma, where the electron and ion scattering processes are determined by pair collisions of charged particles. Two STBS regimes are usually distinguished. In one the electron density perturbation is determined by the ponderomotive action of a bilinear combination of incident and scattered wave fields. In the other the density perturbation is caused by competition between two physical phenomena: nonuniform heating of the electrons as a result of electron–ion collisions in the fields of the incident and scattered waves, and electron heat transfer. In this case we talk of the thermal mechanism for STBS. The perturbation of the electron density then depends strongly on the particular characteristics of heat transfer in a particular plasma state. The theory of thermal STBS has been developed in the greatest detail under conditions when the electron mean free paths are so short compared with the wavelength of the ion-acoustic waves that classical heat transfer theory can be used to describe the heat transfer [10]. Comparatively recently, STBS theory has been developed under conditions of nonlocal electron heat transfer in a laminar plasma [11–13]. The theoretically established strong dependence of thermal STBS on the electron scattering characteristics determining the heat transfer process has led to the observation of qualitatively new scattering laws in a plasma with ion-acoustic turbulence [14]. In this type of plasma the electron heat transfer behavior is determined to a considerable extent by electron scattering at ion-acoustic oscillations of the charge density. In [14] Ovchinnikov *et al.* put forward a theory of STBS for the case of fairly large scattering angles θ_0 when the condition

$$4Zk_0^2 l_t l_{ei} \sin^2(\theta_0/2) \ll 1, \tag{1.1}$$

is violated, where Z is the degree of ionization of the ions, k_0 is the wave number of the fundamental-frequency wave, l_t and l_{ei} are the electron mean free paths for scattering at ion-acoustic oscillations of the charge density and ions. Under conditions when the inequality opposite to (1.1) is satisfied, a correct description of scattering is obtained using concepts of nonlocal heat transfer in a turbulent plasma.

However, scattering at small angles when condition (1.1) is satisfied is also of interest. A theory of small-angle STBS is put forward in the present study. This theory differs substantially from that put forward in [14], first because of the possibility of using a detailed theory of local transport in a turbulent plasma and second, because the change in the turbulence noise as a result of long-wavelength ion-acoustic perturbations of the temperatures and densities of the plasma particles can be taken into account self-consistently. This last factor in particular not only allowed us to give a systematic description of STBS as a parametric instability but also revealed conditions for its forbiddenness as a result of anomalously effective turbulent ion heating.

1063-7761/00/9001-0079$20.00 © 2000 MAIK "Nauka/Interperiodica"

The frequency difference $\Delta\omega$ as a result of scattering is determined by the acoustic frequency ω_s,

$$\Delta\omega = \omega_s = 2k_0 v_s \sin\frac{\theta_0}{2}, \tag{1.2}$$

where v_s is the velocity of sound. Since the condition $4Zk_0^2 l_t l_{ei} > 1$ is usually satisfied, the equation (1.2) and the inequality (1.1) yield the constraint on the relative frequency difference:

$$\frac{\Delta\omega}{\omega_0} \ll \sqrt{\frac{m}{m_i}\frac{\nu_{ei}}{\omega_0}}\sqrt{\frac{l_{ei}}{l_t}} \ll 1, \tag{1.3}$$

where ω_0 is the pump wave frequency, m and m_i are the electron and ion masses, respectively, and ν_{ei} is the electron–ion collision frequency. The inequality (1.3) implies that in order to observe small-angle scattering experimentally we need to use lasers having a fairly narrow emission line width. At the same time, by virtue of the inequality $l_t \ll l_{ei}$, the conditions (1.1) and (1.3) are established more easily in a turbulent plasma compared with similar conditions in a laminar plasma. Hence, the constraint on the range of scattering angles in a turbulent plasma considered subsequently is less stringent than that in a laminar plasma.

In Section 2 we give the necessary information on the hydrodynamic equations for a turbulent plasma in an rf electromagnetic field. In Section 3 we use these equations as the basis to describe the damping of ion-acoustic oscillations having wavelengths much greater than the characteristic scales of the turbulent fluctuations of the charge density. For a plasma without an rf field we specify the conditions under which, as a result of the relatively strong ion damping, acoustic perturbations of the hydrodynamic quantities do not result in the buildup of instability. The STBS effect is considered in Section 4. A description of the scattering in the limit (1.1) is obtained using the field equations supplemented with a system of hydrodynamic equations to find the perturbations of the electron density in a plasma with ion-acoustic turbulence. We discuss the conditions under which a quasi-steady-state turbulence spectrum and electron fluxes can be established within the characteristic time of variation of the ion-acoustic density perturbations. A systematic description of STBS is given under these conditions when the buildup of hydrodynamic instability of the turbulent plasma is suppressed. We determine the dependence of the radiation flux density corresponding to the STBS threshold on the angle θ between the scattering wave vector and the anisotropy axis of the turbulent noise. We describe the dependence of the STBS threshold on the degree of nonisothermicity of the plasma and the turbulent noise distribution in wave vector space, and we show that in a plasma with a high level of turbulent noise, there are directions in which STBS is forbidden. This qualitatively new effect of STBS forbiddenness is caused by anomalously effective turbulent heating of the ions. The spatial gain of the scattered wave is determined. In Section 5 we give estimates to demonstrate the possibility of observing small-angle STBS experimentally.

2. HYDRODYNAMIC TRANSPORT EQUATIONS IN A TURBULENT PLASMA

We consider a nonisothermal plasma with ion-acoustic turbulence in an rf field

$$\frac{1}{2}\mathbf{E}(\mathbf{r}, t)\exp(-i\omega_0 t) + \text{c.c.}, \tag{2.1}$$

where the function $\mathbf{E}(\mathbf{r}, t)$ varies weakly with time over the period of the field $2\pi/\omega_0$. We shall assume that the fundamental frequency is much higher than the plasma frequency of the electrons ω_{Le}. This means that we can neglect the influence of low-frequency turbulent fluctuations of the charge density on the rapidly varying motion of the electrons of frequency ω_0 (see, for example, [15]). At the same time, the slow macroscopic motion of the electrons and ions is determined to a considerable degree by the ion-acoustic turbulence, and to describe this we use a system of hydrodynamic equations for the electron density n, the plasma velocity $\mathbf{u}$, and the temperatures of the electrons T and the ions T_i [16–18]:

$$\frac{\partial n}{\partial t} + \text{div}(n\mathbf{u}) = 0, \tag{2.2}$$

$$\begin{gathered}\left(\frac{\partial}{\partial t} + \mathbf{u}\frac{\partial}{\partial \mathbf{r}}\right)u_k = -\frac{Z}{nm_i}\frac{\partial}{\partial r_k}p + \frac{Z\eta}{nm_i}\left(\Delta u_k + \frac{1}{3}\frac{\partial}{\partial r_k}\text{div}\mathbf{u}\right)\\ -\frac{Ze^2}{4nmm_i\omega_0^2}\left\{n\frac{\partial}{\partial r_k}|\mathbf{E}|^2 - \frac{\partial}{\partial r_j}[n(E_kE_j^* + E_k^*E_j)]\right\},\end{gathered} \tag{2.3}$$

$$\begin{gathered}\left(\frac{\partial}{\partial t} + \mathbf{u}\frac{\partial}{\partial \mathbf{r}}\right)\left(T + \frac{e^2}{6k_Bm\omega_0^2}|\mathbf{E}|^2\right)\\ + \frac{2}{3nk_B}\text{div}\mathbf{q}\frac{2}{3}T\text{div}\mathbf{u} = \nu_{ei}\frac{e^2}{3k_Bm\omega_0^2}|\mathbf{E}|^2 - \nu_T T,\end{gathered} \tag{2.4}$$

$$\left(\frac{\partial}{\partial r} + \mathbf{u}\frac{\partial}{\partial \mathbf{r}}\right)T_i + \frac{2}{3}T_i\text{div}\mathbf{u} = Z\nu_T T. \tag{2.5}$$

In these equations e is the electron charge, $\eta = nk_BT_i/Z\nu_{ii}$ is the coefficient of viscosity, k_B is the Boltzmann constant, ν_{ii} is the ion–ion collision frequency, $p = nk_BT + nk_BT_i/Z$ is the pressure,

$$\nu_{ei} = \frac{4}{3}\frac{\sqrt{2\pi}Z\Lambda e^4 n}{\sqrt{m}(k_BT)^{3/2}} \tag{2.6}$$

is the electron–ion collision frequency, Λ is the Coulomb logarithm, ν_T is the turbulent temperature relaxation frequency, and $\mathbf{q}$ is the electron thermal flux.

Equations (2.2)–(2.5) were written assuming that the electric current density $\mathbf{j}$ is zero and the electroneutrality condition $n = Zn_i$ is satisfied, where n_i is the ion density. The condition $\mathbf{j} = 0$ is frequently used to analyze the hydrodynamics of a laser plasma. When a quasi-steady state of ion-acoustic turbulence is established as a result of competition between Cerenkov radiation from waves at electrons and induced scattering of waves by thermal ions, the explicit form of the heat flux $\mathbf{q}$ in (2.4) and the turbulent frequency ν_T in (2.4), (2.5) depend on the relationship between the density of the turbulence-exciting force $\mathbf{R}$ (A.1) and the quantity [19, 20]

$$R_{NL} = \frac{1}{6\pi} mn v_s \omega_{Li} \frac{r_{De}^2}{r_{Di}^2}, \tag{2.7}$$

where $v_s = \omega_{Li} r_{De}$ is the velocity of sound, $r_{De(i)}$ is the Debye length of the electrons (ions), and ω_{Li} is the ion Langmuir frequency. Using the condition for the absence of current $\mathbf{j} = 0$ and the relations (A.2), (A.4) we eliminate from $\mathbf{R}$ (A.1) the unknown quasi-steady-state electric field. In this case, the density of the force $\mathbf{R}$ and the electron temperature gradient ∇T are directed along the unit vector $\mathbf{n} = \nabla T/|\nabla T|$. From (A.5) for $R \gg R_{NL}$ we then find

$$R = \frac{3}{2} p |\nabla \ln T|, \tag{2.8}$$

$$\mathbf{q} = -\frac{128}{3\pi} \beta_{\parallel} p v_s \sqrt{\frac{R}{R_{NL}}} \mathbf{n} \equiv -\mathbf{n} q, \tag{2.9}$$

where $\beta_{\parallel} = 0.25$. If $R \ll R_{NL}$, the thermal flux (A.3) has the form

$$\mathbf{q} = -p v_s \times \left\{ \frac{25}{4} \left(1 - \beta_{\parallel}(\alpha) + \frac{\alpha}{12} \right) + \frac{128}{3\pi} \beta_{\parallel}(\alpha) \right\} \mathbf{n} \equiv -\mathbf{n} q, \tag{2.10}$$

where $\beta_{\parallel}(\alpha) = \beta_{\parallel} + C_{\beta\alpha}$, $\beta_{\parallel} = 0.177$, $C_{\beta} = 0.062$, and α is a small parameter

$$\alpha = \frac{\ln 2}{\ln(3\pi R_{NL}/8R)} \ll 1. \tag{2.11}$$

The force density R in expressions (2.10) and (2.11) is related to the modulus of the electron temperature gradient by

$$\frac{p|\nabla \ln T|}{R} = \frac{2}{3} + \frac{\pi}{16\beta_{\parallel}(\alpha)} \left(1 - \beta_{\parallel}(\alpha) + \frac{\alpha}{12} \right). \tag{2.12}$$

In these limiting cases for the turbulent frequency of temperature relaxation taking into account (2.8) and (2.12) we have

$$\nu_T = \frac{2}{3} a_T v_s \frac{R}{p} = a_* v_s |\nabla \ln T|, \tag{2.13}$$

where for $R \ll R_{NL}$ we have $a_T = 1$, $a_* = [1 + (3\pi/32\beta_{\parallel})(1 - \beta_{\parallel})]^{-1} = 0.42$, and for $R \gg R_{NL}$ we have $a_* = a_T = 1.3$. Note that in the last expression for ν_T in (2.13) in the limit $R \ll R_{NL}$ only the leading terms with respect to the small parameter (2.11) are retained. Allowing for the relations (2.6)–(2.13), the system of equations (2.2)–(2.5) for a plasma in a given pump field (2.1) is closed and can be used to study the hydrodynamic motion of a current-free plasma.

The order-of-magnitude turbulent frequency of temperature relaxation (2.13) can be written in the following form:

$$\nu_T \sim \frac{Zm}{m_i} \nu_t \sqrt{1 + \frac{R}{R_{NL}}},$$

where ν_t is the turbulent frequency of relaxation of the electron momentum:

$$\nu_t = \sqrt{\frac{9\pi}{8}} \frac{R}{nmv_s}, \quad R \ll R_{NL}, \tag{2.14}$$

$$\nu_t = \sqrt{\frac{9\pi}{8}} \frac{\sqrt{RR_{NL}}}{nmv_s}, \quad R \gg R_{NL}. \tag{2.15}$$

Under anomalous transport conditions we have $\nu_t \gg \nu_{ei}$. Hence the frequency ν_T appreciably exceeds the temperature relaxation frequency in a laminar plasma $\nu_\varepsilon \sim (m/m_i)\nu_{ei}$. The inequality

$$\nu_T \gg \nu_\varepsilon \tag{2.16}$$

means that we can neglect the contribution of electron–ion collisions to the heat exchange between electrons and ions in equations (2.4) and (2.5).

In order to describe STBS in a turbulent plasma as a parametric instability we shall now determine the ground state. We shall assume that in the ground state a plasma in an rf field (2.1) is characterized by the electron density n_0 and also by the electron temperature T_0 and the ion temperature T_{i0}. Let us assume that L_T is the characteristic scale of nonuniformity of the electron temperature in the ground state, as given by

$$\frac{1}{L_T} = |\nabla \ln T_0|. \tag{2.17}$$

We shall use the system of equations (2.2)–(2.5) to describe the interaction between rf radiation and a nonisothermal plasma whose hydrodynamic motion can be neglected, $\mathbf{u} = 0$. The plasma density can then be assumed to be independent of time and the evolution of the electron and ion temperatures in this ground state is described by the equations

$$\frac{\partial}{\partial t}T_0 + \frac{2}{3n_0k_B}\mathrm{div}\mathbf{q}_0 = (\nu_{ei})_0\frac{e^2}{3k_Bm\omega_0^2}|\mathbf{E}_0|^2 - a_*\frac{v_{s0}}{L_T}T_0, \tag{2.18}$$

$$\frac{\partial}{\partial t}T_{i0} = Za_*\frac{v_{S0}}{L_T}T_0, \tag{2.19}$$

which follow directly from (2.4), (2.5), (2.13), and (2.17). In equations (2.18)–(2.19) the values of $\mathbf{q}_0$, $(\nu_{ei})_0$, and v_{S0} are determined by the plasma parameters in the ground state and the amplitude of the pump field $\mathbf{E}_0$ is assumed to be given. It can be seen from equation (2.19) that in a nonisothermal plasma with a quasi-steady-state spectrum of ion-acoustic turbulence the ions undergo comparatively fast heating. In accordance with equation (2.19), the characteristic time for doubling of the ion temperature is $|\partial\ln T_{i0}/\partial t|^{-1} \sim (T_{i0}/ZT_0)L_T/v_{S0}$. As the ions are heated, the degree of nonisothermicity of the plasma ZT_0/T_{i0} decreases and at times of the order L_T/v_{S0} the ion temperature T_{i0} becomes comparable with ZT_0. When describing the state of a plasma with ion-acoustic turbulence for such long times, we need to take into account the comparatively strong Cerenkov attenuation of the sound by the ions, which leads to a substantial change in the electron heat transfer and a difference in the heat flux from that described by expressions (2.9) and (2.10). Subsequently, without analyzing time intervals greater than L_T/v_{S0}, we shall confine our study to STBS in a strongly nonisothermal plasma when Cerenkov attenuation of sound at ions can be neglected.

In addition to equations (2.18) and (2.19) describing the plasma ground state, in order to study STBS we also require equations for the small perturbations of the density $\delta n = n - n_0$, and the electron $\delta T = T - T_0$ and ion temperatures $\delta T_i = T_i - T_{i0}$, which vary on spatial scales much smaller than the scales of nonuniformity of the ground state. In our subsequent analysis of STBS, in order to avoid complicating the formulas we shall omit the subscript 0 from n_0, T_0, and T_{i0} which characterize the plasma ground state. Then for small-scale perturbations of the plasma hydrodynamic parameters we can write the following equations which follow directly from the initial equations (2.2)–(2.5), (2.9), (2.10), and (2.13):

$$\left\{\frac{\partial^2}{\partial t^2} - \frac{4}{3}\frac{Z\eta}{m_in}\Delta\frac{\partial}{\partial t} - v_s^2\Delta\right\}\delta n = \frac{Znk_B}{m_i}\Delta\left\{\delta T + \frac{\delta T_i}{Z}\right\} + \frac{Ze^2}{2m_im\omega_0^2}\left\{\Delta\delta|\mathbf{E}|^2 - \frac{\partial^2}{\partial r_k\partial r_j}\delta(E_kE_j^* + E_k^*E_j)\right\}, \tag{2.20}$$

$$\frac{\partial}{\partial t}\delta T + \frac{2}{3nk_B}\mathrm{div}\delta\mathbf{q} - \frac{2T}{3n}\frac{\partial}{\partial t}\delta n = \nu_{ei}\frac{2e^2\delta|\mathbf{E}|^2}{3k_Bm\omega_0^2} - \delta[\nu_TT], \tag{2.21}$$

$$\frac{\partial}{\partial t}\delta T_i - \frac{2T_i}{3n}\frac{\partial}{\partial t}\delta n = Z\delta[\nu_TT]. \tag{2.22}$$

Here $\delta|\mathbf{E}|^2$ and $\delta(E_kE_j^* + E_k^*E_j)$ are the perturbations of the bilinear combinations of rf field components, $\delta\mathbf{q}$ is the perturbation of the electron thermal flux density:

$$\delta\mathbf{q} = -q\left\{\frac{\nabla\delta T}{|\nabla T|} - \frac{1}{2}\frac{\nabla T(\nabla\delta T\cdot\nabla T)}{|\nabla T|^3} + \frac{3}{4}\left(\frac{\delta n}{n} + \frac{2}{3}\frac{\delta T_i}{T_i}\right)\frac{\nabla T}{|\nabla T|}\right\}, \quad R \gg R_{NL}, \tag{2.23}$$

$$\delta\mathbf{q} = -q\left\{\frac{\nabla T\times[\nabla\delta T\times\nabla T]}{|\nabla T|^3} + \frac{\delta n}{n}\frac{\nabla T}{|\nabla T|}\right\} - pv_s\frac{b}{\ln 2}\alpha^2\left[\frac{\nabla T(\nabla\delta T\cdot\nabla T)}{|\nabla T|} + \frac{\delta T_i}{T_i}\frac{\nabla T}{|\nabla T|}\right], \quad R \ll R_{NL}, \tag{2.24}$$

where the numerical coefficient is $b = (25/4)(1/12 - C_\beta) + (128/3\pi)C_\beta = 0.97$ and

$$\delta[\nu_TT] = a_*v_s\frac{\nabla T\cdot\nabla\delta T}{|\nabla T|} \tag{2.25}$$

is the perturbation of the electron–ion heat exchange. Note that expression (2.24) is written to within terms of the second order with respect to the small parameter α, (2.11) as is required for the following calculations.

3. LONG-WAVELENGTH ION-ACOUSTIC OSCILLATIONS IN A PLASMA WITH SHORT-WAVELENGTH TURBULENCE

In this section we shall analyze long-wavelength sound in a plasma with ion-acoustic turbulence. We shall denote by $\mathbf{k}$ the wave vector of the long-wavelength ion-acoustic oscillations. We assume that the wave vector of the acoustic oscillations satisfies the condition

$$Zk^2l_tl_{ei} \ll 1, \tag{3.1}$$

where $l_{ei} = v_T/\nu_{ei}$ and $l_t = v_T/\nu_t$ are the mean free paths of electrons having the thermal velocity v_T for electron–ion collisions and scattering of electrons by ion-acoustic oscillations of the charge density, respectively.

We know (see, for example, [19]) that turbulent noise is concentrated in a region of comparatively high wave numbers for which

$$k_t > \frac{\nu_{ii}\omega_{Le}r_{Di}^2}{\omega_{Li}^2 r_{De}^3}. \tag{3.2}$$

Instead of the vector **k** for the wave vectors of the turbulent charge density fluctuations in (3.2) we use the notation $\mathbf{k}_t$. It can be seen from a comparison of inequalities (3.1) and (3.2) that in the most interesting case when

$$\frac{\omega_{Le}}{\omega_{Li}}\left(\frac{ZT}{T_i}\right)^{1/2}\left(\frac{Zl_t}{l_{ei}}\right)^{1/2} > 1, \tag{3.3}$$

the long-wavelength perturbations being analyzed do not fall within the turbulence region since $k < k_t$. It should be stressed that turbulent noise with the wave vectors $k_t \sim 1/r_{De}$ from the interval (3.2) determines the transport processes in the plasma [19] and thereby strongly influences the long-wavelength perturbations of the hydrodynamic quantities.

In order to analyze long-wavelength sound we shall use equations (2.2)–(2.5), in which there is no rf field. We shall assume that the long-wavelength perturbation of the electron density has the form

$$\frac{1}{2}\delta n \exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}) + \text{c.c.} \tag{3.4}$$

We shall subsequently analyze those perturbations for which the wave vector **k** satisfies the condition

$$kL \gg 1, \tag{3.5}$$

where L is the smallest scale of nonuniformity of the hydrodynamic quantities n, T, and T_i in the ground state and the period $2\pi/\omega$ is much smaller than the characteristic time of variation of these quantities in the plasma ground state. In addition, we shall assume that the frequency of the perturbations is much lower than the turbulent relaxation frequency of the electron momentum $\omega \ll \nu_t$ and $1/t_s$, which is the reciprocal time taken to establish a quasi-steady-state turbulence spectrum in the region $k_t \sim 1/r_{De}$:

$$\omega \ll t_s^{-1} \sim \frac{\omega_{Li}^2}{\omega_{Le}}\frac{R/R_{NL}}{\sqrt{1 + R/R_{NL}}}. \tag{3.6}$$

Under the conditions (3.1), (3.5), and (3.6), first the distribution of turbulent noise with $k_t \sim 1/r_{De}$ which determines the electron fluxes and the energy relaxation frequency can follow slow changes in the hydrodynamic quantities. Second, in order to describe transport processes we can use expressions for the steady-state charge and heat fluxes (A.2)–(A.5). In connection with the use of hydrodynamic equations for the ions, we note that this approach is meaningful if

$$\nu_{ii} \gg \omega,\ kv_{Ti}, \tag{3.7}$$

where v_{Ti} is the ion thermal velocity.

In accordance with equations (2.2)–(2.5), perturbation of the density is accompanied by perturbations of the remaining hydrodynamic quantities:

$$\begin{gathered}\frac{1}{2}\delta\mathbf{u}\exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}) + \text{c.c.},\\ \frac{1}{2}\delta T_{(i)}\exp(-i\omega t + i\mathbf{k}\cdot\mathbf{r}) + \text{c.c.}\end{gathered} \tag{3.8}$$

Taking into account inequality (3.5) for small perturbations δn, $\delta T_{(i)}$, and $\delta\mathbf{u}$ we have from (2.2)–(2.5) the linearized system of equations

$$\omega\delta n = n(\mathbf{k}\cdot\delta\mathbf{u}), \tag{3.9}$$

$$\begin{aligned}\omega\mathbf{k}\cdot\delta\mathbf{u} &= \frac{k^2k_B}{nm_i}\{Z\delta(nT) + \delta(nT_i)\}\\ &\quad -\frac{4i}{3}\frac{Z\eta}{nm_i}k^2\mathbf{k}\cdot\delta\mathbf{u},\end{aligned} \tag{3.10}$$

$$\omega\delta T - \frac{2}{3nk_B}\mathbf{k}\cdot\delta\mathbf{q} - \frac{2}{3}T\mathbf{k}\cdot\delta\mathbf{u} = -i\delta[\nu_T T], \tag{3.11}$$

$$\omega\delta T_i - \frac{2}{3}T_i\mathbf{k}\cdot\delta\mathbf{u} = iZ\delta[\nu_T T]. \tag{3.12}$$

In equations (3.11) and (3.12) the explicit form of the heat flux perturbation $\delta\mathbf{q}$ and the perturbation of the energy fraction $\delta[\nu_T T]$ transferred from the electrons to the ions depends on the ratio of R to R_{NL} (2.7). Using (2.23) for $R \gg R_{NL}$ we find

$$\begin{aligned}\mathbf{k}\cdot\delta\mathbf{q} &= -ik^2\chi_t(\theta)\delta T\\ &\quad -\frac{32}{\pi}\beta_{\parallel}pv_s k\cos\theta\sqrt{\frac{3}{2}\frac{p}{R_{NL}L_T}}\left(\frac{\delta n}{n} + \frac{2\delta T_i}{3T_i}\right),\end{aligned} \tag{3.13}$$

where θ is the angle between the wave vector of the perturbations **k** and the unit vector $\nabla T/|\nabla T|$ along which the heat flux (2.9) propagates, and $\chi_t(\theta)$ is the effective thermal conductivity

$$\chi_t(\theta) = \frac{32}{\pi}\beta_{\parallel}nk_Bv_s\sqrt{\frac{2}{3}\frac{p}{R_{NL}}}L_T(1 + \sin^2\theta). \tag{3.14}$$

Then, taking into account expression (2.24) for $R \ll R_{NL}$ we have

$$\begin{aligned}\mathbf{k}\cdot\delta\mathbf{q} &= -ik^2\chi_t(\theta)\delta T\\ &\quad - pv_s k\cos\theta\left[(\alpha + b\alpha)\frac{\delta n}{n} + \frac{b}{\ln 2}\alpha^2\frac{\delta T_i}{T_i}\right],\end{aligned} \tag{3.15}$$

$$\begin{aligned}\chi_t(\theta) &= nk_Bv_sL_T\\ &\quad \times\left[(a + b\alpha)\sin^2\theta + \frac{b}{\ln 2}\alpha^2\cos^2\theta\right],\end{aligned} \tag{3.16}$$

where α is determined by formula (2.11), and a is a numerical coefficient:

$$a = \frac{25}{4}(1-\beta_{\parallel}) + \frac{128}{3\pi}\beta_{\parallel} = 7.55. \qquad (3.17)$$

In these limiting cases, to within small corrections of the order $1/kL \ll 1$ or $\alpha^2 \ll 1$ for the perturbation of the term describing energy exchange between electrons and ions (2.25), taking into account (2.13) we have

$$\frac{\delta[\nu_T T]}{\nu_T T} = i\frac{\delta T}{T}kL_T\cos\theta. \qquad (3.18)$$

In accordance with (3.18), the small-scale (3.5) hydrodynamic perturbations (3.4), (3.8) increase the relative change in the energy fraction transferred from the electrons to the ions by a factor of $kL_T|\cos\theta| \ll 1$ for angles θ far from $\pi/2$. It can be seen from the definition (2.13) and relations (2.8) and (2.12) that the additional parameter $kL_T\cos\theta$ appears in (3.18) because the projection of the scale of nonuniformity of the perturbations on the direction of nonuniformity in the unperturbed state is determined by the value of $1/k\cos\theta$ rather than by L_T.

From an analysis of equations (3.9)–(3.18) we can obtain the following dispersion equation linking the complex frequency ω with the perturbation wave vector **k**:

$$\Bigg\{\omega^2 + 2i\omega\gamma_i - k^2 v_{sT}^2 - \omega_s^2\left(1 - \frac{\nu_T}{\omega}kL_T\cos\theta\right)$$
$$\times\left(\omega - \omega_s a_q\frac{q}{pv_s}\cos\theta\right)\left[i\frac{k^2}{nk_B}\chi_t(\theta) \qquad (3.19)\right.$$
$$\left. - \frac{ZT}{T_i}kL_T\frac{\nu_T}{pv_s}\frac{\omega_s}{\omega}\left(\frac{\partial q}{\partial \ln R}\right)\cos^2\theta\right]^{-1}\Bigg\} = 0,$$

where $\omega_s = kv_s$ is the frequency of ion-acoustic oscillations of wavelength much greater than the electron Debye length, $a_q = 1$ for $R \ll R_{NL}$ and $a_q = 13/12$ for $R \gg R_{NL}$,

$$\gamma_i = \frac{2}{3}k^2\frac{Z}{nm_i}\eta \qquad (3.20)$$

is the sound attenuation decrement at the ions. When deriving equation (3.19), we assumed that

$$kL_T \gg 1, \qquad (3.21)$$

$$ZT/T_i \gg 1. \qquad (3.22)$$

Inequality (3.22) corresponds to the conditions for the existence of ion-acoustic turbulence [19, 20]. It should be noted that under the condition (3.22), the difference between the expression for the velocity of sound $v_{sT} = \sqrt{k_B/m_i(ZT + 5T_i/3)}$, which is isothermal for the electrons and adiabatic for the ions [16], and the ion acoustic velocity $v_s = \sqrt{Zk_BT/m_i}$ is small and will subsequently be neglected.

In the limit $R \gg R_{NL}$, inequality (3.21) ensures that equation (3.19) is satisfied for all angles θ. In the opposite limit $R \ll R_{NL}$, and for small angles $\theta \ll \alpha$ equation (3.19) is satisfied if

$$\alpha^2\max(kL_T, ZT/T_i) \gg 1. \qquad (3.23)$$

In the same limit, but for $\theta \gg \alpha$, the condition for validity of (3.19) is less stringent

$$\alpha^2 ZT/T_i + kL_T\sin^2\theta \gg 1. \qquad (3.24)$$

The dispersion equation (3.19) has solutions describing acoustic oscillations in a plasma at frequency ω which differs from the acoustic frequency ω_s by a small imaginary correction $\omega \simeq \omega_s + i\gamma$, $|\gamma| \ll \omega_s$. Neglecting the small variation in the ion-acoustic oscillation frequency in the following analysis, we find from (3.19)

$$\gamma = -\gamma_s(\theta) \equiv -\gamma_i - \gamma_e(\theta), \qquad (3.25)$$

where the function $\gamma_e(\theta)$ characterizing the interaction between the ion-acoustic waves and the electrons, has the form

$$\gamma_e(\theta) = \frac{\nu_t}{2}\frac{\omega_{Li}^2}{\omega_{Le}^2}\psi(\theta)$$
$$\times(1 - a_*\cos\theta)\left(1 - a_q\frac{q}{pv_s}\cos\theta\right). \qquad (3.26)$$

The function $\psi(\theta)$ which determines $\gamma_e(\theta)$ (3.26) depends strongly on the thermal conductivity in the turbulent plasma $\chi_t(\theta)$ (3.14), (3.16) and is positive for all angles θ. The explicit form of the function $\psi(\theta)$ in (3.26) depends on the ratio R/R_{NL}. For $R \ll R_{NL}$

$$\psi(\theta) = \frac{4\sqrt{2}}{9\sqrt{\pi}a_*}\left[a\sin^2\theta + \frac{b}{\ln 2}\alpha^2\cos^2\theta\right]$$
$$\times\left\{\left[a\sin^2\theta + \frac{b}{\ln 2}\alpha^2\cos^2\theta\right]^2 \qquad (3.27)\right.$$
$$\left. + \left[\frac{a_*}{kL_T}\frac{ZT}{T_i}\frac{b}{\ln 2}\alpha^2\cos^2\theta\right]^2\right\}^{-1} > 0,$$

and in the opposite limit $R \gg R_{NL}$

$$\psi(\theta) = \frac{\sqrt{2\pi}}{48\beta_{\parallel}}(2-\cos^2\theta)$$
$$\times\left[(2-\cos^2\theta)^2 + \left(\frac{a_*}{kL_T}\frac{ZT}{T_i}\cos^2\theta\right)^2\right]^{-1} > 0. \quad (3.28)$$

In the range of angles where $\gamma_e(\theta) > 0$, expression (3.26) describes the attenuation of ion-acoustic waves at electrons, and for those angles where $\gamma_e(\theta) < 0$, wave generation takes place.

As has already been noted [18], the existence of anomalous transport means that it is necessary to analyze the stability of the resulting hydrodynamic structures. It can be seen from expression (3.25) that a plasma with advanced ion-acoustic turbulence is stable with respect to the small perturbations (3.4), (3.8) under study if the following inequality is satisfied

$$\gamma_i > \max(-\gamma_e(\theta)). \quad (3.29)$$

In the limit $R \ll R_{NL}$, (3.29) may be expressed in the form

$$\min\left(1, \frac{kL_T}{0.84ZT/T_i}\right)\frac{1.3}{\alpha^2}\frac{\Lambda_i}{\Lambda}\left(\frac{ZT}{T_i}\right)^{5/2}\sqrt{\frac{Zm}{m_i}} < k^2 l_t l_{ei}, \quad (3.30)$$

and conversely, when $R \gg R_{NL}$ condition (3.29) corresponds to the relationship

$$\frac{0.04}{1+\dfrac{0.1}{kL_T}\dfrac{ZT}{T_i}}\frac{\Lambda_i}{\Lambda}\sqrt{\frac{R}{R_{NL}}}\left(\frac{ZT}{T_i}\right)^{5/2}\sqrt{\frac{Zm}{m_i}} < k^2 l_t l_{ei}, \quad (3.31)$$

where Λ_i is the ion Coulomb logarithm. When analyzing inequalities (3.30) and (3.31), we must bear in mind that these should be satisfied together with condition (3.1).

4. STIMULATED BRILLOUIN SCATTERING

Returning to our analysis of the scattering of electromagnetic radiation in a plasma with ion-acoustic turbulence, we express the rf field (2.1) as the sum of the pump field

$$\frac{1}{2}\mathbf{E}_0\exp(-i\omega_0 t + i\mathbf{k}_0\cdot\mathbf{r}) + \text{c.c.} \quad (4.1)$$

and the scattered wave

$$\frac{1}{2}\mathbf{E}_{-1}\exp(-i\omega_{-1} t + i\mathbf{k}_{-1}\cdot\mathbf{r}) + \text{c.c.}, \quad (4.2)$$

whose frequencies ω_0 and ω_{-1} are related to the wave vectors $\mathbf{k}_0$ and $\mathbf{k}_{-1}$ by

$$\omega_0^2 = \omega_{Le}^2 + k_0^2c^2, \quad \omega_{-1}^2 = \omega_{Le}^2 + k_{-1}^2c^2, \quad (4.3)$$

where c is the velocity of light in vacuum. The frequencies and wave vectors of the incident and scattered waves differ by the frequency ω and the wave vector $\mathbf{k}$ of the electron density perturbations

$$\omega_{-1} = \omega_0 - \omega, \quad \mathbf{k}_{-1} = \mathbf{k}_0 - \mathbf{k}. \quad (4.4)$$

The density perturbation itself has the form (3.4).

Assuming that the field $\mathbf{E}_0$ of the fundamental-frequency wave is given, we use the following equation to determine the amplitude $\mathbf{E}_{-1}$ of the scattered wave field

$$\left[\omega_{-1}^2 - k_{-1}^2c^2 - \omega_{Le}^2\left(1 + i\frac{\nu_{ei}}{\omega_{-1}}\right)\right]\mathbf{E}_{-1}^* = \frac{\delta n}{2n}\omega_{Le}^2\mathbf{E}_0^*, \quad (4.5)$$

which contains the amplitude of the electron density perturbations to be determined self-consistently. The incident $\mathbf{E}_0$ and scattered $\mathbf{E}_{-1}$ waves are plane transverse electromagnetic waves for which the conditions $\mathbf{k}_0\cdot\mathbf{E}_0 = \mathbf{k}_{-1}\cdot\mathbf{E}_{-1}$ are satisfied and as can be seen from (4.5) the directions of the vectors $\mathbf{E}_0$ and $\mathbf{E}_{-1}$ are the same. In order to search for δn we shall use the system of hydrodynamic equations (2.20)–(2.22) given in Section 2 which, for the perturbations (3.4) and (3.8), may be written in the form

$$\{\omega^2 + 2i\omega\gamma_i - \omega_s^2\}\delta n$$
$$= \frac{Znk_B}{m_i}k^2\left\{\delta T + \frac{\delta T_i}{Z}\right\} + \frac{Z}{2m_i}k^2\frac{e^2\mathbf{E}_0\cdot\mathbf{E}_{-1}^*}{m\omega_0^2}, \quad (4.6)$$

$$\omega_s\delta T - \frac{2}{3}\omega_s T\frac{\delta n}{n} - \frac{2}{3nk_B}\mathbf{k}\cdot\delta\mathbf{q}$$
$$= i\nu_{ei}\frac{2e^2\mathbf{E}_0\cdot\mathbf{E}_{-1}^*}{3k_Bm\omega_0^2} + a_*kv_s\delta T\cos\theta, \quad (4.7)$$

$$\omega_s\delta T_i - \frac{2}{3}\omega_s T_i\frac{\delta n}{n} = -a_*kv_sZ\delta T\cos\theta. \quad (4.8)$$

In equations (4.7) and (4.8) we neglect the small difference between the frequency ω and the acoustic frequency ω_s. Unlike the system (3.9)–(3.12) which describes long-wavelength acoustic perturbations in a turbulent plasma, the system (4.6)–(4.8) contains terms proportional to $\mathbf{E}_{-1}^*$ and together with equation (4.5), describes the scattering of the pump wave (4.1) at low-frequency long-wavelength acoustic perturbations of the charge density.

We consider the contribution to the pressure perturbation on the right-hand side of equation (4.6)

$$k_Bn(\delta T + \delta T_i/Z), \quad (4.9)$$

which is attributable to perturbations of the electron δT and ion temperatures δT_i. From equation (4.8) we obtain for the ion temperature perturbation

$$\delta T_i = \frac{2}{3}T_i\frac{\delta n}{n} - a_* Z\delta T\cos\theta. \tag{4.10}$$

The first term on the right-hand side (4.10) is responsible for the adiabatic contribution of the ions to the velocity of sound, well known in the hydrodynamics of a laminar plasma. As we showed in the previous section, in the limits of condition (3.22) this contribution to the velocity of sound is small compared with the isothermal contribution of the electrons. We shall subsequently neglect this contribution. Unlike a laminar plasma, in a turbulent plasma the ion temperature increment as a result of energy exchange with electrons described by the second term on the right-hand side of (4.10) is by no means small. This substantial increase in the contribution of electron–ion heat exchange to δT_i is caused first by an increase in the temperature relaxation frequency in a turbulent plasma (2.16) compared with a laminar plasma and second, by an increase in the relative variation of the energy fraction (3.18) transferred from the electrons to the ions when $kL_T|\cos\theta| \gg 1$. Taking into account (4.10) and these observations we obtain for the pressure perturbation

$$k_B n(\delta T + \delta T_i/Z) = k_B n(1 - a_*\cos\theta)\delta T. \tag{4.11}$$

For $R > R_{NL}$ when $a_* = 1.3$ it follows from the right-hand side of (4.11) that angles θ exist for which the contribution to the pressure perturbation (4.9) is in antiphase with the electron temperature perturbation δT. This occurs first when the modulus of the contribution to the ion temperature perturbation δT_i caused by heat exchange with electrons exceeds $|Z\delta T|$ and second, is in antiphase with δT.

The contribution to the last term on the left-hand side of equation (4.7) made by the electron thermal conductivity (3.14), (3.16) which appears in the divergence of the electron thermal flux (3.13), (3.15) may be written (in order of magnitude) in the form

$$\frac{k^2}{nk_B}\chi_t(\theta)\delta T \sim k v_{eff}(k)\delta T. \tag{4.12}$$

Here $v_{eff}(k) \sim (kL_T)\sqrt{1 + R/R_{NL}}\, v_s$ characterizes the effective rate of heat removal by electron heat transfer from the region of action of the heating field. The first term on the left-hand side of (4.7), corresponding to the time variation of δT, is given by $kv_s\delta T$. Allowing for (3.21), this is small compared with (4.12). Then, the last term on the right-hand side of (4.7) describing energy exchange between electrons and ions is also given in order of magnitude by the expression $\sim kv_s\delta T$ and also makes a small contribution to the electron heat balance compared with (4.12). Bearing all this in mind, substituting (4.10) into (4.7), and using (3.13)–(3.16) and (3.26)–(3.28) we obtain the electron temperature perturbation from (4.7)

$$\delta T = \frac{e^2\mathbf{E}_0\cdot\mathbf{E}^*_{-1}}{k_B m\omega_0^2}\frac{\psi(\theta)}{k^2 l_t l_{ei}} - 2i\frac{\gamma_e(\theta)}{\omega_s}T\frac{\delta n}{n}. \tag{4.13}$$

When writing (4.13) we neglected small quantities of the order $1/kL_T$ which make small corrections to the real frequency ω_s. The first term on the right-hand side of (4.13) describes the perturbation of the electron temperature δT caused by mixing of the scattered wave field $\mathbf{E}^*_{-1}$ with the pump field $\mathbf{E}_0$ and corresponds to the balance of the nonuniform heating of the electrons by inverse bremsstrahlung absorption of this interference field and the electron heat transfer. The second term on the right-hand side of (4.13) corresponds to the small dissipative contribution of the electrons to the ion sound increment $\gamma_s(\theta)$.

Eliminating δT and δT_i using (4.10) and (4.13), instead of the system (4.6)–(4.8) we write a single equation to determine the density perturbation created by the bilinear combination of the fields of the incident and scattered waves:

$$[\omega^2 - \omega_s^2 + 2i\omega\gamma_s(\theta)]\frac{\delta n}{n} = \frac{1}{2}\mathbf{E}_0\cdot\mathbf{E}^*_{-1}\left(\frac{e\omega_s}{m\omega_0 v_T}\right)^2 \times\left[1 + 2(1 - a_*\cos\theta)\frac{\psi(\theta)}{k^2 l_t l_{ei}}\right]. \tag{4.14}$$

The terms in the brackets on the right-hand side of formula (4.14) are of a different nature. The first term, one, occurs as a result of the ponderomotive action of the fields. The second, which contains the small parameter $k^2 l_t l_{ei} \ll 1$ in the denominator [see (3.1)] describes the density perturbation as a result of the contribution to the pressure perturbation (4.11) generated by the bilinear interference combination of the fields $\mathbf{E}_0\cdot\mathbf{E}^*_{-1}$.

By jointly analyzing equations (4.5) and (4.14) we can obtain the following dispersion equation which describes parametric instability and links the frequency ω, which is generally complex, with the perturbation wave vector $\mathbf{k}$

$$\begin{aligned} &-(\omega - \Delta + i\gamma_E)[\omega^2 - \omega_s^2 + 2i\omega\gamma_s(\theta)] \\ &= \frac{\omega_s^2}{8\omega_0 r_{De}^2}v_E^2\left[1 + 2(1 - a_*\cos\theta)\frac{\psi(\theta)}{k^2 l_t l_{ei}}\right], \end{aligned} \tag{4.15}$$

where $v_E = |e\mathbf{E}_0/m\omega_0|$ is the amplitude of the electron oscillation velocity in the field of the fundamental-frequency wave,

$$\Delta = \frac{c^2}{2\omega_0}(2\mathbf{k}\cdot\mathbf{k}_0 - k^2), \quad \gamma_E = \nu_{ei}\frac{\omega_{Le}^2}{2\omega_0^2}. \tag{4.16}$$

When deriving equation (4.15) and also when analyzing the acoustic perturbations in the absence of the pump field, we assumed that the conditions (3.21)–(3.24) were satisfied. In addition to these inequalities, we used the following constraint to derive equation (4.15)

$$\omega \ll \omega_0, \tag{4.17}$$

which is known to be satisfied for $\omega \sim \omega_s \ll \omega_0$.

In order to complete the picture, we note that nonrelativistic motion of the plasma at the velocity $u \ll c$ leads to a Doppler shift of the frequency ω and is taken into account by replacing ω with $\omega' = \omega - \mathbf{k} \cdot \mathbf{u}$ in equation (4.15).

We shall also assume that $\omega \simeq \omega_s + i\gamma$ where $|\gamma| \ll \omega_s$ (not to be confused with γ from Section 3). We shall analyze the conditions for which the following inequality holds for the resonance detuning Δ

$$\Delta = \omega_s$$

and the STBS process takes place more efficiently. Then, to determine the imaginary correction γ we have from (4.15)

$$(\gamma + \gamma_E)[\gamma + \gamma_s(\theta)] = \frac{1}{4}W_E(\theta), \tag{4.18}$$

where $\gamma_s(\theta)$ is described by relations (3.25)–(3.28), and

$$W_E(\theta) = \frac{\omega_s}{4\omega_0}\frac{v_E^2}{r_{De}^2}\left[1 + 2(1 - a_*\cos\theta)\frac{\psi(\theta)}{k^2 l_t l_{ei}}\right]. \tag{4.19}$$

As we noted in Section 3, the function $\psi(\theta)$ determining $W_E(\theta)$ (3.27), (3.28) depends strongly on the thermal conductivity in the turbulent plasma $\chi_t(\theta)$ (3.14), (3.16) and is positive for all values of the angle θ.

The expression (4.19) describes the nonlinear interaction of the scattered (4.2) and acoustic (3.4) waves and its sign characterizes the possibility of STBS. When (4.19) is positive, STBS may occur, whereas when it is negative STBS is impossible. Since under the conditions being discussed (3.1) the parameter is $k^2 l_t l_{ei} \ll 1$, the main contribution to the nonlinear interaction is made by the pressure perturbation (4.11) which corresponds to the thermal mechanism of STBS. Thus, for $R > R_{NL}$ for angles θ for which the contribution to the pressure perturbation and the bilinear combination $\mathbf{E}_0 \cdot \mathbf{E}_{-1}^*$ generating this are phase-shifted by 180°, i.e., $(1 - a_*\cos\theta) < 0$, the sign of the nonlinear interaction is negative and STBS cannot develop. It should be noted that for a fairly high degree of nonisothermicity of the plasma when the condition $ZT/T_i > 0.5L_T(l_t l_{ei})^{-1/2} > 1$ is satisfied, even in the limits (3.1) STBS is determined by the ponderomotive mechanism and it is impossible for it to be forbidden.

In equation (4.18) the absolute value of the decrement $|\gamma_s(\theta)|$ is small compared with the frequency ω_s, which is itself low in this particular limit (3.1) of fairly long wavelengths. For these laser plasma parameters the value of $|\gamma_s(\theta)|$ may be either larger or smaller than $2\pi/\tau$ which is the reciprocal duration of the laser pulses used to investigate STBS. Consequently, it is interesting to discuss the two possibilities. We shall first analyze the situation when the pump pulse duration τ is long and the following inequality holds

$$|\gamma_s(\theta)|\tau \gg 2\pi. \tag{4.20}$$

We shall consider the conditions under which the summed sound attenuation decrement at the ions and electrons $\gamma_s(\theta)$ is positive for all values of the angle θ, i.e., inequality (3.29) is satisfied and no hydrodynamic instability develops in the plasma. The square root of equation (4.18) corresponding to the possibility of STBS then has the form

$$\begin{aligned}\gamma &= -\frac{1}{2}[\gamma_E + \gamma_s(\theta)] \\ &+ \frac{1}{2}\sqrt{[\gamma_E - \gamma_s(\theta)]^2 + W_E(\theta)}.\end{aligned} \tag{4.21}$$

In accordance with (4.21), the threshold radiation intensity for STBS is obtained from

$$W_E(\theta) = 4\gamma_E\gamma_s(\theta) > 0. \tag{4.22}$$

Estimates show that for most laser plasmas the condition $|\gamma_s(\theta)| \ll \gamma_E$ holds and the solution of (4.21) near the threshold may be expressed in the following simple form:

$$\gamma \simeq -\gamma_s(\theta) + \frac{W_E(\theta)}{4\gamma_E}. \tag{4.23}$$

Then using relations (3.20), (3.25), (3.27), (3.28), and (4.16) we express condition (4.22) in the form

$$\begin{aligned}\frac{(v_E^2)_{th}}{v_T^2} &= 8\frac{\nu_{ei}}{\omega_0}\frac{\gamma_i + \gamma_e(\theta)}{\omega_s} \\ &\times \frac{1}{1 + 2(1 - a_*\cos\theta)\psi(\theta)/k^2 l_t l_{ei}} > 0.\end{aligned} \tag{4.24}$$

We shall discuss expression (4.24) which determines the pump radiation intensity at which STBS occurs. As usual in STBS theory, the presence of the two small parameters ν_{ei}/ω_0 and $\gamma_s(\theta)/\omega_s$ on the right-hand side of (4.24) indicates that STBS occurs at a comparatively low radiation intensity when the amplitude of the electron oscillation velocity in the fundamental-frequency wave field is considerably lower than the electron thermal velocity. When discussing the relation (4.24), it is convenient to introduce the scattering angle θ_0 between the wave vectors of the incident $\mathbf{k}_0$ and scattered waves $\mathbf{k}_{-1}$, which has already been used in formulas (1.1) and (1.2). For the wave vector $\mathbf{k}$ of the acoustic waves which determines the scattering we then have

$$k = 2k_0\sin\frac{\theta_0}{2}.$$

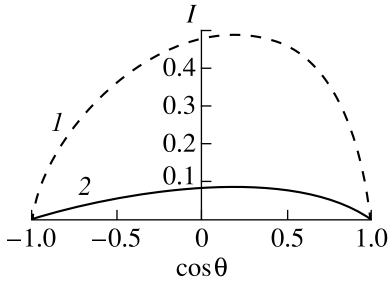


Fig. 1. Dependence of the STBS threshold on the angle between the vector of the ion acoustic waves and the force density vector **R**. The curves $I(\theta)$ were obtained for $\alpha = 0.3$, $ZT/T_i = 10$, $kL_T = 50$, $((\nu_t/2\gamma_i)(\omega_{Li}^2/\omega_{Le}^2) = 0.02$, and $k^2 l_t l_{ei} = 0.1$ (*1*) and 0.01 (*2*).

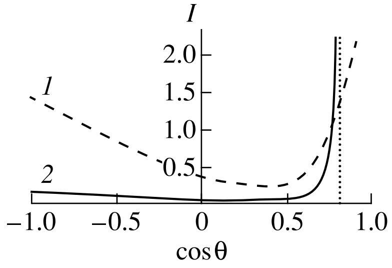


Fig. 2. Form of the function $I(\theta)$ for $R = 16R_{NL}$, $kL_T = 10$, $(\nu_t/2\gamma_i)(\omega_{Li}^2/\omega_{Le}^2) = 2$. The other parameters are the same as in Fig. 1.

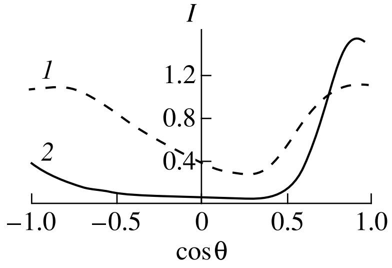


Fig. 3. Form of the function $I(\theta)$ for $R = 16R_{NL}$. Unlike Fig. 2, the nonisothermicity of the plasma is five times higher, $ZT/T_i = 50$. The other parameters are the same.

Bearing in mind this definition and considering the case when conditions (3.30) and (3.31) are satisfied, we have approximately from (4.24)

$$\frac{(v_E^2)_{th}}{v_T^2} = \frac{32}{3} \times \frac{(T_i/ZT)(\nu_{ei}/\nu_{ii})(k_0 v_s/\omega_0)\sin(\theta_0/2)}{1 + (1 - a_*\cos\theta)\psi(\theta)/(2k_0^2 l_t l_{ei}\sin^2(\theta_0/2))} > 0. \tag{4.25}$$

It follows from (4.25) that as the significant dependence corresponding to the threshold of the radiation flux density as a function of the angle θ decreases, where θ is the angle between the scattering vector and the anisotropy axis of the turbulent noise, the STBS forbiddenness is lifted. This dependence is observed particularly clearly in the limit $R > R_{NL}$ when, as has already been noted in the discussion of equation (4.14) and relation (4.19), the phase shift between the contribution to the pressure perturbation (4.9) and the bilinear field combination $\mathbf{E}_0 \cdot \mathbf{E}_{-1}^*$ changes by 180°. The change in the phase matching for small θ_0 has the result that instead of a buildup of ion-acoustic oscillations of the density, these are suppressed. As a result of this suppression of the ion-acoustic oscillations STBS is forbidden for those angles θ where expression (4.25) is negative. The functions

$$I(\theta) = \frac{(v_E^2)_{th}}{v_T^2}\frac{\omega_0}{8\nu_{ei}}\frac{\omega_s}{\gamma_i}$$

over the entire range of angles θ are plotted in Figs. 1–3 for values of the parameter $(\nu_t/2\gamma_i)(\omega_{Li}^2/\omega_{Le}^2)$ satisfying condition (3.29). In accordance with Fig. 1, when a small force acts on the electrons where $R \ll R_{NL}$ and $(\nu_t/2\gamma_i)(\omega_{Li}^2/\omega_{Le}^2) = 0.02$, the STBS threshold is lowest for those electron density perturbations whose wave vector **k** is directed along **R**. For vectors **k** oriented in the transverse directions to **R** the STBS threshold is considerably higher. For example, for $R \gg R_{NL}$ and $(\nu_t/2\gamma_i)(\omega_{Li}^2/\omega_{Le}^2) = 2$, and comparatively weak nonisothermicity when, for example $kL_T = ZT/T_i = 10$, the STBS threshold is lowest for vectors **k** directed at large angles to **R** (see Fig. 2). For directions **k** close to **R** the STBS threshold increases sharply. This increase is particularly appreciable for $k^2 l_t l_{ei} \ll 1$ and as follows from (4.19), (4.24), (4.25), and Fig. 2, may lead to STBS forbiddenness for those angles where $W_E(\theta) < 0$. It should be stressed that formula (4.25) is only meaningful for those angles θ where its denominator is positive. The condition that the denominator of (4.25) vanishes corresponds to the limiting angle θ_* which separates the θ angle space into the region where STBS can develop and the region where STBS is forbidden. It follows from (4.23) that for vectors **k** falling within the range of angles for which $W_E(\theta) < 0$, the decrement γ is negative and no perturbations grow. The dependences plotted in Fig. 2 change significantly as the plasma nonisothermicity increases. It can be seen from Fig. 3 that for the selected plasma parameters, an increase in the nonisothermicity by a factor of 5 lifts the STBS forbiddenness for **k** oriented along **R** and the lowest values of the STBS threshold are obtained for **k** directed at right angles to **R**.

We shall now analyze the STBS spatial gain which is of interest for various applications. For this purpose, when analyzing the dispersion equation (4.15), we shall assume that the frequency ω is real and $\omega = \omega_s$, $\Delta = \omega_s$ and the wave vector has a small imaginary part $-i\mathbf{G}$, $G \ll k$. Then for the spatial gain in the direction of propagation of the scattered wave $G_{-1} = \mathbf{k}_{-1} \cdot \mathbf{G}/k_{-1}$ we find

$$G_{-1} = \frac{\nu_{ei}\omega_{Le}^2}{2\omega_0 k_0 c^2}\left\{\frac{v_E^2}{(v_E^2)_{th}} - 1\right\}, \qquad (4.26)$$

where the dependence of the function $(v_E^2)_{th}$ on the angle θ is described by expression (4.24). In accordance with (4.26), the spatial gain of the scattered wave is most efficient for those angles θ for which the largest excess over the threshold is achieved. At the same time, since we assumed that $G_{-1} \ll k_{-1}$ (or $G \ll k$) to derive relation (4.26), under the conditions for its validity the excess over the STBS threshold should not be anomalously large.

We shall now discuss STBS in the limit opposite to (4.20) when the pulse duration of the pump field (4.1) is short compared with the characteristic time of variation of the amplitudes of the hydrodynamic perturbations in the plasma,

$$|\gamma_s(\theta)|\tau \ll 2\pi. \qquad (4.27)$$

We shall determine the threshold pump field intensity, as is usually assumed for short pulses, using the relation

$$\gamma\tau = 2\pi, \qquad (4.28)$$

where γ is given by formula (4.21). We then obtain the following expression for the threshold intensity of the pump field:

$$\begin{aligned}\frac{(v_E^2)_{th,\tau}}{v_T^2} &= 8\frac{\nu_{ei}}{\omega_0}\frac{2\pi}{\omega_s\tau}\left[1+\frac{2\pi}{\gamma_E\tau}\right] \\ &\times\left[1+2(1-a_*\cos\theta)\frac{\psi(\theta)}{k^2 l_t l_{ei}}\right]^{-1} > 0.\end{aligned} \qquad (4.29)$$

As in the case (4.24), it follows from formula (4.29), the conditions $|\gamma| \ll \omega_s$, and (4.28) that the amplitude of the electron oscillation velocity corresponding to the threshold pump field intensity is also low compared with the electron thermal velocity. It should be noted that the comparatively low threshold energy flux density (4.29) of a pump field having the pulse duration (4.27) exceeds the corresponding value (4.24) for a long pulse

$$\frac{(v_E^2)_{th,\tau}}{(v_E^2)_{th}} = \frac{2\pi}{\max|\gamma_s(\theta)|\tau}\left[1+\frac{2\pi}{\gamma_E\tau}\right] \gg 1 \qquad (4.30)$$

times. Thus, in order for STBS to develop in short laser pulses, the energy flux densities of these fields must be higher than those required for long pulses. As for the case of long pump pulses, the threshold value of (4.29) depends strongly on the angle between the vectors **n** and **k**. As a result of (3.29) and (4.27) the corresponding dependences are similar to those plotted in Figs. 1–3.

To conclude this Section, we note that in the limit of short pump field pulses (4.27) formula (4.26) holds for the spatial STBS gain near the threshold where $(v_E^2)_{th}$ should be replaced by $(v_E^2)_{th,\tau}$ as given by (4.29).

5. DISCUSSION

We shall discuss the conditions under which STBS may be observed in a plasma with ion-acoustic turbulence. According to inequality (3.5), the characteristic scale of the hydrodynamic quantities determining the perturbation should be smaller than the spatial scale of variation of these quantities. This implies that the scattering angle θ_0 should not be too small,

$$2\sin\frac{\theta_0}{2} \approx \theta_0 \gg 1.5\times10^{-3}\left[\frac{10^{-2}}{L_T[\mathrm{cm}]}\right]\left[\frac{2\times10^{15}}{\omega_0[\mathrm{s}^{-1}]}\right], \qquad (5.1)$$

where it is assumed that $L \sim L_T$. Generally, another lower constraint on θ_0 is imposed by inequalities (3.30) and (3.31). However, in the most interesting case of short laser pulses when $|\gamma_s(\theta)|\tau \ll 2\pi$ no such constraint exists. At the same time, using the hydrodynamic description of the perturbations presupposes that the parameter $Zk^2l_tl_{ei}$ (3.1) is small compared with unity, which gives

$$\begin{aligned}\theta_0 &< 0.1\left(\frac{\Lambda}{5}\right)^{1/2}\left[\frac{10^{-2}}{L_T[\mathrm{cm}]}\right]^{1/4}\left[\frac{2\times10^{15}}{\omega_0[\mathrm{s}^{-1}]}\right] \\ &\times\left[\frac{ZT}{10T_i}\right]^{1/4}\left[\frac{R}{p}L_T\right]^{1/4}\left[\frac{100}{T[\mathrm{eV}]}\right]^{9/8}\left[\frac{n[\mathrm{cm}^{-3}]}{10^{20}}\right]^{5/8}.\end{aligned} \qquad (5.2)$$

Another constraint of the theory put forward above involves the assumption that the characteristic time for the establishment of a quasi-steady-state spectrum of turbulent noise is shorter than the time of variation of the perturbations of the hydrodynamic quantities (see (3.6)). In the limits of inequalities (3.1) and (3.6), the spectrum of turbulent noise and the electron fluxes can follow the comparatively slow variation of the perturbations. Inequality (3.6) imposes another upper constraint on the angle θ_0:

$$\begin{aligned}\theta_0 &< 0.2\left[\frac{10^{-2}}{L_T[\mathrm{cm}]}\right]^{1/2}\left[\frac{2\times10^{15}}{\omega_0[\mathrm{s}^{-1}]}\right] \\ &\times\left[\frac{10T_i}{ZT}\right]^{1/2}\left[\frac{R}{p}L_T\right]^{1/2}\left[\frac{100}{T[\mathrm{eV}]}\right]^{1/4}\left[\frac{n[\mathrm{cm}^{-3}]}{10^{20}}\right]^{1/4}.\end{aligned} \qquad (5.3)$$

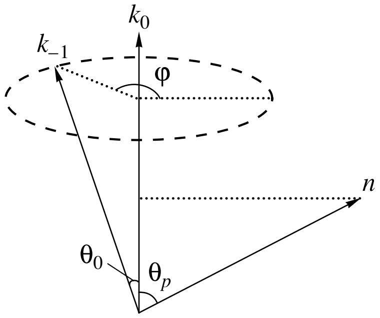


Fig. 4. Volume scattering diagram.

We shall consider the constraints (5.1)–(5.3) for an iron laser plasma under conditions typical of experiments [21, 22] where $T = 200$ eV, $n = 10^{20}$ cm^{-3}, $Z = 5$, $L_T = 100$ μm, and $ZT/T_i = 20$. For these plasma parameters the inequalities (5.1) and (5.2) impose the following constraints on the scattering angle in degrees:

$$0.2° < \theta_0 < 4°, \tag{5.4}$$

and inequality (5.3) is weaker than (5.2). Note that for $k \approx k_0\theta_0$ the inequalities (3.7) are also satisfied, which means that the ions can be described hydrodynamically. Consequently, the anisotropy characteristics of the STBS threshold in a turbulent plasma observed above should be observed for comparatively small scattering angles. For such small angles θ_0 the scattering wave vector $\mathbf{k}$ is almost perpendicular to the wave vector of the incident wave $\mathbf{k}_0$. This implies that for waves propagating along $\mathbf{n}$, which is the direction of anisotropy of the ion-acoustic turbulence, the scattering characteristics will be determined by the vectors $\mathbf{k}$ almost orthogonal to $\mathbf{n}$. Conversely, if $\mathbf{k}_0 \perp \mathbf{n}$, the small-angle scattering will be determined by the vectors $\mathbf{k}$ oriented along $\mathbf{n}$.

To complete the picture we note the following. Outside the range of angles (5.4) the scattering wave vector is so large that the following inequality is satisfied

$$Zk^2 l_t l_{ei} \gg 1. \tag{5.5}$$

In order to describe scattering using relation (5.5), we need to allow for the nonlocal nature of the electron heat transport. A theory of STBS under the conditions (5.5) was given in a recent publication [14] for the case when the turbulent noise distribution is defined.

To conclude our discussion of the characteristics of small-angle scattering we also note that, as can be seen from relation (4.29), for the plasma parameters given in this section and radiation pulses with $2\pi/\gamma_E \ll \tau \ll 2\pi/|\gamma_s|$, the STBS threshold does not exceed $\sim 10^{10}$–$10^{11}(2\pi/|\gamma_s|\tau)$ W/cm^2. In particular, for $\gamma_s \sim 10^8$ s^{-1} and $\tau \sim 1$ ns the threshold energy flux density is $\sim 10^{12}$–10^{13} W/cm^2.

The conditions for STBS forbiddenness have not been discussed so far. We shall make a special analysis of this qualitatively new effect. It has been shown that for $Zk^2 l_t l_{ei} \ll 1$ or (5.2), STBS forbiddenness occurs in the limit of high values of the effective density of the turbulence-exciting force $\mathbf{R}$ when $R \gg R_{NL}$. In this case, forbiddenness occurs when the angle θ between the direction of the unit vector $\mathbf{n} = \nabla T/|\nabla T|$ (which determines the axis of anisotropy of the ion-acoustic turbulence) and the difference $\mathbf{k}_0 - \mathbf{k}_{-1}$ (the wave vectors of the pump field $\mathbf{k}_0$ and the scattered wave $\mathbf{k}_{-1}$ is equal to or smaller than θ_*. The angle θ_* is determined from the condition that the denominator of formula (4.25) vanishes and separates the region of angles θ into two parts, where STBS is allowed when $\theta_* < \theta < \pi$ and forbidden when $0 \le \theta \le \theta_*$. Let us assume that θ_p is the angle between the vectors $\mathbf{n}$ and $\mathbf{k}_0$ which we can call the angle of incidence. We select the polar axis of the spherical coordinates in the direction of the vector $\mathbf{k}_0$. Then the direction of the vector $\mathbf{k}_{-1}$ which defines the direction of observation of the scattered wave will be determined by the polar angle θ_0 (the scattering angle) and the azimuthal angle φ measured from the projection of the vector $\mathbf{n}$ on the plane perpendicular to $\mathbf{k}_0$ (see Fig. 4). In order to observe STBS forbiddenness experimentally, the following condition must be satisfied:

$$\cos\varphi \le \frac{\cos\theta_p \sin(\theta_0/2) - \cos\theta_*}{\sin\theta_p \cos(\theta_0/2)}, \tag{5.6}$$

which ensures that the inequality $0 \le \theta \le \theta_*$ is satisfied.

Since the STBS theory constructed by us for (3.1) is valid for small scattering angles (5.2), in the zeroth approximation with respect to the small contribution of quantities of the order $\theta_0 \ll 1$, condition (5.6) can be written in the form

$$\pi - \arccos\left(\frac{\cos\theta_*}{\sin\theta_p}\right) \le \varphi \le \pi + \arccos\left(\frac{\cos\theta_*}{\sin\theta_p}\right), \tag{5.7}$$

where the angle of incidence θ_p is in the range

$$\pi/2 - \theta_* \le \theta_p \le \pi/2 + \theta_*. \tag{5.8}$$

In particular, under the conditions of thermal STBS being discussed

$$\theta_* = \arccos\frac{1}{a_*} \simeq 40°, \tag{5.9}$$

and inequality (5.8) has the form

$$50° \le \theta_p \le 130°. \tag{5.10}$$

In Fig. 5 the shaded region illustrates the range of azimuthal angles φ of the vector $\mathbf{k}_{-1}$ in which (5.7) is satisfied and STBS is forbidden. Here $\mathbf{k}_{-1\perp}$ and $\mathbf{n}_\perp$ are the components of the vectors $\mathbf{k}_{-1}$ and $\mathbf{n}$ perpendicular to $\mathbf{k}_0$. It follows from (5.7) that the maximum width of the

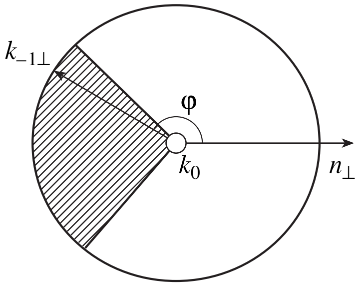


Fig. 5. Diagram of scattering in the plane perpendicular to the vector $\mathbf{k}_0$. In the shaded area of azimuthal angles φ of the vector $\mathbf{k}_{-1}$ STBS is forbidden.

region of azimuthal angles φ of the vector $\mathbf{k}_{-1}$ in which STBS is forbidden is

$$\pi - \theta_* \le \varphi \le \pi + \theta_* \qquad (5.11)$$

and is achieved when the angle of incidence is $\theta_p \simeq \pi/2$, i.e., the pump wave is incident almost perpendicular to the axis of anisotropy of the turbulence. Under the conditions of thermal STBS inequalities (5.11) have the form

$$140° \le \varphi \le 220°.$$

By determining experimentally the limiting values of the angles θ_p and φ for which STBS forbiddenness occurs, we can find θ_* and check, not only the numerical value of the constant a_* but also, most importantly, we can check experimentally the recent statement made in the literature that the ion heating in a plasma with ion acoustic turbulence is anomalously effective.

This work was supported financially by the Russian Foundation for Basic Research (projects nos. 96-02-16779, 99-02-18075), the program for State Support of Leading Scientific Schools (project no. 96-15-96750) and the Federal Target Scientific-Technical Program "Research and Development of Priority Directions in the Development of Science and Technology for Civilian Purposes," subprogram "Physics of Quantum and Wave Processes."

APPENDIX

In a highly nonisothermal plasma, the number of ions interacting with ion-acoustic waves by the Cerenkov scattering mechanism is comparatively small and their influence on the turbulence spectrum can be neglected. In this plasma, a quasi-steady state of ion acoustic turbulence is established under the influence of various processes such as Cerenkov emission of waves by electrons and induced scattering of waves at thermal ions. Under these conditions the laws governing electron transport depend on the relationship between the effective density of the turbulence-exciting force

$$\mathbf{R} = en\mathbf{E}_q - \frac{\partial}{\partial \mathbf{r}}p \qquad (A.1)$$

and the force density R_{NL} characterizing the influence of the induced scattering of waves by ions (see (2.7)) where $\mathbf{E}_q$ is the quasi-steady state electric field in the plasma. When $R \ll R_{NL}$, the density of the current and the thermal electron flux are described by [19]

$$\begin{aligned}\mathbf{j} = env_s\bigg\{&\left[\frac{3}{2}\left(1 - \beta_\parallel(\alpha) + \frac{\alpha}{12}\right) + \frac{16}{\pi}\beta_\parallel(\alpha)\right]\mathbf{n}\\ &- \frac{24}{\pi}(\beta_\parallel(\alpha)\mathbf{n}\cdot\boldsymbol{\zeta}\mathbf{n} + \beta_\perp\mathbf{n}\times[\boldsymbol{\zeta}\times\mathbf{n}])\bigg\},\end{aligned} \qquad (A.2)$$

$$\begin{aligned}\mathbf{q} = pv_s\bigg\{&\left[\frac{15}{4}\left(1 - \beta_\parallel(\alpha) + \frac{\alpha}{12}\right) + \frac{64}{\pi}\beta_\parallel(\alpha)\right]\mathbf{n}\\ &- \frac{160}{\pi}(\beta_\parallel(\alpha)\mathbf{n}\cdot\boldsymbol{\zeta}\mathbf{n} + \beta_\perp\mathbf{n}\times[\boldsymbol{\zeta}\times\mathbf{n}])\bigg\},\end{aligned} \qquad (A.3)$$

where $\mathbf{n} = \mathbf{R}/R$, $\boldsymbol{\zeta} = (p/R)\partial \ln T/\partial \mathbf{r}$, $\beta_\perp = 0.02$. In the opposite limiting case when $R \gg R_{NL}$, the electron fluxes have the form [19]

$$\begin{aligned}\mathbf{j} &= \frac{16}{\pi}env_s\sqrt{\frac{R}{R_{NL}}}\\ &\times\left\{\beta_\parallel\mathbf{n} - \frac{3}{2}(\beta_\parallel\mathbf{n}\cdot\boldsymbol{\zeta}\mathbf{n} + \beta_\perp\mathbf{n}\times[\boldsymbol{\zeta}\times\mathbf{n}])\right\},\end{aligned} \qquad (A.4)$$

$$\begin{aligned}\mathbf{q} &= \frac{64}{\pi}pv_s\sqrt{\frac{R}{R_{NL}}}\\ &\times\left\{\beta_\parallel\mathbf{n} - \frac{5}{2}(\beta_\parallel\mathbf{n}\cdot\boldsymbol{\zeta}\mathbf{n} + \beta_\perp\mathbf{n}\times[\boldsymbol{\zeta}\times\mathbf{n}])\right\},\end{aligned} \qquad (A.5)$$

where $\beta_\perp = 0.80$. Expressions (A.2)–(A.5) can be used to analyze electron charge and heat fluxes parallel and perpendicular to the axis of anisotropy of the turbulent noise.

REFERENCES

1. H. A. Baldis and C. J. Walsh, Phys. Fluids **26**, 3426 (1983).
2. K. Tanaka, L. M. Goldman, W. Seka, *et al.*, Phys. Fluids **27**, 2960 (1984).
3. P. E. Young, K. G. Estabrook, *et al.*, Phys. Fluids B **2**, 1907 (1990).
4. G. P. Banfi, K. Eidmann, and R. Sigel, Opt. Commun. **52**, 35 (1984).

5. J. Handke, S. A. H. Rizvi, and B. Kronast, Appl. Phys. **25**, 109 (1981).
6. S. J. Karttunen and R. R. E. Salomaa, Phys. Lett. A **88**, 350 (1982).
7. J. F. Drake, P. K. Kaw, Y. C. Lee, *et al.*, Phys. Fluids **17**, 778 (1974).
8. W. M. Manheimer and D. G. Colombant, Phys. Fluids **24**, 2319 (1981).
9. W. L. Kruer, Phys. Fluids **23**, 1273 (1980).
10. L. Spitzer, Jr. and R. Harm, Phys. Rev. **89**, 977 (1953).
11. R. W. Short and E. M. Epperlein, Phys. Rev. Lett. **68**, 3307 (1992).
12. P. K. Shukla, Phys. Fluids B **5**, 4253 (1993).
13. A. V. Maximov and V. P. Silin, Phys. Lett. A **192**, 67 (1994).
14. K. N. Ovchinnikov, V. P. Silin, and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **113**, 629 (1998).
15. V. P. Silin and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **98**, 117 (1990).
16. A. V. Maksimov, V. P. Silin, and M. V. Chegotov, Fiz. Plazmy **16**, 575 (1990).
17. V. P. Silin and S. A. Uryupin, Fiz. Plazmy **22**, 790 (1996).
18. V. Yu. Bychenkov and V. P. Silin, Fiz. Plazmy **13**, 1097 (1987).
19. V. Yu. Bychenkov, V. P. Silin, and S. A. Uryupin, Phys. Rep. **164**, 119 (1988).
20. V. Yu. Bychenkov, V. P. Silin, and S. A. Uryupin, Comments Plasma Phys. Contr. Fusion **13**, 239 (1990).
21. L. L. Losev and V. I. Soskov, Opt. Commun. **135**, 71 (1997).
22. A. A. Antipov, A. Z. Grasyuk, S. V. Efimovskiĭ, *et al.*, Kvantovaya Élektron. **25**, 31 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 1, 2000, pp. 93–101.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 1, 2000, pp. 105–114.
Original Russian Text Copyright © 2000 by Belov, A. Ivanov, D. Ivanov, Pal', Starostin, Filippov, Dem'yanov, Petrushevich.

**PLASMA,
GASES**

Coagulation of Charged Particles in a Dusty Plasma

I. A. Belov*, A. S. Ivanov*, D. A. Ivanov*, A. F. Pal', A. N. Starostin**, A. V. Filippov**, A. V. Dem'yanov**, and Yu. V. Petrushevich****

**Kurchatov Institute Russian Science Center,
Moscow, 123182 Russia*

***Russian State Science Center, Troitsk Institute of Innovative and Thermonuclear Research (TRINITI),
Troitsk, Moscow oblast, 142092 Russia;
e-mail: afpal@fly.triniti.ru, afpal@mics.msu.su*

Received June 3, 1999

Abstract—An investigation is made of characteristic features in the behavior of small particles in a dusty plasma attributable partly to the suppression of coagulation as a result of monopolar charging for particle sizes smaller than the Debye shielding length and partly to the reduction in the effect of charging for larger particles. Similarity relations linking the plasma composition and particle charge with the parameters of the dust component are used to determine the range of parameters for which the linear approximation of the particle charge as a function of their sizes holds. A modified classical theory of coagulation in the diffusion approximation is used to study some anomalies in the behavior of the particle size distribution. It is established that unlike an ordinary aerosol, in a dusty plasma the dispersion of the distribution and the average particle size may decrease with time. It is shown for the first time that a long-lived "quasi-liquid" state of a dusty plasma may be established as a result of the anomalous behavior of the size distribution function of coagulating charged particles. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of small particles in a dusty plasma has attracted interest because of its fairly unique and not always understandable behavior in laboratory experiments. Particular mention should be made of the observation of quasi-crystalline, levitating, dusty structures whose appearance is usually attributed to the strongly nonideal properties of a dusty plasma and the establishment of long-range particle interaction under these conditions [1]. Since a dusty plasma forms spontaneously in many cases during the dissociation and subsequent nucleation of the initial material in a gas discharge, particle coagulation is important for the self-organization processes of a dusty plasma. Studies of particle coagulation in a dusty plasma are also topical in relation to material transport and surface contamination during the fabrication of semiconductor devices [2]. Similar topics arise in problems associated with plasma-chemical technologies for fabricating powder materials [3]. In this last case, coagulation is one of the main mechanisms for dust particle growth. Depending on the experimental conditions, a wide range of different objects have been observed during the coagulation process, ranging from particles having a complex "cauliflower" structure to fractal clusters having branching dendritic structures [4, 5].

The phenomenon of coagulation has been studied in fairly great detail in investigations of the behavior of various types of aerosols, including charged ones [6]. In situations encountered in practice, such as, for example, atmospheric clouds, bipolar charging is generally observed, sometimes displaced by the inclusion of various particle charging mechanisms. In particular, if diffusion charging takes place in a weakly ionized atmosphere, an average, negligible, almost symmetric charging is observed with a Boltzmann particle charge distribution. This is attributable to the low concentration of electrons in an atmosphere containing electronegative gases. In this case, the charging is accomplished by fluxes of positive and negative ions having similar mobilities. The incorporation of other mechanisms such as thermionic or photoemission, or electrification accompanying dispersion, naturally shifts the charge distribution in a particular direction.

Particle charging in a plasma having a fairly high electron density has some important differences. As a result of the appreciable difference between the mobilities of the electrons and the positive ions, monopolar charging takes place where the dispersed particles have extremely large negative charges [1].

The subject of the present investigation is a plasma created by the ionization of gases (such as xenon) by decay products of radioactive dust particles having the following characteristic parameters: rate of creation of electron–ion pairs 10^{15}–10^{16} cm^{-3} s^{-1}, pressure ~1 atm, electron density ~10^{10}–10^{11} cm^{-3}, particle charging times 10^{-5}–10^{-6} s, dust particle sizes up to 100 μm, and concentrations up to 10^{7} cm^{-3}. This plasma may be of interest for the development of a nuclear battery using radioactive waste [7]. In this case, the particle charge distribution is closely related to their sizes. Hence, the particle size distribution function is one of the most important characteristics describing the behavior and properties of a dusty plasma. The distribution function is formed mainly as a result of the coagulation and deposition of dust particles. In some studies [8, 9] the authors have postulated that a quasi-steady-state particle size distribution is established. When this state is attained, the growth of particles of a particular size is compensated by their losses as a result of coagulation and deposition. For very small particles the deposition losses are negligible whereas for comparatively large ones coagulation can be neglected.

We know [10] that the rate of coagulation may increase or decrease depending on the polarities and magnitudes of the charges at the particles. The ratio of the coagulation constants for charged and uncharged particles was first calculated by Fuks [6]. It was shown that for a weakly charged bipolar aerosol the increase in coagulation as a result of attraction is compensated to a considerable extent by the decrease caused by repulsion. At the same time, for a strongly charged bipolar aerosol the increase in coagulation as a result of attraction is considerably greater than its decrease as a result of repulsion and the rate of coagulation increases.

According to [10], for a monopolar charged aerosol, electrostatic repulsion leads to the separation of like charged particles, lowers the rate of coagulation, and causes a drop in the aerosol concentration. Note that this statement only holds completely when the integral electroneutrality of the system is destroyed.

Characteristics of the particle behavior in a dusty plasma associated, on the one hand, with the suppression of coagulation as a result of monopolar charging for particle sizes smaller than the Debye shielding length R_d and, on the other hand, with the reduction of charging for particle sizes larger than R_d, result in these dependences being impaired.

In the present paper we investigate the influence of various anomalies in the coagulation process on the particle size distribution function in a dusty plasma. We show for the first time that a long-lived "quasi-liquid" dusty plasma state may be achieved as a result of the anomalous behavior of the size distribution function of the coagulating charged particles.

2. PLASMA COMPOSITION AND CHARGING OF SMALL PARTICLES

First, we need to analyze the charge state of the small particles and establish a relationship between the magnitude of the charge, the particle size, and the plasma composition for given conditions (for example, for a given ionization source power). In the present study we consider a plasma without negative ions. As has been noted, when dispersed particles enter a plasma, they become negatively charged, mainly because the electron mobility is substantially higher than the ion mobility. The characteristic particle charging time is 10^{-5}–10^{-6} s [10]. It should be noted that the characteristic discharging time determined by the ion diffusion is considerably greater than these figures, but still considerably shorter than the characteristic times of variation of the particle size distribution function as a result of coagulation over a wide range of particle sizes and concentrations. In view of this, the problems of particle charging and the formation of the distribution function may be separated and solved systematically. Estimates of particle charging in plasmas have been made in various studies [7, 11]. The steady-state charge is determined from the condition that the electron and ion fluxes to a particle are equal. It is easy to show that the particle charge q for low dust particle concentrations N_d is a linear function of the radius:

$$eq = r_d U,$$

where U is the floating potential, r_d is the particle radius, and e is the elementary charge. However, as N_d increases, this relationship is destroyed. Since the dependence of the charge on the particle radius is basic for studying the coagulation process, we need to make a more detailed study of this dependence and estimate the range of parameters where it is valid.

To be specific, we shall assume the number of ion and electron pairs formed per unit plasma volume per unit time is Q. This plasma is created by an external polarizer such as an electron beam or radioactive dust particles. We place aerosol particles in the plasma and estimate their charge in the diffusion approximation. We shall assume that the mean free path of the ions and electrons is much smaller than the particle sizes. For atmospheric pressure this condition is satisfied for particle sizes $\geq 10^{-4}$ cm.

The plasma composition is determined by the charge conservation equation (the plasma quasineutrality condition) and the equation of continuity for the electrons and ions:

$$\begin{gathered}
\frac{\partial N_-}{\partial t} + \operatorname{div}\mathbf{j}_- = Q - \beta N_- N_+, \\
\frac{\partial N_+}{\partial t} + \operatorname{div}\mathbf{j}_+ = Q - \beta N_- N_+, \\
\sum_k q_k n_{dk} + N_+ = N_-, \\
\operatorname{rot}\mathbf{E} = 0, \\
\mathbf{j}_- = -D_-\nabla N_- - K_- N_- \mathbf{E}, \\
\mathbf{j}_+ = -D_+\nabla N_+ + K_+ N_+ \mathbf{E}.
\end{gathered} \tag{1}$$

Here $N_\pm$, $\mathbf{j}_\pm$, $D_\pm$, and $K_\pm$ are the concentrations, current densities, diffusion coefficients, and mobilities of the ions ("+") and the electrons ("–"), $\mathbf{E}$ is the electric field strength, β is the electron–ion recombination coefficient, and $n_{dk} = n_d(r_k)$ is the concentration of particles having sizes between r_k and r_{k+1} (see below). The evolution of the particle size distribution function is considered in the following section.

In the steady-state and spherically symmetric case, the currents $J_\pm$ are given by [11]:

$$J_\pm = 4\pi e r^2\left(\pm D_\pm \frac{\partial N_\pm}{\partial r} - K_\pm E N_\pm\right). \tag{2}$$

The mobilities $K_\pm$ and diffusion coefficients $D_\pm$ of the ions and electrons are related by the Einstein relationships

$$kT_\pm K_\pm = eD_\pm, \tag{3}$$

where k is the Boltzmann constant, and $T_\pm$ is the temperature.

We assume that in the region of greatest variation of the ion and electron density the particle charge is not shielded by the plasma, and the electric field in this region is determined by the Coulomb law. This assumption is mainly valid when the condition

$$L_+ \ll a_d, \tag{4}$$

is satisfied where

$$L_+ = \frac{N_+^\infty}{|(\partial N_+/\partial r)|_{r = r_d}}$$

is the characteristic size of the region of greatest variation of the ion density and

$$a_d = \frac{1}{\sqrt[3]{(4/3)\pi N_d}}$$

is the mean interparticle distance. Note that condition (4) is known to be satisfied if the simpler condition for estimates: $r_d \ll a_d$ is satisfied. If the particle charge is $|q| \sim 10^2$ and the radius is $r_d \sim 10^{-4}$ cm, the field is $E \approx 1500$ V cm^{-1}. The reduced electric field strength at atmospheric pressure is fairly high $E/N_0 \approx 5 \times 10^{-17}$ V cm^2 (N_0 is the number of atoms per unit volume) and an electron temperature difference occurs, which may reach 1–5 eV in rare gases.

In the steady state when condition (4) is satisfied, the following relation holds for the currents

$$J_+ = -J_- \approx \text{const}. \tag{5}$$

By integrating relation (5) for the case of a monodisperse aerosol $n_{dk} = N_d$, $q_k = q$, $r_k = r_d$ for all particles with the boundary conditions

$$N_\pm|_{r=\infty} = N_\pm^\infty, \quad N_\pm|_{r = r_d} = 0, \tag{6}$$

we obtain an expression linking the ion and electron densities far from the dust particle:

$$\frac{N_-^\infty}{N_+^\infty} = \frac{K_+}{K_-}\frac{1 - \exp\{-qe^2/kT_- r_d\}}{\exp\{qe^2/kT_+ r_d\} - 1}. \tag{7}$$

This expression agrees with the expression to determine the charge of a dust particle obtained in the bounded orbital approximation (see [1]) where the ratio of the ion and electron thermal velocities is naturally replaced by the ratio of the drift velocities or, ultimately, the mobilities in the diffusion approximation. Bearing in mind that $T_- \gg T_+$ and $K_- \gg K_+$ for the case $N_+^\infty \approx N_-^\infty$, we have from (7)

$$q = -\frac{kT_- r_d}{e^2}\ln\left(1 + \frac{K_-}{K_+}\right). \tag{8}$$

Thus, the particle charge is negative and is determined by the electron temperature. In particular, for xenon having the radius $r_d = 10^{-4}$ cm and $kT_- = 2$ eV, we have $|q| \approx 10^4$. The mobility ratio in xenon, according to [12, 13], is $K_-/K_+ \approx 10^3$ under these conditions.

The condition for equal ion and electron densities at infinity is satisfied for comparatively low aerosol particle concentrations and fairly high-intensity ionization of the gas when the electron density is considerably higher than the total charge collected at the dust particles per unit volume. However, according to our estimate of the charge this condition is not satisfied even for dust concentrations $N_d \sim 10^7$ cm^{-3} and $N_- \sim 10^{11}$ cm^{-3}.

We shall attempt to estimate the influence of changes in the electron density on the charging of dust particles. In the proposed approach this problem can be solved by redefining the electron and ion density at infinity. If condition (4) is satisfied, we can assume that these quantities are the same as the averages over the plasma volume $\bar{N}_\pm$. The plasma composition is determined by the equation of quasineutrality and the balance equation for the rates of generation and loss of electrons and ions:

$$\begin{aligned} \bar{N}_+ &= \bar{N}_- - qN_d, \\ Q &= \beta \bar{N}_+ \bar{N}_- + J_+ N_d/e. \end{aligned} \tag{9}$$

Having supplemented these equations with relation (7), which links the electron and ion densities far from the particle, after simple transformations we obtain a system of equations to determine the charge q and the densities $\bar{N}_+$ and $\bar{N}_-$:

$$\begin{aligned} \bar{N}_+ &= \bar{N}_- - qN_d, \\ Q &= (\beta - \beta_{id})\bar{N}_+\bar{N}_- + \beta_{id}\bar{N}_+^2, \\ \frac{\bar{N}_-}{\bar{N}_+} &= \frac{K_+}{K_-}\frac{1 - \exp\{-qe^2/r_d kT_-\}}{\exp\{qe^2/r_d kT_+\} - 1}, \end{aligned} \tag{10}$$

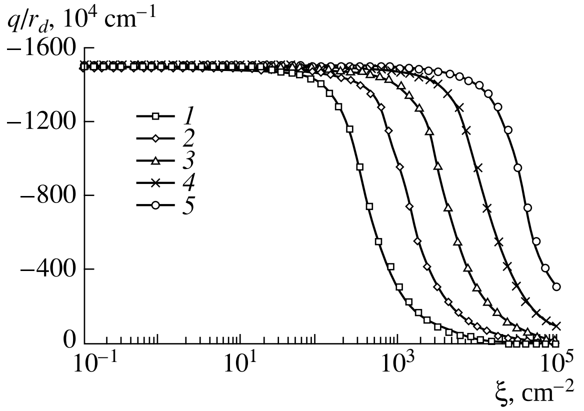


Fig. 1. Ratio of charge to radius of dust particles as a function of the parameter ξ for various ionization rates. Curves *1–5* correspond to ionization rates of 10^{13}–10^{17} cm^{-3} s^{-1}.

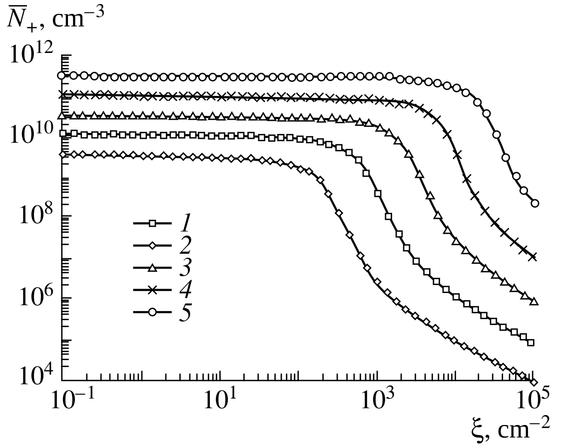


Fig. 2. Average electron density $\bar{N}_-(\xi)$ for various ionization rates. The correspondence between the curves and the ionization rates is as in Fig. 1.

where β_{id} is the effective coefficient of ion recombination at dust particles, which is given by

$$\beta_{id} = \frac{4\pi e K_+}{1 - \exp\{qe^2/r_d kT_+\}}. \tag{11}$$

When condition (4) is satisfied, the expression for β_{id} is simplified and has the form

$$\beta_{id} = 4\pi e K_+. \tag{12}$$

The expression for ion–ion recombination in the Langevin theory [13] has the same form when one of the ions (in our case a negatively charged dust particle) has almost zero mobility.

It is easy to see that the dependence of $\bar{N}_+$ and $\bar{N}_-$ on the parameters of the plasma dust component is determined by the ratio q/r_d. Thus, according to (10), this ratio q/r_d is a function of the product $N_d r_d = \xi$. In this context the following similarity relations hold:

$$\begin{aligned} \bar{N}_- &= f_1(\xi, Q), \\ \bar{N}_+ &= f_2(\xi, Q), \\ q/r_d &= f_3(\xi, Q). \end{aligned} \tag{13}$$

If we assume that $q \sim r_d$, i.e., $f_3(\xi, Q) = \text{const}$, then ξ is proportional to the total charge collected on the aerosol particles. Then the relations (13) have the fairly simple meaning that the plasma composition is determined only by the source power and the total charge collected by the aerosol dispersed in it. Note that if $f_3 \neq \text{const}$, the parameter ξ no longer characterizes the total charge and it is clearly difficult to obtain a simple interpretation of the relations (13). At the same time, the similarity relations hold and may be used to analyze the experimental results. Note that the similarity relations for the solution of the system of equations (10) may be expressed in the simpler form:

$$\begin{aligned} \bar{N}_-/\sqrt{Q} &= f_4(\sqrt{Q}/\xi), \\ \bar{N}_+/\sqrt{Q} &= f_5(\sqrt{Q}/\xi), \\ q/r_d &= f_6(\sqrt{Q}/\xi). \end{aligned} \tag{14}$$

Figures 1–3 give numerical solutions of the system (10) for various rates of ionization of xenon at atmospheric pressure. The electron temperature is taken to be 0.25 eV and the ion temperature 300 K. The mobilities reduced to atmospheric pressure are: $K_- = 3000$ cm^2 V^{-1} s^{-1} [12] and $K_+ = 0.55$ cm^2 V^{-1} s^{-1} [13]. The coefficient of dissociative recombination of Xe_2^+ (the dominant ions in a xenon plasma at atmospheric pressure) at the electron temperature given above is $\beta = 0.9 \times 10^{-6}$ cm^3 s^{-1} [14].

According to these calculations, the curve $q(\xi)/r_d$ (Fig. 1) has the form of a slightly broadened Fermi step. The floating potential of the particles $U = eq/r_d$ remains almost constant over a wide range of parameters of the plasma dust component in the range $\xi \le \xi_0$. The value of the parameter ξ_0 is determined by the ionization source power. In particular, for $Q = 10^{15}$ cm^{-3} s^{-1} this statement holds fairly accurately as far as $\xi = \xi_0 \approx 10^3$ cm^{-2}. On the plane (r_d, N_d) the region $\xi \le \xi_0$ is positioned between the coordinate axes and the hyperbola $N_d = \xi_0/r_d$ (see Fig. 4). The particle floating potential remains constant provided that the rate of volume loss of electrons and ions as a result of electron–ion recombination considerably exceeds their rate of loss at the surface of the dust particles. From expressions (10) and (12) we find that this condition is satisfied in the range of parameters of the dust component where

$$\xi \ll \frac{e^2}{T_-}\frac{\sqrt{Q/\beta}}{\ln(1 + K_-/K_+)}. \tag{15}$$

It is interesting to note that if the initial concentrations and sizes of the aerosol particles satisfy the condition $\xi \le \xi_0$, during the coagulation process, which is accompanied by a drop in the particle concentration and an increase in their sizes, the inequality ξ still holds. This allows us to investigate the coagulation of dust particles under the condition U = const, defining the initial value of the parameter ξ to be inside the shaded region (Fig. 4). In this case, as can be seen from the system of equations (1), the charging of a polydisperse system of dust particles having the size distribution function n_{dk} will be the same as that for a monodisperse system with the parameters $N_d = \sum_k n_{dk}$ and $r_d = \sum_k n_{dk} r_k / N_d$. Thus, the results obtained in this section for monodisperse particles can also be applied to polydisperse ones provided that the condition $\xi \le \xi_0$ is satisfied.

3. PARTICLE SIZE DISTRIBUTION FUNCTION

We shall now investigate the coagulation of particles in a dusty plasma. The theory of aerosol coagulation usually starts from the assumption that the particles coagulate, i.e., merge or coalesce, on each contact. This can be confirmed in numerous experiments [10]. However, if the particles have fairly high, like charges, this assumption is not completely justified because of electrostatic repulsion. However, it should be borne in mind that aerosol particles are not points. In many cases of practical interest the particles possess fairly high electrical conductivity so that at distances comparable with the particle sizes polarization plays an important role in their interaction. The polarization effect may be so appreciable that it leads to the attraction of like charged dust particles. In addition, shielding of the charges in the plasma also plays an important role in the interaction of charged dust particles. Thus, in order to solve the coagulation problem systematically, we generally need to make self-consistent calculations of the interaction of at least two nonpoint particles in an ionized gas.

In the most primitive model of the coagulation of a monodisperse aerosol, to simplify the problem it is assumed that one of the particles is fixed and the frequency of contact between this particle and others undergoing thermal motion is determined. The shape and size of the fixed particle are assumed to be unchanged. This approach can be used to estimate the change in the concentration of aerosol particles with time. Assuming that each contact reduces the number of particles by one, we write the fundamental coagulation equation in the form [6]

$$\frac{dN_d}{dt} = -GN_d^2. \qquad (16)$$

The coagulation constant G is given by

$$G = 8\pi r_d D, \qquad (17)$$

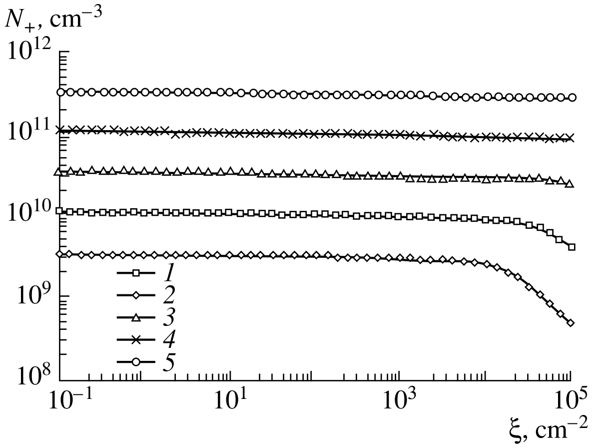


Fig. 3. Average ion density $\overline{N}_+(\xi)$ for various rates of ionization. The correspondence between the curves and the ionization rates is as in Fig. 1.

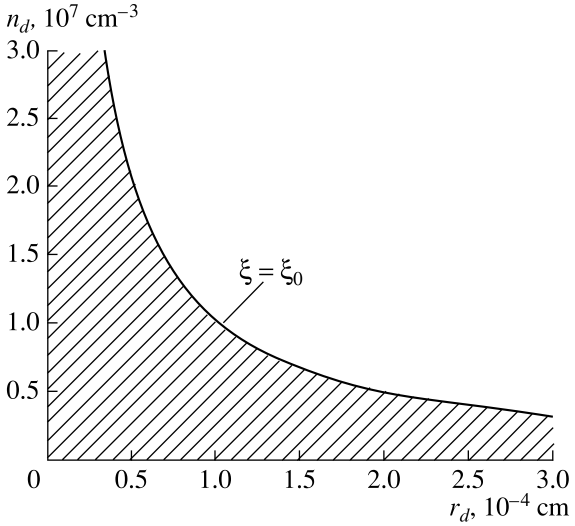


Fig. 4. Region where the condition $\xi \le \xi_0$ is satisfied on the plane (r_d, N_d) where the particle floating potential remains almost constant.

where D is the diffusion coefficient of the dust. For example, in air the coagulation constant for particles of 10^{-4} cm is 3.44×10^{-10} cm^3 s^{-1} [6]. According to formula (14), the time taken for the aerosol concentration to fall to half of $N_d = 3 \times 10^8$ cm^{-3} is $t_{1/2} = 1/GN_d \approx 10$ s.

The description of the coagulation of a polydisperse system having an arbitrary initial particle size distribution is based on solving the system of transport equations

$$\frac{\partial n_d(r_k, t)}{\partial t} = -\sum_i (1-(1/2)\delta_{ik}) G(r_i, r_k) n_d(r_i, t) n_d(r_k, t) \qquad (18)$$

$$+(1/2)\sum_i\sum_j G(r_i, r_j)\theta_{ij}^k n_d(r_i, t)n_d(r_j, t),$$

where $n_d(r_k, t)$ is the concentration of particles having radii between r_k and r_{k+1} at time t, $G(r_i, r_j)$ are the coagulation constants for particles having radii r_i and r_j, respectively, δ_{ik} is the Kronecker delta, θ_{ij}^k is the interpolation coefficient ($i, j = 1 - N$), where (V_i is the volume of a particle of radius r_i) [15]:

$$\theta_{ij}^k = \frac{V_{k+1} - (V_i + V_j)}{V_{k+1} - V_k}, \quad \text{if } (V_i + V_j) \in [V_k, V_{k+1}],$$

$$\theta_{ij}^k = \frac{V_{k-1} - (V_i + V_j)}{V_{k-1} - V_k}, \quad \text{if } (V_i + V_j) \in [V_{k-1}, V_k], \tag{19}$$

$$\theta_{ij}^k = 0 \text{ otherwise}.$$

For the case of diffusive coagulation the constants are given by [6]

$$G(r_i, r_j) = \frac{kT}{3\eta}(r_i + r_j)\left(\frac{1}{r_i} + \frac{1}{r_j}\right), \tag{20}$$

where η is the viscosity of the gas.

The system of equations (18) can be solved numerically using a range of computer programs, in particular the NAUA code [15]. The NAUA program can analyze the main parameters of a coagulating aerosol and study its deposition in a given geometry. However, particle charging has not yet been included in the calculations of this program. Generally, modifying the program to make systematic allowance for the charging of particles in a plasma is fairly complex. However, preliminary estimates of the degree of influence of particle charging on the coagulation process and the deformation of the particle size distribution function were made subject to various simplifying assumptions.

We shall estimate the coagulation constant of spherical particles having radii r_i and r_k and charges q_i and q_k following the logic put forward in [6]. Clearly we shall not incur a large error by using the method [6] for particles of various sizes since the rate constant for diffusive coagulation in this range depends weakly on the particle sizes [10]. We denote the force of the electrostatic particle interaction by $F(r)$ where r is the distance between their centers. In the coagulation process we are dealing with the diffusion of Brownian particles toward an absorbing medium in the presence of a radial electrostatic force under whose action the charged particles acquire the velocity $V = BF$, where B is the mobility of the charged dust particles. In the steady-state regime we have

$$D\Delta n_d = B\,\mathrm{div}(\mathbf{F}n_d). \tag{21}$$

Taking into account the spherical symmetry of the problem we find

$$4\pi r^2\left(D\frac{\partial n_d}{\partial r} - BFn_d\right) = J = \text{const}. \tag{22}$$

Note that this equation is quite similar to relation (2) used above. The first term on the left-hand side is equal to the number of particles passing across the spherical surface per unit time as a result of diffusion and the second term gives the number passing across this surface as a result of ordered motion in an electric field. In sum these give the rate of particle deposition on the sphere. The function $n_d(r)$ should satisfy the boundary conditions $n_d = n_{d0}$ for $r = \infty$ and $n_d = 0$ for $r = r_i + r_k$, where n_{d0} is the initial concentration of dust particles. Using the boundary conditions and taking into account the Einstein relation $D = BkT$, we find the ratio of the coagulation constants of the charged and uncharged particles ($\gamma = J/J_0$):

$$\gamma_{ik} = \left(\int_0^1 \exp\left\{\frac{1}{kT}\Psi_{ik}\left(\frac{r_i + r_k}{x}\right)\right\}dx\right)^{-1}, \quad \Psi_{ik}(r) = eq_k\varphi_i(r), \tag{23}$$

where φ_i is the electrostatic potential and $x = (r_i + r_k)/r$.

In the case of monopolar charging of aerosol particles, the polarization effect may play a significant role particularly if the difference in the charges is large. However, in the present study we neglect this influence in the calculations and only take into account the shielded Coulomb contribution

$$\varphi_i(r) = \frac{eq_i}{r}\exp\left\{-\frac{r}{R_d}\right\}, \tag{24}$$

where R_d is the Debye shielding radius. In the absence of shielding, i.e., in the limiting case $R_d \longrightarrow \infty$, relation (23) is easily integrated analytically:

$$\gamma_{ik} = \frac{\lambda_{ik}}{e^{\lambda_{ik}} - 1}, \quad \lambda_{ik} = \frac{q_i q_k e^2}{(r_i + r_k)kT}. \tag{25}$$

We shall estimate the value of γ_{ik} for the case $q_i = q_k = 10$, $r_i = r_k = 5 \times 10^{-5}$ cm, $T = 300$ K.[1] In this case we have $\gamma \sim 1.5 \times 10^{-2}$. If the particle charge increases to 50, then $\gamma \sim 10^{-63}$. In this last case, coagulation is evidently almost completely suppressed. According to these formulas, the ratio of the coagulation constants under the condition $r_i = r_k$ (in the region of practical

---

[1] Note that the range of variation of the particle sizes is broadened appreciably toward smaller sizes by introducing the Cunningham correction [6]. Allowance for this correction reduces the coagulation constants of small particles. However, the decrease in coagulation as a result of electrostatic repulsion of the charged particles is several orders of magnitude greater than this effect and its contribution in this range of sizes leads to small quantitative changes.

interest $\lambda_{ik} > 1$) is a monotonically decreasing function of particle size. However, allowance for shielding leads to qualitatively different behavior of the coagulation constant for the charged particles. We shall analyze in greater detail the situation where $U = \text{const}$, i.e., $eq_i = r_iU$ and $eq_k = r_kU$, and then

$$\Psi_{ik}\left(\frac{r_i + r_k}{x}\right) = xU^2\frac{r_ir_k}{r_i + r_k}\exp\left\{-\frac{r_i + r_k}{xR_d}\right\}. \qquad (26)$$

Accordingly the matrix of the coagulation constants for charged particles having radii r_i and r_k is described by

$$G(r_i, r_k) = \frac{kT}{3\eta}\frac{(r_i + r_k)^2}{r_ir_k} \times \left[\int_0^1 \exp\left\{\frac{1}{kT}\Psi_{ik}\left(\frac{r_i + r_k}{x}\right)\right\}dx\right]^{-1}. \qquad (27)$$

The results of a numerical integration of relations (27) are plotted in Fig. 5 ($U = 0.15$ V, $R_d = 8.4 \times 10^{-4}$ cm, $T = 300$ K). According to these results the behavior of the coagulation constants in a dusty plasma changes radically. The rate of coagulation of very small charged particles with larger ones is extremely high. This is first because of the weak charge at these particles and second because of the comparatively high Brownian mobility of these particles. As the size increases, the particle coagulation constants decrease sharply as a result of the electrostatic repulsion of the charges. However, if the size of at least one of the two colliding particles exceeds the Debye length, charge shielding begins to play an important role. In this case, the function $G(r_i, r_k)$ passes through a minimum with increasing particle size, forming a deep dip at whose outer edge the coagulation rate constant approaches the value typical of an uncharged aerosol. The characteristic sizes and shape of the dip are determined by the plasma parameters and thus in real situations their spectrum of variation is extremely extensive. This leads to various effects in the behavior of dusty structures during their self-organization process.

It should be noted that an additional stimulus to the combining of large particles with small ones is provided by the polarization effect noted above, which leads to the attraction of like-charged particles at distances of the order of the radius of the larger particle. If the Debye length is of the same order of magnitude, the electrostatic interaction between the particles may change sign, leading to an increase in the rate of coagulation even compared with its value in an uncharged aerosol. The nonmonotonic dependence of the rate constant on the particle size substantially changes the behavior of the particle size distribution function.

The formulas obtained can be used to modify the NAUA module describing aerosol coagulation fairly simply by renormalizing the coagulation constant in equation (18) in accordance with relations (26) and (27). In the calculation scheme the range of variation of the particle sizes ($r_{\min}$, $r_{\max}$) is divided in accordance with

$$(m-1)\ln(r_k/r_{k-1}) = \ln(r_{\max}/r_{\min}),$$

where m is the number of points ($m = 60$). In order to ensure that the largest particles do not build up in the system and distort the distribution function, we introduced their deposition.

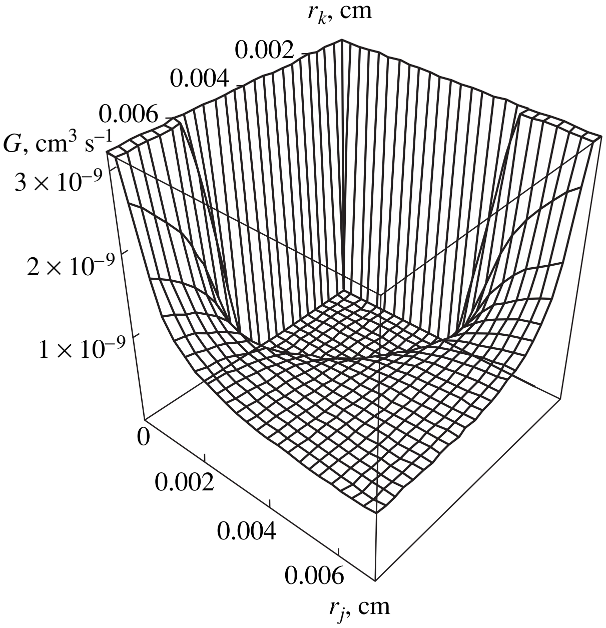


Fig. 5. Matrix of coagulation constants of a polydisperse aerosol in a dusty plasma as a function of the colliding particle radii.

Our calculations of the dependences of the concentrations on the particle sizes taking this factor into account are plotted in Figs. 6–8. Graphs of $n_d(r_k, t)$ for an uncharged aerosol at various times are plotted for comparison (Figs. 6a–8a). The initial distribution in Figs. 6a and 6b was taken to be a fairly wide log-normal particle size distribution with a standard geometric deviation of 2, and mean radius $\bar{r}_k = 10^{-4}$ cm (the initial total concentration of dust particles is 10^8 cm^{-3}). The calculations were made for a plasma having the Debye length $R_d = 8.4 \times 10^{-4}$ cm and floating potential $U = 0.15$ V. The range of variation of the particle sizes is ($r_{\min} = 10^{-6}$ cm, $r_{\max} = 3 \times 10^{-3}$ cm). As was predicted for the uncharged aerosol (Fig. 6a), the particle concentration initially decreases rapidly and their average size increases. Figure 6b shows curves describing the distribution of like-charged particles in a dusty plasma. A comparison of Figs. 6a and 6b shows that charging of the particles leads to a general slowing of the coagulation process. At the same time, the behavior of the distribution function also changes qualitatively. Initially a small fraction which undergoes substantially less slowing of the coagulation than the larger parti-

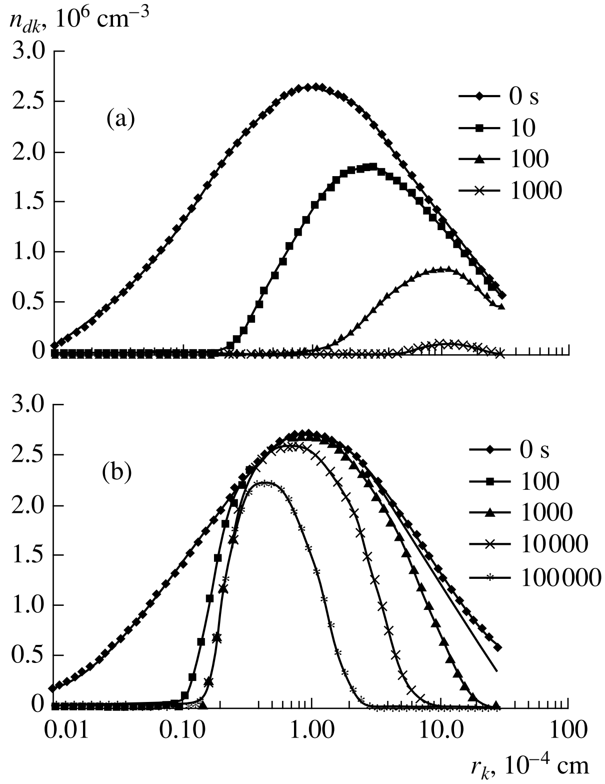


Fig. 6. Dynamics of variation of a wide particle size distribution in an uncharged polydisperse aerosol (a) and in a dusty plasma (b) with $\bar{r}_k = 1 \times 10^{-4}$ cm.

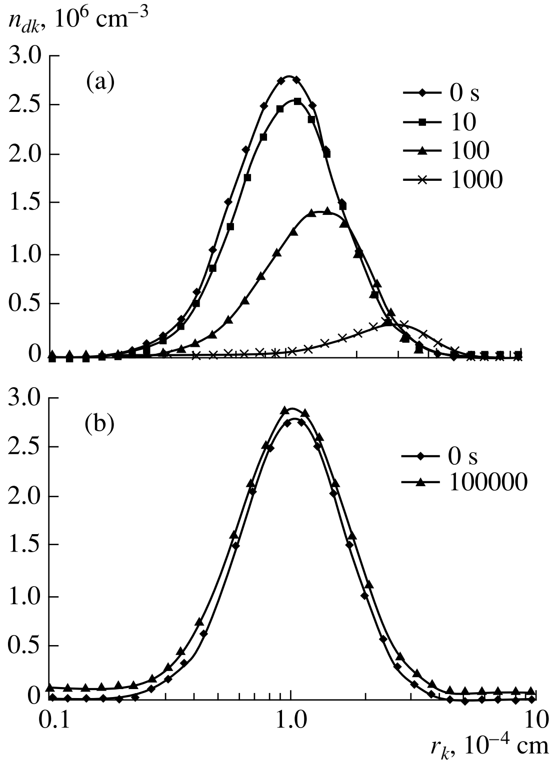


Fig. 7. Dynamics of variation of a narrow particle size distribution in an uncharged polydisperse aerosol (a) and in a dusty plasma (b) with $\bar{r}_k = 1 \times 10^{-4}$ cm. The distribution functions for $t = 0$ and 10^5 s are almost the same hence the last curve is shifted by one division upward for convenience of representation.

cles is rapidly "eaten away." As a result, clusters with a "cauliflower" structure appear, consisting of numerous small particles. Collisions between the larger particles are impeded by the forming Coulomb barrier. As it moves toward increasing particle size, the distribution function, as it were, "encounters" the Coulomb barrier. However, for particles having sizes comparable with or greater than the Debye length the electrostatic repulsion effect is reduced and their coagulation becomes quite probable. The distribution becomes narrower and the dispersion decreases. The deposition of larger particles plays a decisive role here. In addition, the form of the distribution more closely resembles a normal one. Another interesting feature of the behavior of the distribution function under these conditions is the possibility of a decrease in the average particle size with time.

We have so far analyzed the behavior of a particle size distribution with a comparatively broad spectrum. If a fairly narrow distribution function with $\bar{r}_k < R_d$ is taken as the initial distribution and deposition caused mainly by contact with the walls and by gravity is eliminated, a highly charged aerosol may exist for a long time without any perceptible changes in the particle concentration. In order to illustrate this, Figs. 7a and 7b show particle size distributions at various times for $\bar{r}_k = 10^{-4}$ cm substantially smaller than the Debye length ($R_d = 8.45 \times 10^{-4}$ cm). Here and subsequently the standard geometric deviation of the initial distributions was taken to be 0.5 and the range of variation of the particle sizes ($r_{\min} = 10^{-6}$ cm, $r_{\max} = 3 \times 10^{-2}$ cm). Figure 7a corresponds to an uncharged aerosol and Fig. 7b to a charged aerosol in the absence of deposition. The calculations show that in a dusty plasma (Fig. 7b) no appreciable changes in the distribution profile occur for time intervals up to 10^7 s. For the uncharged aerosol the concentration decreases by an order of magnitude within ~10^3 s (Fig. 7a) as a result of coagulation. Figure 8 shows graphs of particle size distributions with $\bar{r}_k = 2 \times 10^{-3}$ cm in an uncharged aerosol (Fig. 8a) and in a dusty plasma with $R_d = 8.45 \times 10^{-4}$ cm (Fig. 8b). In this case ($\bar{r}_k > R_d$) the influence of charging on the rate of coagulation of the dust particles is not as appreciable as in the previous case. However, attention is drawn to the fact that in the absence of deposition during the coagulation process the graph of $n_d(r_k)$ has an additional maximum in the region of larger sizes (Fig. 8b).

Thus, the behavior of the particle size distribution in a dusty plasma is anomalous and has various interesting features which may lead to some unusual physical effects such as the levitation of a long-lived quasi-liq-

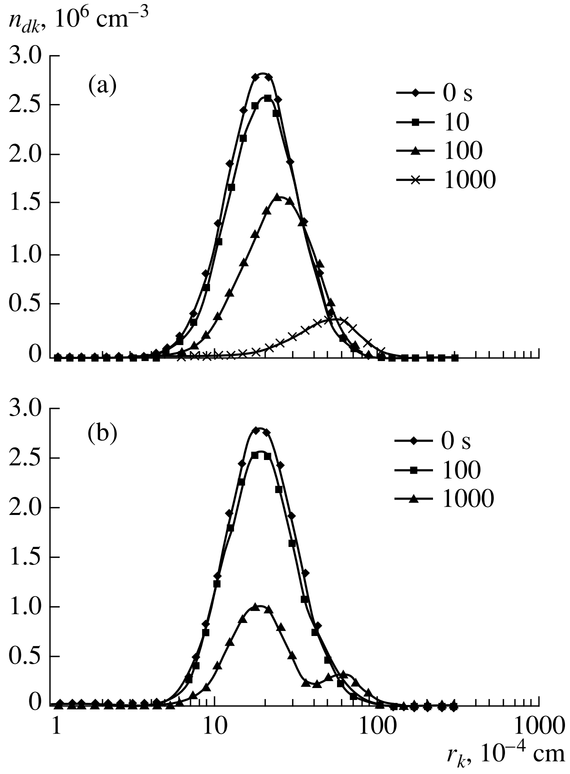


Fig. 8. Dynamics of variation of a narrow particle size distribution in an uncharged polydisperse aerosol (a) and in a dusty plasma (b) with $\bar{r}_k = 2 \times 10^{-3}$ cm.

uid dusty structure in an electrostatic trap. The lifetime of an uncharged aerosol having a particle concentration higher than 10^9 cm^{-3} is extremely limited and its properties have been little studied. Thus, the creation of a stable aerosol with such a high particle density is not only of applied interest but also of purely scientific interest.

To conclude, it should be noted that this investigation was carried out in the diffusion approximation. However, qualitatively similar results are also obtained for low pressures and smaller particles, i.e. in cases of molecular flow. In this context, the influence of these effects on structural rearrangement processes in cosmic dust formations is of particular interest. However, this aspect requires special study.

ACKNOWLEDGMENTS

The authors are deeply grateful to V.Yu. Baranov, V.E. Fortov, and H. Hora for their interest in the work and useful discussions.

REFERENCES

1. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997).
2. S. J. Choi and M. J. Kushner, J. Appl. Phys. **74**, 853 (1993).
3. S. A. Krapivina, *Plasma-Chemical Technological Processes* (Khimiya, Leningrad, 1981).
4. S. R. Forrest and T. A. Witten, J. Phys. A: Math. Gen. **12**, L109 (1979).
5. F. J. Huang and M. J. Kushner, J. Appl. Phys. **81**, 5960 (1997).
6. N. A. Fuks, *Mechanics of Aerosols* (Md., Army Chemical Center, 1958; Akad. Nauk SSSR, Moscow, 1955).
7. V. Yu. Baranov, I. A. Belov, A. V. Dem'yanov, *et al.*,. Preprint No. IAE-6105/6 (Moscow, Institute of Atomic Energy, 1998); I. A. Belov, A. S. Ivanov, D. A. Ivanov, *et al.*, Pis'ma Zh. Tekh. Fiz. **25**, 89 (1999).
8. G. M. Hidy and J. R. Brock, J. Coll. Sci. **20**, 123 (1965).
9. G. M. Hidy and J. R. Brock, *The Dynamics of Aerocolloidal Systems* (Pergamon, Oxford, 1970).
10. P. C. Reist, *Introduction to Aerosol Science* (Mir, Moscow, 1987; Macmillan, London, 1984).
11. B. M. Smirnov, *Aerosols in Gases and Plasmas* (Inst. Vysokh. Temp. Akad. Nauk SSSR, Moscow, 1990).
12. J. J. Dutton, J. Phys. Chem. Ref. Data **4**, 557 (1975).
13. B. M. Smirnov, *Ions and Excited Atoms in Plasmas* (Atomizdat, Moscow, 1974).
14. J. B. A. Mitchell, Phys. Rep. **186**, 215 (1990).
15. H. Bunz, M. Kouro, and W. Schock, *NAUA-MOD4–A Code for Calculating Aerosol Behaviour in LWR Core Melt Accidents, Code Description and Users Manual, KfK-3554* (Kernforschungszentrum, Karlsruhe, 1983).

Translation was provided by AIP