DNA molecule as an elastic Heisenberg chain

V. L. Golo

Department of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

E. I. Kats

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia
(Submitted 11 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 1833–1844 (May 1997)

A DNA molecule is simulated by an anisotropic elastic fiber which defines the configuration of the molecule central line and is supplemented with a chain of quantum two-level systems imitating hydrogen bonds between two polynucleotide chains in the DNA double helix. The system Hamiltonian consists of Kirchhoff's classical elastic energy and the energy of a quantum anisotropic chain of ''spins'' 1/2. The two-level systems and macroscopic vector variables which determine the conformation of the central line are coupled by a classical vector field $\mathbf{q}$, which is introduced to take into account the existence of two polynucleotide strands. Averaging over fast (microscopic) variables yields an effective potential $U(\mathbf{q})$. In the approximation of weak coupling between the systems, the spectrum of elementary excitations and effective potential $U(\mathbf{q})$ have been calculated in explicit form. The relation between elementary excitations in the ''magnetic'' subsystem and so-called breathing modes [C. Mandel, N. R. Kallenbach, and S. W. Englander, J. Mol. Biol. **135**, 391 (1980); G. Manning, Biopolymers **22**, 689 (1983)] corresponding to low-frequency excitations in DNA molecules is discussed.
 [S1063-7761(97)02105-7]

1. INTRODUCTION

According to Crick and Watson's model (see the monograph[3] and the review[4], and references therein), a DNA molecule consists of two polynucleotide chains (strands) intertwined in a double helix. The strands are connected to each other by relatively weak hydrogen bonds, and elements of the chains are connected to one another by much stronger covalent bonds. This structure is characterized by three spatial scales, namely (1) a microscale no larger than 10 Å (this is the diameter of the double helix or the distance of 3.4 Å between two neighboring bases in a chain); (2) a mesoscopic scale of the order of the persistence length of a DNA molecule (this is the length over which the elastic energy is comparable to the temperature) of about 10^3 Å; and (3) a macroscopic scale equal to the total molecule length, which can be up to 10^{10} Å.

Given these large differences among the scales, one can adopt a variety of approaches in describing DNA properties on different scales. For example, hydrogen bonds must be analyzed using the quantum-mechanical approach, the properties on the mesoscale are determined by the elastic energy[1)], and on the macroscopic scale a DNA molecule is similar to a conventional polymer molecule, and most of its features can be interpreted using well-known scaling laws[6] and are largely determined by the conformational entropy.

On the other hand, the differences among these scales do not mean that the respective properties are fully independent. In some phenomena, two of the scales or even all three can be important. One of them is the well-known and biologically very important division (replication) of two complementary polynucleotide chains of a DNA molecule. At higher temperatures (or in an appropriate chemical environment) hydrogen bonds holding two strands together are broken, and two single-chain molecules are formed from one double helix. As noted above, lengths of hydrogen bonds are microscopic, hence the description of their rupture is a quantum-mechanical problem. Furthermore, the two complementary chains should separate in the process of replication, therefore they should bend with respect to each other, and this process is, obviously, controlled by elastic energy, i.e., processes on the mesoscopic scale should be taken into consideration. Models taking into account such interaction between microscopic (hydrogen bonds) and mesoscopic (elastic energy) scales have been studied by several authors.[7–10] Dauxois *et al.*[7–9] described this interaction as softening of the elastic bending constant at high amplitudes of normal oscillations, whereas Marky and Manning[10] ascribed this softening to the rupture of hydrogen bonds.

In our previous brief report,[11] we described this behavior in terms of a classical vector field $\mathbf{q}$ introduced to account for the existence of two polynucleotide chains. The stability of a DNA molecule against replication was provided by an external potential $U(\mathbf{q})$ described in the harmonic approximation, i.e., an additional parameter was introduced to the theory. The aim of the present work is to calculate the potential $U(\mathbf{q})$ using the microscopic approach and to study the effect of hydrogen bonds on the mesoscopic elastic parameters of DNA.

This paper is organized as follows. In Sec. 2 a model of a DNA-like molecule is formulated. The molecule is described as an anisotropic elastic fiber which defines the configuration of the double-helix backbone. Hydrogen bonds between two polynucleotide chains are simulated by quantum two-level systems, i.e., we assume that a hydrogen bond can be only in two states, closed and open. A bilinear spin–orbit interaction leads to coupling between the two subsystems.

Thus, we consider the DNA molecule to be a combination of two systems, one of which is classical and determines the molecule elasticity, and the other of which is a quantum system responsible for binding the strands in the double helix. We assume that the spatial configuration of the molecule is largely determined by the classical system, while the quantum system is ''tuned up'' to this configuration, i.e., an analog of the Born–Oppenheimer approximation can be used.

In Sec. 3 the proposed model is investigated. Averaging over the microscopic variables, which are faster than mesoscopic ones, yields the effective potential $U(\mathbf{q})$ stabilizing the DNA molecule. An explicit expression for $U(\mathbf{q})$ has been obtained in the limit of weak ''spin–orbit'' coupling.

In Sec. 4 the spectrum of elementary excitations for the model is calculated. Finally, in the last section, physical consequences of our results are discussed, particularly the relationship between elementary excitations of two-level systems and the so-called breathing modes of the DNA molecule, which were discussed in Ref. 2 in order to interpret the experimental data[1] on fluctuations of hydrogen bonds.

2. TWO-LEVEL SYSTEMS AND HYDROGEN BONDS BETWEEN STRANDS IN A DNA MOLECULE

Following Marky and Manning,[10] we consider the central line of the DNA molecule to be an elastic fiber or rod. As noted above, the conformation of this fiber on the intermediate (mesoscopic) scale is largely controlled by its elastic energy. The latter can be expanded, as usual, in terms of the strain tensor, and these expansions can be conveniently expressed in a form which is a generalization of the solution to Kirchhoff's classic problem of equilibrium of an elastic rod[12–14]:

$$E_0=\int_0^L ds\left(\frac{1}{2}a_{ij}\omega_i\omega_j+b_i\omega_i\right). \tag{1}$$

Here L is the length of the elastic fiber (the condition that the mechanical model apply limits this length to several persistence lengths), s is the distance measured along the fiber axis, a_{ik} is the matrix of elastic constants of the rod (the anisotropy of a_{ik} reflects the existence of two nonidentical helices), and the vector $\mathbf{b}$ describes spontaneous deformation of the DNA molecule, which leads to its superhelical structure. A physical cause of spontaneous deformation can be, for example, adsorption of a DNA molecule on a nucleosome, which is usually described as a cylindrical surface (see Ref. 10, where the equilibrium configuration of a nucleosome is studied using the models of an elastic rod). Introduction of the vector $\boldsymbol{\omega}$ requires some clarification.

In order to describe the conformation of the DNA molecule central line, one can introduce the moving Frenet frame[15] $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, where the $\mathbf{v}_1$ is the tangent vector, and $\mathbf{v}_2$ and $\mathbf{v}_3$ are aligned with the principal deformation axes of the elastic rod,[16] i.e.,

$$\mathbf{v}_1=\frac{d\mathbf{r}}{ds},\quad \mathbf{v}_i\cdot\mathbf{v}_j=\delta_{ij}. \tag{2}$$

The vector $\boldsymbol{\omega}$ in Eq. (1) describes variation of the local Frenet frame along the fiber:

$$\frac{d}{ds}\mathbf{v}_j=\boldsymbol{\omega}\times\mathbf{v}_j. \tag{3}$$

Equation (3) means that all admissible deformations of the fiber can be described in terms of rotations of the Frenet frame, i.e., the elastic fiber is incompressible, which is a fairly good approximation for a DNA molecule, whose torsional and bending elastic constants are several orders of magnitude less than the compressibility constant.[4,17]

Note another important factor. The term in Eq. (1) linear in ω_i indicates that there is a preferred alignment of a DNA helix adsorbed on a nucleosome. Indeed, a DNA molecule adsorbed on a protein cylinder composed of histons is a DNA section of 146 base pairs (bp) wound in the shape of a left-handed helix of 1.8 turns on the cylinder.[10] Therefore the expression for the elastic energy containing a homogeneous quadratic form in ω_i, which is used by some authors,[12] is probably an oversimplification.

At constant a_{ij} and $\mathbf{b}$, the minimum of the energy described by Eq. (1) corresponds to a constant value of $\boldsymbol{\omega}$:

$$\omega_i=\left(\frac{1}{a}\right)_{ij}b_j,$$

which describes the helical conformation of the molecule.

Thus far, we have not taken into account the existence of two strands forming the DNA double helix and hydrogen bonds between them; we have only discussed the conformation of the central axis. A minimal extension of the elastic model that facilitates the simulation of hydrogen bonds and double helix might be as follows. We assume that hydrogen bonds between two polynucleotide chains can be modeled as two-level systems. The two states of such a system correspond to open (broken) and closed hydrogen bonds. The strength of the hydrogen bond is about 5 kcal/mol, and under natural physiological conditions about 1% of such bonds are virtually broken.[1,2]

Thus, imagine that two-level systems are associated with points on the central axis, whose conformation is determined by the energy defined in Eq. (1). This compound system of the elastic fiber plus the two-level systems is a minimal generalization of the model described by Eq. (1), taking hydrogen bonds into account, and it allows us to calculate the DNA conformation on the mesoscopic scale. Our goal is to describe Kirchhoff's elastic fiber supplemented with the two-level systems.

It is convenient to replace Kirchhoff's continuous fiber with a discrete chain (note that Klenin *et al.*[17] and Chirico and Langowski[18] studied a discrete version of the elastic model (1)). Thus, there is a two-level system, which can be either in an open (excited) or in a closed (ground) state, $|1\rangle$ or $|0\rangle$, at each site of the chain, $n=0,\pm1,\pm2\ \ldots$. In this Hilbert space, at each site we define a basis:

$$|0\rangle=\begin{pmatrix}0\\1\end{pmatrix},\quad |1\rangle=\begin{pmatrix}1\\0\end{pmatrix}, \tag{4}$$

and the Hilbert space of the entire system is defined by the basis which is a direct product of all eigenvectors defined in Eq. (4) at all sites.

The following simple relationships are obviously valid: $\sigma_-|1\rangle=|0\rangle$; $\sigma_+|0\rangle=|1\rangle$, where the operators σ_-, σ_+ in the basis defined by Eq. (4) take the form

$$\sigma_-=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \quad \sigma_+=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}.$$

The Hamiltonian of the two-level systems can be also presented in the same basis (its reference frame is defined by the conformation of hydrogen bonds in the molecule and therfore is local, i.e., it depends on the position of a specific nucleotide pair):

$$H_m=-\varepsilon\sum_n \sigma_n^{(3)}+\sum_n \nu_1(\sigma_n^+\sigma_{n+1}^-+\sigma_n^-\sigma_{n+1}^+)$$

$$+\nu_2\sum_n (\sigma_n^+\sigma_{n+1}^++\sigma_n^-\sigma_{n+1}^-). \tag{5}$$

The meaning of components of the Hamiltonian (5) is quite clear. The first term determines the energy of two-level systems, ε is the energy required to break a hydrogen bond, and the operator $\sigma^{(3)}$ has the form

$$\sigma^{(3)}=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

whereas the second and third terms determine the transition rates to neighboring sites with and without a change in the states of the two-level systems (ν_1 and ν_2 are the corresponding matrix elements, and we assume, for simplicity, that only nearest neighbors interact).

In this form, the Hamiltonian (5) is too complicated, even in the one-dimensional case. It corresponds to the so-called XYZ-model in an external field, and although this model enables one to obtain an exact solution,[19] it is too cumbersome and will not be used in our study, though excitations like Bethe's spin complexes can be interesting for studies of nonlinear effects in DNA conformation (these phenomena were studied by some authors in the framework of a nonlinear DNA model[7–9,17]).

We assume for simplicity that $\nu_2=0$, i.e., the matrix element of the transition between sites with, for example, two broken hydrogen bonds, is zero. In this case, our model transforms to the so-called XXZ-model, which was first studied by Bethe.[20] Thus, in this case ($\nu_1\equiv\nu$, $\nu_2=0$) we have

$$H_m=-\varepsilon\sum_n \sigma_n^{(3)}+\nu\sum_n (\sigma_n^+\sigma_{n+1}^-+\sigma_n^-\sigma_{n+1}^+). \tag{6}$$

Let us determine how the operators $\sigma_n^{(3)}$, $\sigma_n^{\pm}$, $n=0, 1, 2, \ldots, N-1$ are related to the DNA conformation. To this end, let us take into account the fact that the breaking of a hydrogen bond at site n, which is described via $\sigma_n^{(3)}$ as a transition to the excited state, has, from an outsider's viewpoint, a certain preferred direction defined by a unit vector $\mathbf{u}_n$ in the laboratory frame. Therefore, the states of hydrogen bonds are described, from an outsider's viewpoint, by the operators $\mathbf{S}_n$, which are related to the local operators $\sigma_n^{(i)}$ by the canonical transformation[21]

$$\mathbf{S}_n=\mathbf{u}_n\sigma_n^{(3)}+\mathbf{A}_n\sigma_n^++\mathbf{A}_n^*\sigma_n^-. \tag{7}$$

Here $\sigma_n^{(2)}\equiv(1/2i)(\sigma_n^+-\sigma_n^-)$, $\sigma_n^{(1)}\equiv(1/2)(\sigma_n^++\sigma_n^-)$, and $\mathbf{u}_n$ is the unit vector of the ''quantization'' axis, i.e., the direction in which the hydrogen bond at site n can be broken (opened). The complex vector $\mathbf{A}_n$ is orthogonal to $\mathbf{u}_n$ ($\mathbf{A}_n\cdot\mathbf{u}_n=0$, $\mathbf{A}_n^*\cdot\mathbf{u}_n=0$). Furthermore, if

$$\mathbf{A}_n\cdot\mathbf{A}_n=0, \quad \mathbf{A}_n\cdot\mathbf{A}_n^*=\frac{1}{2},$$

the vector operators $\mathbf{S}_n$ obey the usual commutation relations for spin operators:

$$[S_n^{(i)},S_m^{(j)}]=i\varepsilon_{ijk}S_n^{(k)}\delta_{mn},$$

which justifies our using the terminology usually applied to magnetic materials.

The choice of the vector $\mathbf{A}_n$ is somewhat arbitrary, which enables us to introduce a reference frame $\mathbf{W}_n^{(1)}$, $\mathbf{W}_n^{(2)}$, $\mathbf{W}_n^{(3)}$ determined by the hydrogen bonds,

$$\mathbf{u}_n\equiv\mathbf{W}_n^{(3)}, \quad \frac{1}{2}(\mathbf{A}_n+\mathbf{A}_n^*)\equiv\mathbf{W}_n^{(1)}, \quad \frac{1}{2i}(\mathbf{A}_n-\mathbf{A}_n^*)\equiv\mathbf{W}_n^{(2)}, \tag{8}$$

and, in addition, to choose the vector $\mathbf{W}_n^{(1)}$, which coincides with the tangent to the central line, i.e.,

$$\mathbf{W}_n^{(1)}\equiv\mathbf{v}_n^{(1)}, \tag{9}$$

while $\mathbf{v}_n$ is defined in accordance with the discrete version of Eq. (2).

Thus, as a result of the calculations in this section, we have introduced two local reference frames, namely, $\{\mathbf{v}_n^{(i)}\}$, defined by the deformation of the molecule central line, and $\{\mathbf{W}_n^{(j)}\}$, determined by hydrogen bonds. Notably, in our ''gauge'' defined by Eq. (9), $\mathbf{W}_n^{(1)}$ coincides with the vector tangent to the central line $\mathbf{v}_n^{(1)}$.

Both these references (or, what is the same, Kirchhoff's Hamiltonian of the elastic fiber in Eq. (1) and the Hamiltonian (6) of the two-level systems) have so far been independent. Recall that we are dealing with the DNA model in which the two-level systems are defined in the reference frame $\{\mathbf{W}_n^j\}$, and deformation of the central line is described by Kirchhoff's vector field $\boldsymbol{\omega}$ in the reference frame $\{\mathbf{v}_n^{(i)}\}$. But a DNA molecule consists of two polynucleotide chains. As noted in our previous publication,[11] a classical vector field $\mathbf{q}$ should be included in Kirchhoff's model of a single elastic fiber. This field also acts as a mediator between the Kirchhoff and ''magnetic'' subsystems. In the simplest manner, this interaction can be described by the following bilinear contribution to the energy:

$$H_{\text{int}}=\gamma\sum_n U_n^{-1}\mathbf{q}_n\cdot\boldsymbol{\sigma}_n. \tag{10}$$

Here γ is the effective coupling constant, $\mathbf{q}_n$ is the vector defining the direction in which a bond is broken in Kirchhoff's reference frame $\{\mathbf{v}_n^{(i)}\}$, $\boldsymbol{\sigma}_n$ is defined by Eq. (7) in the

reference frame $\{\mathbf{W}_n^{(i)}\}$, and the rotation matrix U_n, which transforms one reference frame into the other, is

$$U_n=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_n & -\sin\varphi_n \\ 0 & \sin\varphi_n & \cos\varphi_n \end{pmatrix},$$

where φ_n is the local angle of rotation about the tangent vector of the curve, which is a vector common to both references. Note that the interaction described by Eq. (10) has a structure similar to that of the spin–orbit coupling, and contains $\mathbf{q}_n$ and $\boldsymbol{\sigma}_n$ in the lowest powers allowed by the symmetry of the system.

3. CONFORMATION ENERGY

The model described in the previous section thus includes the energy of elastic deformation of the DNA central line given by Eq. (1), the Hamiltonian (6) of the two-level systems on this curve, and ''spin–orbit'' coupling (10). Generally speaking, it must also include the energy due to inhomogeneity of the vector field $\mathbf{q}$, which is expressed in the harmonic approximation as

$$E_g=\frac{K}{2}\sum_n (\nabla\mathbf{q}_n)^2, \tag{11}$$

where ∇ is a covariant derivative, since the change in the vector $\mathbf{q}_n$ due to motion along the line can be ascribed both to the rotation of the Kirchhoff reference frame and the change in $\mathbf{q}_n$ with respect to the local reference. In the continuous limit

$$\nabla\mathbf{q}\equiv\frac{\partial}{\partial s}\mathbf{q}+\boldsymbol{\omega}\times\mathbf{q}$$

(in the discrete version, we should use $\boldsymbol{\omega}_n$ and $\mathbf{q}_n$ and replace the derivative $\partial/\partial s$ with a finite difference).

Here we will not fully investigate the model that takes all the contributions to the energy given by Eqs. (1), (6), (10), and (11) into account, but limit our study to the case of a closed non-superhelical conformation supplemented by two-level systems. In this case, the Hamiltonian contains only the terms defined by Eqs. (6) and (10), which can be expressed as

$$\begin{aligned} H=\gamma\sum_{n=0}^{N-1}&[(\cos\varphi_n q_n^{(2)}+\sin\varphi_n q_n^{(3)})\sigma_n^{(2)} \\ &+(-\sin\varphi_n q_n^{(2)}+\cos\varphi_n q_n^{(3)})\sigma_n^{(3)}]-\varepsilon\sum_{n=0}^{N-1}\sigma_n^{(3)} \\ &+\nu\sum_{n=0}^{N-1}(\sigma_n^+\sigma_{n+1}^-+\sigma_n^-\sigma_{n+1}^+). \end{aligned} \tag{12}$$

All the notations in Eq. (12) have been described above, and the chain of N sites is closed.

It is convenient to transform the operators in Eq. (12) into Bose operators of approximate second quantization[21]:

$$b_n=\sigma_n^+,\quad b_n^+=\sigma_n^-,\quad \sigma_n^{(3)}=\frac{1}{2}-b_n^+b_n. \tag{13}$$

The operators $b_n^{\pm}$ are Bose operators in the approximation of low density of excited bonds (which means, from the physical viewpoint, that the temperature is much lower than the DNA denaturation energy). Note also that these exitations are formally similar to excitons in dielectrics.[22]

Substituting Eq. (13) into (12) and taking a Fourier transform, with

$$b_n=\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\exp\left(i\frac{2\pi nk}{N}\right)b_k,$$

$$b_n^+=\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\exp\left(-i\frac{2\pi nk}{N}\right)b_k^+$$

we obtain the following Hamiltonian in the k-representation:

$$\begin{aligned} H=&\frac{1}{2}\gamma\sum_{n=0}^{N-1}(-\sin\varphi_n q_n^{(2)}+\cos\varphi q_n^{(3)}) \\ &+\sum_{k_1k_2}\varepsilon_{k_1k_2}b_{k_1}^+b_{k_2}+\sum_k (D_kb_k^++D_k^*b_k). \end{aligned} \tag{14}$$

Here we have used the notation

$$\begin{aligned} \varepsilon_{k_1k_2}=&\delta_{k_1k_2}\left(\varepsilon+2\nu\cos\frac{2\pi k_1}{N}\right)-\frac{\gamma}{N}\sum_{n=0}^{N-1}(-\sin\varphi_n q_n^{(2)} \\ &+\cos\varphi_n q_n^{(3)})\exp\left(i\frac{2\pi(k_2-k_1)n}{N}\right) \end{aligned} \tag{15}$$

and

$$\begin{aligned} D_k=&\frac{i}{2}\frac{\gamma}{\sqrt{N}}\sum_{n=0}^{N-1}\exp\left(-i\frac{2\pi nk}{N}\right)(\cos\varphi_n q_n^{(2)} \\ &+\sin\varphi_n q_n^{(3)}). \end{aligned} \tag{16}$$

In order to diagonalize the Hamiltonian in Eq. (14), we have to perform a canonical transformation

$$b_k=B_k+l_k,\quad b_k^+=B_k^++l_k^*, \tag{17}$$

where l_k are not operators but c-numbers, which should be chosen so as to exclude the Hamiltonian components linear in the operators b_k and b_k^+. This condition yields

$$l_k=\sum_{k'}(\varepsilon_{kk'}^{-1}D_{k'}), \tag{18}$$

and yields the following result, apart from inessential constants:

$$\begin{aligned} H=&\frac{1}{2}\gamma\sum_{n=0}^{N-1}(-\sin\varphi_n q_n^{(2)}+\cos\varphi_n q_n^{(3)}) \\ &+\sum_{k_1k_2}\varepsilon_{k_1k_2}B_{k_1}^+B_{k_2}+3\sum_{k_1k_2}D_{k_1}(\varepsilon^{-1})_{k_1k_2}D_{k_2}. \end{aligned} \tag{19}$$

The last term in Eq. (19) determines the vacuum energy, and the canonical transformation given by Eq. (17) corresponds to a transition to coherent states. The first term on the right-hand side of Eq. (19) describes collective excitations (''spin waves'') in the two-level systems associated with the elastic

fiber. At temperatures sufficiently low in comparison with the temperature T_m of total denaturation of DNA, one can ignore excited spin waves, and Eq. (19) yields the conformation energy $E_{\rm conf}$ of the molecule.

In order to make our calculations easily understandable, we suppose that the ''spin–orbit'' coupling is weak, i.e., $\gamma \ll 1$. Then we can derive from Eq. (19) the following expression to order γ^2:

$$E_{\rm conf}=\frac{1}{2}\gamma\sum_{n=0}^{N-1}(\sin\varphi_n q_n^{(2)}+\cos\varphi_n q_n^{(3)}) + \frac{3}{4}\sum_{n=0}^{N-1}\frac{1}{\varepsilon+2\nu\cos(2\pi k/N)} \times\frac{1}{N}\sum_{n,m=0}^{N-1}\exp\left(i\frac{2\pi(m-n)k}{N}\right)(\cos\varphi_n q_n^{(2)} +\sin\varphi_n q_n^{(3)})(\cos\varphi_n q_n^{(2)}+\sin\varphi_m q_m^{(3)}). \quad (20)$$

In order to make further progress, we make one more natural assumption: $|\nu/\varepsilon|\ll 1$, which means that the transition matrix element between two neighboring sites is smaller than the energy of a broken hydrogen bond.

To the lowest order in this small parameter, one can easily derive from Eq. (20) the following formula:

$$E_{\rm conf}=\frac{\gamma}{4}\sum_{n=0}^{N-1}(e^{i\varphi_n}q_n+e^{-i\varphi_n}q_n^*)+\frac{3\gamma^2}{64\varepsilon}\sum_{n=0}^{N-1}2|q_n|^2 +2[(q_n^{(2)})^2-(q_n^{(3)})^2]\cos 2\varphi_n +4q_n^{(2)}q_n^{(3)}\sin 2\varphi_n. \quad (21)$$

Here we have introduced the notation

$$q_n\equiv q_n^{(2)}+iq_n^{(3)}.$$

Finally, we take into account the large difference between the mesoscopic and microscopic scales mentioned in the introduction. This means that the variable φ_n is rapidly oscillating on the mesoscopic scale, and Eq. (21) can be averaged over this variable, which yields

$$E_{\rm conf}=\frac{3\gamma^2}{32\varepsilon}\sum_{n=0}^{N-1}q_n^2. \quad (22)$$

Note that a similar harmonic potential ($\propto \int_0^L ds\ \mathbf{q}^2$) was used in our previous work,[11] where a model of DNA splitting was studied. The calculations in this section can be considered a derivation of this effective potential using the microscopic approach.

Note that for stability of a DNA molecule with respect to the conformation energy given by Eq. (22), the conditions $\varepsilon>0$ and $\nu>0$ (the latter derives from Eq. (20)) should hold. Both conditions are quite natural.

4. ELEMENTARY EXCITATIONS AND TOPOLOGICAL STRUCTURE OF THE MODEL

Elementary excitations are due to variations over time and space in the variables in Eqs. (1), (5), (10), and (11), which determine the total energy. In the general case, with arbitrary parameters of the subsystems and energy of interaction between them, the spectrum of elementary excitations can be calculated only numerically. In the limit of weak spin–orbit coupling, however, it can be derived approximately by analytic methods, the Kirchhoff elastic energy then being much higher than the energy of hydrogen bonds. Therefore, when spin–orbit coupling is weak, the vector field **q** due to the Kirchhoff energy can be considered an external field acting on the magnetic system (Hamiltonians in Eqs. (5) and (10)), i.e., the Born–Oppenheimer approximation can be used.

In this approximation, the magnetic subsystem can be described by the one-dimensional XYZ-model (or XXZ) in an inhomogeneous (in the general case) external field, which should be determined by solving the classical problem of an elastic Kirchhoff fiber supplemented with the vector field **q**, i.e., by minimizing the energies determined by Eqs. (1) and (11). If the field configuration is arbitrary, an exact solution for the XXZ-model is too cumbersome, even in one dimension, so we analyze the special case of a nonsupercoiled closed DNA molecule, for which $\mathbf{q}=\text{const}$, i.e., a one-dimensional XXZ-chain in a constant external field. An exact solution in this case can be obtained using the Bethe ansatz.[20]

As follows from a more accurate solution, the spectrum of elementary excitations consists of a branch of one-particle excitations (magnons) with the dispersion relation νk^2, where k is the quasimomentum or, in the continuous limit, the wave vector, as well as more complicated excitations (so-called complexes), which are propagating bound states of two or more broken bonds (with opposite spins). The latter excitations always have higher frequencies than single-particle excitations, and at temperatures far from the denaturation temperature they can be ignored. Here we do not give exact expressions for the spectrum of elementary excitations but, as in the previous section, limit our discussion to the mean-field approximation.

The spectrum of hydrogen-bond excitations derived from Eq. (15) contains topological information about the state of the molecule. Note that in deriving the spectrum, we have to diagonalize the matrix $\varepsilon_{k_1k_2}$. In first-order perturbation theory with respect to γ, we derive from Eq. (15)

$$\varepsilon_k=\varepsilon+2\nu\cos\frac{2\pi k}{N}-\frac{\gamma}{N}\sum_{n=0}^{N-1}(-\sin\varphi_n q_n^{(2)} +\cos\varphi_n q_n^{(3)}).$$

The last term is none other than the scalar product $\mathbf{q}_n\cdot\mathbf{u}_n$, which can be interpreted as follows. Since $N\gg 1$, we can consider the continuous model of the molecule as a closed curve in space, and the vector field **u** as a continuous field of vectors normal to the curve. Then the scalar product $\mathbf{q}\cdot\mathbf{u}$ describes rotation of the field **q**, which, as noted above, defines the two-chain structure of the molecule in the normal-coordinate reference frame determined by the field **u** in the neighborhood of the curve.

Note that the dispersion relation for ε_k in first-order perturbation theory with respect to γ yields only the average value of this rotation. The last term in Eq. (15) is the Fourier

coefficient of the periodic function of the coordinate on the curve, $\mathbf{q}\cdot\mathbf{u}$, and yields complete information about rotation of the system of two strands. An important point is that Eq. (15) takes account of rotation of the field $\mathbf{q}$ in the normal-coordinate reference frame defined by the conformation of hydrogen bonds between the chains of the double helix.

A change in the conformation or topology of the DNA molecule leads to a change in the spectrum of excitations ε_k if it changes the value of rotation Nq of the field $\mathbf{q}$. Note the relationship between the rotation Nq of the field $\mathbf{q}$ and the so-called linking number Lk, which has been studied by several authors.[4,5,17] In our model, the field $\mathbf{q}$ simulates the structure of the double helix, i.e., the presence of two chains, therefore Nq describes the linkage between these chains. Hence, the change in the spectrum ε_k reflects a change in the topology of the DNA conformation. The behavior of breathing modes introduced by Manning[2] to describe the dynamics of hydrogen bonds is probably also related to the topology of DNA conformation. At present, the connection between DNA topology and mobility has been established with certainty (see Ref. 4 and references therein). An experimental discovery of a relationship between DNA topology and the spectrum of elementary excitations would undoubtedly be of great interest.

5. CONCLUSIONS

Manning[2] published detailed data about two types of elementary excitations observed in real DNA molecules. These are so-called bending and breathing modes. The former are due to bending of a DNA molecule (i.e., a transition from the state with zero mean curvature to a state with a definite rms curvature), whereas the latter are due to the breaking of a hydrogen bond and unwinding of the double helix to the next closed bond.

The bending energy of a segment of length $l=260$ Å (the mean distance between open hydrogen bonds at temperatures close to the denaturation point) for the rms curvature of a DNA molecule ($\sqrt{\langle 1/r^2\rangle}\approx 1/200$ Å^{-1}) is $\Delta G\approx 0.59$ kcal/mol.[1,2,10]

Using standard concepts of bending in elastic rods, we can estimate the bending elastic constant a in the Kirchhoff energy given by Eq. (1):

$$a\approx l\Delta G\approx 10^{-20}-10^{-19}\ \text{erg}\cdot\text{cm}.$$

Hence, the dispersion relation for bending modes is

$$\omega\approx\sqrt{\frac{a}{\rho_1}}q^2.$$

Here ω is the frequency, q is the wave vector, and ρ_1 is the molecular mass per unit length:

$$\rho_1=\rho d^2\approx 10^{-15}\ \text{g/cm},$$

where d is the molecular diameter (~ 10 Å).

For wave vectors q of the order of the inverse segment length l, we have a mode of very high frequency, $\omega\sim 10^{10}-10^{11}$ s^{-1}.

Thus, bending modes can be identified with eigenmodes determined by the Kirchhoff component of the energy in Eq. (1). As for breathing modes, they are naturally associated with the low-frequency part of the elementary excitation spectrum (spin waves) due to the magnetic energy (Eq. (6)). The parameter ε (the bond rupture energy) in this energy is well known[1,2,10] and equals about 5 kcal/mol (i.e., about 10^2 K per bond).

It is more difficult to determine the parameter ν (transition matrix element between neighboring sites). Given the temperature coefficient of the length of the closed-bonds segment[2] (-1.1 ± 0.1 bp/deg, where bp is the separation between neighboring base pairs along the DNA central line equal to 3.4 Å), one can estimate ν as follows. Suppose that the temperature has changed by one degree. In accordance with the temperature coefficient given above, the boundary of the closed-bonds segment has shifted through a length equal to the distance to the next bond. On the other hand, this small change in the temperature (much smaller than the bond energy, which equals, as noted above, $\sim 10^2$ K) does not change the gap in the spectrum of excitations, and this allows us to estimate ν. Comparison with the dispersion relation in the previous section yields $\nu\simeq 10^{-17}-10^{-16}$ erg.

From this estimate of ν, we can derive the group velocity v_b of the breathing mode at a wave vector of approximately the inverse length l of the closed-bond segment:

$$v_b\sim\frac{\nu d^2 q}{\hbar}\approx 1-10\ \text{cm/s}.$$

For comparison, the group velocity of the bending mode at the same wavelength is

$$v_{\text{ben}}\sim 10^5\text{cm/s}.$$

Thus, we have come to the conclusion that spin waves of the magnetic subsystem described by the Hamiltonian in Eq. (6) describe breathing modes, which are really observed in DNA molecules, and the fast bending modes are elastic waves determined by the Kirchhoff energy. Note that the existence of fast and slow modes with very different velocities may be important, in view of the role DNA plays in the cell biophysics.

This work was performed as a part of the *Statistical Physics* state-sponsored program, and was supported by INTAS (Grant No. 94-40-78) and the Russian Fund for Fundamental Research (Grant No. 96-02-16207a).

[1)] In a recent publication[5] a detailed analysis of the DNA molecule conformation in terms of an elastic model taking into account thermal fluctuations was described.

[1] C. Mandel, N. R. Kallenbach, and S. W. Englander, J. Mol. Biol. **135**, 391 (1980).
[2] G. Manning, Biopolymers **22**, 689 (1983).
[3] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson, *Molecular Biology of the Cell*, Garland Publishers, New York (1989).
[4] A. V. Vologodskii, S. D. Leven, K. V. Klenin *et al.*, Ann. Rev. Biophys. Biomol. Struct. **23**, 609 (1992).
[5] B. Fain, J. Rudnik, and S. Ostlund, E-prints Archive, Cond. Mat./96 10 126.
[6] P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University, Ithaca (1979).
[7] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[8] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[9] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[10] N. L. Marky and G. S. Manning, Biopolymers **31**, 1543 (1991).

[11] V. L. Golo and E. I. Kats, JETP Lett. **62**, 627 (1995).
[12] J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).
[13] V. L. Golo and E. I. Kats, JETP Lett. **60**, 679 (1994).
[14] G. Kirchhoff, *Mechanik*, Tenbnir, Berlin (1897).
[15] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Contemporary Geometry* [in Russian], Nauka, Moscow (1979).
[16] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, New York (1980).
[17] K. V. Klenin, H. D. Frank-Kamenetskiĭ, and J. Langowski, Biophys. J. **68**, 81 (1995).
[18] G. Chirico and J. Langowski, Biopolymers **34**, 415 (1994).
[19] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems* [in Russian], Nauka, Moscow (1987).
[20] H. Bethe, Z. Phys. **71**, 205 (1931).
[21] S. V. Tyablikov, *Methods in the Quantum Theory of Magnetism* [in Russian], Nauka, Moscow (1965).
[22] V. M. Agranovich, *Theory of Excitons* [in Russian], Nauka, Moscow (1968).

Translation provided by the Russian Editorial office.

Cutoff of long-wave phonons in a nanocrystal due to a nonuniform strain field

V. V. Meshcheryakov[1)]

Moscow Institute of Steel and Alloys, 117936 Moscow, Russia
(Submitted 19 December 1995; resubmitted 19 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 1845–1857 (May 1997)

This paper considers the effect of extended monopole and dipole strain fields on the low-frequency boundary of the phonon spectrum in a crystal of finite dimensions. The boundary shift depends on the dynamical volume of the nonuniform strain region, which is determined by the parameters of the crystal and the sources of stress. An increase in the volume of the deformed region leads to a decrease in the undistorted part of the crystal, where a phonon with the largest wavelength can be produced. A monopole strain field is more efficient in cutting off long-wave phonons than a dipole strain field, and can ''soften'' the phonon spectrum. If a source generates stresses of the order of those on an interatomic scale, these effects can be the strongest and most diverse in crystals or phase precipitates with dimensions of less than 10^{-6} cm.
 [S1063-7761(97)02205-1]

1. INTRODUCTION

The phonon spectrum of an ideal macroscopic crystal extends from $\omega_L = 2\pi c/L$, where c is the speed of sound and L is the sample length, to the Debye frequency $\omega_D = \pi c/a$, where a is the interatomic spacing.[1] For example, if $L \approx 1$ cm, $a \approx 10^{-8}$ cm, and $c \approx 10^5$ cm/s, then $\omega_D \approx 10^{13}$ s^{-1} and $\omega_L \approx 10^6$ s^{-1}.

If L is reduced to nanocrystal dimensions that can be achieved by existing technologies, low-frequency phonons should be cut off from the spectrum, and the low-frequency boundary of the phonon spectrum should shift towards higher frequencies. For example, in a crystal with $L \approx 10^{-6}$ cm, the cutoff frequency is $\omega_L \approx 10^{12}$ s^{-1}.

A real crystal always contains strain sources that generate long-range static fields, in which atoms are displaced from their equilibrium positions. For example, vacancies, interstitial atoms, and impurities generate strain fields Q typical of dipole centers,[2] decaying over a range of several nanometers as a power of the distance, $Q \propto r^{-2}$. Such a source of strain in a macroscopic crystal shifts phonon modes by the energy of static lattice deformation,[3] hence it does not change the shape of the phonon spectrum. If the crystal size is in the nanometer range, the strain field region may occupy an appreciable part of its volume. Does the above-cited estimate of the lower spectrum boundary apply to this case? Should the phonon spectrum shift as a whole in the presence of a dipole-like point defect?

These questions are also appropriate in the case of a monopole static source, especially because no research, it seems, has been performed in this field, whereas regions with strain fields similar to those produced by a monopole source should be generated by external forces acting on a crystal (for example, in an atomic-force microscope[4]), by forces on grain boundaries in polycrystals (for example, during plastic deformation and destruction of materials[5]), by forces on boundaries between different phases (in alloys in pre-martensitic states[6]), etc. In an isotropic infinite medium, a monopole source generates a field[7] $Q \propto r^{-1}$, i.e., the monopole field has a longer range than the dipole field, so it can lead to greater changes in material properties.

The present paper demonstrates that in nanocrystals containing static strain sources, the estimate of the lower boundary of the phonon spectrum given above is not valid. The reason is the cutoff of long-wave phonons by nonuniform strain fields occupying finite volumes, which reduce the volume of the undistorted crystal, where harmonic lattice vibrations with the lowest frequecies occur. The problem is solved using a phonon Hamiltonian with terms due to point-like monopole and dipole sources of strain. Only if both of these strain sources are taken into account one can describe the nonuniform strain in terms of a superposition of harmonic oscillators after some canonical transformations of the Hamiltonian. The condition that the displacements and momenta associated with these oscillators be real is used in estimating cutoff frequencies.

The cutoff of the phonon spectrum is interpreted in terms of the stationary volume where a steady nonuniform strain occurs. It has been proved that the stationary volumes due to arbitrary strain sources in a quantized medium are finite. The relation of this result to the classical description of deformation fields is discussed.

This paper also describes an attempt to explain the damping property of alloys in terms of a band gap in the phonon spectrum, and suggests a feasible reason for a ''soft'' mode in structural transitions in solids.

2. STEADY STATES OF EXTENDED NONUNIFORM STRAIN IN A DYNAMIC CONTINUUM

Let us modify the previously known description of nonuniform strain in a medium[8] for solving the quantum problem of vibrations in a crystal by introducing a static strain distribution, taking into account the configuration of the lattice static deformation.

For a medium with given density ρ and elastic constant κ, the Hamiltonian describing oscillations of its particles and their linear displacements due to given nonuniform strain can be expressed as

$$H = H_f + H_d, \tag{1}$$

$$H_f=\sum_{\alpha=1}^{3}\int_0^{L^3}d\mathbf{r}\left[\frac{\kappa}{2}(\nabla Q_\alpha(\mathbf{r},t))^2+\frac{P_\alpha^2(\mathbf{r},t)}{2\rho}\right], \tag{2}$$

$$H_d=-\sum_{\alpha=1}^{3}\int_0^{L^3}d\mathbf{r}\ F_\alpha(\mathbf{r},t)Q_\alpha(\mathbf{r},t). \tag{3}$$

The function $P_\alpha(\mathbf{r},t)$ in Eq. (2) is the momentum density, and $Q_\alpha(\mathbf{r},t)$ is the displacement of a point of the continuum. The function H_f describes the motion of points with harmonic interactions among them, and H_d characterizes the nonuniform strain of the continuum generated by the force density $F_\alpha(\mathbf{r})$.

Let us express the force density in the form of a static multipole expansion[9]:

$$F_\alpha(\mathbf{r})=F_{1\alpha}\delta(\mathbf{r})-F_{2\alpha}(\mathbf{R}\cdot\nabla)\,\delta(\mathbf{r}), \tag{4}$$

where $F_{1\alpha}$ is the component of the monopole force, $F_{2\alpha}$ is the component of the dipole force, and $\mathbf{R}$ is the dipole separation vector.

The monopole term in Eq. (4) determines the density of a stress monopole or a ''single force,'' which is applied to a point of the continuum in a given direction and deforms the continuum without changing its volume. This force does not satisfy the condition of crystal equilibrium, and it can only be an external force applied to a crystal from the outside.

The dipole term in Eq. (4) determines the density of a stress dipole or a ''double force''[2] which is applied to the points defined by the vectors $\pm\mathbf{R}$ with respect to the origin and generate bulk deformation of the continuum.

The expansion given by Eq. (4) provides an exhaustive description of an arbitrary nonuniform strain due to a point-like source, since the force density in a physical system is bounded by the first two multipoles.

Thus, our study addresses a nonuniform strain defined by Eq. (4), generating a potential component in the Hamiltonian in Eq. (1), which belongs to the class of problems concerning the effect of local lattice irregularities on free oscillations of the atoms.[10] This distinguishes the paper from previous investigations of kinetic nonuniformities.[11,12]

Let us define the canonically conjugate momentum density and displacement in the form of an expansion in terms of plane monochromatic waves:

$$Q_\alpha(\mathbf{r},t)=\sum_{\mathbf{k}}\frac{1}{V^{1/2}}[A_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}e^{i\mathbf{k}\cdot\mathbf{r}}+A_{\mathbf{k}\alpha}^*e^{i\omega_{\mathbf{k}}t}e^{-i\mathbf{k}\cdot\mathbf{r}}],$$

$$P_\alpha(\mathbf{r},t)=\sum_{\mathbf{k}}\frac{i\rho\omega_{\mathbf{k}}}{V^{1/2}}[-A_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}e^{i\mathbf{k}\cdot\mathbf{r}}+A_{\mathbf{k}\alpha}^*e^{i\omega_{\mathbf{k}}t}e^{-i\mathbf{k}\cdot\mathbf{r}}], \tag{5}$$

where V is the crystal volume.

By substituting Eq. (5) into Eq. (2) and using the orthogonality condition

$$\frac{1}{V}\int e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\mathbf{r}=\delta_{\mathbf{k}\mathbf{k}'},$$

where $k_a=2\pi n_a/L_a$ is a wave vector component and n_a is an integer, we obtain the dispersion relation

$$\kappa\mathbf{k}^2-\rho\omega^2(\mathbf{k})=0$$

and the Hamiltonian

$$H_f=\sum_{\mathbf{k}\alpha}\frac{\hbar\omega_{\mathbf{k}}}{2}(a_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha}^*+a_{\mathbf{k}\alpha}^*a_{\mathbf{k}\alpha}). \tag{6}$$

The dimensionless amplitude $a_{\mathbf{k}\alpha}$ in Eq. (6) is related to the amplitude in the expansion (5) via

$$a_{\mathbf{k}\alpha}=(2\rho\omega_{\mathbf{k}}/\hbar)^{1/2}A_{\mathbf{k}\alpha}.$$

Let us discuss in detail the calculation of H_d. Substituting the expansion (5) into Eq. (3), we obtain

$$\begin{aligned}H_d=\sum_{\mathbf{k}\alpha}\left(\frac{\hbar}{2\rho V\omega_{\mathbf{k}}}\right)^{1/2}\Bigg[&F_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}\int d\mathbf{r}\ e^{i\mathbf{k}\cdot\mathbf{r}}\delta(\mathbf{r})\\&+F_{\mathbf{k}\alpha}^*a_{\mathbf{k}\alpha}^*e^{i\omega_{\mathbf{k}}t}\int d\mathbf{r}\ e^{-i\mathbf{k}\cdot\mathbf{r}}\delta(\mathbf{r})-F_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}\\&\times\int d\mathbf{r}\ e^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{R}\cdot\nabla)\,\delta(\mathbf{r})-F_{\mathbf{k}\alpha}^*a_{\mathbf{k}\alpha}^*e^{i\omega_{\mathbf{k}}t}\\&\times\int d\mathbf{r}\ e^{-i\mathbf{k}\cdot\mathbf{r}}(\mathbf{R}\cdot\nabla)\,\delta(\mathbf{r})\Bigg].\end{aligned}$$

Since

$$\int d\mathbf{r}\ e^{\pm i\mathbf{k}\cdot\mathbf{r}}\delta(\mathbf{r})=1,$$

$$\int d\mathbf{r}\ e^{\pm i\mathbf{k}\cdot\mathbf{r}}(\mathbf{R}\cdot\nabla)\,\delta(\mathbf{r})=\mp i\mathbf{k}\cdot\mathbf{r},$$

the function H_d can be expressed in the form

$$-\sum_{\mathbf{k}\alpha}\left(\frac{\hbar}{2M\omega_{\mathbf{k}}}\right)^{1/2}[F_{\mathbf{k}\alpha}(t)a_{\mathbf{k}\alpha}+F_{\mathbf{k}\alpha}^*(t)a_{\mathbf{k}\alpha}^*]. \tag{7}$$

This formula contains new variables, namely the crystal mass $M=\rho V$ and generalized forces

$$F_{\mathbf{k}\alpha}(t)=(F_{1\alpha}+i\mathbf{k}\cdot\mathbf{r}F_{2\alpha})e^{-i\omega_{\mathbf{k}}t},$$

$$F_{\mathbf{k}\alpha}^*(t)=(F_{1\alpha}-i\mathbf{k}\cdot\mathbf{r}F_{2\alpha})e^{i\omega_{\mathbf{k}}t}.$$

Expression (7) demonstrates that the static nonuniformity defined by Eq. (4) is transformed into steady-state oscillations in a dynamic continuum. This justifies our attempt to transform the sum (7) to a canonical Hamiltonian in terms of a set of harmonic oscillators.

With this end in view, let us introduce new amplitudes

$$q_{0\mathbf{k}\alpha}=-F_{1\alpha}/M\omega_{\mathbf{k}}^2,\quad p_{0\mathbf{k}\alpha}=-F_{2\alpha}\mathbf{k}\cdot\mathbf{R}/\omega_{\mathbf{k}}, \tag{8}$$

which transform the sum (7) to

$$\begin{aligned}\sum_{\mathbf{k}\alpha}\left(\frac{\hbar\omega_{\mathbf{k}}}{2M}\right)^{1/2}[&(M\omega_{\mathbf{k}}q_{0\mathbf{k}\alpha}+ip_{0\mathbf{k}\alpha})a_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}\\&+(M\omega_{\mathbf{k}}q_{0\mathbf{k}\alpha}-ip_{0\mathbf{k}\alpha})a_{\mathbf{k}\alpha}^*e^{i\omega_{\mathbf{k}}t}].\end{aligned} \tag{9}$$

Introducing the functions

$$\gamma_{\mathbf{k}\alpha}=\left(\frac{1}{2M\hbar\omega_{\mathbf{k}}}\right)^{1/2}(M\omega_{\mathbf{k}}q_{0\mathbf{k}\alpha}+ip_{0\mathbf{k}\alpha})e^{-i\omega_{\mathbf{k}}t},$$

$$\gamma_{\mathbf{k}\alpha}^*=\left(\frac{1}{2M\hbar\omega_{\mathbf{k}}}\right)^{1/2}(M\omega_{\mathbf{k}}q_{0\mathbf{k}\alpha}-ip_{0\mathbf{k}\alpha})e^{i\omega_{\mathbf{k}}t}$$

enables us to write the sum (9) in a well-known form[8] and, with due account of Eq. (6), we obtain

$$H=\sum_{\mathbf{k}\alpha}\frac{\hbar\omega_{\mathbf{k}}}{2}(a_{\mathbf{k}\alpha}a^{*}_{\mathbf{k}\alpha}+a^{*}_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha})$$

$$+\sum_{\mathbf{k}\alpha}\hbar\omega_{\mathbf{k}}[\gamma_{\mathbf{k}\alpha}(t)a_{\mathbf{k}\alpha}+\gamma^{*}_{\mathbf{k}\alpha}(t)a^{*}_{\mathbf{k}\alpha}]. \tag{10}$$

Let us use the transformations

$$a_{\mathbf{k}\alpha}=\tilde{a}_{\mathbf{k}\alpha}-\gamma^{*}_{\mathbf{k}\alpha},\quad a^{*}_{\mathbf{k}\alpha}=\tilde{a}^{*}_{\mathbf{k}\alpha}-\gamma_{\mathbf{k}\alpha},$$

which allow us to get rid of the constant factors $a_{\mathbf{k}\alpha}$ and $a^{*}_{\mathbf{k}\alpha}$ in the second term on the right-hand side of Eq. (10) and transform the Hamiltonian to a form convenient for introducing quantization operators:

$$H=\sum_{\mathbf{k}\alpha}\frac{\hbar\omega_{\mathbf{k}}}{2}(\tilde{a}_{\mathbf{k}\alpha}\tilde{a}^{*}_{\mathbf{k}\alpha}+\tilde{a}^{*}_{\mathbf{k}\alpha}\tilde{a}_{\mathbf{k}\alpha})$$

$$-\sum_{\mathbf{k}\alpha}\hbar\omega_{\mathbf{k}}\gamma_{\mathbf{k}\alpha}(t)\gamma^{*}_{\mathbf{k}\alpha}(t). \tag{11}$$

In order to quantize continuum vibrations, let us introduce Bose second quantization operators:

$$\tilde{a}_{\mathbf{k}\alpha}e^{-i\omega_{\mathbf{k}}t}\rightarrow\tilde{b}_{\mathbf{k}\alpha},\quad \tilde{a}^{*}_{\mathbf{k}\alpha}e^{i\omega_{\mathbf{k}}t}\rightarrow\tilde{b}^{+}_{\mathbf{k}\alpha}, \tag{12}$$

which satisfy the commutation relations

$$[\tilde{b}_{\mathbf{k}\alpha},\tilde{b}^{+}_{\mathbf{k}'\alpha'}]=\delta_{\mathbf{k}\mathbf{k}'}\delta_{\alpha\alpha'},\quad [\tilde{b}_{\mathbf{k}\alpha},\tilde{b}_{\mathbf{k}'\alpha'}]=[\tilde{b}^{+}_{\mathbf{k}\alpha},\tilde{b}^{+}_{\mathbf{k}'\alpha'}]=0. \tag{13}$$

Using Eq. (13), substituting Eq. (12) into Eq. (11), and expressing $\gamma_{\mathbf{k}\alpha}$ and $\gamma^{*}_{\mathbf{k}\alpha}$ in explicit form, we obtain the operator

$$\hat{H}=\sum_{\mathbf{k}\alpha}\hbar\omega_{\mathbf{k}}\left(\tilde{b}_{\mathbf{k}\alpha}\tilde{b}^{+}_{\mathbf{k}\alpha}+\frac{1}{2}\right)-\sum_{\mathbf{k}\alpha}\left(\frac{M\omega^{2}_{\mathbf{k}}q^{2}_{0\mathbf{k}\alpha}}{2}+\frac{p^{2}_{0\mathbf{k}\alpha}}{2M}\right). \tag{14}$$

The first term on the right-hand side of Eq. (14) is the Hamiltonian in the second quantization representation, which describes a superposition of noninteracting oscillators.[8] The second term looks like the Hamiltonian of a classical harmonic oscillator, but it is not, since the condition that displacements and momenta are real, which is implied by the expansion in Eq. (5), means that the displacements and momenta squared should be positive definite, and the functions $-q^{2}_{0\mathbf{k}\alpha}$ and $-p^{2}_{0\mathbf{k}\alpha}$ do not satisfy this condition.

One of the ways to obtain a Hamiltonian of noninteracting harmonic oscillators in canonical form is to use the transformation

$$q^{2}_{\mathbf{k}\alpha}=a\hbar/M\omega_{\mathbf{k}}-q^{2}_{0\mathbf{k}\alpha},$$
$$p^{2}_{\mathbf{k}\alpha}=b\hbar\omega_{\mathbf{k}}M-p^{2}_{0\mathbf{k}\alpha}, \tag{15}$$

where a and b are unspecified constant factors that satisfy the condition $a+b=1$. After this transformation, we derive from Eq. (14)

$$\hat{H}=\hat{\tilde{H}}_{f}+\tilde{H}_{d}=\sum_{\mathbf{k}\alpha}\hbar\omega_{\mathbf{k}}\tilde{b}_{\mathbf{k}\alpha}\tilde{b}^{+}_{\mathbf{k}\alpha}+\sum_{\mathbf{k}\alpha}\left(\frac{M\omega^{2}_{\mathbf{k}}q^{2}_{\mathbf{k}\alpha}}{2}+\frac{p^{2}_{\mathbf{k}\alpha}}{2M}\right). \tag{16}$$

Let us discuss a plausible interpretation of the Hamiltonian $\tilde{H}_{d}$ in Eq. (16). The form of the transformation given by Eqs. (15) shows that a region of nonuniform strain in a dynamic medium can be described in terms of a superposition of harmonic oscillators governed by the Hamiltonian $\tilde{H}_{d}$. But the motion described by $\tilde{H}_{d}$ is not purely classical. Let us reconsider the amplitudes defined by Eqs. (15). Using the definitions of $q_{0\mathbf{k}\alpha}$ and $p_{0\mathbf{k}\alpha}$ in Eq. (8), we obtain

$$q_{\mathbf{k}\alpha}=(a\hbar/M\omega_{\mathbf{k}}-F^{2}_{1\alpha}/M^{2}\omega^{4}_{\mathbf{k}})^{1/2},$$
$$p_{\mathbf{k}\alpha}=(b\hbar M\omega_{\mathbf{k}}-F^{2}_{2\alpha}(\mathbf{k}\cdot\mathbf{R})^{2}/\omega^{2}_{\mathbf{k}})^{1/2}. \tag{17}$$

These amplitudes determine real displacements and momenta only if the ground-state energy of crystal vibrations, which has a purely quantum nature, is taken into account. For $F_{1\alpha}=F_{2\alpha}$ the product $p_{\mathbf{k}\alpha}q_{\mathbf{k}\alpha}=\hbar/2$ is consistent with the uncertainty principle, according to which a material point in the phonon mode characterized by the vector $\mathbf{k}$ cannot have a definite position and momentum at the same time. We therefore conclude that steady states of a region with nonuniform strain are related to collective oscillations of a deformed continuum.

3. ENERGY SPECTRUM OF THE DEFORMED CONTINUUM

The spectrum of eigenvalues of the operator $\hat{H}$ is

$$\varepsilon_{\mathbf{k}\alpha}(\omega,\theta)=\varepsilon_{f\alpha}(\omega)+\varepsilon_{0}(\omega)+\varepsilon_{1\alpha}(\omega)+\varepsilon_{2\alpha}(\theta), \tag{18}$$

where $\varepsilon_{f\alpha}=\hbar\omega_{\mathbf{k}}n_{\mathbf{k}\alpha}$ is the energy of $n_{\mathbf{k}\alpha}$ phonons in the state with wave vector $\mathbf{k}$, $\varepsilon_{0}=\hbar\omega_{\mathbf{k}}/2$ is the ground-state energy of lattice vibrations with vector $\mathbf{k}$, $\varepsilon_{1\alpha}=-F^{2}_{1\alpha}/2M\omega^{2}_{\mathbf{k}}$ is the Fourier component of the monopole strain energy, and $\varepsilon_{2\alpha}=-F^{2}_{2\alpha}R^{2}\cos\theta/2Mc^{2}$ is the Fourier component of the dipole strain energy. In calculating $\varepsilon_{2\alpha}$, we have taken into account the phonon dispersion relation and introduced the polar angle θ between the plane-wave vector $\mathbf{k}$ and the dipole vector $\mathbf{R}$.

The requirement that the displacements and momenta in Eq. (17) be real imposes constraints on the spectrum defined by Eq. (18). These are related to the boundary frequencies that derive from the conditions $q^{2}_{\mathbf{k}\alpha}=0$ and $p^{2}_{\mathbf{k}\alpha}=0$:

$$\omega_{1\alpha}=(F^{2}_{1\alpha}/\hbar M)^{1/3}, \tag{19}$$

$$\omega_{2\alpha}=F^{2}_{2\alpha}R^{2}\cos^{2}\theta/\hbar Mc^{2}, \tag{20}$$

where we have put $a=b=1$. Having thus determined the coefficients a and b, we assume that the crystal contains either a monopole or dipole source. Then, after eliminating the unobservable zero-point energy, we derive from Eqs. (18)–(20) the energy of the excited phonon states:

$$\varepsilon_{\mathbf{k}\alpha}=\hbar\omega_{\mathbf{k}}n_{\mathbf{k}}-\frac{\hbar\omega^{3}_{1\alpha}}{2\omega^{2}_{\mathbf{k}}} \tag{21}$$

in the frequency band $\omega_{D}\geqslant\omega\geqslant\omega_{1\alpha}$ for a crystal containing a stress monopole, and

$$\varepsilon_{\mathbf{k}\alpha}=\hbar\omega_{\mathbf{k}}n_{\mathbf{k}}-\frac{\hbar\omega_{2\alpha}}{2} \tag{22}$$

in the frequency band $\omega_{D}\geqslant\omega\geqslant\omega_{2\alpha}$ for a crystal with a stress dipole.

It follows from Eq. (22) that the energy of phonon modes shifts by a constant value $-\hbar\omega_{2\alpha}/2$ owing to a stress

dipole, i.e., the energy of a mode with a given absolute value of the wave vector does not change. This result is known from the dynamics of a linear crystal containing a point defect.[3] But Eq. (22) also yields another result. Owing to the frequency bound $\omega \geqslant \omega_{2\alpha}$, modes with energies ranging between zero (or $\varepsilon_L = \hbar\omega_L$ due to finite crystal size) and $\hbar\omega_{2\alpha}/2$ are cut off.

In addition to the frequency cutoff, the spectrum defined by Eq. (21) contains a term that lowers energies of phonon modes by $\Delta\varepsilon_{\mathbf{k}}$ because the bottom of the phonon band is a function of ω. This effect may be responsible for ''softening'' of the phonon spectrum near its long-wave boundary. For $n_{\mathbf{k}}=1$ and $\omega_{\mathbf{k}}=\omega_{1\alpha}$ the softening is $\Delta\varepsilon_{\mathbf{k}}=-\hbar\omega_{1\alpha}/2$.

We now try to understand the nature of the frequency limitations in the spectra decribed by Eqs. (21) and (22).

4. ESTIMATES OF MASS LIMITS AND SOME STATIC CHARACTERISTICS OF STRAIN REGIONS

Note that the crystal mass M in Eqs. (19) and (20) takes certain limiting values if we assume that all frequencies below the Debye frequency are cut off, and we adopt typical values of parameters F_1, F_2, and R for point-like sources. In order to calculate the limiting masses, we determine forces F_1 and F_2 which to order of magnitude are no greater than the force f_0 needed to eject an atom from its crystal cell. The justification is that the medium is continuous or, in other words, interatomic bonds are not broken. This force can be estimated[9] to be $f_0 \approx \kappa\Omega^{2/3}$, where Ω is the atomic volume. By taking typical values for condensed matter, $\kappa \approx 10^{12}$ dyn/cm^2 and $\Omega \approx 10^{-23}$ cm^3, we obtain $f_0 \approx 10^{-3}$ dyn. This estimate is within the range of experimentally observed interatomic forces in condensed matter, which are usually between 10^{-2} dyn for ionic bonds and 10^{-6} dyn for Van der Waals bonds.[13] The value $F_1 \approx f_0$ yields the limiting mass of a crystal deformed by a monopole source:

$$M_{1\alpha} = \frac{F_{1\alpha}^2}{\hbar\omega_D^3} \approx \frac{f_0^2\Omega}{\pi^3\hbar c^3} \approx 10^{-19}\text{g}. \tag{23}$$

By taking $F_2 \approx f_0$ and the dipole arm R equal to the interatomic distance $\Omega^{1/3}$, we obtain the highest value (for $\cos^2\theta = 1$, i.e., for a plane wave propagating along $\mathbf{R}$) of the limiting mass of a crystal deformed by a dipole source:

$$M_{2\alpha} = \frac{F_{2\alpha}^2 R^2}{\hbar\omega_D c^2} \approx \frac{f_0^2\Omega}{\pi\hbar c^3} \approx 10^{-18}\text{g}. \tag{24}$$

These estimates indicate that the cutoff of phonon frequencies should be important in crystals with typical dimensions of the order of nanometers. This value is close to typical dimensions of regions deformed by point-like sources. Thus, the cutoff frequency of the phonon spectrum can be estimated by comparing the parameters of a small crystal and deformed regions. One must first estimate the change in volume of the crystal due to deformation, but this approach does not apply to a monopole source, since the crystal volume is not changed by the monopole deformation, although the local density can be changed. This change, however, cannot be described in terms of Eq. (1).

On the other hand, the phonon spectrum cutoff due to either type of strain source is similar. This is consistent with the proximity between the estimates of the limiting masses in Eqs. (23) and (24). Therefore the cutoff should be described using a common approach in the two cases.

One can try to calculate the displacement field around sources and estimate the volume of the deformed region using the displacement as a function of the distance from the source. This can be easily done by replacing the Hamiltonian in Eq. (1) by the Lagrangian

$$L = \sum_{\alpha=1}^{3} \int d\mathbf{r}\ \dot{Q}_\alpha(\mathbf{r}) P_\alpha(\mathbf{r}) - H,$$

where the generalized velocities are defined as the variational derivatives $\dot{Q}_\alpha(\mathbf{r}) = \delta H/\delta P_\alpha(\mathbf{r})$. Calculating these derivatives and the function L and deriving the Lagrange equations for the generalized coordinates and velocities, we obtain the equations for displacements of material points of the medium due to given nonuniform stress:

$$\rho\frac{\partial^2\mathbf{Q}(\mathbf{r})}{\partial t^2} - \kappa\nabla^2\mathbf{Q}(\mathbf{r}) = \mathbf{F}_1\delta(\mathbf{r}) - \mathbf{F}_2(\mathbf{R}\cdot\nabla)\,\delta(\mathbf{r}).$$

These equations can be solved by using the Fourier transform or Green's function for the linear differential equation. As a result, we have static fields of displacements due to a monople source

$$\mathbf{Q}_1(\mathbf{r}) = \mathbf{F}_1/4\pi\kappa r \tag{25}$$

and a dipole source

$$\mathbf{Q}_2(\mathbf{r}) = \mathbf{F}_2(\mathbf{R}\cdot\mathbf{r})/4\pi\kappa r^3. \tag{26}$$

The configuration of the vector field defined by Eq. (25) is determined by the vector $\mathbf{F}_1$, which defines the unidirectional field of the continuum displacement, whose amplitude drops with distance from the application point. The displacements defined by Eq. (26) form a dipole field and describe compressive, tensile, or shear deformation in the vicinity of the source, depending on the mutual alignment of the vectors $\mathbf{F}_2$ and $\mathbf{R}$.

The derivation of these well-known results has been described to demonstrate that the fields defined by Eqs. (25) and (26) are static, although the medium was originally modeled by Eq. (1) as a dynamic continuum. In the solution of classical Lagrange equations, the time derivative vanishes when the force density is time-independent.

One can say that in solving the classical problem, possible relationships between the dynamic parameters of the medium and nonuniform strain field are lost. In the simplest model introduced by Eq. (1), this shows up as the absence of the density ρ in the strain fields defined by Eqs. (25) and (26). Mathematically, this difficulty is overcome by using the equations for the eigenvalues (i.e., by reverting to the quantum Hamiltonian) and quantization of continuum vibrations. In this case, as follows from Sec. 2, the static nonuniformity is described in terms of a nonuniform steady state and is thus related to the system dynamics.

According to Eq. (26), the displacement at a distance $r \approx 10^{-7}$ from the source is $Q_2(10^{-7})\cdot 100\%/a \approx 1\%$. Ac-

cording to Eq. (25), we have $Q_1(10^{-7})\cdot 100\%/a \approx 10\%$ at the same distance. These estimates indicate that the typical dimension of regions deformed by sources with forces no greater than the interatomic force is $l \approx 10^{-7}$–10^{-6} cm. Hence their volume is $V=l^3 \approx 10^{-21}$–10^{-18} cm^3 and mass $M \approx 10^{-20}$–10^{-17} g.

One can see that the previous estimates of the masses of deformed crystals, in which all phonon modes except the Debye mode are excluded, are quite close to the masses corresponding to the volumes of deformed regions. This indicates that the boundary frequencies or wavelengths must be sought as functions of the volumes related to the boundary masses M_1 and M_2. But first, we note that the range of monopole deformation is larger than that of dipole deformation. This difference shows up in the formation energy ΔE of strain fields. Substituting the amplitudes in Eq. (18) into Eq. (14), we obtain

$$\Delta E = -\sum_{\mathbf{k}\alpha} \frac{F_{1\alpha}^2 + F_{2\alpha}^2 (\mathbf{k}\cdot\mathbf{R})^2}{2M\omega_{\mathbf{k}}^2}.$$

For example, in the Debye approximation, we obtain after changing from summation to integration ($\Sigma_{\mathbf{k}\alpha} \to V \int d\mathbf{k}/(2\pi)^3$):

$$\Delta E = \Delta E_1 + \Delta E_2 = -\frac{k_D}{(2\pi)^2 c^2 \rho}\left[F_1^2 + \frac{2}{9}\mathbf{F}_2^2 (k_D R)^2\right]. \tag{27}$$

Substituting $F_1 \approx F_2 \approx f_0$, $k_D \approx \pi/\Omega^{1/3} \approx 10^8$ cm^{-1}, and $R \approx \Omega^{1/3}$ into Eq. (27), we obtain $\Delta E_1 \approx -10$ eV and $\Delta E_2 \approx -1$ eV.

The value of ΔE_2 is in agreement with measurements of the energy required to produce a point defect, such as a vacancy or an impurity.[2,3,14,15] ΔE_1 is in qualitative agreement with the deformation energy needed for structural fluctuations of small particles with dimensions of 10^{-7} cm.[16,17]

It follows from these estimates that, given the longer range of monopole strain fields and their formation energy larger than that of dipole fields, the monopole strain should have a stronger effect on the phonon spectrum cutoff frequency.

5. DYNAMICAL VOLUMES OF NONUNIFORM STRAIN FIELDS IN CRYSTALS

Given the frequencies determined by Eqs. (19) and (20) and with the intent of expressing the final result in terms of volumes of deformed regions, we obtain

$$\omega_{1\alpha} = \omega_D \left(\frac{V_{1\alpha}}{V}\right)^{1/3}, \quad \text{where} \quad V_{1\alpha} = \frac{F_{1\alpha}^2 \Omega}{\hbar \pi^3 c^3 \rho}, \tag{28}$$

$$\omega_{2\alpha} = \omega_D \frac{V_{2\alpha}}{V}, \quad \text{where} \quad V_{2\alpha} = \frac{F_{2\alpha}^2 \Omega^{1/3} R^2 \cos^2\theta}{\hbar \pi c^3 \rho}. \tag{29}$$

With the parameters adopted above, $V_{1\alpha} \approx 10^{-20}$ cm^3 and $V_{2\alpha} \approx 10^{-19}$ cm^3, and at crystal volumes $V=V_{1\alpha}$ and $V=V_{2\alpha}$, all the phonon modes up to the Debye frequency are cut off. The respective crystal masses are $M_{1\alpha}=\rho V_{1\alpha}$ and $M_{2\alpha}=\rho V_{2\alpha}$. If $V_{1\alpha}=V_{2\alpha} \ll V$, it turns out that $\omega_{2\alpha}=\omega_{1\alpha}^3/\omega_D^2$ because of the cube root in Eq. (28), so that $\omega_{1\alpha} \gg \omega_{2\alpha}$. This means that in relatively large crystals the long-wave cutoff frequency is determined by strain fields generated by monopole sources. On the other hand, if $\omega_{1\alpha}=\omega_{2\alpha}$, we have $V_{1\alpha}=V_{2\alpha}^3/V^2$ and $V_{1\alpha} \ll V_{2\alpha}$, i.e., the same cutoff frequency requires a larger volume of deformed regions due to dipole sources than due to monopole sources. This is in agreement with the conclusion of the previous section that monopole sources of strain are more efficient.

Thus, unlike the long-wave boundary frequency ω_L of the phonon spectrum of an undistorted crystal, the boundary frequencies $\omega_{1\alpha}$ and $\omega_{2\alpha}$ of a deformed crystal depend on the ratios $V_{1\alpha}/V_{2\alpha}$, $V_{1\alpha}/V$, and $V_{2\alpha}/V$. This leads us to believe that the long-wave cutoff results from a decrease in the volume of the undistorted crystal region due to both monopole and dipole strain fields. An increase in the volumes $V_{1\alpha}$ and $V_{2\alpha}$, which have the sense of α-wave orientation of stationary volumes in steady-state extended objects, leads to a decrease in the volume of the part of the crystal where the phonon with the longest wavelength can be formed. This cutoff should shift the lower edge of the phonon spectrum to the high-frequency side without shifting its high-frequency edge. If we use the model of a ''rigid'' band, which shifts as a whole without interchanges between phonon modes, depending on the formation energy of a static defect,[3] one can say that in addition to the band shift, a steady-state nonuniform strain field should also cut off its long-wave edge. Moreover, the position of the phonon-band bottom is a function of ω near the long-wave edge of the spectrum of a crystal deformed by a monopole strain field. This ''softening'' of the spectrum due to the force applied to a crystal and its small dimension is reminiscent of the results reported in Refs. 18–21, where the disappearance of structural instability of small particles on substrates and localized regions in aged alloys as their size increased was described. It is possible that the mechanism based on monopole strain sources allows formation of the ''soft'' mode phenomenologically introduced in the theory of structural transformations.[22] On the other hand, the high damping capability of some alloys, such as copper–manganese,[21] can be interpreted in terms of the long-wave cutoff of the phonon spectrum. Empirical studies[23] indicate that the damping is caused by irregularity and high defect concentration in materials. In fact, if there is a low-frequency gap in the acoustic phonon spectrum of a dispersive (especially decaying) material with a large number of strain sources, vibrations at frequencies below some boundary frequency are forbidden. This conclusion applies to vibrations of a crystal. This is probably also true when elastic vibrations are generated by an external source, the only difference being the damping of forced vibrations, which leads to the damping effect of the material.

More experimental data are needed for a more detailed analysis of anticipated frequency cutoffs and softening of the phonon spectrum in materials distorted by nonuniform strain fields.

Let us again discuss the volumes $V_{1\alpha}$ and $V_{2\alpha}$ in Eqs. (28) and (29) and some of their features.

$V_{1\alpha}$ and $V_{2\alpha}$ are the dynamical volumes of steady-state extended entities formed by forces acting on a crystal. Being

geometrical characteristics alternative to the elastic displacement fields described by Eqs. (25) and (26), these volumes, unlike displacement fields, are expressed in terms of all dynamic characteristics of the medium included in Eq. (1).

The finite values of $V_{1\alpha}$ and $V_{2\alpha}$ indicate that the strain field generated by a point-like source is also bounded. This is a consequence of continuum quantization and can easily be generalized: an arbitrary static force acting on a crystal without breaking its interatomic bonds generates a finite region where atoms are displaced from their equilibrium positions. On the contrary, if $\hbar \to 0$, the volumes $V_{1\alpha}$ and $V_{2\alpha}$ tend to infinity, i.e., in the classical approach to an extended body the strain field should extend to infinity, which is in agreement with results in the theory of elasticity (Eqs. (25) and (26)). This conclusion holds even when Kanzaki's method and its varied atomistic modifications[2,3] are used. Thus far, although improved to some extent by microscopic calculations of interatomic potentials, these methods have belonged to the field of classical theories of deformation of solids.

The cutoff frequency determined by the field due to a dipole source is replaced by the value due to a monpole source at a crystal volume

$$V_{0\alpha}=(V_{2\alpha}^3/V_{1\alpha})^{1/2}=F_{2\alpha}^2R^3/\hbar F_{1\alpha}c^3\rho.$$

In the case of sources with forces $F_1 \approx F_2 \approx f_0$ and the dipole arm $R \approx \Omega^{1/3}$, we have $V_{0\alpha} \approx 10^{-19}$ cm and $L_{0\alpha} \approx 10^{-6}$ cm^3. This result demonstrates that the cutoff frequency as a function of the shape of the strain field can be nontrivial in crystals or small sections with dimensions in the nanometer range.

Note also that in the case of purely shear dipole deformation, the dynamical volume $V_{2\alpha}$ vanishes. A tentative explanation of this is that atoms driven by a plane wave across the deformation region have additional displacements of opposite sign when the deformation is generated by a shear dipole, whereas all displacements have the same sign in the strain field generated by a monopole or a longitudinal dipole.

In conclusion, given the small dimensions of the crystal, let us estimate the effect of the nanocrystal surface on its bulk properties, hence on the cutoff frequency and softening of the phonon spectrum calculated neglecting surface effects. Since the volume $V_{0\alpha}$ contains $N_{0\alpha}=V_{0\alpha}/\Omega$ atoms, this number equals about 10^3 for the value of Ω given above. This means that the surface effect on a sample with a volume of 10^{-19} cm^3 can be measured by the parameter $(N_{0\alpha}^{2/3}/N_{0\alpha})\cdot 100\% \approx 10\%$, which does not affect the qualitative results of our study in the sense that the discussed effects are controlled by the bulk properties of the crystal.

At the same time, we cannot rule out, on the base of our results, the possibility that surface or other effects can play an important role, for example, in restructuring small particles,[24] and the effect of the surface on the phonon-spectrum cutoff remains a current problem. As concerns the latter statement, it is noteworthy that, first, relaxation of surface atomic layers should also lead to a freqency shift in bulk phonon modes. It seems probable that the cutoff will take place at the highest frequency, owing to the superposition of atomic diplacements due to point-like sources and relaxation of surface layers. Second, frequencies of surface modes can also shift owing to various stress sources. It is clear, however, that dedicated investigation needed to address these issues would be labor-consuming.

The author is grateful to O. A. Kazakov and M. A. Shtremel for heated discussions (exhaustive information on damping alloys provided by M. A. Shtremel in due time is also acknowledged), and to I. Ya. Polishchuk for his friendly and constructive criticism of the manuscipt.

[1)]E-mail: Meshcheryakov@trf.misa.ac.ru

[1] N. B. Brandt and S. M. Chudinov, *Electronic Structure of Metals* [in Russian], Moscow State University, Moscow (1973).
[2] G. Leibfried and N. Brauer, *Point Defects in Metals*, Springer-Verlag, Heidelberg (1978).
[3] A. M. Stoneham, *Theory of Defects in Solids*, Clarendon Press, Oxford (U.K.) (1975).
[4] V. M. Svistunov, M. A. Belogolovskiĭ, and A. N. D'yachenko, Usp. Fiz. Nauk **154**, 153 (1988) [Sov. Phys. Usp. **31**, 86 (1988)].
[5] O. V. Klyavin, *Physics of Crystal Plasticity at Helium Temperatures* [in Russian], Nauka, Moscow (1987).
[6] V. S. Boĭko, R. I. Garber, and A. M. Kosevich, *Reversable Plasticity of Materials* [in Russian], Nauka, Moscow (1991).
[7] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, New York (1986).
[8] H. Haken, *Quantenfeldtheorie des Festkörpers* [in German], B. G. Teubner, Stuttgart (1973).
[9] V. V. Meshcheryakov, Fiz. Tverd. Tela (St. Petersburg) **37**, 43 (1995) [Phys. Solid State **37**, 20 (1995)].
[10] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to Theory of Disordered Systems* [in Russian], Nauka, Moscow (1982).
[11] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **17**, 1076 (1947).
[12] I. Ya. Polishchuk, A. P. Zhernov, and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **94**, 259 (1988).
[13] G. Binnig, C. F. Quate, and Ch. Gerber, Phys. Rev. Lett. **56**, 930 (1986).
[14] A. M. Bratkovskiĭ, I. E. Zeĭn, Fiz. Tverd. Tela (Leningrad) **26**, 2561 (1984) [Sov. Phys. Solid State **26**, 1553 (1984)].
[15] G. Jacucei and R. Taylor, J. Phys. F **9**, 1487 (1979).
[16] J. Dundurs, L. D. Marks, and P. M. Ajayan, Phil. Mag. A **57**, 605 (1988).
[17] S. M. Komarov, JETP Lett. **58**, 553 (1993).
[18] J.-O. Bovin, R. Wallenberg, and D. J. Smith, Nature **317**, 47 (1985).
[19] S. Iijima and T. Ichihashi, Phys. Rev. Lett. **56**, 616 (1986).
[20] L. D. Marks, P. M. Ajayan, and J. Dundurs, Ultramicroscopy **20**, 78 (1986).
[21] D. M. Farkas, T. Yamashita, and J. Perkins, Acta Met. Mat. **38**, 1883 (1990).
[22] A. Brause and R. Cowly, *Structural Phase Transitions*, Taylor and Francis, London (1981).
[23] Yu. K. Favstov, Yu. N. Shul'ga, and A. G. Rakhshtadt, *Material Properties of Highly Damping Alloys* [in Russian], Metallurgiya, Moscow (1980).
[24] É. L. Nagaev, Usp. Fiz. Nauk **162**, 49 (1992) [Sov. Phys. Usp. **35**, 747 (1992)].

Translation provided by the Russian Editorial office.

Atomic structures of gallium-rich GaAs(001)-4×2 and GaAs(001)-4×6 surfaces

R. Z. Bakhtizin[1)]

Bashkir State University, 450074 Ufa, Russia

Qikun Xue[2)] and T. Sakurai[2)]

Institute for Materials Research, Tohoku University, Sendai 980-77, Japan

T. Hashizume[3)]

Hitachi Advanced Laboratory, Hitachi Ltd., Saitama 350-03, Japan

(Submitted 13 November 1998)
Zh. Éksp. Teor. Fiz. **111**, 1858–1868 (May 1997)

Scanning tunneling microscopy is applied for the first time to an atomic-resolution investigation of the 4×2 and 4×6 phases on a gallium-rich GaAs(001) surface obtained by molecular-beam epitaxy and migration-enhanced epitaxy. A unified structural model is proposed with consideration of the results of experiments and first-principles calculations of the total energy. In this model the 4×2 phase consists of two Ga dimers in the top layer and a Ga dimer in the third layer, and the 4×6 phase is matched to periodically arranged Ga clusters at the corners of a 4×6 unit cell on top of the 4×2 phase.
[S1063-7761(97)02305-6]

1. INTRODUCTION

The polar GaAs(001) surface is widely employed in the fabrication of semiconductor devices and often serves as a substrate in molecular-beam epitaxy.[1–3] Its atomic structure is of considerable interest, since most devices based on GaAs are grown on this surface by just such a method. The GaAs(001) surface itself exhibits a large number of phases and structural phase transitions between them that are intimately related to the stoichiometric composition, which can be controlled by varying the temperature of the sample and the ratio between the atomic concentrations of Ga and As in the beam, i.e., the surface treatment conditions.[3] The mechanism of coherent growth during molecular-beam epitaxy on a GaAs(001) surface was first proposed in Ref. 4.

While the structure of the As-stabilized GaAs(001)-2×4 surface grown by molecular-beam epitaxy has been studied quite extensively,[5–19] there are only a few known successful attempts at investigating the phases on the gallium-rich GaAs(001) surface, such as the GaAs(001)-4×2 and GaAs(001)-4×6 phases, because of the formidable difficulties in preparing such surfaces and studying them in detail. For example, it is essentially impossible to obtain a GaAs surface with a freshly sputtered Ga layer under the conditions of standard molecular-beam epitaxy at the high partial pressures of As vapor typical of this method, $10^{-6}-10^{-4}$ Torr, although the growth of such phases under nearly equilibrium conditions is of considerable interest.[15–19] We were able to overcome the experimental difficulties indicated by employing migration-enhanced epitaxy,[20] which has made it possible to create a universal approach to the preparation of such surfaces and to perform the first successful *in situ* investigations (in the same vacuum system) of gallium-rich GaAs(001)-4×2 and GaAs(001)-4×6 phases by scanning tunneling microscopy (STM). Migration-enhanced epitaxy significantly increases (by tenfold) the migration distance of the Ga atoms in comparison with the case of traditional molecular-beam epitaxy, making it possible to obtain a smooth gallium-rich surface and to observe reversible phase transitions between different phases ($c(4\times4)$, 2×4, 2×6, 4×2, and 4×6) on it simply by regulating the $[As_4]/[Ga]$ concentration ratio in the beam.[21–24]

In the present paper we report the experimental discovery of two types of 4×6 phases, viz., the so-called 4×6 pseudophase, i.e., a structure with a smaller gallium content on the surface in comparison with the 4×2 phase, and a true 4×6 phase, i.e., a phase with a larger gallium content. Both phases consist of a subcell with two Ga dimers and two dimeric Ga vacancies at the top of the surface layer, as well as another Ga dimer in the third layer, i.e., they appear as the mirror image of the unit cell of the As-rich 2×4 phase, whose structural model was proposed in Ref. 24. We interpret the significant differences between the STM images of the 4×2 and 4×6 phases on the basis of models of charge transfer between the second-layer As atoms, the top-layer Ga dimers, and the Ga adatoms. These models were a result of the performance of theoretical calculations and computer simulation of the images.[25] A new structural model of the 4×6 and 4×2 phases is proposed.

2. METHOD

All the experiments were performed in the ultrahigh-vacuum (with a residual pressure in the measuring chamber equal to 5×10^{-11} Torr) scanning tunneling microscope of Tohoku University,[26] which is connected directly to a specially developed molecular-beam epitaxy chamber with a built-in reflection high-energy electron diffractometer for monitoring and controlling the substrate surface growth process.[27] The tunneling microscope was also equipped with an additional chamber with analyzers for low-energy electron diffraction and Auger electron spectroscopy. The quality of the scanning tips was monitored on the atomic level using a miniature field-ion microscope, which is an integral part of

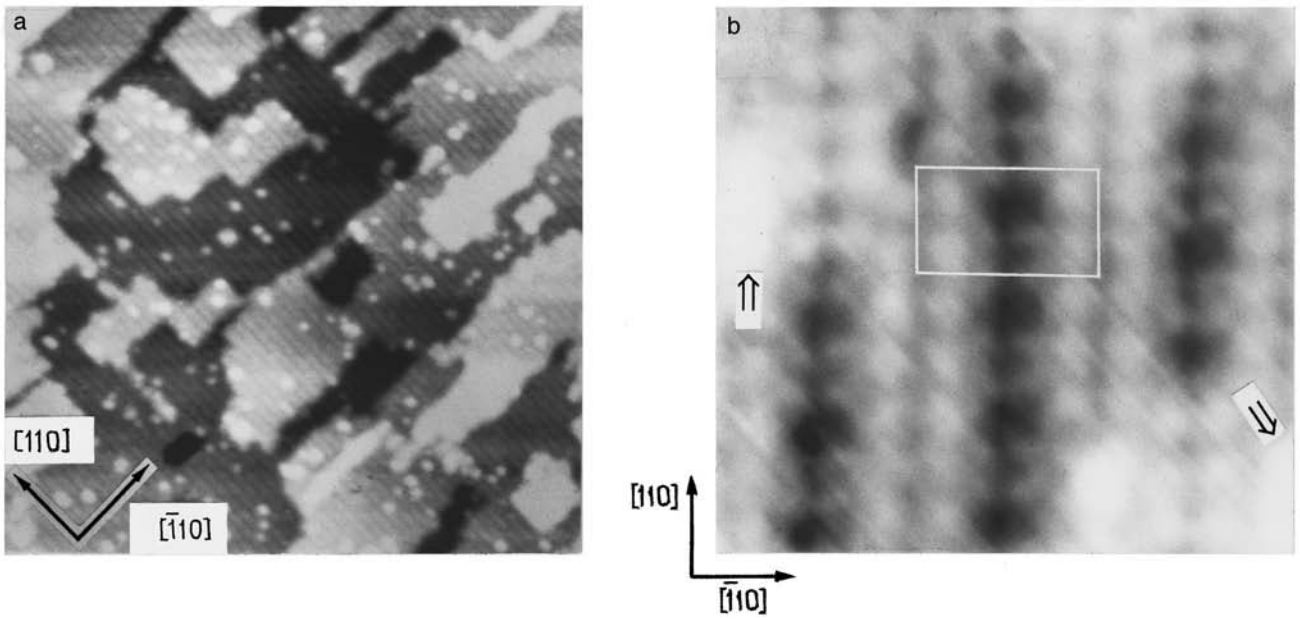

FIG. 1. a) STM image of a GaAs(001)-4×2 surface of large area (900×900 Å^2). b) Magnified STM image of a portion of the same surface showing its detailed structure (the rectangle demarcates the unit cell). The rate of deposition of the layer was 0.25 μm/h, and the growth temperature was 500 °C. The migration-enhanced epitaxy cycle was carried out with $[As_4]/[Ga]=8$ and $V_s=-1.8$ V.

the microscope.[26,28] The high cooling rate of the apparatus permitted *in situ* STM investigations of different surface phases during their growth.[22–24]

Type-*n* GaAs(001) substrates, which were doped with silicon to $1\times10^{18}-2\times10^{18}$ cm^{-3} and measured 4×10 mm, were degreased and etched using standard cleaning procedures before being placed in the molecular-beam epitaxy chamber. After the oxide layer on the substrate was removed in a beam of As_4, a buffer layer of *n*-type GaAs with a thickness of 0.5–1.0 μm (which was also doped with silicon to 2×10^{18} cm^{-3}) was grown using the traditional method of molecular-beam epitaxy at 600 °C under optimal conditions. Before the growth process was completed, all the shields were closed, the heating of the substrate was stopped, and the As and Ga sources (Knudsen cells) were brought to the temperatures needed to obtain the $[As_4]/[Ga]$ concentration ratio in the beam required for low-temperature growth under the conditions of migration-enhanced epitaxy.[21] Afterwards, the sample was annealed at 500 °C in a beam of As_4 until a well ordered 2×4 reflection high-energy electron diffraction (RHEED) pattern appeared, and 12 growth cycles were carried out (usually) under the conditions of migration-enhanced epitaxy. After the concluding cycle with the deposition of gallium was performed, the sample was rapidly (within 2 s) cooled to room temperature and moved into the chamber of the tunneling microscope. All the images were obtained at room temperature at a constant tunneling current $I_t=4\times10^{-11}$ A with a negative bias V_s on the sample. After the STM investigations were performed, the structure of the sample surface was repeatedly monitored by RHEED.

3. RESULTS AND DISCUSSION

Figure 1a presents a typical filled-states STM image of a 4×2 surface prepared using migration-enhanced epitaxy with a concentration ratio $[As_4]/[Ga]=8$ in the beam and exposures to As and Ga alternating at 2 s intervals. As long as the growth temperature was relatively low, only three terraces with an almost ideal degree of ordering, which were separated from one another by steps with a height *H* equal to the thickness of a double GaAs monolayer ($H=2.8$ Å), formed within an area of 900×900 Å^2. An attentive examination of the image readily reveals bright lines arranged periodically in the [110] direction and regular spaces between them with a width of ~16 Å in the [110] direction, which correspond to a fourfold (4×) increase in the periodicity of the surface, as was observed in other studies.[16,19] This can be attributed to the formation of well ordered domains of a 4×2 structure. Thus, the structure of the 4×2 surface does not appear excessively complicated in comparison to the structure of the 2×4 phase.[29] The islands on this essentially defect-free surface are most likely isotropic and form smooth steps of both types (see A and B in Fig. 2), in contrast to the 2×4 surface, where the islands are highly anisotropic.[22–24,29]

Figure 1b presents the image of the same 4×2 surface, but with a far higher resolution, on which the unit cell of the 4×2 structure is demarcated. It is clearly seen that each bright line observed in Fig. 1a actually consists of a pair of rows separated by a 5.1 Å gap, and in the [110] direction the surface has a 2× periodicity. We note that this image is similar as a whole to the one obtained by Skala *et al.*. (Fig. 1 in Ref. 19).

Although the STM images obtained are similar to the ones previously observed, in this paper we propose another model, totally different from Skala's As model,[19] which has been used hitherto to describe the 4×2 phase. On the basis of a careful analysis of the images, we arrived at the fundamentally different conclusion that the bright spots are images of the second-layer As atoms, rather than the first-layer Ga

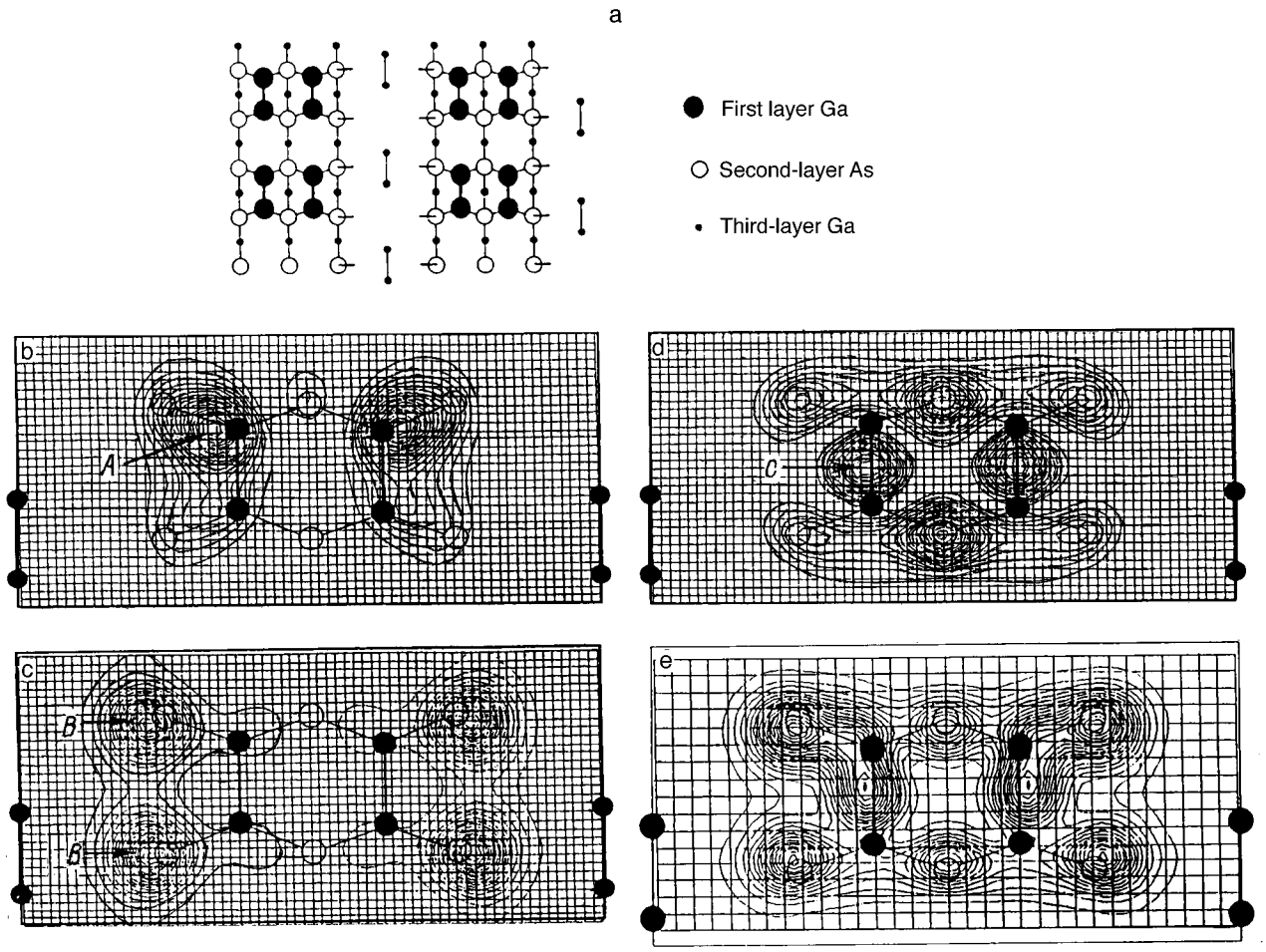

FIG. 2.

atoms, and that the darker spots are images of the first-layer Ga dimers. The most acceptable atomic model for this phase is the two-layer Ga model, which was first proposed by Biegelsen[16] (Fig. 2a). The basic arguments supporting this model for interpreting the images obtained are as follows.

1) The surface-preparation procedure that we used favors a gallium-rich surface, since the final exposure was to the Ga beam.

2) The difference that we observed between the heights of the bright and dark rows is only 0.6 Å. In this case As atoms with completely filled dangling bonds will appear in the filled-states STM images as bright spots, while the Ga atoms with unfilled dangling bonds will be relatively dark. Thus, if the brighter spots correspond to the first-layer As atoms, much greater contrast (>1.4 Å$=H/2$) is expected between the first-layer As atoms and the second-layer Ga atoms, which have considerably less charge, and only the first-layer As dimers should actually have bright images, as was observed in the case of the arsenic-rich 2×4 surface in Ref. 24.

3) According to the As model, the first-layer As atoms form dimers. For this reason, a structure in the form of clusters or a structure in the form of a corona was expected for the As dimers on the filled-states images.[8,16–18,24,25] However, the STM images presented above (as well as the images obtained in Ref. 19) clearly exhibit two isolated peaks separated by a 5.1 ± 0.3 Å gap in the [110] direction (Fig. 1b). Therefore, the As model should be categorically ruled out.

However, there is also a problem in interpreting the STM images with the Ga model. The STM images are clearly inconsistent with the Ga model, because the spots comprising the dark rows are located in the gaps between the spots forming the brighter rows along the [110] direction, while, according to the Ga model, the Ga and As atoms should be aligned in the rows (Fig. 2a).

To account for this inconsistency, we performed first-principles calculations of the total energy. The details of the approach used were described in Refs. 25 and 30, and the results of the calculations themselves are presented in Figs. 2b-e. A comparison of the theoretically calculated surface energy bands and the STM images obtained in the filled-states regime ($V_s=-1.8$ V) revealed that the main contribution to the STM image shown in Fig. 1b is made by states located between the 71th and 76th bands (Fig. 2b-d). Band 75 is the band of the HOMO (the highest occupied molecular orbital), and it contributes to the charge localized in the second-layer As atoms (peak *B* in Fig. 2c). The distribution of the charge density from the 74th, 73rd, and 72nd bands, generally overlaps the distribution of the charge density from the HOMO band. The contribution of the Ga dimer located at

the top of the layer becomes appreciable in the 71th band, and a maximum of the charge-density distribution appears at the midpoint of the Ga dimer (peak *C* in Fig. 2d). Since there is considerable charge transfer from the surface defects and the adsorbate atoms (which are indicated by arrows in Fig. 1b), the 76th band (i.e., the band of the LUMO – the lowest unoccupied molecular orbital) turns out to be partially filled at the dangling-bond level (peak *A* in Fig. 2b) and makes a contribution to the image. It is strongly bent due to the presence of the third-layer Ga dimers. The size of the contribution of the LUMO band to the filled-states STM image depends on the amount of charge transferred to the Ga dimers. Since the calculated density of states presented in Figs. 2b-e actually reflects the charge distribution at a distance of only 0.9 Å from the surface, while the STM image reproduces the distribution of the electron density at a distance of approximately 10 Å from the surface of the sample, maxima *A* and *B* can be shifted by a certain distance from the true atomic positions. When this circumstance is taken into account, the STM image in Fig. 1b agrees quite well with the two-layer Ga model (Fig. 2a).

Additional weighty evidence supporting the conclusion drawn above that the Ga(001)-4×2 phase is described well by the Ga model[16] was provided by the following experiment. When a beam with a higher Ga content (with $[As_4]/[Ga]=6$ instead of the previous value of 8) was used in the migration-enhanced epitaxy cycle or when the 2×6 phase prepared by the migration-enhanced epitaxy technology[20] was annealed in a vacuum for $\geqslant$15 min, both the STM images and the RHEED patterns exhibited a 4×6 phase with an even greater Ga content. A typical STM image of this ''true'' 4×6 phase can be seen in lower right-hand corner of Fig. 3a, whose central portion displays a 2×6 phase.

It is interesting that the 4×6 phase is the only phase of its kind and is characterized by rows of large oval spots, which are regularly arranged at each corner of the unit cell, as is clearly seen in the high-quality high-resolution STM image (Fig. 3b). The features of oval shape are $\sim$0.1 Å (i.e., $\sim H/2$) above the Ga dimers and 0.9 Å below the As-(2×6) rows. Comparing the images in Figs. 1b and 3b, we can assign the rows of bright pairs of points oriented in the [110] direction in Fig. 3b to the first-layer Ga dimers (they are marked by two black arrows), rather than the second-layer As atoms, in contrast to the case of the 4×2 phase. The large bright features of oval shape occupy the midpoints of the As rows (in Fig. 3b these rows are marked by three black arrows), frequently overlapping them (an instance of this is indicated by a white arrow). We note here that each individual Ga dimer is clearly resolved in Fig. 3b and that the appearance of such high contrast in their images can be attributed to charge transfer from the oval spots to the Ga dimers. An example of a similar situation can be seen in Fig. 1b. The profile presented in Fig. 3b of a scan along the white AA' line clearly demonstrates that the arrangement of the Ga atoms is consistent with the underlying GaAs lattice. In the cases in which the original 4×2 surface has one oval spot on the As rows (which is identified by a white arrow), the neighboring Ga dimers become brighter, and the As rows become darker. As the number of such oval features increases, the contrast in the image of the As and Ga atoms on the 4×2 surface varies and ultimately becomes completely reversed, giving the same picture as the 4×6 surface as a result.

To determine what the large clusters localized at each corner of the 4×6 unit cell represent, we performed another series of experiments to observe the evolution of the 4×6 phase as the annealing temperature is gradually increased from 500 °C and a larger amount of Ga is simultaneously deposited on the 4×6 surface. As the temperature was increased, the bright features mentioned above became increasingly more distinct on the flat terraces, and even individual drops of Ga could be observed. Figure 4a presents the STM image obtained after the annealing of the 4×6 phase for 20 min at 600 °C. The RHEED patterns from this surface displayed fourfold symmetry in the [110] direction, but did not display low symmetry in the [110] direction. When Ga was deposited alone on the As-(2×6) surface at 500 °C in an amount equal to three or more monolayers, the same STM image as in Fig. 4a was generally observed. When the substrate temperature was increased to 600 °C, it was logical to expect the zigzag rows of As on the 2×6 surface to vanish completely as a result of the desorption of a considerable amount of As. During the further deposition of excess Ga, it was no longer possible to accommodate the entire quantity of Ga on the surface with the formation of stable Ga–As bonds. This provided a convincing argument for the resultant binding of the excess Ga atoms to one another with the formation of clusters over the As layer. They ultimately join to form large drops of Ga (some of them are indicated by arrows in Fig. 4a).

On the basis of the foregoing results of a systematic STM investigation of the detailed structure of the (001) surface of a GaAs crystal grown by molecular-beam epitaxy, we developed the first atomic model of the 4×6 phase in the form presented in Fig. 4b. Taking into account all the difficulties involved in the preparation of the new 4×6 phase under the conditions of strong enrichment with gallium, we ascribe the characteristic oval features to clusters of Ga adatoms. Accordingly, the 4×6 phase is a gallium-rich form of the 4×2 phase with additional Ga clusters, which are arranged in an ordered fashion on the GaAs(001) surface. At the same time, we cannot determine the exact size of the Ga cluster, but we assume that it must consist of six or eight atoms and must have a structure similar to the structure of the arsenic-rich $c(4\times4)$ subcell.

It was presumed in Refs. 10–13 that the 4×6 phase is simply a result of the superposition of 4×1 and 1×6 domains or that the domain of the 4×1 structure is essentially a 4×2 phase that has been disordered in the [110] direction. Similarly, the idea that a 4×12 superstructure exists was advanced in some earlier investigations[10,11] to account for the streaked 1/4 or 1/6nY* lines on the 1×6 RHEED patterns. Our STM investigations (see Fig. 3) clearly demonstrated that these structures are mixed 2×6 and 4×2 (and/or 4×6) phases, which are observed under the conditions of a smaller Ga content than the 4×2 phase and that there is a clearcut ''true'' 4×6 phase, which is obtained under the

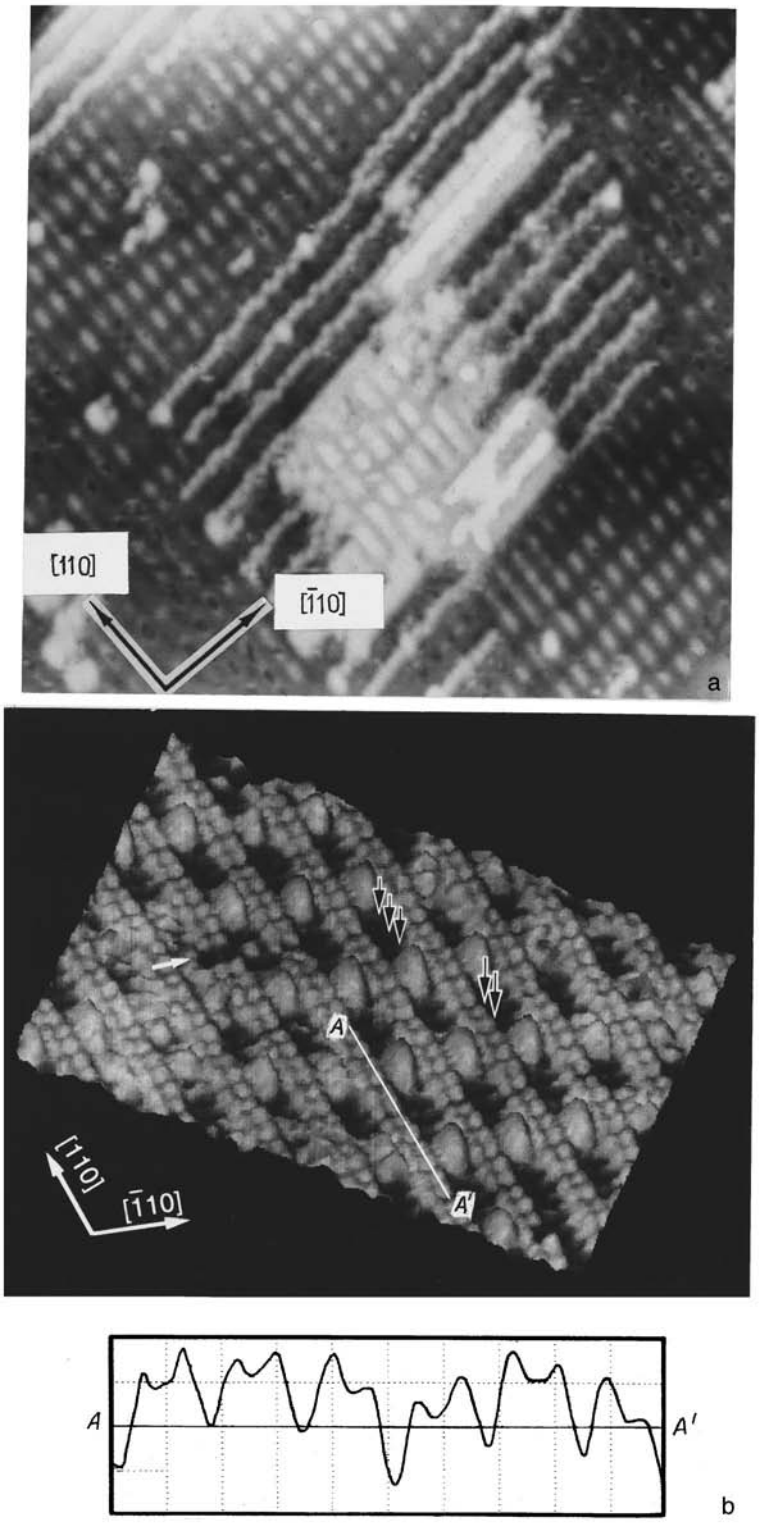

FIG. 3. Filled-states STM images ($V_s=-1.8$ V): a) 4×6 phases with residues of a locally arranged 2×6 phase (380×380 Å^2); b) high-resolution image of a surface with a 4×6 structure, showing detailed features of this phase (140×140 Å^2).

conditions of migration-enhanced epitaxy with a larger Ga content than the 4×2 phase.

Thus, in accordance with our proposed structural model, the ''true'' 4×6 phase, which has 4×6 translational symmetry, is described by the deposition of a well ordered periodic array of Ga clusters at each corner of the 4×6 unit cell at the top of the Ga double layer of the 4×2 surface. The 4×6 pseudophase with a smaller Ga content is, in fact, a mixture of the arsenic rich 2×6 phase and the gallium-rich $4\times2/c(8\times2)$ reconstruction or the 4×6 phase, so that the Ga content can vary over a broad range and depends strongly on the procedure used to prepare the surface.

The surface stoichiometry of different gallium-rich GaAs(001) surfaces has been debated for a long time.[9–12] The idea of superimposing different superstructures to form a 4×6 phase suggests that the surface stoichiometry should be the same for the 1×6, 4×6, and 4×2 phases.[12] This, of course, is untrue, and on the basis of the results of the STM

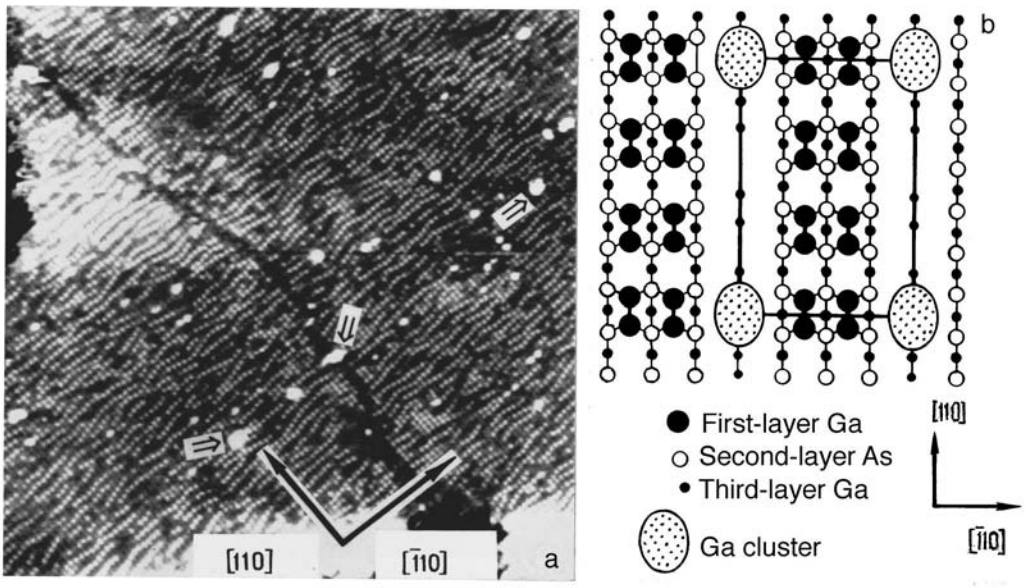

FIG. 4. a) STM image of a surface of large area (1300×1300 Å^2) with an unordered 4×6 phase formed after the annealing of a well ordered 4×6 phase at 600 °C for 20 min in an ultrahigh vacuum, showing many Ga clusters; $V_s=-1.8$ V. b) Structural model of the 4×6 phase.

investigation presented above, we concluded that the surface concentration of Ga increases along the sequence consisting of the 2×4, 4×2, and 4×6 phases.

4. CONCLUSIONS

The 4×2 and 4×6 phases on the gallium-rich GaAs(001) surface have been investigated at the atomic level by scanning tunneling microscopy.

The only satisfactory model of the 4×2 phase is Biegelsen's Ga model.

As for the 4×6 phase, we have established the atomic structural model of a separate 4×6 phase for the first time, which is characterized by Ga clusters regularly arranged on the 4×2 surface.

We thank Prof. T. Ohno (National Research Institute of Metals, Tsukuba, Japan) and Prof. J.-M. Zhou (Institute of Physics, Academy of Sciences of the Peoples' Republic of China, Peking, China) for useful discussions and valuable comments.

This work was performed with partial support from the ''Surface Atomic Structures'' Russian State Program of the State Committee for Science and Technology (Project No. 96-2.27).

[1)]E-mail: raouf@online.ru
[2)]E-mail: xue@apfim.imr.tohoku.ac.jp; sakurai@apfim.imr.tohoku.ac.jp
[3)]E-mail: tomi@harl.hitachi.co.jp

[1] A. Madhukar and S. V. Ghaisas, CRC Crit. Rev. Solid State Mater. Sci. **14**, 1 (1988).
[2] M. A. Herman and H. Sitter, *Molecular Beam Epitaxy: Fundamental and Current Status*, Springer-Verlag, New York–Berlin–Heidelberg (1996).
[3] *Molecular Beam Epitaxy*, A. Cho (ed.), AIP Press, New York (1994).
[4] H. H. Farrell, J. P. Harbison, and L. D. Peterson, J. Vac. Sci. Technol. B **5**, 1482 (1987); D. J. Frankel, C. Yu, J. P. Harbison, and H. H. Farrell, J. Vac. Sci. Technol. B **5**, 1113 (1987).
[5] J. H. Neave and B. A. Joyce, J. Cryst. Growth **44**, 387 (1978).
[6] A. Y. Cho, J. Appl. Phys. **42**, 2074 (1971).
[7] P. K. Larsen and D. J. Chadi, Phys. Rev. B **37**, 8282 (1988).
[8] M. D. Pashley, K. W. Haberern, W. Friday *et al.*, Phys. Rev. Lett. **60**, 2176 (1988); M. D. Pashley and K. W. Haberern, Phys. Rev. Lett. **67**, 2697 (1991).
[9] A. J. van Rommel, J. E. Chrombeen, and T. G. van Dirschot, Surf. Sci. **72**, 95 (1978).
[10] W. Ranke and K. Jacobi, Prog. Surf. Sci. **10**, 1 (1981).
[11] P. Drathen, W. Ranke, and K. Jacobi, Surf. Sci. **77**, L162 (1978).
[12] J. R. Creighton, Surf. Sci. **234**, 287 (1990).
[13] H. Qi, P. E. Gee, and R. F. Hicks, Phys. Rev. Lett. **72**, 250 (1994).
[14] C. Deparis and J. Massies, J. Cryst. Growth, **108**, 157 (1991).
[15] I. Kamiya, D. E. Aspnes, H. Tanaka *et al.*, Phys. Rev. Lett. **68**, 627 (1992); I. Kamiya, D. E. Aspnes, L. T. Florez, and J. P. Harbison, Phys. Rev. B **46**, 15 894 (1992).
[16] D. K. Biegelsen, R. D. Bringans, J. E. Northrup, and L.-E. Swartz, Phys. Rev. B **41**, 5701 (1990).
[17] J. E. Northrup and S. Froyen, Phys. Rev. Lett. **71**, 2276 (1993); Phys. Rev. B **50**, 2015 (1994).
[18] J. Falta, R. M. Tromp, M. Copel *et al.*, Phys. Rev. Lett. **69**, 3068 (1992).
[19] S. L. Skala, J. S. Hubacek, J. R. Tucker *et al.*, Phys. Rev. B **48**, 9138 (1993).
[20] Y. Horikoshi, M. Kawashima, and H. Yamaguchi, Jpn. J. Appl. Phys. **25**, L868 (1986).
[21] Q. K. Xue, T. Hashizume, J. M. Zhou *et al.*, Appl. Surf. Sci. **87/88**, 364 (1995).
[22] J. M. Zhou, Q. K. Xue, H. Chaya *et al.*, Appl. Phys. Lett. **64**, 583 (1994).
[23] Q. K. Xue, J. M. Zhou, T. Hashimme, and T. Sakurai, J. Appl. Phys. **75**, 5021 (1994).
[24] T. Hashizume, Q. K. Xue, J. M. Zhou *et al.*, Phys. Rev. Lett. **73**, 2208 (1994); Phys. Rev. B **51**, 4200 (1995).
[25] T. Ohno, Phys. Rev. Lett. **70**, 631 (1993).
[26] T. Sakurai, T. Hashimme, I. Kamiya *et al.*, Prog. Surf. Sci. **33**, 3 (1990); T. Hashizume, I. Sumita, Y. Murata *et al.*, J. Vac. Sci. Technol. A **9**, 742 (1991).
[27] S. Miwa, Y. Haga, E. Morita *et al.*, Jpn. J. Appl. Phys. **32**, 1508 (1993).
[28] R. Z. Bakhtizin, C. Park, T. Hashizume, and T. Sakurai, Zh. Éksp. Teor. Fiz. **108**, 977 (1995) [J. Exp. Theor. Phys. **81**, 538 (1995)].
[29] E. J. Heller, Z. Y. Zhang, and M. G. Lagally, Phys. Rev. Lett. **71**, 743 (1993); E. J. Heller and M. G. Lagally, Appl. Phys. Lett. **60**, 2675 (1992).
[30] T. Ohno, Phys. Rev. Lett. **73**, 460 (1994).

Translated by P. Shelnitz

Helical instability of a straight Abrikosov vortex in an anisotropic superconductor

I. M. Dubrovskiĭ

Institute of Metal Physics, Ukrainian National Academy of Sciences, 252142 Kiev, Ukraine
(Submitted 25 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1869–1878 (May 1997)

Because of attraction of the parallel currents forming an Abrikosov vortex, the vortex energy per unit length decreases, under bending of the vortex, by a quantity proportional to the square of the curvature. Solving the London equation in an approximation allowing for this effect makes it possible to calculate the energy of an Abrikosov vortex in the form of a helix whose length and pitch are much larger than the correlation length, whose curvature is small compared to the reciprocal London length, and whose slope in relation to an axis coinciding with the direction in which the vortex energy is the highest is also small. When the anisotropy is large, which is characteristic of high-T_c superconductors, the energy of such an Abrikosov vortex is lower than that of a straight Abrikosov vortex. Certain consequences of the fact that the Abrikosov vortices in a high-T_c superconductor are helical are discussed. Among these is a phase transition that breaks the symmetry between Abrikosov vortices shaped like right- and left-hand helixes in relation to the magnetic field. © *1997 American Institute of Physics.* [S1063-7761(97)02405-0]

1. INTRODUCTION

Several researchers[1–3] have examined the problem of whether the helical configuration of an Abrikosov vortex in a uniaxial anisotropic London superconductor is an equilibrium configuration when, on the average, the vortex is directed along the anisotropy axis, in which case its specific energy is the highest. The single factor taken into account was that the helix is deflected from its average direction by a constant angle θ. Then, as demonstrated by Brandt[3], the energy of a helical Abrikosov vortex per unit length along the average direction is higher than that of a straight Abrikosov vortex. Indeed, in the logarithmic approximation,[4] the specific energy of an Abrikosov vortex deflected by a small angle from the axis is

$$\varepsilon=\left(\frac{\Phi_0}{4\pi\lambda}\right)^2\left[1-\frac{\theta^2}{2}(1-\gamma^2)\right]\ln\,\kappa$$
$$=\varepsilon_0\left[1-\frac{\theta^2}{2}(1-\gamma^2)\right]\ln\,\kappa. \tag{1}$$

Here Φ_0 is the quantum of magnetic flux, λ is the London length along the anisotropy axis, $\gamma\leqslant 1$ is the anisotropy constant ($\gamma=1$ in the isotropic case), and $\kappa\gg 1$ is the Ginzburg–Landau constant. The relative elongation of the helix in relation to the straight line is $1+\theta^2/2$. Then the variation of the energy per unit length along the average direction due to the transition from a straight Abrikosov vortex to a helical Abrikosov vortex is

$$\left(1+\frac{\theta^2}{2}\right)\varepsilon-\varepsilon_0\,\ln\,\kappa=\frac{\theta^2\gamma^2}{2}\varepsilon_0\,\ln\,\kappa. \tag{2}$$

Here, however, $\gamma^2\ln\,\kappa\ll 1$ may be much less than unity (e.g., for YBCO we have $\gamma=1/8$ and $\kappa=50$), so that the logarithmic approximation may be inadequate, i.e., the corrections to this approximation, while being small, exhibit a different dependence on the position and shape of the Abrikosov vortex and the parameters of the superconductor and can affect the result. A general model of Abrikosov vortices was employed in Ref. 5 to calculate the corrections to the logarithmic approximation for the energy of a straight oblique vortex in an anisotropic superconductor. The same approach is adopted here to allow for corrections to the energy of an Abrikosov vortex that are associated with bending of the vortex axis.

The dependence of the specific energy of an Abrikosov vortex on the curvature of the vortex axis has not been studied. Usually it is assumed that no such dependence exists, and the increase in the energy of an Abrikosov vortex caused by the bending of the vortex axis in which the average direction of the axis remains unchanged is related only to elongation of the axis. De Gennes[6] mentions the fact that an Abrikosov vortex has a rigid core, i.e., the specific energy increases under bending, but does not justify the statement and does not use it. Qualitative considerations point to an opposite effect: bending drives the specific energy of a vortex down. Under axis bending, the parallel circular currents forming an Abrikosov vortex condense at the curvature center and become rarefied on the opposite side. Since the force of attraction of parallel currents is inversely proportional to the distance between the currents, the energy must decrease as a result of axis bending. These ideas can be corroborated by calculating the energy of a circular Abrikosov vortex by expanding in powers of the curvature (see Appendix). When the radius is large, the energy of the interaction of two diametrically opposite sections of the circular vortex is exponentially low, so that the dependence on the curvature can be explained only by the presence of bending. The factor of the second power of the curvature (the factor of the first power is zero), i.e., the bending modulus of the vortex, proves to be negative.

Not only is the slope of the helix constant in relation to the average direction, but so is the curvature. The negative correction to the Abrikosov-vortex energy, which is related to the curvature, exhibits only a slight dependence on anisot-

ropy, with the result that for small values of γ the helical configuration of an Abrikosov vortex may prove to be more energeticallyfavored than the straight configuration.

The helical configuration of Abrikosov vortices was also examined in Refs. 7–9 in situations when an external magnetic field and a transport current act on a superconductor simultaneously. However, the estimates of the energy of a helical Abrikosov vortex done in these papers are not sufficiently accurate to solve the problem. To calculate the energy more accurately, we introduce the model of a helical vortex in a London superconductor. If the axis of an Abrikosov vortex is a helix of small curvature and slope, this can be done by introducing a special system of curvilinear coordinates. In this system we first solve the inhomogeneous London equation, which determines the magnetic field of the vortex, a solution in the form of an expansion in powers of the curvature, and then calculate the energy of the Abrikosov vortex.

2. HELICAL COORDINATES

Helical coordinates can be considered the result of deformation of cylindrical coordinates whose axis becomes a helix. Suppose that the helix is specified in Cartesian coordinates by the following equations:

$$x_\nu = a\cos\psi, \quad y_\nu = a\sin\psi, \quad z_\nu = b\psi, \quad \psi = \frac{l}{\sqrt{a^2+b^2}}. \tag{3}$$

Here l is the length of the arc of the helix measured from the point of its intersection with the plane $z=0$, and a and b are the parameters of the helix, with a the radius of the cylinder formed by the helix, and $2\pi b$ the pitch of the helix.

Let us introduce dimensionless cylindrical coordinates ρ, φ, and ζ:

$$\rho = \frac{\sqrt{x^2+y^2}}{\lambda}, \quad \zeta = \frac{z}{\lambda}; \tag{4}$$

we also introduce the dimensionless quantities

$$\alpha = \frac{\lambda a}{b^2}, \quad \beta = \frac{b}{\lambda}. \tag{5}$$

The point $\mathbf{r}_\nu$ at which the plane passing through $\mathbf{r}$ perpendicular to the helix intersects the helix is determined by

$$(\mathbf{r}-\mathbf{r}_\nu)\cdot\frac{d\mathbf{r}_\nu}{d\psi} = 0. \tag{6}$$

In terms of the variables ρ, φ, ζ, and ψ, Eq. (6) assumes the form

$$\alpha\beta\rho\,\sin(\varphi-\psi)+\zeta-\beta\psi = 0. \tag{7}$$

Let us see when the solution of this equation in ψ for fixed $0\leqslant\varphi<2\pi$ and $0\leqslant\zeta<2\pi\beta$ is unique. All the chords of the sinusoid

$$y(\psi) = \alpha\beta\rho\,\sin(\varphi-\psi)+\zeta \tag{8}$$

have a slope whose absolute value obeys the condition

$$\left|\frac{y_2-y_1}{\psi_2-\psi_1}\right| = \alpha\beta\rho\left|\frac{\sin(\varphi-\psi_2)-\sin(\varphi-\psi_1)}{\psi_2-\psi_1}\right| \leqslant \alpha\beta\rho. \tag{9}$$

Hence, for $\alpha\rho<1$, the straight line $y(\psi)=\beta\psi$ intersects the sinusoid only at one point, and the condition that the solution of Eq. (7) be unique can be written as

$$\rho < \alpha^{-1}. \tag{10}$$

In this region of space, planes perpendicular to the helix can be interpreted as level surfaces $\psi=$ const of the new curvilinear system of coordinates. In what follows, α is assumed small, so that the region $\rho\geqslant\alpha^{-1}$, i.e., $\sqrt{x^2+y^2}>\lambda\alpha^{-1}$, is unimportant.

We calculate the magnetic induction field and the energy of a helical Abrikosov vortex in the form of series expansions in powers of α to within terms proportional to α^2. We also assume that

$$\alpha\beta = \frac{a}{b} = \tan\theta \approx \theta \ll 1, \quad \alpha\beta^2 = \frac{a}{\lambda} < 1. \tag{11}$$

Then, to the same accuracy, the dimensionless curvature is

$$\lambda C = \frac{\lambda a}{a^2+b^2} \approx \alpha. \tag{12}$$

In the plane $\psi=$ const, we introduce polar coordinates centered at the point where the plane intersects the helix. Then the radius $\tau=|\mathbf{r}-\mathbf{r}_\nu|$ is determined by the equation

$$\tau^2 = \rho^2 - 2\alpha\beta^2\rho\,\cos(\varphi-\psi)+\alpha^2\beta^4 + \alpha^2\beta^2\rho^2\sin^2(\varphi-\psi). \tag{13}$$

The azimuthal angle ω is measured from the direction opposite the principal normal to the helix. Then

$$\cos\omega = \frac{(\mathbf{r}-\mathbf{r}_\nu)d^2\mathbf{r}_\nu/d\psi^2}{|\mathbf{r}-\mathbf{r}_\nu||d^2\mathbf{r}_\nu/d\psi^2|} = \frac{1}{\tau}[\rho\,\cos(\varphi-\psi)-\alpha\beta^2],$$
$$\sin\omega = \frac{\rho}{\tau}\sin(\varphi-\psi)\sqrt{1-\alpha^2\beta^2}. \tag{14}$$

In deriving Eqs. (14) we used Eqs. (7) and (13). The sign of $\sin\omega$ in defined in such a way that it coincides with the sign of the mixed product

$$\left((\mathbf{r}-\mathbf{r}_\nu)\cdot\frac{d^2\mathbf{r}_\nu}{d\psi^2}\right)\mathbf{z} = \rho\,\sin(\varphi-\psi), \tag{15}$$

where $\mathbf{z}$ is the unit vector along the z axis. Solving Eqs. (7), (13), and (14) in the adopted approximation, we arrive at the following expressions of the cylindrical coordinate ρ, φ, ζ in the region where $\rho<\alpha^{-1}$, $0\leqslant\zeta<2\pi\beta$, and $0\leqslant\varphi<2\pi$ in terms of the new curvilinear coordinates τ, ω, ψ:

$$\rho = \tau+\alpha\beta^2\cos\omega+\frac{\alpha^2\beta^2}{2\tau}(\beta^2-\tau^2)\sin^2\omega,$$
$$\varphi = \omega+\psi-\frac{\alpha\beta^2}{\tau}\sin\omega+\frac{\alpha^2\beta^4}{2\tau^2}\sin 2\omega-\frac{\alpha^2\beta^2}{4}\sin 2\omega, \tag{16}$$

$$\zeta=\beta\psi-\alpha\beta\tau\sin\omega.$$

To use the notation of tensor analysis we introduce the following indices:

$$\tau\to 1,\quad \omega\to 2,\quad \psi\to 3. \tag{17}$$

The fundamental tensor of this coordinate system has the following form:

$$\begin{aligned} &g_{11}=1,\quad g_{13}=g_{31}=g_{12}=g_{21}=0,\quad g_{22}=\tau^2,\\ &g_{33}=\tau^2+\beta^2+2\tau\alpha\beta^2\cos\omega+\alpha^2\beta^4-\alpha^2\beta^2\tau^2\sin^2\omega,\\ &g_{23}=g_{32}=\tau^2-\frac{\alpha^2\beta^2\tau^2}{2}. \end{aligned} \tag{18}$$

Thus, the new system of coordinates is not orthogonal, which means one must distinguish between covariant and contravariant components of a vector.

Equation (3) describe a right-handed helix in relation to the direction of the z axis. Since here the direction of the z axis is fixed by the physics of the problem (it coincides with the direction of the average magnetic induction field of the Abrikosov vortex), one must distinguish between right- and left-handed helices. If the above calculations are done for a left-handed helix, the sign of the components $g_{23}=g_{32}$ of the fundamental tensor is reversed.

3. MODEL OF A HELICAL ABRIKOSOV VORTEX AND THE VORTEX ENERGY

The magnetic induction field of an Abrikosov vortex for $\kappa\gg 1$ is described by the equation

$$\mathbf{h}+\nabla\times(\hat{\mu}\,\nabla\times\mathbf{h})=\frac{\Phi_0}{2\pi\lambda^2}\Delta(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}). \tag{19}$$

Here the coordinates are measured in units of λ, the tensor $\hat{\mu}$ in terms of the major axes has the form

$$\mu_{11}=\mu_{22}=1,\quad \mu_{33}=\gamma^{-2}, \tag{20}$$

and the anisotropy axis is selected as axis 3. The vector field $\Delta(\mathbf{R})\cdot\mathbf{n}(\mathbf{R})$ is determined by the complete system of Ginzburg–Landau equations. For $\kappa\gg 1$ and a straight Abrikosov vortex this field is parallel to the vortex axis, occupies a region near the axis of a size of order κ^{-1}, and its flux through a plane perpendicular to the axis is $2\pi\lambda^2$.

The principal assumption of the proposed model can be formulated as follows: if the parameters of the helix are such that $a/\lambda=\alpha\beta^2\gg\kappa^{-1}$ and $b/\lambda=\beta\gg\kappa^{-1}$, the structure of the Abrikosov vortex core does not change, i.e., the value of $\Delta(\mathbf{R})$ depends only on τ, is negligible when $\tau>\kappa^{-1}$, and

$$\int_0^\infty\Delta(\tau)\tau\,d\tau=1, \tag{21}$$

while the vector $\mathbf{n}(\mathbf{R})$ is always perpendicular to the coordinate surface $\psi=$const (a plane) and is equal to unity in absolute value. Then $\mathbf{n}(\mathbf{R})$ in helical coordinates has only one covariant component:

$$n_3=\sqrt{\frac{g}{g_{22}}}\approx\beta\left(1+\alpha\tau\cos\omega+\frac{\alpha^2\beta^2}{2}\right). \tag{22}$$

The initial formula for calculating the specific energy (per unit length along the average direction) of a helical Abrikosov vortex is

$$\varepsilon=\frac{\lambda^3}{16\pi^2 b}\int_V(\nabla\times\mathbf{h})\cdot(\hat{\mu}\nabla\times\mathbf{h})dV. \tag{23}$$

where V is the volume of a $2\pi b$-thick layer, i.e., one turn of the helix. Integrating the second term by parts (see Ref. 6) and allowing for (19), we get

$$\varepsilon=\frac{\Phi_0}{32\pi^3\beta}\int_V\Delta(\tau)(\mathbf{n}\cdot\mathbf{h})dV. \tag{24}$$

If we now express the scalar product in terms of the covariant components of the vectors $\mathbf{n}$ and $\mathbf{h}$ via (18) and (22) and integrate (24) with respect to ψ, we get

$$\varepsilon=\frac{\Phi_0}{16\pi^2\beta}\int_0^{2\pi}\int_0^\infty\Delta(\tau)\left[h_3-\left(1-\frac{\alpha^2\beta^2}{2}\right)h_2\right]\tau\,d\tau d\omega. \tag{25}$$

To integrate with respect to τ, we reason along the same lines as in Ref. 5, i.e., we note that $\Delta(\tau)\tau$ has a sharp peak at $\tau=\kappa^{-1}$. Taking the other functions outside the integral with respect to τ and allowing for (21), we get

$$\varepsilon=\frac{\Phi_0}{16\pi^2\beta}\int_0^{2\pi}\left[h_3(\kappa^{-1},\omega)-\left(1-\frac{\alpha^2\beta^2}{2}\right)h_2(\kappa^{-1},\omega)\right]d\omega. \tag{26}$$

By writing Eqs. (19) in the form of equations for the covariant components of $\mathbf{h}$ and expanding in powers of α, we find a solution of the form

$$\begin{aligned} h_1(\tau,\omega)&=\alpha h_1^{(1)}(\tau)\sin\omega,\\ h_2(\tau,\omega)&=\alpha h_2^{(1)}(\tau)\cos\omega+\alpha^2[h_2^{(2)}(\tau)\\ &\quad+\tilde{h}_2^{(2)}(\tau)\cos 2\omega],\\ h_3(\tau,\omega)&=h_3^{(0)}(\tau)+\alpha h_3^{(1)}(\tau)\cos\omega+\alpha^2[h_3^{(2)}(\tau)\\ &\quad+\tilde{h}_3^{(2)}(\tau)\cos 2\omega]. \end{aligned} \tag{27}$$

Plugging (27) into (26) and integrating with respect to ω, we obtain

$$\varepsilon=\frac{\Phi_0}{8\pi\beta}\{h_3^{(0)}(\kappa^{-1})+\alpha^2[h_3^{(2)}(\kappa^{-1})-h_2^{(2)}(\kappa^{-1})]\}. \tag{28}$$

The system of ordinary differential equations for the $h_i^{(m)}(\tau)$ simplifies considerably if we introduce the functions

$$u^{(1)}(\tau)=\frac{d}{d\tau}h_2^{(1)}-h_1^{(1)},\quad v^{(2)}(\tau)=h_3^{(2)}-\hbar_2^{(2)}. \tag{29}$$

Then

$$\begin{aligned} &\frac{d^2h_3^{(0)}}{d\tau^2}+\frac{1}{\tau}\frac{dh_3^{(0)}}{d\tau}-h_3^{(0)}=-f,\\ &\frac{d^2u^{(1)}}{d\tau^2}-\frac{1}{\tau}\frac{du^{(1)}}{d\tau}-\eta^2u^{(1)} \end{aligned}$$

$$= -\tau \frac{d}{d\tau}[(1-\eta^2)h_3^{(0)} - (1-\gamma^2)f], \tag{30}$$

$$\frac{d^2 h_3^{(1)}}{d\tau^2} + \frac{1}{\tau}\frac{dh_3^{(1)}}{d\tau} - \delta^2 h_3^{(1)} - \frac{h_3^{(1)}}{\tau^2}$$

$$= \tau(h_3^{(0)} - f) - \frac{\tau}{\gamma^2}\frac{d^2 h_3^{(0)}}{d\tau^2} + \frac{1+\gamma^2}{\gamma^2 \tau}u^{(1)} - \frac{1-\gamma^2}{\gamma^2}\frac{du^{(1)}}{d\tau},$$

$$\frac{d^2 v^{(2)}}{d\tau^2} + \frac{1}{\tau}\frac{dv^{(2)}}{d\tau} - v^{(2)}$$

$$= \frac{1}{2\tau}\frac{d}{d\tau}(\tau h_3^{(1)}) - \frac{\tau}{2}\frac{dh_3^{(0)}}{d\tau} - \frac{u^{(1)}}{2}$$

$$+ \beta^2\left(\frac{1}{\gamma^2} - 1\right)\left(f - \frac{1}{2\tau}\frac{du^{(1)}}{d\tau} - \frac{h_3^{(0)}}{2}\right),$$

where we have introduced the following notation:

$$\frac{\Phi_0 \beta}{2\pi\lambda^2}\Delta(\tau) = f, \quad \delta^2 = 1 + \beta^{-2}, \quad \eta = \gamma\delta. \tag{31}$$

The boundary conditions for all the functions are the following: the functions tend to zero as $\tau \to \infty$ and are finite and continuous. The fundamental solutions of the corresponding homogeneous equations are well-known, the right-hand side of the first equation is a known function, and the right-hand side of every successive equation can be expressed in terms of the solutions of the preceding equation. Now the solution of the system (30) can easily be obtained. After certain transformations via integration by parts are performed, the solution can be expressed in terms of the modified Bessel functions $I_n(\xi\tau)$ and $K_n(\xi\tau)$ and the following integrals of these functions:

$$L_n(\xi,\tau) = \int_0^\tau I_n(\xi x)\Delta(x)x^{n+1}dx,$$

$$M_n(\xi,\tau) = \int_\tau^\infty K_n(\xi x)\Delta(x)x^{n+1}dx. \tag{32}$$

Here ξ can be γ, δ, η, and $n = 0, 1$. When this solution is plugged into (28) and the integrals (32) are to be evaluated, we employ the fact that $\tau\Delta(\tau)$ has a sharp peak at $\tau = \kappa^{-1}$. For $\beta \gg \kappa^{-1}$ the parameters δ and η cannot be large, $\gamma < 1$, so that for the modified Bessel functions we can take their approximate expressions for a small argument. Then, if we allow for (21), we arrive at the final expression for the energy of a helical Abrikosov vortex per unit length along the axis of the helix:

$$\varepsilon = \varepsilon_0\left\{\ln\frac{2\kappa}{e^C} + \frac{\alpha^2\beta^2}{2}\left[\gamma^2 \ln\frac{2\kappa}{e^C\gamma}\right.\right.$$

$$\left.\left. + \frac{1}{2}(\beta^2 - \gamma^2)\ln(1+\beta^{-2}) - \frac{1}{2}\right]\right\}, \tag{33}$$

where C is Euler's constant.

Let us estimate the contribution ε_n to the Abrikosov-vortex energy per unit length caused by a decrease in the absolute value of the order parameter in a region of radius $\lambda\kappa^{-1}$ near the vortex line. If we assume that the order parameter is zero inside this region and equal to its equilibrium value outside, we get

$$\varepsilon_n \cong \left(\frac{\Phi_0}{8\pi\lambda}\right)^2 = \frac{1}{4}\varepsilon_0. \tag{34}$$

This is the upper bound—actually ε_n is somewhat smaller. Allowing for the fact that $1 + \alpha^2\beta^2/2$ of the vortex line fits into a unit length of the helix, the term $(1+\alpha^2\beta^2/2)\varepsilon_0 d$, with $d < 1/4$, should be added to (33). This, however, does not alter the qualitative conclusions that follow from (33).

4. CONCLUSIONS AND PHYSICAL EFFECTS

Equation (33) shows that for all $\gamma \leqslant 1$ there exists a range of values of β sufficiently small for the energy of a helical Abrikosov vortex to be lower than that of a straight Abrikosov vortex. Note, however, that Eq. (33) was derived on the assumption that $\beta \gg \kappa^{-1}$. For instance, at $\gamma = 1$ the second term in (33) is negative for $\beta \leqslant \kappa^{-1}$, with the result that for an isotropic superconductor the above conclusion is invalid. At $\gamma = 1/8$ and $\kappa = 50$ a helical Abrikosov vortex is energetically favored over the straight configuration for $\beta < 1.4$.

If we ignore the bending energy, then the energy of a helical Abrikosov vortex coincides with that of a straight Abrikosov vortex inclined to the anisotropy axis by an angle $\theta = \alpha\beta$. To calculate this energy, we must let α go to zero and β to infinity in (33), but must not alter the product $\alpha\beta$. This result coincides with the one obtained from the respective formula in Ref. 5. The energy is always higher than that of a straight Abrikosov vortex directed along the anisotropy axis, as is the case in Ref. 3. Thus, the lowering of the Abrikosov vortex energy in the transition to the helical configuration reveals itself only if we allow for the negative bending energy.

The above suggests that when the anisotropy is strong, a straight Abrikosov vortex directed along the anisotropy axis is unstable. Its true configuration can be found by solving the corresponding variational problem, but such a problem has yet to be formulated. If, on the average, the induction is directed along the anisotropy axis, symmetry considerations suggest that the helical configuration is a stable one. To find the values of its parameters α and β that minimize the free energy, we must calculate $\varepsilon(\alpha,\beta)$ without resorting to perturbation techniques. When the average direction of induction deviates from the anisotropy axis, the equilibrium Abrikosov vortex configuration differs from the helical. This, possibly, brings together the results of the anisotropic London model and those of the step model of Abrikosov vortices in a layered superconductor examined in Ref. 10.

The fact that the axial line of an Abrikosov vortex is curvilinear may be important for quantitative calculations of the interaction of this vortex with various inhomogeneities in the superconductor and with the transport current, i.e., for analyzing the mechanisms of pinning and the penetration of the superconductor by the vortex. New phenomena involving the system of Abrikosov vortices are also related to this fact.

As noted earlier, an Abrikosov vortex can be either a left- or right-handed helix in relation to the direction of magnetic induction. In the absence of a transport current along the direction of induction, the energies of these left- and right-handed vortices are equal. The repulsion of unlike Abrikosov vortices must be stronger than that of like vortices, since similarly directed circular vortices lying in the same plane attract each other. Right- and left-handed Abrikosov vortices can transform into each other by passing through the stage of a topological soliton. When the Abrikosov vortex number density is low, the number of right- and left-handed vortices must be the same, to within fluctuations, while as the number density increases, a phase transition occurs with spontaneous breaking of left–right symmetry. The driving field for this transition is a transport current parallel to the magnetic induction, since it changes the energies of right- and left-handed vortices differently.

A lattice consisting of helical Abrikosov vortices cannot be interpreted as a deformation of a lattice of straight vortices. When the Abrikosov vortex number density is low, such a lattice differs little, both in its energy and elastic properties, from a lattice of straight vortices, since left- and right-handed Abrikosov vortices are mixed at random and the energy of interaction of their circular components decreases with increasing distance much faster than that of straight Abrikosov vortices.

At the phase transition to an ordered state the dependence of the lattice's elastic moduli on the magnetic field strength experiences a singularity. In a lattice ordered in the direction of the helix the interaction energy is lower, due to the attraction of the circular components of helical Abrikosov vortices, than in a lattice of straight vortices of the same number density. The elastic properties are also different (in relation to bending, in particular). The appearance of new singularities can be expected when the lattice constant becomes equal to the radius of the helix. The overlap of helices must lead to an enhancement of repulsion, which, possibly, is balanced by the straightening of the vortices.

APPENDIX

The energy of a circular Abrikosov vortex of large radius R in an isotropic superconductor can be calculated by the same method as that of a helical vortex (see the main text). Here toroidal coordinates are the natural system of coordinates (see Ref. 11). If the axis of the cylindrical coordinates r, φ, z passes through the center of the circumference of the axial line of the vortex perpendicular to its plane, these coordinates can be expressed in terms of the toroidal coordinates t,φ,ω as follows:

$$r=R\frac{1-t^2}{1-2t\,\cos\,\omega+t^2},\quad z=\frac{2Rt\,\sin\,\omega}{1-2t\,\cos\,\omega+t^2}. \tag{A1}$$

The single component of the induction field is $h_\varphi(t,\omega)$. The equation for the function $H(t,\omega)=h_\varphi\sqrt{r/R}$ has the form

$$H-\frac{\lambda^2}{4R^2t^2}(1-2t\,\cos\,\omega+t^2)^2\left[t^2\frac{\partial^2H}{\partial t^2}+t\frac{\partial H}{\partial t}+\frac{\partial^2H}{\partial\omega^2}-\frac{3t^4}{(1-t^2)^2}H\right]$$
$$=\frac{\Phi_0}{2\pi\lambda^2}\Delta\left(\frac{2R}{\lambda}t\right)n(t,\omega)\sqrt{\frac{r}{R}}, \tag{A2}$$

where $\Delta(\tau)$ is the same function as in the main text, and $n(t,\omega)$ can be determined from the condition that the magnetic flux through the half-plane $\varphi=\text{const}$ is equal to the flux quantum Φ_0. The dimensionless curvature $\alpha=\lambda/R$ is the small parameter needed for solving this equation. Substituting $\tau\alpha/2$ for t in Eq. (A2) and employing perturbation-theory techniques, we can calculate the magnetic induction in the second order and the specific energy of a circular Abrikosov vortex. The result is

$$\varepsilon=\varepsilon_0\left(\ln\frac{2\kappa}{e^C}-\frac{3}{8}\alpha^2\right). \tag{A3}$$

Thus, the bending modulus for an Abrikosov vortex in an isotropic London superconductor is negative.

[1] T. Koyama and M. Tachiki, Solid State Commun. **79**, 1949 (1991).
[2] M. L. Kulič and A. Kramer, Solid State Commun. **79**, 1949 (1991).
[3] E. H. Brandt, Phys. Rev. Lett. **69**, 1105 (1992).
[4] A. V. Balatskiĭ, L. I. Burlachkov, and L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **90**, 1478 (1986) [Sov. Phys. JETP **63**, 866 (1986)].
[5] I. M. Dubrovskiĭ, Zh. Éksp. Teor. Fiz. **111**, 954 (1997) [JETP **84**, 525 (1997)].
[6] P. G. de Gennes, *Superconductivity of Metals and Alloys*, Benjamin, New York (1966).
[7] Yu. A. Genenko, JETP Lett. **59**, 841 (1994).
[8] Yu. A. Genenko, Physica C **235**–**240**, 2709 (1994).
[9] Yu. A. Genenko, J. Appl. Phys. **76**, 7144 (1994).
[10] L. N. Bulaevskii, M. Ledvij, and V. G. Kogan, Phys. Rev. B **46**, 366 (1992).
[11] E. Madelung, *The Mathematical Apparatus of Physics* [Russian translation], Fizmatgiz, Moscow (1961), [translation of *Die Mathematischen Hilfsmittel des Physikers*, 5th ed., Springer, Berlin (1954)].

Translated by Eugene Yankovsky

Phase transitions in a system of spatially separated electrons and holes

Yu. E. Lozovik*) and O. L. Berman

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 14 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 1879–1895 (May 1997)

The formation of a superfluid exciton liquid in a system of spatially separated electrons and holes in a system of two coupled quantum wells is predicted and its properties are investigated. The ground-state energy and the equilibrium density of the exciton liquid are calculated as functions of distance D between the quantum wells. The properties of a rarefied exciton gas with dipole–dipole repulsions are considered, where this gas is the metastable phase for $D<1.9a^*$ and the stable phase for $D>1.9a^*$ (a^* is the radius of the two-dimensional exciton). The gas–liquid quantum transition is examined for increasing D. The Berezinskiĭ–Kosterlitz–Thouless transition temperatures, at which superfluidity arises in the system, are found for different values of D. Possible experimental manifestations of the predicted effects are discussed. © *1997 American Institute of Physics.* [S1063-7761(97)02505-5]

1. INTRODUCTION

Systems with spatially separated electrons and holes in systems of double quantum wells have attracted a great deal of attention recently,[1–4] particularly in connection with the predicted superfluidity in such systems,[5] and in connection with quasi-Josephson phenomena[5,6] and the unusual properties such systems manifest in strong magnetic fields.[7] There exist a number of physical realizations of systems with spatially separated electrons and holes. The electrons and holes can be created by laser radiation, they enter into a state of partial thermodynamic equilibrium in the system of coupled quantum wells and are found in the equilibrium state at times substantially longer than their energy relaxation times but shorter than their recombination times. Their recombination times can be quite long due to the spatial separation of the electron and hole wave functions (but for some coupled quantum wells[3]—due to localization of the electrons and holes in different regions of momentum space), which makes the quasi-equilibrium state of the electron–hole system easily accessible. In this case the electrons and holes are characterized by non-coincident quasi-equilibrium chemical potentials. It is also possible for spatially separated electrons and holes to be found in a state of thermodynamic equilibrium such that their chemical potentials coincide. This can happen in second-order coupled quantum wells (e.g., for structures based on InAs/GaSb).

In connection with the experiments which have been undertaken to date, the greatest interest has lain in the as-yet unexamined properties of the double-quantum-well structures, in particular the question of condensed phases of the excitons in the given system, the form of the phase diagram, etc. These questions are the subject of the present study, which considers a two-layer system in the absence of a magnetic field. Such a system was recently investigated experimentally.[1,2] The behavior of a spatially separated electron–hole system in a strong magnetic field is discussed in Ref. 8.

The present study predicts the properties of the liquid phase of excitons with spatially separated electron and holes and discusses the gas–liquid quantum transition as a function of the distance D between the wells. In addition, it predicts the temperature of the topological transition to the superfluid state, which depends on D.

2. HAMILTONIAN OF THE SYSTEM. HARTREE–FOCK APPROXIMATION

To determine the existence conditions (at zero temperature) of a liquid consisting of excitons with spatially separated electron and holes, it is necessary to calculate the dependence of the ground-state energy E on the (non-equilibrium) concentration n and determine the minimum of $E(n)$ for various values of the distance D between the layers.

The Hamiltonian of a system of spatially separated electrons and holes can be written in the second-quantization representation

$$\hat{H}=\sum_{p=0}^{\infty}\left[\left(\frac{p^2}{2m_e}-\mu_e\right)a_p^+a_p+\left(\frac{p^2}{2m_h}-\mu_h\right)b_p^+b_p\right]+\frac{1}{2}\sum_{pp'k}\{V(k)[a_p^+a_{p'}^+,a_{p'+k}a_{p-k}+b_p^+b_{p'}^+,b_{p'+k}b_{p-k}]-2\tilde{V}(k)a_p^+b_{p'}^+,b_{p'+k}a_{p-k}\},\quad(1)$$

where a_p^+ and b_p^+ are the creation operators of the electron and hole; $m_e=m_h=m$ is the effective mass of the electron and hole; $V(k)=2\pi e^2/\varepsilon k$ is the Coulomb interaction in one layer; $\tilde{V}(k)=V(k)e^{-kD}$ is the interaction between an electron and hole located in different layers; D is the distance between the layers of electrons and holes; ε is the static dielectric constant; μ_e and μ_h are the chemical potentials, which are governed by the normalization conditions (we assume the electron and hole concentrations to be equal, $N_e=N_h$)

$$\sum_p\langle a_p^+a_p\rangle=\sum_p\langle b_p^+b_p\rangle=\frac{1}{2}N,$$

 1063-7761/97/051027-09$10.00

where $N=N_e+N_h$ is the total number of particles in the system; $n=N/S$ is the surface concentration of the particles in the system; and S is the area of the system.

To calculate the ground-state energy of an exciton liquid consisting of spatially separated electrons and holes, it is necessary to take their pairing into account.[5,9] Toward this end, we use the Gor'kov technique for the normal and anomalous Green's functions:[10,11]

$$G_{\alpha\beta}(x,x')=-i\langle T(\tilde{\psi}_\alpha(x)\tilde{\psi}^+_\beta(x'))\rangle,$$

$$F_{\alpha\beta}(x-x')=e^{i\mu t}\langle N|T(\tilde{\psi}_\alpha(x)\tilde{\psi}_\beta(x'))|N+2\rangle,$$

$$F^+_{\alpha\beta}(x-x')=e^{-i\mu t}\langle N+2|T(\tilde{\psi}^+_\alpha(x)\tilde{\psi}^+_\beta(x'))|N\rangle,$$

where $G_{\alpha\beta}(x,x')$ and $F_{\alpha\beta}(x-x')$ are respectively the normal and anomalous Green's functions, and $\tilde{\psi}_\alpha(x)$ and $\tilde{\psi}^+_\alpha(x)$ are the single-particle Fermi operators: $\mu=\mu_e+\mu_h$.

To start with, let us consider the Hartree–Fock approximation, which we will treat as an initial approximation. In a homogeneous system the diagrams of the Hartree approximation give zero contribution to the energy, as follows from electrical neutrality. Let us consider the exchange diagrams of the Hartree–Fock approximation.

With the intention of departing from the framework of the BCS approximation (for weak coupling), we will solve the equations for the normal and anomalous Green's functions in the general case without assuming that the quasiparticle momenta are small in comparison with the Fermi momentum. The normal and anomalous Green's functions in the generalized Hartree–Fock approximation (with possible spontaneous symmetry breaking and gap $\Delta\neq 0$) are solutions of the equations for the Gor'kov–Nambu matrix function $\hat{G}$:

$$\hat{G}=\begin{pmatrix} G & F^+ \\ -F & G\end{pmatrix},\quad G(p)=\frac{\omega+\xi}{\omega^2-\epsilon^2(p)},$$

$$F^+(p)=\frac{|\Delta(p)|^2}{\omega^2-\epsilon^2(p)}, \tag{2}$$

Here $\xi=p^2/2m-\mu$,

$$\epsilon(p)=\frac{1}{z_p^2+1}\left\{(z_p^2-1)(p^2-r^2\mu p_0^2)\,\frac{\sqrt{2}p_0 r}{\sqrt{\pi}}\int_0^\infty dp'\int_0^{2\pi}d\phi\right.$$

$$\times\frac{(z_p^2-1-2z_pz_{p'}\,\exp[-Dp_F\sqrt{p^2-2pp'\cos\phi+p'^2}])p'}{(z_{p'}^2+1)\sqrt{p^2-2pp'\cos\phi+p'^2}}$$

$$\times\frac{p_F^2}{4}, \tag{3}$$

$r=1/\sqrt{\pi n}$ is the mean distance between the electron and hole layers, $z_p=u_p/v_p$, and the functions u_p^2 and v_p^2 are defined by the relations

$$u_p^2=\frac{1}{2}\left[1+\frac{\xi}{\epsilon(p)}\right],\quad v_p^2=\frac{1}{2}\left[1-\frac{\xi}{\epsilon(p)}\right].$$

We have introduced the notation

$$p_0^2=2\int_0^\infty\frac{pdp}{1+z_p^2},\quad p_F=\sqrt{2\pi\hbar^2 n};$$

all momenta are expressed in units of p_F/p_0. In the Hartree–Fock approximation the gap Δ is determined from the self-consistency condition. In the approximation we are using for an intermediate type of coupling the gap is determined from a variational calculation with allowance for the correlation energy (see also Refs. 9 and 12). As the function being varied it is convenient to use the function z_p (see above).

The energy can be expressed in terms of the Gor'kov–Nambu function $\hat{G}$ of the system

$$E=-\mathrm{Tr}\int_0^{e^2}\frac{de'^2}{e'^2}\int_0^\infty\frac{d\omega}{2\pi}\int\frac{d^2p}{(2\pi)^2}$$

$$\times[\hat{G}^{(0)}(\omega,p)]^{-1}[\hat{G}(\omega,p)-\hat{G}^{(0)}(\omega,p)]e^{i\omega t}, \tag{4}$$

where $t\to+0$; $\hat{G}^{(0)}(\omega,p)=G^{(0)}(\omega,p)\hat{1}$ is the free-particle Green's function, and $\hat{1}$ is the unit matrix (in the integrand we make use of the standard substitution of the physical charge $e^2\to e'^2$ with subsequent integration over e'^2 up to the physical value e^2).

In the Hartree–Fock approximation the Green's function $\hat{G}$ is given by Eq. (2). Taking Eqs. (2)–(4) into account, we find the functional $E_{HF}\{n\}$ in the Hartree–Fock approximation:

$$2\left(\frac{E_{HF}\{n\}}{n}+\mu\right)=\frac{4}{r^2p_0^4}\int_0^\infty\frac{p^3dp}{1+z_p^2}-\frac{\sqrt{2}}{\pi^2r^2p^3}\int_0^\infty pqdpdq$$

$$\times\int_0^{2\pi}\frac{|V(p-q)+\tilde{V}(p-q)z_pz_q|d\phi}{(1+z_p^2)(1+z_q^2)}. \tag{5}$$

As will be shown below, at large $D\geqslant D_{sp}=1.9a^*$ the liquid phase is absent and the electron and hole concentrations are free parameters of the problem. In this case, at high concentrations $na^{*2}\gg 1$ the quasiparticle spectrum $\epsilon(p)=\sqrt{\xi^2+\Delta^2(p)}$, where $\Delta(p)$ is the gap in the quasiparticle spectrum found by minimizing the Hartree–Fock functional. In particular, for $D\gg a^*$ we have (see Ref. 5)

$$\Delta(p)=\exp\left(-\frac{16D^2p_0}{\pi a^*}\right). \tag{6}$$

However, at $D\lesssim a^*$ correlation effects turn out to be important.

3. ALLOWANCE FOR CORRELATION EFFECTS

We consider the results obtained in the Hartree–Fock approximation for the exciton phase as an initial approximation. We use the Green's functions obtained in Sec. 2 in the Hartree–Fock approximation to calculate the correlation energy $E^{\mathrm{cor}}(z_p)$. We wish to know what classes of diagrams for the correlation energy are important.

For small momentum transfer we estimate the ratio of the minimum momentum transfer $k\geqslant\Delta/v_F$ (Δ is the gap obtained by a variational calculation; see below) to the Fermi

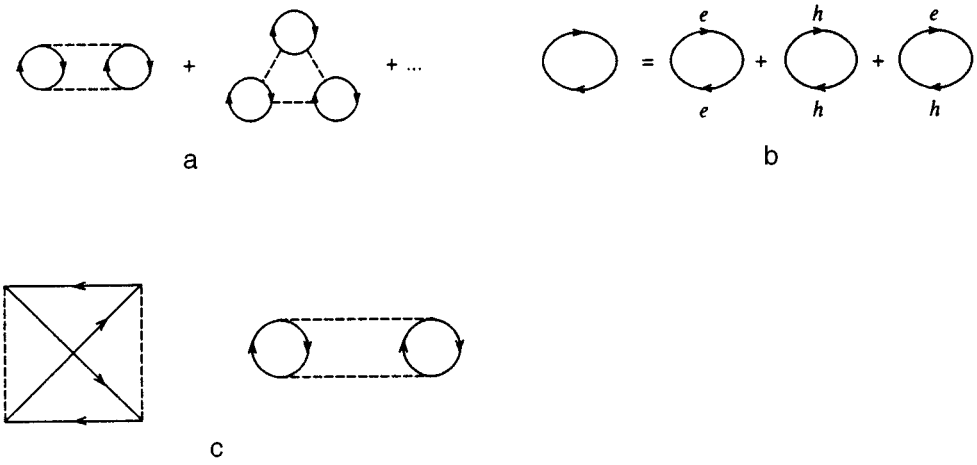

FIG. 1. Diagrams for the correlation energy of an electron–hole system: *a*—diagrams of the random-phase approximation, taken into account for $k/p_F \ll 1$; *b*—polarization operator in the random-phase approximation; *c*—second-order diagrams, taken into account for $k/p_F \gg 1$.

radius p_F. At the equilibrium concentration point for the interlayer distance $D=0$, for which the gap is greatest, using the results obtained below, we have

$$\frac{k}{p_F} \geqslant \frac{\Delta}{v_F p_F} = \frac{\Delta m}{2\pi\hbar^2 n} \approx 0.07 \ll 1.$$

Consequently, the separability condition for the diagrams of the random-phase approximation[13] (Fig. 1a) holds for small values of the momentum transfer k $(k/p_F \ll 1)$ and for our system.

For the correlation energy in the random-phase approximation we have

$$E_1^{\mathrm{cor}} = -\frac{1}{n}\int \frac{d^2k}{(2\pi)^2}\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}\{\ln[1-V_k\Pi(k,\omega)] - V_k\Pi(k,\omega)\}, \tag{7}$$

where $\Pi(k,\omega)$ is the polarization operator of a two-component electron–hole system with allowance for electron–hole pairing (Fig. 1b):

$$\Pi(k,\omega) = \Pi_{ee}(k,\omega) + \Pi_{hh}(k,\omega) + \Pi_{eh}(k,\omega). \tag{8}$$

For the electron–electron and hole–hole polarization operators we have

$$\Pi_{ee(hh)}(k,\omega) = -2\int \frac{d^2p}{(2\pi)^2}\int_{-\infty}^{+\infty}\frac{d\epsilon}{2\pi i} \times\{[G_{HF}(p,\epsilon)G_{HF}(-p+k,-\epsilon+\omega) + G_{HF}(p,\epsilon)G_{HF}(-p-k,-\epsilon-\omega)] + [F_{HF}(p,\epsilon)F_{HF}(p+k,\epsilon+\omega) + F_{HF}(p,\epsilon)F_{HF}(p-k,\epsilon-\omega)]\}, \tag{9}$$

where F_{HF} and G_{HF} are the anomalous and normal Green's functions in the Hartree–Fock approximation (2). For the electron–hole polarization operator we obtain

$$\Pi_{eh}(k,\omega) = -2\int \frac{d^2p}{(2\pi)^2}\int_{-\infty}^{+\infty}\frac{d\epsilon}{2\pi i}[F_{HF}(p,\epsilon)F_{HF}(p + k,\epsilon+\omega) + F_{HF}(p,\epsilon)F_{HF}(p-k,\epsilon-\omega)], \tag{10}$$

In the absence of pairing and neglecting tunneling between the layers the mixed operator $\Pi_{eh}=0$.

The correlation energy E_1^{cor} for small momentum transfer $(k/p_F \ll 1)$ is given by relations (7)–(10) and depends on the function z_p, which we take as the variational function.

The correlation energy E_2^{cor} for large momentum transfer $(k/p_F \gg 1)$ is given by relation (4), where the Green's function $G(\omega,p)$ is approximated by the sum of a ring diagram and an exchange diagram, both of second order (See Fig. 1c).

We will use an interpolation for the correlation energy E^{cor} (successfully employed by Brinkman and Rice to calculate a metallic electron–hole liquid[14] and also used in Refs. 9 and 12) for all values of the momentum transfer:

$$E^{\mathrm{cor}} = \frac{E_1^{\mathrm{cor}}E_2^{\mathrm{cor}}}{E_1^{\mathrm{cor}}+E_2^{\mathrm{cor}}}. \tag{11}$$

Next, for various values of the interlayer distance D and the quasiparticle momentum n the total energy

$$E_t = E_{HF} + E^{\mathrm{cor}}, \tag{12}$$

is defined as a functional of the function $z_p = u_p/v_p$.

Some remarks are in order at this point regarding the relationship between the present approach and Bogolyubov's principle of cancellation of dangerous diagrams. The appearance in the diagrams of parts connected to the reset of the diagram by only one pair of lines corresponding to an electron and hole with zero total momentum leads to a divergence of the corresponding diagram. This pair of lines can be replaced by a total two-particle Green's function G_2 that describes the bound states of the electron and hole (an exciton). The exciton energy levels are poles of the function G_2 as a function of the total energy of the pair. For an electron–hole pair created from the vacuum, the total momentum is zero and the total energy is $\mu = \mu_e + \mu_h$. The energy of an exciton with zero momentum is by definition

equal to μ. That means that an electron–hole pair with zero total momentum and energy μ corresponds to a pole of the function G_2. Consequently, diagrams incorporating virtual processes of creation from vacuum of electron–hole pairs with zero total momentum diverge.

To eliminate the divergence, it is necessary to ensure the mutual cancellation of diagrams that lead to virtual creation from the vacuum of electron–hole pairs with zero total momentum. The application of this cancellation principle with allowance for only the first order in the interaction is equivalent to solving the self-consistent equations in the Hartree–Fock approximation (with allowance for spontaneous symmetry breaking, and $\Delta \neq 0$) for the function z_p, which can also be found in this approximation from the variational principle.

If we take account of diagrams corresponding to dynamic correlation processes between the quasiparticles, then the cancellation of divergent diagrams will correspond in our approximation to taking account of the diagrams in Fig. 1. The minimization of the total energy of the system relative to the function z_p carried out below (with allowance for the correlation energy) is an approximation to the solution of the equation for cancellation of divergent diagrams.

4. VARIATIONAL CALCULATION

The variational calculation of the total energy E_t was performed numerically. First, all expressions were reduced to dimensionless form using the radius and energy of the two-dimensional exciton, respectively, as our units of length and energy:

$$a^*=\frac{\varepsilon\hbar^2}{2me^2}, \quad E_x=\frac{2me^4}{\varepsilon\hbar^2},$$

where $m=m_e m_h/(m_e+m_h)$ is the reduced mass.

To minimize E_t we used various test functions z_p. The results did not differ greatly, but the best value was obtained using a test function of the form

$$z_p=A\left(1+\frac{p^2}{4}\right)^{3/2}+B, \tag{13}$$

where A and B are the variational parameters.[12] The test function (13) was chosen in such a way that for $B=0$ the function z_p^{-1} coincides with the Fourier transform of the wave function of the Wannier–Mott two-dimensional exciton ($v_p=0$, i.e., $z_p\to\infty$, corresponds to the semimetallic state).

The two-dimensional integrals over momentum for finding the polarization operators, energy, and Green's functions were calculated by the Monte Carlo method. Inside the square from 0 to p_F over both coordinates 2000 points were randomly chosen, the values of the functions at these points were summed, and the sum was divided by the area of the square in momentum space.

Minimization with respect to the variational parameters A and B was also performed by the Monte Carlo method. The energy of the system was calculated for randomly chosen values of A and B. The increments ΔA and ΔB in the parameters A and B were assigned. In the next step, the parameters A and B were assigned the new values $A+k\Delta A$ and $B+k\Delta B$, where the random number k was set equal to 0.1 or -1. The new values of the parameters A and B were accepted if the corresponding value of the energy was less than the value corresponding to the previous values of the parameters. Otherwise, the new values were rejected. The quantities ΔA and ΔB (periods of the grid in which the randomly varying parameters were chosen) were chosen to be equal to 10^{-3}–10^{-2} times the values of the corresponding parameters A and B.

By means of these variational calculations we evaluated the ground-state energy of system E_t for different values of the parameter $r_s=1/\sqrt{\pi n}$ and various distances D between the electron and hole layers.

For $D=0$ in the limit $r_s\to\infty$ (concentration $n\to 0$) the ground-state energy (in the calculation, per electron–hole pair) tends to the energy of the two-dimensional exciton. The authors of Ref. 15 calculated the dependence of the energy of one exciton in a system of spatially separated electron and holes on the distance between the planes. The results obtained in our study (see Fig. 2) in the limit of large r_s are in good agreement with the results of Ref. 15. The variationally calculated values of A and B indicate that the exciton phase is stable for the above-considered isotropic electron–hole system for all D.

A gap appears in the spectrum of new quasiparticles for all n and D, which grows with decreasing n and increasing D. As $n\to 0$, this gap becomes equal to the binding energy of the two-dimensional exciton. This means that the above-considered isotropic electron–hole system is an insulator (for all n and D).

To conclude this section, note that taking anisotropy into account leads to a more complicated phase diagram of the system. Specifically, in an anisotropic system at concentrations greater than some concentration that depends on the distance between the layers of electrons and holes, $n_{cr}(D)$, a transition from the exciton (insulating) phase to the metallic phase should take place. The latter phase, in contrast to he exciton phase, is no superfluid. A treatment within the framework of a model of Keldysh–Kopaev type[9] for an anisotropic, dense electron–hole system shows that the metal–insulator transition is continuous, while $n_{cr}(D)$ is a monotonically decreasing function and $n_{cr}(D)\propto 1/D$ for $D\gg a^*$.

5. LOW-DENSITY SYSTEM OF EXCITONS IN A TWO-LAYER STRUCTURE

At low electron and hole concentrations $na^2\ll 1$ and low temperatures, the system is a weakly non-ideal exciton gas[1)] with dipole moments in the ground state that are perpendicular to the layers $d=eD$ and grow with increasing interlayer distance D ($a(D)$ is the radius of the excitons in the layers; $a\approx a^*$ for $D\ll a^*$ and $a\approx a^{*1/4}D^{3/4}$ for $D\gg a^*$). In this case the approximation described in Sec. 1 must be replaced by an approximation that takes account to first order of the exchange interaction and (in the ladder approximation) the direct interaction between the two-dimensional excitons. In contrast to ordinary electron–hole systems,[9] in a spatially

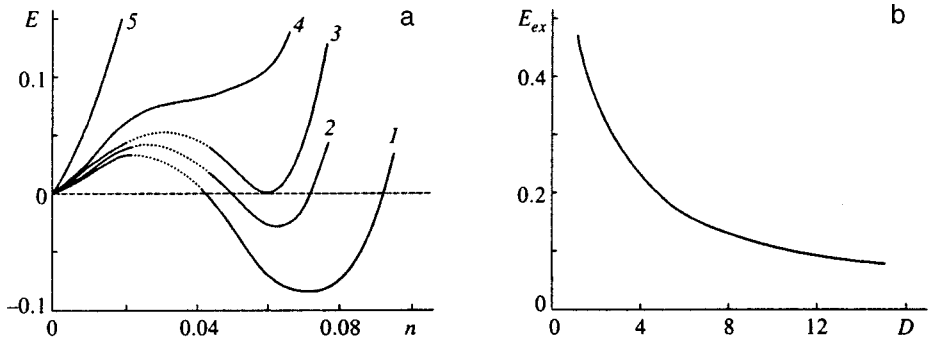

FIG. 2. a) Energy E of a system of spatially separated electrons and holes as a function of concentration; E in units of the binding energy of a two-dimensional exciton $\mathrm{Ry}_2^*=2me^4/\hbar^2\varepsilon$, n in units of $(a^*)^{-2}$, $a^*=\hbar^2\varepsilon/2me^2$. Curves *1–5* correspond to D equal to 0 (*1*), 0.5 (*2*), 1.1 (*3*), 1.9 (*4*), and 5 (*5*). The point $E=0$ for all five curves corresponds to the exciton energies $E_{\mathrm{ex}}(D)$ for the corresponding value of D (see Fig. 2b). b) The binding energy $E_{\mathrm{ex}}(D)$ of an exciton with spatially separated electrons and holes as a function of distance D between the layers; E_{ex} in units of Ry_2^*, D in units of a^*.

separated electron–hole system, calculations show that the main contribution to the energy comes from the direct dipole–dipole repulsion $U_d=e^2D^2/\epsilon r^3$, while the van der Waals attraction of the excitons and the exchange interaction are already negligible at $D\geqslant 0.5$—less than one-hundredth of the contribution of the direct dipole–dipole interaction (in a narrow region of small n the dipole–dipole repulsion leads to crystallization; see Sec. 6). The smallness of the exchange interaction in a spatially separated electron–hole system is related to the smallness of the tunneling exponential for penetration of quasiparticles through the dipole–dipole interaction barrier.

To calculate the exchange interaction energy, we take account of pairing of the electrons and holes by way of the canonical Bogolyubov transformation over the electron and hole operators, which is described by the unitary operator $\hat{S}$ (see Ref. 12):

$$\hat{S}=\exp\left[\sum_p \phi_p(a_p^+ b_{-p}^+ - b_{-p}a_p)\right], \tag{14}$$

$$\hat{S}a_p\hat{S}^+ = u_p a_p + v_p b_{-p}^+, \tag{15}$$

$$\hat{S}b_p\hat{S}^+ = u_p b_p + v_p a_{-p}^+, \tag{16}$$

where

$$u_p=\cos\,\phi_p, \quad v_p=\sin\,\phi_p, \quad u_p^2+v_p^2=1. \tag{17}$$

Using the anticommutation relations for the Fermi operators in the standard way, we obtain the transformed Hamiltonian

$$\hat{H}=\hat{H}_0+\hat{H}'+U, \tag{18}$$

where U is a numerical functional of u and v, and $\hat{H}_0$ and $\hat{H}'$ are respectively the quadratic Hamiltonian and the quadratic Hamiltonian in the new operators.

The function v_p obeys the normalization condition

$$\sum_p v_p^2=\frac{n}{2}, \tag{19}$$

which follows from the relations

$$\sum_{p,s}\langle \hat{S}a_{p,s}^+ a_{p,s}\hat{S}^+\rangle=\sum_{p,s}\langle \hat{S}b_{p,s}^+ b_{p,s}\hat{S}^+\rangle=2\sum_p v_p^2=n,$$

where n is the dimensionless surface charge concentration (na^2 in ordinary units).

In the expressions for $\hat{H}_0$ the functional of u and v standing in front of the combinations $a_p^+a_p$ and $b_p^+b_p$ is the energy spectrum of the transformed quasiparticles (quasielectrons and quasiholes):

$$E_p^{e,h}=(u_p^2-v_p^2)\left[\epsilon^{e,h}(p)-\mu_{e,h}-\sum_{p'} V_{p-p'}v_{p'}^2\right]$$

$$+2u_pv_p\sum_{p'}\tilde{V}_{p-p'}v_{p'}u_{p'}. \tag{20}$$

To calculate the exchange interaction energy, we can take advantage of the fact that k_p vanishes in pairwise combinations of operators $a_p^+b_{-p}^+ + b_{-p}a_p$ (which yield singular contributions to the energy) in the modified Hamiltonian $\hat{H}_0$ of (18) that takes electron–hole pairing into account. We neglect tunneling transitions between the bands of paired electrons and holes in different layers. It follows from the above condition $k_p=0$ that

$$\left[\epsilon(p)-\mu-2\sum_{p'} V_{p-p'}v_{p'}^2\right]u_pv_p$$

$$-(u_p^2-v_p^2)\sum_{p'}\tilde{V}_{p-p'}u_{p'}v_{p'}=0. \tag{21}$$

For a low-density system, u_p is of order unity, while v_p is small in the dimensionless parameter $\sqrt{n}$ (na^2 in ordinary units; see Eq. (19)). To lowest order in v_p, i.e., accurate to terms of order $\sqrt{n}$, Eq. (21) reduces to

$$[\epsilon(p)-\mu_0]v_p-\int \tilde{V}_{p-p'}v_{p'}\frac{d^2p'}{(2\pi)^2}=0. \tag{22}$$

Equation (22) together with the normalization condition (19) has the form

$$v_p=\sqrt{n}\psi_0(p), \quad \mu_0=-\epsilon_0, \tag{23}$$

where ϵ_0 and $\psi_0(p)$ are the binding energy and wave function of the ground state of an isolated exciton with spatially separated electrons and holes. The wave function $\psi_0(p)$ and energy ϵ_0 were calculated for various interlayer distances D in Ref. 15. The next order in v_p in Eq. (21) contains exchange interaction effects. The exchange correction μ_{ex} to the chemical potential can be found from Eq. (21) via ordinary perturbation theory if we substitute the zeroth approximation for v_p from Eq. (23) into the terms of order $n^{3/2}$,

dropped in Eq. (22), and treat them as a small perturbation. As a result, for the contribution of the exchange interaction to the chemical potential μ_{ex}, at low concentrations n and small interlayer distances D we obtain

$$\mu_{ex}=4.71ne^{-D}. \tag{24}$$

For large D the electron–hole interaction has the form

$$\tilde{V}(r)=-\frac{e^2}{\sqrt{D^2+r^2}}\approx-\frac{e^2}{D}+\frac{e^2r^2}{2D^3},$$

so that the ground state of the exciton can be described by the wave function of an oscillator with radius[15]

$$\rho=(4a^*)^{1/4}D^{3/4}. \tag{25}$$

As a result, for the contribution of the exchange interaction μ_{ex} at low n and large D we obtain

$$\mu_{ex}=11.31D^{17/4}n\ \exp\left(-\frac{D^{1/4}}{\sqrt{2}}\right). \tag{26}$$

The exchange interaction in a spatially separated system is suppressed in comparison with a two-dimensional electron–hole system in one layer. This has to do with the small tunneling exponential for penetration through the dipole–dipole interaction barrier.

The contribution of the dipole interaction can be represented by the sum of diagrams shown in Fig. 3. For small values of the momentum transfer $q\to0$, which are the most important at low concentrations, the integrand corresponding to the diagram in Fig. 3c is proportional to $1/q^3$, and in Fig. 3d to $1/q^2$. Consequently, the diagrams depicted in Fig. 3c make a larger contribution to the energy than the diagrams in Fig. 3d, and to account for the direct dipole–dipole interaction between excitons we may apply the theory of a two-dimensional Bose gas.[17] The relation between the vertex Γ and the two-dimensional scattering amplitude $f_0(\kappa)$ for an interaction potential of the form $U(r)=Ar^{-n}$ $(n>2)$ has the form

$$\Gamma=-2\,f_0(\kappa)\sqrt{\frac{2\pi\kappa}{i}}, \tag{27}$$

where the two-dimensional scattering amplitude is given by

$$f_0(\kappa)=\frac{\sqrt{\pi i/2\kappa}}{\ln(\kappa A^{1/(n-2)})}. \tag{28}$$

The contribution of the dipole interaction to the chemical potential is $\mu_d=\Gamma n=\kappa^2/2$.

We estimate the contribution of the dipole–dipole interaction to the chemical potential under the condition $\ln[(\kappa A)^{-1}]\gg1$ to be

$$\mu_d\approx\frac{4\pi n}{\ln(1/8\pi nA^{2/(n-2)})}, \tag{29}$$

where $n=3$ and $A=D^2$ for the dipole–dipole interaction.

The specifics of a two-dimensional Bose system include divergence of the exciton scattering amplitude at low energies (see Eq. (28)) and a corresponding deviation from the linear dependence of the chemical potential (29) on the concentration n at low n characteristic of the three-dimensional case; in the two-dimensional system the potential is proportional to $n/\ln n$ in the limit $n\to0$ (the specific region of applicability of the ladder approximation in the two-dimensional case is related to this[5]).

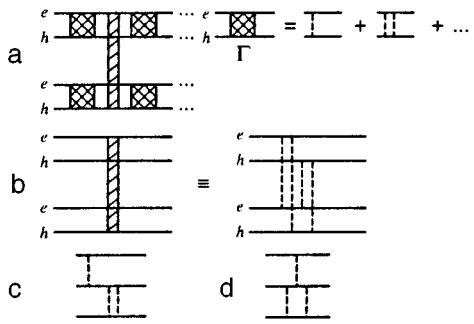

FIG. 3. a) Important diagrams for a rarefied exciton system; b) Ladder diagrams for one exciton; c) Interaction V between excitons ($V\sim D^2/r^3$ for $r\gg D$); d) Unimportant diagram in the limit $n\to0$.

In the estimate of the contribution of the van der Waals interaction to the chemical potential μ_W, the arguments are completely analogous to the above analysis of the dipole–dipole interaction. As a result, (29) turns out to be valid for the contribution of the van der Waals interaction, with $n=6$ and $A=C_6$—the coefficient of the van der Waals interaction. Let us find the coefficient C_6 for small and large D. For a two-dimensional system the coefficient C_6 is related to the polarizability α of the ground state:

$$C_6=-\frac{3}{2}\,\alpha\rho^2, \tag{30}$$

where ρ is the exciton radius.

The polarizability is

$$\alpha=-2e^2{\sum_k}'\ \frac{|x_{0k}|^2}{E_0-E_k}=\frac{2ime^2}{\hbar^2}\,(x\hat{b})_{00}, \tag{31}$$

where x is the coordinate along the film; E_k is the energy of the unperturbed levels, and $\hat{b}$ is an auxiliary operator (see Ref. 18):

$$\hat{z}=\frac{m}{\hbar}\frac{d\hat{b}}{dt}. \tag{32}$$

The function $f(r)$, introduced via the equality $\hat{b}\psi_0=f(r)\psi_0\cos\phi$, obeys the equation

$$ir=\frac{1}{2}\,f''+\frac{1}{2r}\,f'-\frac{1}{2r^2}\,f+\frac{\psi_0'}{\psi_0}\,f', \tag{33}$$

where ψ_0 is the ground-state wave function of the exciton. The unperturbed function ψ_0 for small D is calculated in first-order perturbation theory for $U(r)=2e^2D^2/\varepsilon r^3$ in the wave functions of the two-dimensional exciton (for $D=0$). For large D the ground state of the exciton is described by the wave function of an oscillator with radius given by (25).

Employing the solution of the differential equation for $f(r)$, we find α from Eq. (31). For small D we have

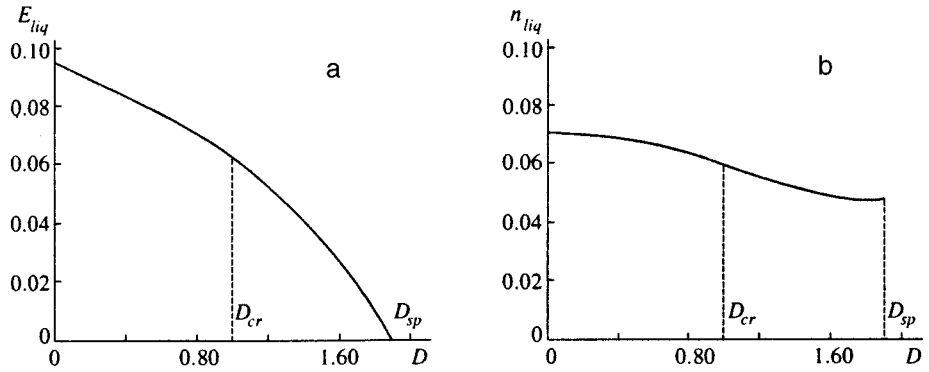

FIG. 4. a) Energy of an exciton liquid $E_{\rm liq}$ with spatially separated electrons and holes as a function of distance D between the layers; $E_{\rm liq}$ in units of Ry_2^*, D in units of a^*. For $D=D_{cr}=1.1$ a first-order transition in D occurs, for $D_{cr}<D<D_{sp}$ the exciton liquid is metastable, for $D=D_{sp}=1.9$ the exciton liquid becomes unstable. b) Equilibrium concentration $n_{\rm liq}$ of an exciton liquid (in units of $(a^*)^{-1}$) as a function of distance D between the layers (in units of a^*).

$$\alpha=\frac{21}{16}\,a^{*3}\left(1+\frac{64\pi^2m^2e^4D^4}{\hbar^4a^{*2}}\right), \tag{34}$$

and for large D

$$\alpha=0.93D^{9/4}a^{*3/4}. \tag{35}$$

Substituting (34) and (35) into (30) and (29), we find the contribution of the van der Waals interaction to the chemical potential.

The chemical potential of the system $\mu=dE/dn$ is obtained as the sum of the above-considered contributions of the exchange, dipole–dipole, and van der Waals interactions of the excitons:

$$\mu=\mu_{\rm ex}+\mu_d+\mu_W. \tag{36}$$

Employing the chemical potential (36) and Eqs. (24), (26), (29), (34), and (35), we calculate the velocity c of acoustic excitations in the system $c=\sqrt{(n/m)d\mu/dn}$ as a function of n and D.

To calculate the ground-state energy $E(n)$ over the entire range of equilibrium concentrations n, we combined the results of the calculations at high and intermediate concentrations (see Sec. 4) with the results obtained for low concentrations. Figure 4 shows that an exciton liquid can exist for $D<D_{sp}=1.9$ while for $D\geqslant D_{sp}$ the liquid phase is absent.

Calculated values of the energy $E(n)$ that allow for the contribution of the dipole–dipole interaction and the van der Waals and exchange interactions are plotted in Fig. 2 (see the range of low n, where $dE/dn>0$ by virtue of the dipole–dipole repulsion of the excitons).

6. GAS–LIQUID QUANTUM TRANSITION IN A TWO-LAYER EXCITON SYSTEM

The calculated curves $E(n)$ (see Fig. 2) have a minimum whose depth corresponds to the binding energy of the exciton liquid. The binding energy decreases with increasing D, the minimum becomes shallower, and the equilibrium concentrations $n_{\rm liq}(D)$ decrease. For $D<D_{cr}=1.1$ the energy of the liquid phase is greater than the energy $E(0)$ of an isolated exciton with spatially separated electrons and holes, i.e., in this range of D the stable phase is the exciton liquid, and the metastable phase is the exciton gas.

The binding energy of an isolated biexciton with spatially separated electrons and holes[8] (decreasing with increasing D) is less than the binding energy of the exciton liquid, i.e., the liquid phase in the system under consideration is also stable against decay into isolated biexcitons. For $D=D_{cr}\approx 1.1$ the binding energy of the exciton liquid becomes comparable to the energy of an isolated exciton, and at this point the discontinuous quantum phase transition ''quantum liquid–exciton gas'' takes place with an increase or decrease in D. However, over the narrow range $D_{cr}<D<D_{sp}$ ($D_{sp}=1.9$, $D_{cr}=1.1$) the minimum corresponding to the exciton liquid phase still exists, i.e., the liquid phase remains metastable. At larger D ($D>1.9$) the liquid phase is absolutely unstable and only the gas phase is stable, with the exciton density being determined only by external conditions. In the latter case all regimes are possible—from a low-density Bose gas to a dense exciton phase of BCS type.

We emphasize that the presence of a quantum transition with respect to the parameter D and the instability of the liquid for $D>D_{sp}$ are a specific feature of excitons with spatially separated electrons and holes associated with their dipole–dipole repulsion at large distances.[2)]

The formation of a liquid exciton phase can be detected by a photoluminescence line shift or by discontinuous variation of the exciton diffusion coefficient. The onset of the liquid phase is also characterized by the formation of liquid exciton droplets that can be detected experimentally from the giant fluctuations, corresponding to the motion of the droplets, of the photoluminescence by means of local observation of the photoluminescence using an opaque mask with pinholes or with the help of an optical fiber. It should also be possible to observe exciton droplets from large fluctuations of the current along the layers in a system of two quantum wells. Interlevel resistance due to the entrainment of electrons and holes may also be a sensitive indicator of a gas–liquid phase transition and a transition to the superfluid and other phases of an electron–hole system.

7. TRANSITION TO THE SUPERFLUID STATE AND SUPERFLUID PHASES OF A TWO-LAYER EXCITON SYSTEM

As was shown above for $T=0$, the system under consideration has a gap in its energy spectrum. For this reason, at

$T=0$ the system is a superfluid. With increasing temperature the superfluidity of the excitons in the quasi-two-dimensional system under consideration is preserved up to the temperature T_c of the Kosterlitz–Thouless topological phase transition,[19] at which depairing of the vortex pairs takes place in the exciton system. This temperature is given by

$$T_c=\frac{0.45\pi\hbar^2 n_s}{k_B m}, \tag{37}$$

where n_s is the concentration of the superfluid component, k_B is the Boltzmann constant, and m is the reduced mass of the quasiparticle.

The concentration n_s depends closely on temperature, and the value of n_s that enters into Eq. (37) is its value at the Kosterlitz–Thouless transition temperature T_c, i.e., Eq. (37) is an equation for finding T_c.

The function $n_s(T)$ can be found from the relations $n_s=n-n_n$ (n and n_n are the total density and the density of the normal component); $n_n=j_n/v_n$ (j_n and v_n are the current and velocity of the normal component). Using these relations, the superfluid density $n_s(T)$ can be expressed in terms of the electron and hole concentrations $n_e=n_h=n$ and the interaction parameters in the electron–hole system. Let all the quasiparticles move with velocity u. In this case the argument of the distribution function will be not $\epsilon(p)$, but $\epsilon(p)-pu$. Motion of the quasiparticles will lead to a momentum

$$P=2\int pn_F(\epsilon-pu)\,\frac{d^2p}{(2\pi\hbar)^2} \tag{38}$$

(n_F is the Fermi function). Taking the velocity u to be low and expanding in u, it is possible to find the coefficient n_n in the expression $P=n_n mu$, which is the concentration of the normal component.

To start with, we consider high and intermediate densities. As our estimate of the contribution of single-particle excitations at high and intermediate densities, we use the relation between $n_s(T)$ and T_c^0 in the BCS approximation[20] for $T_c^0-T_c\ll T_c^0$:

$$\frac{n_s}{n}=\frac{2(T_c^0-T)}{T_c^0}, \tag{39}$$

and the relation

$$\Delta(0)=1.76T_c^0. \tag{40}$$

For $\Delta(0)$ we substitute the value of the gap at zero temperature obtained by variational calculation.

It is also possible to take account of the contribution of collective excitations to the concentration of the normal component. In contrast to superconductors, where, as a consequence of the charge of the Cooper pairs, instead of an acoustic spectrum of collective oscillations a plasma branch of oscillations arises, in the exciton phase the pairs are neutral and the acoustic branch exists. At low temperatures the contribution of the elementary excitations in thermodynamic equilibrium can be described in the approximation of an ideal Bose gas. In this case the Bose distribution function must be substituted in the integral (38) in place of the Fermi distribution function. We then have for n_s

$$n_s=n-n\,\frac{2T-T_c^0}{T_c^0}-\frac{\pi^2T^3}{30\hbar^2c^4m}-\frac{2mT\pi^2}{3(2\pi\hbar)^2}, \tag{41}$$

where $c=\sqrt{(n/m)d\mu/dn}$ is the speed of sound and μ is the chemical potential of the system determined for intermediate densities by variational calculation (see Sec. 4).

For a low-density exciton gas the overall approach is analogous to that at intermediate densities, but we take account of the contribution to the normal density only from collective acoustic oscillations. Then

$$n_s=n-\frac{\pi^2T^3}{30\hbar^2c^4m}-\frac{2mT\pi^2}{3(2\pi\hbar)^2}, \tag{42}$$

where c is the speed of sound for a low-density exciton gas, and the chemical potential of the system μ (in contrast to the previous case of intermediate densities) is defined as the sum (36) of contributions of the exchange, dipole–dipole, and van der Waals interactions of the excitons.

As a result, substituting (41) or (42) into (37), it is possible to estimate the temperature of the Kosterlitz–Thouless phase transition (see Table I). The magnitude of T_c decreases appreciably as the distance between the wells D increases and the exciton concentration decreases. The transition to the superfluid state for $D<D_{cr}\approx 1.1$ in the metastable gas phase, and in the stable gas phase for $D>D_{cr}$, will depend on the concentration, which is determined by external conditions and given by the formula derived above for the superfluid density. For $T=T_c$ the global superfluid density disappears with a (universal[19]) jump, and at somewhat higher temperatures the fluctuational local superfluid density $n_s(L)$, determined by the renormalizations at all lower intermediate scales L, disappears.

The local superfluid density above T_c can be manifested, for example, in observations of exciton diffusion to intermediate distances (with the help of local measurements of exciton photoluminescence at two points using optical fibers or pinholes (in experiments like those in Ref. 3)). Discontinuous growth of the coefficient of mutual Coulomb entrainment of the electrons and holes in a two-layer system should correspond to the appearance of a global superfluid density at $T=T_c$. In contrast to ordinary superconductors, nonzero (and equal) electric fields are created in the superfluid phase by the entrainment of particles of one layer by the other. The superfluid state at $T<T_c$ is manifested in the existence of non-decaying (''superconducting'') oppositely directed electric currents in each layer. Taking tunneling into account in an equilibrium electron–hole system (in a second-order double quantum well) leads to interesting Josephson phenomena in the system: a transverse Josephson current, inhomogeneous longitudinal currents,[6] diamagnetism in a magnetic field H parallel to the junction (where this magnetic field is less than some critical value H_{c1} depending on the tunneling coefficient), and a mixed state with Josephson vortices for $H>H_{c1}$ (Ref. 8) (in addition, taking tunneling into

TABLE I.

D	$-E_{ex}$ [Ry$_2^*$]	$-E_t$ [Ry$_2^*$]	E_b [Ry$_2^*$]	r_s^{min}	$k_BT_c^0\times10^{-3}$ [Ry$_2^*$]	$k_BT_c\times10^{-3}$ [Ry$_2^*$]
0	1.0	1.06	0.06	2.2	1.7	1.3
0.5	0.80	0.84	0.04	2.6	1.3	1.1
1.0	0.50	0.51	0.01	3.2	0.8	0.7
5.0	0.26	-	-	-	-	0.2

Note. The dependence on the interlayer distance D is given here in tabular form for the energy of an isolated exciton E_{ex}, the total energy E_t per electron–hole pair, the binding energy $E_b = E_{ex} - E_t$ of the liquid phase per pair, the equilibrium distance between the excitons in the ground state r_s^{min} ($r_s^{min} = 1/a^*\sqrt{\pi n_{min}}$; $a^* = \hbar\varepsilon/2me^2$; n_{min} is the equilibrium density), and the Kosterlitz–Thouless (T_c) and BCS (T_c^0) critical temperatures. Regarding notation, m is the reduced mass and ε is the dielectric constant. The values of T_c and T_c^0 in the first three rows correspond to the equilibrium concentrations of the liquid phase at the given distances D; the fourth row corresponds to the fixed value $r_s = 100.0$ (the liquid phase does not exist at $D \geqslant D_{sp} \simeq 1.9$); Ry$_2^* = 2me^2/\hbar\varepsilon$.

account leads to a loss of symmetry of the order parameter and to a change in the character of the phase transition).

ACKNOWLEDGMENTS

This work was supported by grants from the Russian Foundation for Basic Research, the program ''Solid-State Nanostructures,'' and the International Science Foundation.

One of the authors (Yu. E. L.) is grateful to A. H. MacDonald, U. Sivan, E. Burstein, and the participants of the Conference on Low-Dimensional Systems (Trieste) for their interest in this work and for helpful discussions of the results of this work. Yu. E. L. thanks Yu. Lu and E. Tosatti for their hospitality at the International Center for Theoretical Physics (Trieste), where part of this work was carried out.

*)e-mail: lozovik@isan.msk.su

1)As was first pointed out in Ref. 5, in a narrow $n-T$ region, dipole excitons form a crystalline phase (see also Ref. 16 and the references therein).

2)Similarly, a liquid–gas quantum transition is possible in a two-dimensional exciton system with variation of the electric field.[18]

[1] U. Sivan, P. M. Solomon, and H. Strikman, Phys. Rev. Lett. **68**, 1196 (1992).

[2] T. Fuzuzawa, E. E. Mendez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990); J.-P. Cheng, J. Kono, B. D. McCombe *et al.*, Phys. Rev. Lett. **74**, 450 (1995).

[3] L. V. Butov, A. Zrenner, G. Abstreiter *et al.*, Phys. Rev. Lett. **73**, 304 (1994); A. Zrenner, L. V. Butov, M. Hang, G. Abstreiter, G. Bohm, and G. Weigmann, Phys. Rev. Lett. **72**, 3382 (1994).

[4] M. Bayer, V. B. Timofeev, F. Faller *et al.*, Phys. Rev. B **54**, 8799 (1996).

[5] Yu. E. Lozovik and V. I. Yudson, JETP Lett. **22**, 11 (1975); Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)]; Yu. E. Lozovik and V. I. Yudson, Solid State Commun. **18**, 628 (1976); **21**, 211 (1977); Physica A **93**, 493 (1978); Yu. E. Lozovik, *Abstracts of the First All-Union Conference on Insulator Electronics*, Tashkent (1973), p. 59.

[6] A. V. Klyuchnik and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **76**, 670 (1979) [Sov. Phys. JETP **49**, 355 (1979)]; Yu. E. Lozovik and A. V. Klyuchnik, J. Low Temp. Phys. **38**, 761 (1980); J. Phys. C **11**, L483 (1978); Yu. E. Lozovik and V. I. Yudson, JETP Lett. **25**, 14 (1977); I. O. Kulik and S. I. Shevchenko, Solid State Commun. **21**, 409 (1977); Yu. E. Lozovik and V. I. Yudson, Solid State Commun. **22**, 117 (1977).

[7] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)]; **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)]; **82**, 1188 (1982) [Sov. Phys. JETP **55**, 691 (1982)]; A. B. Dzyubenko and Yu. E. Lozovik, Fiz. Tverd. Tela **25**, 1519 (1983) [Sov. Phys. Solid State **25**, 874 (1983)]; **26**, 1540 (1984) [Sov. Phys. Solid State **26**, 938 (1984)]; A. B. Dzyubenko and Yu. E. Lozovik, J. Phys. A **24**, 415 (1991); D. Paquet, T. M. Rice, and K. Ueda, Phys. Rev. B **32**, 5208 (1985); S. M. Dikman and S. V. Iordanskiĭ, JETP Lett. **63**, (1996).

[8] Yu. E. Lozovik, *Abstracts of the Adriatic Conference on Electron Quantum Liquids in Systems of Reduced Dimensions*, International Center for Theoretical Physics, Trieste (1996), p. 51.

[9] L. V. Keldysh and A. N. Kozlov, Zh. Éksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys. JETP **27**, 521 (1968)]; L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela **6**, 2791 (1964) [Sov. Phys. Solid State **6**, 2219 (1964); A. N. Kozlov and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **48**, 1184 (1965) [Sov. Phys. JETP **21**, 799 (1965)]; L. V. Keldysh and A. P. Silin, Brief Reports of the Physical Institute of the Academy of Sciences [FIAN], No. 8, 33 (1975); B. I. Halperin and T. M. Rice, Solid State Phys. **21**, 115 (1968); Yu. E. Lozovik and V. I. Yudson, Fiz. Tverd. Tela **17**, 1613 (1975) [Sov. Phys. Solid State **17**, 1054 (1975)].

[10] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dayaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood Cliffs, N.J. (1963).

[11] J. R. Schrieffer, Theory of Superconductivity, Benjamin, New York (1964).

[12] A. V. Klyuchnik, Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **20**, 625 (1978) [Sov. Phys. Solid State **20**, 364 (1978)].

[13] D. Pines and P. Nozières, Theory of Quantum Liquids, Benjamin, New York (1966).

[14] W. F. Brinkman and T. M. Rice, Phys. Rev. B **7**, 1508 (1973).

[15] Yu. E. Lozovik and V. N. Nishanov, Fiz. Tverd. Tela (Leningrad) **18**, 3267 (1976) [Sov. Phys. Solid State **18**, 1905 (1976)].

[16] Yu. E. Lozovik and O. L. Berman, JETP Lett. **64**, (1996); Yu. E. Lozovik and O. L. Berman, Physica Scripta (in print) (1996).

[17] Yu. E. Lozovik and V. I. Yudson, Physica A **93**, 493 (1978).

[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).

[19] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); D. R.Nelson and J. M. Kosterlitz, Phys. Rev. Lett. **39**, 1201 (1977).

[20] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, North-Holland, Amsterdam (1988).

Translated by Paul F. Schippnick

Density of states near the Anderson transition in a space of dimensionality $d=4-\epsilon$

I. M. Suslov

P.L. Kapitsa Institute of Physical Problems, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 5 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 1896–1914 (May 1997)

Asymptotically accurate results have been obtained for the average Green's function and the density of states in a Gaussian random potential for dimensionality of space $d=4-\epsilon$ over the entire energy region, including the vicinity of the mobility threshold. For $N\sim 1$ (N is the order of the perturbation theory) only parquet terms corresponding to higher terms in $1/\epsilon$ are taken into account. For large N all powers of $1/\epsilon$ are taken into account with their coefficients calculated in the main asymptotic limit in N. This calculation is performed by combining the condition of renormalization theory with the Lipatov asymptotic limit. © *1997 American Institute of Physics.* [S1063-7761(97)02605-X]

1. INTRODUCTION

According to generally accepted thinking,[1,2] the single-electron density of states does not have a singularity at the Anderson transition, in contrast to the conductivity and the localization radius of the wave functions.[3–6] Nevertheless, its calculation is of fundamental significance since all known methods break down in the vicinity of the transition. In addition, the density of states and the conductivity, defined respectively by the average Green's function $\langle G(x,x')\rangle$ and the correlator $\langle G^R G^A\rangle$, are not completely independent. A study in the parquet approximation shows[7] that the mathematical difficulties in both cases are of the same nature and are connected with the ''ghost'' pole problem. On the other hand, to satisfy the Ward identity linking the eigenenergy part with the irreducible vertex in the Bethe–Salpeter equation[8] would require exact agreement of the diagrams taken into account in the calculation of the conductivity and density of states; this circumstance is not dealt with in any of the presently existing theories[7] with the exception of the theory recently proposed in Ref. 9.

For weak disorder the mobility threshold lies in the vicinity of the starting boundary of the spectrum, at which the random potential can be taken to be Gaussian by virtue of the possibility of averaging over scales that are small in comparison with the wavelength of the electron, but large in comparison with the distance between scatterers (the so-called Gaussian segment of the spectrum[10]). Calculation of the average Green's function for the Schrödinger equation with Gaussian random potential reduces to the problem of a second-order phase transition with an n-component order parameter $\varphi=(\varphi_1,\varphi_2,\ldots,\varphi_n)$ in the limit $n\to 0$.[11,12] In this case the coefficients in the Ginzburg–Landau Hamiltonian

$$H\{\varphi\}=\int d^dx\left(\frac{1}{2}\,c|\nabla\varphi|^2+\frac{1}{2}\,\kappa_0^2|\varphi|^2+\frac{1}{4}\,u|\varphi|^4\right) \qquad (1)$$

are linked with the parameters of the disordered system by the relations

$$c_0=1/2m,\quad \kappa_0^2=-E,\quad u=-a_0^dW^2/2, \qquad (2)$$

where d is the dimensionality of the space, m and E are the mass and energy of the particle, a_0 is the lattice constant, and W is the amplitude of the random potential (in what follows we take $c_0=1$). The ''incorrect'' sign of the coefficient of $|\varphi|^4$ leads to the inapplicability of the usual mean-field theory and the necessity of a fluctuational treatment[11,13] over the entire parameter space; the functional integrals for $u<0$ are understood in the sense of an analytic continuation from positive u, which for a retarding Green's function is carried out through the lower half-plane.[12]

The present paper completes the program of constructing a $(4-\epsilon)$-expansion initiated in Refs. 14–16. The dimensionality of the space $d=4$ is singled out for the Hamiltonian (1) from considerations of renormalizability: for $d>4$ the theory is not renormalizable and the discreteness of the lattice is of fundamental significance, ensuring the existence of a cutoff parameter $\Lambda\sim a_0^{-1}$ at high momenta[14]; for $d=4$ a logarithmic situation holds sway, admitting the existence of both non-renormalizable[15] and renormalizable models[16]; for $d<4$ the theory is renormalizable with the help of one subtraction, and passage to the continuum limit $a_0\to 0$, $a_0^dW^2\to$ const is possible. The use of simplifications arising at high dimensionalities to construct a $(4-\epsilon)$-dimensional theory requires the successive consideration of all four types of theories; this was done in Refs. 14–16 and in the present work. The results of this work have already been published in a brief exposition in Ref. 17.

2. STRUCTURE OF THE APPROXIMATION

The calculation of the average Green's function $\langle G(p,\kappa)\rangle$ (p is the momentum and κ is the renormalized value of κ_0) reduces in the standard way to a calculation of the eigenenergy $\Sigma(p,\kappa)$, for which the structure of the perturbation-theory series in four-dimensional space at $p=0$ has the form[15]

$$\Sigma(0,\kappa)-\Sigma(0,0)=\kappa^2\sum_{N=1}^{\infty}u^N\sum_{K=0}^{N}A_N^K\left(\ln\frac{\Lambda}{\kappa}\right)^K. \qquad (3)$$

Reference 16 established the structure of the approximation which allows one to obtain asymptotically accurate results (in the limit of weak disorder) for a renormalizable class of models, this being the zeroth approximation for the $(4-\epsilon)$-dimensional theory. For $N\sim 1$ it is sufficient to take

account of the coefficients A_N^N corresponding to the ''leading logarithms;'' for large N this approximation is insufficient in light of the higher rate of growth with respect to N of the coefficients of the lower-order logarithms: therefore it is generally speaking necessary to take account of all the coefficients A_N^K, but it suffices to calculate them in the leading asymptotic limit in N. The latter is possible by combining the condition of renormalizability of the theory with the Lipatov asymptotic limit.[18]

The sum of the high-order terms of the perturbation-theory series gives a nontrivial contribution associated with the divergence of the series and is important only for negative u; this latter result explains why in the usual theory of phase transitions it is possible to restrict the calculation to the leading logarithmic approximation.[19,20]

For $d=4-\epsilon$ the expansion analogous to (3) has the form

$$\kappa^2+\Sigma(0,\kappa)-\Sigma(0,0)\equiv\kappa^2 Y(\kappa) = \kappa^2\sum_{N=0}^{\infty}(u\Lambda^{-\epsilon})^N\sum_{K=0}^{N}A_N^K(\epsilon) \times\left[\frac{(\Lambda/\kappa)^\epsilon-1}{\epsilon}\right]^K, \tag{4}$$

where the coefficients $A_N^K(\epsilon)$ are finite in the limit $\epsilon\to 0$ and $A_0^0(\epsilon)\equiv 1$. Expansion (4) follows from the fact that the quantity Y in Nth-order perturbation theory is a homogeneous polynomial of degree N built up from $\Lambda^{-\epsilon}$ and $\kappa^{-\epsilon}$: indeed, in the transition from the Nth-order diagram to the $(N+1)$-th–order diagram the dimensionality in the momentum decreases by ϵ (Ref. 21), which gives the factor $\Lambda^{-\epsilon}$ or $\kappa^{-\epsilon}$ depending on whether high or low momenta determine the corresponding contribution. Separating out the factor ϵ^{-K} ensures the correct limit in expansion (3) as $\epsilon\to 0$.

The standard procedure for carrying out the ϵ-expansion[11,13] consists in expanding the coefficients $A_N^K(\epsilon)$ in powers of ϵ

$$A_N^K(\epsilon)=\sum_{L=0}^{\infty}A_N^{K,L}\epsilon^L \tag{5}$$

and preserving in each order of the perturbation theory some of the higher orders in $1/\epsilon$; the first ϵ-approximation corresponds to taking account of only the coefficients $A_N^{N,0}$, which coincide with the coefficients of the leading logarithms in expansion (3). As is the case for $d=4$, such an approximation is insufficient for $u<0$ due to the higher rate of growth with N of the coefficients of the lower terms in $1/\epsilon$: limiting the expansion to the coefficients $A_N^{N,0}$ is possible only for $N\sim 1$, whereas for larger N it is necessary to take into account all the coefficients $A_N^{K,L}$, calculating them in the leading asymptotic limit in N.

According to Eq. (4) the quantity Y is a function of $g_0\equiv u\Lambda^{-\epsilon}$ and Λ/κ; it satisfies the Callan–Symanzik equation

$$\left(\frac{\partial}{\partial\ln\Lambda}+W(g_0,\epsilon)\frac{\partial}{\partial g_0}+V(g_0,\epsilon)\right)Y=0, \tag{6}$$

which expresses the condition of renormalizability of the theory, and Eq. (15) of Ref. 16, which was obtained in an analogous way. The functions $W(g_0,\epsilon)$ and $V(g_0,\epsilon)$ can be expanded in the following series:

$$W(g_0,\epsilon)=\sum_{M=1}^{\infty}W_M(\epsilon)g_0^M=\sum_{M=1}^{\infty}\sum_{M'=0}^{\infty}W_{M,M'}g_0^M\epsilon^{M'},$$
$$V(g_0,\epsilon)=\sum_{M=1}^{\infty}V_M(\epsilon)g_0^M=\sum_{M=1}^{\infty}\sum_{M'=0}^{\infty}V_{M,M'}g_0^M\epsilon^{M'}, \tag{7}$$

whose first coefficients were calculated in Ref. 21:

$$W_1(\epsilon)=-\epsilon,\quad W_{2,0}=K_4(n+8),$$
$$W_{3,0}=-3K_4^2(3n+14),\quad V_{1,0}=-K_4(n+2) \tag{8}$$

(according to Ref. 16 the function $V(g_0,\epsilon)$ coincides with the function $\eta_2(g_0,\epsilon)$ introduced in Ref. 21); the quantity K_4 is defined in Eqs. (14). Substituting expansions (4) and (7) into Eq. (6) leads to a system of equations for the coefficients $A_N^K(\epsilon)$:

$$(K+1)A_N^{K+1}(\epsilon)=(N-K)\epsilon A_N^K(\epsilon) -\sum_{M=1}^{N-K}[(N-M)W_{M+1}(\epsilon) +V_M(\epsilon)]A_{N-M}^K(\epsilon), \tag{9}$$

or for the coefficients $A_N^{K,L}$:

$$(K+1)A_N^{K+1,L}=(N-K)A_N^{K,L-1}(1-\delta_{L,0}) -\sum_{M=1}^{N-K}\sum_{M'=0}^{L}[(N-M)W_{M+1,M'} +V_{M,M'}]A_{N-M}^{K,L-M'}. \tag{10}$$

Wilson's method[11,13] is based on the fact that in the nth ϵ-approximation one needs to know the coefficients $A_N^{N-K,L}$ for $K+L\leqslant n-1$, for which Eqs. (10) yield the closed system of difference equations

$$-Nx_N=[W_{2,0}(N-1)+V_{1,0}]x_{N-1},$$
$$-(N-1)y_N=[W_{2,0}(N-1)+V_{1,0}]y_{N-1}+[W_{3,0}(N-2) +V_{2,0}]x_{N-2},$$
$$-Nz_N=[W_{2,0}(N-1)+V_{1,0}]z_{N-1}+[W_{2,1}(N-1) +V_{1,1}]x_{N-1}-y_N, \tag{11}$$

(where $x_N\equiv A_N^{N,0}$, $y_N\equiv A_N^{N-1,0}$, $z_N\equiv A_N^{N,1}$, ...), which is solvable by the method of variation of parameters;[22] assigning the initial conditions and determining the quantities $W_{2,0}$, $V_{1,0}$, ... requires the calculation of some lower orders of the perturbation theory. In particular, for the coefficients $A_N^{N,0}$ we easily obtain

$$A_N^{N,0}=(-W_{2,0})^N\frac{\Gamma(N-\beta_0)}{\Gamma(N+1)\Gamma(-\beta_0)},$$
$$\beta_0=-\frac{V_{1,0}}{W_{2,0}}=\frac{n+2}{n+8}. \tag{12}$$

To investigate the higher orders in ϵ, the Wilson method turns out to be ineffective, and it is more convenient to start with (9). Information about the coefficients $A_N^K(\epsilon)$ for $N \gg 1$ can be obtained by the Lipatov method,[18] according to which the later coefficients of the expansion in u of the functional integrals with the Hamiltonian (1) are determined by the saddle-point configurations—instantons—and have factorial growth in N. For factorial series there exists a simple algebra that enables one to manipulate them as simply as finite expressions,[15] which in turn enables one to find the expansion coefficients of arbitrary M-point Green's functions, proceed from them to the eigenenergy and the vertex parts, etc. According to Sec. 6, the Nth coefficient of the expansion of $\Sigma(p,\kappa)$ in powers of u has the form

$$[\Sigma(p,\kappa)]_N = c_2\Gamma(N+b) \times a^N \int_0^\infty d\ \ln R^2 R^{-2} \langle \phi_c^3 \rangle_{Rp} \langle \phi_c^3 \rangle_{-Rp} \times \exp\left(-Nf(\kappa R) + N\epsilon \ln R + 2K_d I_4(\kappa R)\ \frac{1-(\Lambda R)^{-\epsilon}}{\epsilon} \right), \tag{13}$$

where

$$a = -3K_4, \quad b = \frac{d+2}{2}, \quad c_2 = c(3K_4)^{7/2},$$

$$f(x) = -\frac{\epsilon}{2}\,(C+2+\ln\,\pi) - 3x^2\left(C + \frac{1}{2} + \ln\frac{x}{2} \right),$$

$$\langle \phi_c^3 \rangle_p = 8\sqrt{2}\pi^2 p K_1(p),$$

$$I_4(x) = \widetilde{I}_4\ \exp(f(x)), \quad \widetilde{I}_4 = \frac{16}{3}\,S_4,$$

$$S_d = 2\pi^{d/2}/\Gamma(d/2), \quad K_d = S_d(2\pi)^{-d}, \tag{14}$$

C is Euler's constant, $K_1(x)$ is the modified Bessel function, and the constant c is defined below in Sec. VI. Re-expanding series (4)

$$\kappa^2 + \Sigma(0,\kappa) - \Sigma(0,0) = \kappa^2 \sum_{N=0}^{\infty} (u\kappa^{-\epsilon})^N \sum_{K=0}^{N} B_N^K(\epsilon) \times \left[\frac{1-(\Lambda/\kappa)^{-\epsilon}}{\epsilon} \right]^K, \tag{15}$$

in such a way that the coefficients $B_N^K(\epsilon)$ are related to the coefficients $A_N^K(\epsilon)$ by

$$A_N^K(\epsilon) = \sum_{K'=0}^{K} C_{N-K'}^{N-K} B_N^{K'}(\epsilon)\epsilon^{K-K'}, \tag{16}$$

setting $p=0$ in Eq. (13), making the substitution $R \to R/\kappa$, and transforming the exponential

$$\exp\left\{ 2K_d I_4(R)\ \frac{1-(\Lambda R/\kappa)^{-\epsilon}}{\epsilon} \right\} = \exp\left\{ 2K_d I_4(R)\ \frac{1-R^{-\epsilon}}{\epsilon} \right\} \times \sum_{K=0}^{\infty} \frac{\{2K_d I_4(R) R^{-\epsilon}\}^K}{K!} \left[\frac{1-(\Lambda/\kappa)^{-\epsilon}}{\epsilon} \right]^K, \tag{17}$$

we obtain for the coefficients $B_N^K(\epsilon)$ at large N

$$B_N^K(\epsilon) = \widetilde{c}_2\Gamma(N+b)a^N\,\frac{1}{K!}\int_0^\infty d\ \ln R^2 R^{-2} \times (2K_d I_4(R) R^{-\epsilon})^K \times \exp\left(-Nf(R) + N\epsilon \ln R + 2K_d I_4(R)\ \frac{1-R^{-\epsilon}}{\epsilon} \right), \tag{18}$$

where $\widetilde{c}_2 = c_2\langle \phi_c^3 \rangle_0^2$. By analogy with the case $d=4$ (Ref. 16), the Lipatov method reproduces the coefficients $B_N^K(\epsilon)$ well only for $K \ll N$, which is connected with their rapid falloff with K and the limited accuracy ($\sim 1/N$) of the leading asymptotic behavior. Substituting (18) into Eq. (16), we obtain the following result for the coefficients $A_N^K(\epsilon)$ with $N \gg 1$:

$$A_N^K(\epsilon) = \widetilde{c}_2\Gamma(N+b)a^N C_N^K \int_0^\infty d\ \ln R^2 R^{-2} \times \left(\epsilon + \frac{2K_d\widetilde{I}_4}{N}\ e^{f(R)-\epsilon\ln R} \right)^K \times \exp\left(-Nf(R) + N\epsilon \ln R + 2K_d I_4(R)\ \frac{1-R^{-\epsilon}}{\epsilon} \right), \tag{19}$$

which follows from Eq. (18) under the condition that the sum in Eq. (16) is determined by values of $K' \ll N$. Retaining only the term with $M=1$ in the sum (9), it is easy to convince oneself that the equation so obtained is satisfied by the result (19) for $K \ll N$ in the case $N\epsilon \lesssim 1$ and for all K for $N\epsilon \gg 1$. The latter has to do with the fact that for $N\epsilon \gg 1$, the sum in Eq. (16) is determined by values of $K' \sim K/\epsilon N \ll N$ for all K in the region of applicability of formula (18). The indicated reduction of Eq. (9) is possible at large N by virtue of the factorial growth of $A_N^K(\epsilon)$ under the assumption that $W_N(\epsilon)$ and $V_N(\epsilon)$ grow more slowly than $A_N^0(\epsilon)$. This latter result can be assumed to be a consequence of the validity of formula (19) for $K=0, 1, 2$ (see Ref. 16 for a more detailed exposition).

The system of equations (9) determines the coefficients $A_N^K(\epsilon)$ with $K>0$ for prescribed $A_N^0(\epsilon)$. Since Eq. (19) is valid for the latter for all $N \gg 1$, it can be used as a boundary condition on system of equations (9), which enables one to determine all the $A_N^K(\epsilon)$ with large N. Thus, retaining only the leading order in $1/\epsilon$ for $N \sim 1$, defined by the coefficients (12), it is not hard to find the sum of series (4).

3. STUDY OF THE COEFFICIENTS $A_N^K(\epsilon)$

We will limit the sum (9) to terms with $M=1$ and $M=2$:

$$KA_N^K(\epsilon)=(N-K+1)\epsilon A_N^K(\epsilon)-W_2(\epsilon)[N-1-\beta(\epsilon)]\times A_{N-1}^{K-1}(\epsilon)-W_3(\epsilon)NA_{N-2}^{K-1}(\epsilon), \tag{20}$$

Here

$$\beta(\epsilon)=-\frac{V_1(\epsilon)}{W_2(\epsilon)} \xrightarrow{\epsilon\to 0} \beta_0. \tag{21}$$

We set $A_N^{N+1}(\epsilon)=0$ by definition in order to account for the absence of the last term in Eq. (20) with $K=N$. The last term in Eq. (20) is of order $\sim 1/N$ in comparison with the previous term and is taken to lowest order in $1/N$; the need to take it into account has to do with the fact that to calculate $A_N^K(\epsilon)$ with $K\sim N$ from the assigned values of $A_N^0(\epsilon)$ requires $\sim N$ iterations, and for an accuracy of each iteration of $\sim 1/N$ the errors build up. In what follows we will drop the argument ϵ in the intermediate formulas.

Making the substitution

$$A_N^K=(-W_2)^K\frac{\Gamma(N-\beta)}{\Gamma(K+1)\Gamma(N-K-\beta)}A_{N-K}^0X_{N,N-K} \tag{22}$$

in Eq. (20) and introducing the notation

$$h_M=-\frac{\epsilon}{W_2}\frac{A_{M+1}^0}{A_M^0}\frac{M+1}{M-\beta},$$

$$f_M=\frac{W_3}{W_2}\frac{A_{M-1}^0}{A_M^0}(M-1-\beta), \tag{23}$$

we obtain

$$X_{N,M}=h_MX_{N,M+1}+X_{N-1,M}+\frac{f_M}{N}X_{N-2,M-1} \tag{24}$$

with boundary condition

$$X_{N,N}=1. \tag{25}$$

Rewriting Eq. (24) in the form

$$X_{N,M}=(\hat{l}_M+\hat{\delta}_M)X_{N,M+1}, \tag{26}$$

where

$$\hat{l}_M\equiv h_M+e^{-i\hat{p}},\quad \hat{\delta}_M\equiv\frac{f_M}{N}e^{-2i\hat{p}}, \tag{27}$$

$e^{-i\hat{p}}$ is the shift operator by -1, which operates on both arguments, and invoking the boundary condition (25), it is easy to obtain

$$\begin{aligned}X_{N,M}&=(\hat{l}_M+\hat{\delta}_M)(\hat{l}_{M+1}+\hat{\delta}_{M+1})\ldots(\hat{l}_{N-1}+\hat{\delta}_{N-1})X_{N,N}\\&=\hat{l}_M\hat{l}_{M+1}\ldots\hat{l}_{N-1}1+\sum_{p_1=M}^{N-1}\hat{l}_M\ldots\hat{l}_{p_1-1}\hat{\delta}_{p_1}\hat{l}_{p_1+1}\ldots\hat{l}_{N-1}1\\&\quad+\sum_{p_1=M}^{N-2}\sum_{p_2=p_1+1}^{N-1}\hat{l}_M\ldots\hat{l}_{p_1-1}\hat{\delta}_{p_1}\hat{l}_{p_1+1}\ldots\hat{l}_{p_2-1}\\&\quad\times\hat{\delta}_{p_2}\hat{l}_{p_2+1}\ldots\hat{l}_{N-1}1+\ldots.\end{aligned} \tag{28}$$

We do not indicate the argument N of the operator $\hat{\delta}$, which is shown on the left side of the equation. The main contribution to the sum comes from terms with a small number of operators $\hat{\delta}$, which are not difficult to calculate. The result

$$\hat{l}_M\hat{l}_{M+1}\ldots\hat{l}_{M'-1}=\sum_{L=0}^{M'-M}C_{M'-M}^Lh_Mh_{M+1}\ldots h_{M'-L-1}e^{-iL\hat{p}}, \tag{29}$$

which determines the zeroth-order term in $\hat{\delta}$, follows by induction. For products with one $\hat{\delta}$ operator we have

$$\begin{aligned}&\hat{l}_M\ldots\hat{l}_{p_1-1}\hat{\delta}_{p_1}\hat{l}_{p_1+1}\ldots\hat{l}_{M'-1}\\&=\sum_{L_1=0}^{p_1-M}\sum_{L_2=0}^{M'-p_1-1}C_{p_1-M}^{L_1}C_{M'-p_1-1}^{L_2}h_M\ldots h_{p_1-L_1-1}\\&\quad\times\frac{f_{p_1-L_1}}{N-L_1}h_{p_1-L_1-1}\ldots h_{M'-L_1-L_2-3}e^{-i(L_1+L_2+2)\hat{p}}.\end{aligned} \tag{30}$$

Noting that by virtue of (23)

$$f_{p_1-L_1}h_{p_1-L_1-1}=\left(-\frac{\epsilon W_3}{W_2^2}\right)(p_1-L_1), \tag{31}$$

we reduce the result (30) to the form

$$\begin{aligned}&\left(-\frac{\epsilon W_3}{W_2^2}\right)\sum_L h_M\ldots h_{M'-L-3}e^{-i(L+2)\hat{p}}\\&\quad\times\sum_{L_1}C_{p_1-M}^{L_1}C_{M'-p_1-1}^{L-L_1}\frac{p_1-L_1}{N-L_1}.\end{aligned} \tag{32}$$

The sum over L_1 has a saddle-point at $L_c=L(p_1-M)/(M'-M-1)$; replacing L_1 by L_c in the last fraction in (32) and making use of the addition theorem for binomial coefficients (Ref. 22, p. 745) we obtain

$$\begin{aligned}&\left(-\frac{\epsilon W_3}{W_2^2}\right)\sum_L C_{M'-M-1}^Lh_M\ldots h_{M'-L-3}e^{-i(L+2)\hat{p}}\\&\quad\times\left.\frac{p_1-(p_1-M)\tau}{N-(p_1-M)\tau}\right|_{\tau=L/(M'-M-1)}.\end{aligned} \tag{33}$$

Result (33) has the same structure as (29), and by induction it is not hard to find products with a moderate number s of $\hat{\delta}$ operators; from Eq. (28) we obtain

$$\begin{aligned}X_{N,M}&=\sum_{s=0}^{\infty}\left(-\frac{\epsilon W_3}{W_2^2}\right)^s\sum_{L=0}^{\min\{N-M-s,N-2s\}}\\&\quad\times C_{N-M-s}^Lh_M\ldots h_{N-L-2s-1}\\&\quad\times\sum_{p_1=M}^{N-s}\frac{p_1-(p_1-M)\tau}{N-(p_1-M)\tau}\sum_{p_2=p_1+1}^{N-s+1}\\&\quad\times\frac{p_2-2-(p_2-M-1)\tau}{N-2-(p_2-M-1)\tau}\sum_{p_s=p_{s-1}+1}^{N-1}\\&\quad\times\left.\frac{p_s-2s+2-(p_s-M-s+1)\tau}{N-2s+2-(p_s-M-s+1)\tau}\right|_{\tau=L/(N-M-s)}.\end{aligned} \tag{34}$$

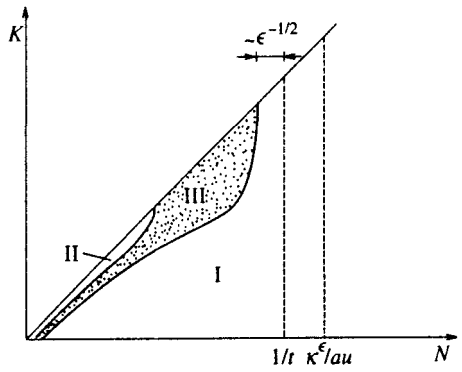

FIG. 1. Regions I and II, which give nonperturbative and quasiparquet contributions to the sum (4); the parameter $t\sim\epsilon$ is defined by formula (42). The nonperturbative contribution is estimated in effect for $N=\kappa^\epsilon/au$; the inequality $\kappa^\epsilon/au>1/t$ corresponds to a positive value of Δ (see Eq. (45)).

Calculating the sum over $p_1,p_2,\ldots,p_s$ in the two overlapping regions of parameter space, we obtain the following results for it ($\tau'\equiv1-\tau$):

$$\frac{1}{s!}\left\{\frac{M+\tau'(N-M)}{(1-\tau')^2}\ln\frac{N}{\tau'N+(1-\tau')M}-\frac{\tau'(N-M)}{1-\tau'}\right\}^s,\quad \max\{\tau'N,M\}\gg s, \tag{35}$$

$$\frac{1}{s!}\frac{\Gamma(M+\tau'N+1)}{\Gamma(M+\tau'N-s+1)}\left\{\ln\frac{N}{\tau'N+M-s}-\frac{\tau'N}{\tau'N+M}\right\}^s,\quad \tau'N\sim M\sim s. \tag{36}$$

In the first case it is possible to neglect quantities $\sim s$ in the fractions within the summation range, and transform from sums to integrals; in the second case it is possible to calculate the sums systematically by separating out the two highest powers of the large logarithms. Formula (36) is valid literally for $s\gg1$, whereas for $s\sim1$ the difference between the expression in braces and $\ln N$ exceeds the accuracy of the calculation.

The product

$$h_Mh_{M+1}\ldots h_{N-L-2s-1}=\left(-\frac{\epsilon}{W_2}\right)^{N-L-2s-M}\frac{A^0_{N-L-2s}}{A^0_M}\times\frac{\Gamma(M-\beta)}{\Gamma(M+1)}\frac{\Gamma(N-L-2s+1)}{\Gamma(N-L-2s-\beta)} \tag{37}$$

entering into expression (34) depends on the coefficients A^0_N, which are assumed to be known. By analogy with the case $d=4$ (Ref. 16), in the (N,K) plane it is possible to distinguish two regions in Fig. 1: region I, in which the sum in (34) is determined by indices $N-L-2s\gg1$, such that the Lipatov asymptotic limit is valid for the coefficients A^0_N, and region II ($M\ll\ln N$, $N\epsilon\ll1$), ''controlled'' by the trivial coefficient $A^0_0=1$. Between regions I and II lies the region of non-universality—region III ($M\sim\ln N$, $N\epsilon\lesssim1$), in which information about the coefficients A^0_N with $N\sim1$ is important. Region III does not make a substantial contribution to the sum (4).

The conditions $N-L-2s\gg1$, $\max\{\tau'N,M\}\gg1$, and $N-M-L\gg s$ are satisfied in region I for $N\epsilon\gtrsim1$. This enables one to use (19) for A^0_N and (35) for the sum over p_i, and neglect the magnitude of s in the slowly varying functions within the summation range in (34) and to sum over s. After substituting the result back into (22), we obtain

$$A_N^{N-M}=\frac{\epsilon^{N-M}}{M!}\tilde c_2\Gamma(N-\beta)a^N\sum_{L=0}^{N-M}\frac{(N-L)!}{L!(N-L-M)!}\times\left(-\frac{W_2}{a\epsilon}\right)^L J(N-L)(N-L)^{b+\beta}e^{S(L)}, \tag{38}$$

where

$$J(N)=\int_0^\infty d\,\ln R^2R^{-2}\exp\left(-Nf(R)+N\epsilon\ln R+2K_dI_4(R)\frac{1-R^{-\epsilon}}{\epsilon}\right), \tag{39}$$

$$S(L)=\frac{W_3}{\epsilon a^2L}\left(\frac{N-L-M}{N-L}\right)^2\times\left[1+\frac{(N-L)(N-M)}{L(N-M-L)}\ln\frac{N-L}{N}\right]. \tag{40}$$

For $N-M\ll N$ or $N\epsilon\gg1$, the sum over L in (38) is determined by values $L\ll N$, and (38) goes over to (19). For $M\sim1$, (38) becomes

$$A_N^{N-M}(\epsilon)=\frac{1}{M!}\epsilon^{N-M}\tilde c_2\Gamma(N-\beta)a^N\sqrt{t/2\pi}\exp\left[f_\infty(Nt\ln N-1)+\frac{1}{t}\right]\times\int_0^\infty dx\exp\left[-\frac{t}{2}\left(N-\frac{1}{t}-x\right)^2\right]\times x^{M+b+\beta-f_\infty Nt}J(x), \tag{41}$$

where

$$t=-\frac{\epsilon a}{W_2}\xrightarrow{\epsilon\to0}\frac{3\epsilon}{n+8},\quad f_\infty=\frac{W_3}{aW_2}\xrightarrow{\epsilon\to0}\frac{3n+14}{n+8}. \tag{42}$$

The assumptions made in the derivation of (41) are fulfilled in the region $Nt>1$ or $1-Nt\ll\epsilon^{1/2}$.

For $N\epsilon\ll1$, the sum over L in (34) is determined by the neighborhood of the upper limit of the sum, so that $\tau'\ll1$; for $M\gg\ln N$ and $M\sim\ln N$, Eqs. (35) and (36) apply, respectively. For $M\ll\ln N$, terms with $s\gtrsim M$, $L=N-2s$ dominate, and by virtue of the equality $A^0_0=1$ we have the following result for region II:

$$A_N^{N-M}(\epsilon)=(-W_2)^N\frac{\Gamma(N-\beta)}{\Gamma(N+1)\Gamma(-\beta)}\sum_{L=0}^\infty\epsilon^L\frac{1}{M!L!}$$

$$\times\left(-\frac{W_3}{W_2^2}\right)^{M+L}(N\ln N)^{M+L}, \tag{43}$$

which can be obtained from the system of equations (11) by separating out the leading asymptotic behavior in N for $A_N^{N-K,L}$. For $N\epsilon\ln N\ll 1$, terms with $s\leqslant M$ dominate, and for arbitrary M we have the result

$$X_{N,M}=\frac{\Gamma(M-\beta)}{A_M^0}\sum_{s=0}^{M}\frac{A_{M-s}^0}{\Gamma(M-s-\beta)}\frac{1}{s!}\times\left(\frac{W_3}{W_2}\ln\frac{N}{M-s+1}\right)^s e^{t(N-M)M}, \tag{44}$$

whose region of applicability expands without limit as $\epsilon\to 0$, and this result transforms to Eqs. (42) and (43) of Ref. 16 for $d=4$.

4. ENERGY RENORMALIZATION AND DECAY

As in the case $d=4$ (Ref. 16), there are two important contributions to the sum (4)—a nonperturbative and a quasiparquet contribution, arising respectively from regions I and II (Fig. 1). We restrict the discussion to the continuum limit $\Lambda\to\infty$, in which only the coefficients $A_N^N(\epsilon)$ remain in the sum (4). The quasiparquet contribution is calculated on the basis of formula (43), and has the form

$$[Y(\kappa)]_{\text{quasiparq}}=\left[\Delta+\frac{W_3(\epsilon)}{W_2(\epsilon)}u\kappa^{-\epsilon}\ln\Delta\right]^{\beta(\epsilon)},$$
$$\Delta\equiv 1+W_2(\epsilon)u\frac{\kappa^{-\epsilon}}{\epsilon}, \tag{45}$$

where the coefficients $W_2(\epsilon)$, $W_3(\epsilon)$, and $\beta(\epsilon)$ can be taken to zeroth order in ϵ. Within the limits of accuracy of the calculations, the argument Δ of the logarithm can be replaced by its minimum value $\tilde{\Delta}\sim\epsilon\ln\epsilon$ (defined by Eqs. (51) and (52) below), since for $\Delta\gg\tilde{\Delta}$ the logarithmic term is unimportant. Therefore (45) can be rewritten in the form

$$[Y(\kappa)]_{\text{quasiparq}}=[1+W_{2,0}\tilde{u}\kappa^{-\epsilon}/\epsilon]^{\beta_0},$$
$$\tilde{u}\equiv u\left[1+\frac{W_{3,0}}{W_{2,0}^2}\epsilon\ln\tilde{\Delta}\right], \tag{46}$$

which differs from the parquet form[23] only by the substitution of $\tilde{u}$ for u.

To calculate the nonperturbative contribution, we set

$$A_N^N(\epsilon)=\tilde{c}_2\Gamma(N+b)\epsilon^N a^N F(N) \tag{47}$$

and sum (4) from some large N_0 to infinity according to Eq. (46) in Ref. 16:

$$[\Sigma(0,\kappa)]_{\text{nonpert}}\equiv i\Gamma_0(\kappa^2)=i\pi\tilde{c}_2\kappa^2(\kappa^\epsilon/au)^b e^{-\kappa^\epsilon/au}F(\kappa^\epsilon/au). \tag{48}$$

The nonperturbative contribution is associated with the divergence of the series, and formally arises from the region of arbitrarily large N. However, it must be calculated on the basis of (41), not (19), since the correction factor distinguishing (41) from (19) is evaluated in effect for $N=\kappa^\epsilon/au$ and turns out to be substantial. I did not recognize this circumstance in Ref. 17; therefore, Eqs. (22) and (23) in Ref. 17 differ from Eqs. (52), (53), and (55) below.

Approximating the series (4) by the sum of contributions (46) and (48), we obtain

$$\kappa_0^2-\kappa_c^2=\kappa^2[1+8K_4\tilde{u}\kappa^{-\epsilon}/\epsilon]^{1/4}+i\Gamma_0(\kappa^2),$$
$$\kappa^2=-E-i\Gamma, \tag{49}$$

where $\kappa_c^2=\Sigma(0,0)$ and we have allowed for the fact that $\kappa_0^2=\kappa^2+\Sigma(0,\kappa)$. Equation (49) is solved like Eq. (93) in Ref. 15. Setting

$$\kappa^2=|\kappa|^2e^{-i\varphi},\quad x=\frac{2}{\epsilon}\left[\left(\frac{|\kappa|^2}{\Gamma_c}\right)^{\epsilon/2}-1\right],\quad \Gamma_c=\left(\frac{8K_4|\tilde{u}|}{\epsilon}\right)^{2/\epsilon} \tag{50}$$

and separating the real and imaginary parts of (49), we obtain a connection between the decay Γ and the renormalized energy E with the unrenormalized energy $E_B=-\kappa_0^2$ in parametric form:

$$\Gamma=\Gamma_c\left(1+\frac{\epsilon x}{2}\right)^{2/\epsilon}\sin\varphi,\quad E=-\Gamma_c\left(1+\frac{\epsilon x}{2}\right)^{2/\epsilon}\cos\varphi,$$
$$-E_B+E_c=\Gamma_c\left(1+\frac{\epsilon x}{2}\right)^{2/\epsilon}\left(\frac{\epsilon x/2}{1+\epsilon x/2}\right)^{1/4}\left[\cos\left(\varphi+\frac{\varphi}{4x}\right)-\tan\frac{\varphi(1+2\epsilon x)}{3}\sin\left(\varphi+\frac{\varphi}{4x}\right)\right], \tag{51}$$

where E_c is defined by Eq. (108) in Ref. 15, and $x(\varphi)$ is a single-valued function in the interval $0<\varphi<\pi$, analogous to the function shown in Fig. 2 of Ref. 15, and implicitly defined by the equation

$$\sin\left(\varphi+\frac{\varphi}{4x}\right)=\frac{e^{-4x/3}}{x^{1/4}}I(x)\cos\frac{\varphi(1+2\epsilon x)}{3}, \tag{52}$$

where

$$I(x)=\tilde{c}_2\left(\frac{3}{4}\right)^{1/4}\left(\frac{\pi t}{2}\right)^{1/2}\times\exp\left\{-f_\infty+f_\infty\left(1+\frac{\epsilon x}{2}\right)\ln\left[\tilde{\Delta}\left(1+\frac{\epsilon x}{2}\right)\Big/\ t\right]\right\}\times\int_0^\infty dz\ \exp\left[-\frac{t}{2}\left(\frac{\epsilon x}{2t}-z\right)^2\right]z^{b+\beta-f_\infty(1+\epsilon x/2)}J(z). \tag{53}$$

Equations (51) and (52) simplify substantially in two overlapping regions. For $x\gg\ln(1/\epsilon)$, i.e., at high $|E|$, where the right-hand side of Eq. (52) is small and the quantity φ is near 0 or π, we obtain the asymptotic behavior of $\Gamma(E)$,

$$\Gamma(E)=\begin{cases}\dfrac{1}{8}\pi\epsilon E[(E/\Gamma_c)^{\epsilon/2}-1]^{-1}, & E\gg\Gamma,\\ \Gamma_0(E)[1-(|E|/\Gamma_c)^{-\epsilon/2}]^{-1/4}, & -E\gg\Gamma,\end{cases} \tag{54}$$

which produce the illusion of a ghost pole[7] ($\Gamma_0(E)\equiv\Gamma_0(|\kappa|^2)$). For large positive E the result of the kinetic equation is reproduced; for large negative E the decay becomes purely nonperturbative.

At low energies, $x\lesssim\epsilon^{-1/2}$, we have

$$\sin\left(\varphi+\frac{\varphi}{4x}\right)=I(0)\,\frac{e^{-4x/3}}{x^{1/4}}\cos\frac{\varphi}{3},$$

$$I(0)\sim\epsilon^{-7/12}\left(\ln\frac{1}{\epsilon}\right)^{17/12}, \tag{55}$$

which describes the neighborhood of the ghost pole and has the same functional form as the four-dimensional equation (see Eq. (51) in Ref. 16 for $x\ll x_0$ and Eq. (100) in Ref. 15). The minimum values of Δ and x are reached simultaneously, and to logarithmic accuracy they are

$$\Delta_{\min}\equiv\tilde{\Delta}\approx\frac{7}{8}\,\epsilon\,\ln\frac{1}{\epsilon},\quad x_{\min}\approx\frac{7}{16}\ln\frac{1}{\epsilon}, \tag{56}$$

so that the detour about the pole must be taken at a distance of the order of $\epsilon\ln(1/\epsilon)$.

5. DENSITY OF STATES

To calculate the density of states requires a knowledge of the eigenenergy $\Sigma(p,\kappa)$ for finite momenta[15]; like $p=0$, this quantity consists of a nonperturbative and a quasiparquet contribution. The quasiparquet contribution is given by the parquet equations (Ref. 15, Sec. 7) with the substitution $u\to\hat{u}$; the proof of this is completely analogous to the situation $d=4$ (Ref. 16, Sec. 5). The nonperturbative contribution turns out to be important only in the region of large negative E, where it is directly determined by the Lipatov asymptotic behavior, and can be calculated on the basis of formula (13) (for $N=\kappa^\epsilon/au\gg1/\epsilon$ the correction factor distinguishing results of the type (41) and (19) is equal to unity)

$$\begin{aligned}[\Sigma(p,\kappa)]_{\text{nonpert}}=i\pi c_2\kappa^2&\left(\frac{\kappa^\epsilon}{au}\right)^b e^{-\kappa^\epsilon/au}\\ &\times\int_0^\infty d\,\ln R^2R^{-2}\langle\phi_c^3\rangle_{pR/\kappa}\langle\phi_c^3\rangle_{-pR/\kappa}\\ &\times\exp\left\{-\frac{\kappa^\epsilon}{au}\,[f(R)-\epsilon\ln R]\right.\\ &\left.+\frac{2K_dI_4(R)}{\epsilon}\right\}.\end{aligned} \tag{57}$$

For $p=0$ the integral is governed by the neighborhood of the saddle point R_0, which is a root of the equation

$$\epsilon=6R_0^2(-\ln R_0+\ln 2-C-1), \tag{58}$$

so that $R_0\approx\sqrt{\epsilon/3\ln(1/\epsilon)}$. For $p\lesssim\kappa R_0^{-1}$, Eq. (57) does not depend on p; for $p\gtrsim\kappa R_0^{-1}$ it falls off rapidly with increasing p. By virtue of the logarithmic accuracy of the following calculations (Ref. 15, Sec. 8) the result

$$[\Sigma(p,\kappa)]_{\text{nonpert}}\approx[\Sigma(0,\kappa)]_{\text{nonpert}}\theta(\kappa R_0^{-1}-p). \tag{59}$$

suffices. Taking the above into account, the final expression for $\Sigma(p,\kappa)$ has the form

$$\begin{aligned}\Sigma(p,\kappa)-\Sigma(0,\kappa)=\kappa^2&\left\{1-\frac{3}{2}\left[\frac{t(x)}{t(x_\infty)}\right]^{-1/4}+\frac{1}{2}\left[\frac{t(x)}{t(x_\infty)}\right]^{-3/4}\right\}\\ &-i\Gamma_0(\kappa^2)\,\theta(p-\kappa R_0^{-1})\end{aligned} \tag{60}$$

(cf. Eq. (116) in Ref. 15), where

$$t(x)=1+8K_4\tilde{u}x/\epsilon,\quad x=p^{-\epsilon},\quad x_\infty=\kappa^{-\epsilon}. \tag{61}$$

Substituting (60) into Eqs. (117) and (118) of Ref. 15 for $d=4-\epsilon$, we obtain

$$\begin{aligned}\nu=\frac{\Gamma_c}{4\pi|\tilde{u}|}&\left(1+\frac{\epsilon x}{2}\right)^{2/\epsilon}\left\{\left(1+\frac{2}{\epsilon x}\right)^{-1/4}\left(1-\frac{R_0^\epsilon}{2+\epsilon x}\right)\right.\\ &\left.\times\sin\left(\varphi+\frac{\varphi}{4x}\right)-\left(1+\frac{2}{\epsilon x}\right)^{-3/4}\sin\left(\varphi+\frac{3\varphi}{4x}\right)\right\},\end{aligned} \tag{62}$$

which together with (51) and (52), determines the density of states $\nu(E)$ in parametric form.

Let us now turn our attention to the presence of scaling: for the energy measured in units of Γ_c and the density of states in units of $\Gamma_c/|\tilde{u}|$, all dependences are determined by universal functions that are independent of the degree of disorder. For $|E|\gg\Gamma$, we have the asymptotic behavior

$$\nu(E)=\begin{cases}\dfrac{1}{2}\,K_4E^{(d-2)/2}\left[1-\left(\dfrac{E}{\Gamma_c}\right)^{-\epsilon/2}\right]^{-1/4}, & E\gg\Gamma,\\ \dfrac{\Gamma_0(E)}{4\pi|\tilde{u}|}\left\{1-\dfrac{R_0^\epsilon}{2}\left(\dfrac{|E|}{\Gamma_c}\right)^{-\epsilon/2}-\left[1-\left(\dfrac{|E|}{\Gamma_c}\right)^{-\epsilon/2}\right]^{1/2}\right\}, & -E\gg\Gamma,\end{cases} \tag{63}$$

indicating a ghost pole. For large positive E, the function $\nu(E)$ transforms into the density of states of an ideal system, and at large negative energies E we obtain the following result for the fluctuation tail:

$$\begin{aligned}\nu(E)&=\frac{K_4}{\pi}\,\Gamma_0(E)|E|^{-\epsilon/2}\ln\frac{1}{R_0}\\ &=\tilde{c}_2K_4\left(\frac{2\pi}{3}\ln\frac{1}{R_0}\right)^{1/2}R_0^{-3}|E|^{(d-2)/2}\left[\frac{\tilde{I}_4|E|^{\epsilon/2}}{4|u|}\right]^{(d+1)/2}\\ &\quad\times\exp\left(\frac{2K_dI_4(R_0)}{\epsilon}-\frac{I_4(R_0)|E|^{\epsilon/2}}{4|u|R_0^\epsilon}\right),\end{aligned} \tag{64}$$

whose energy dependence coincides with that obtained in Refs. 25–27, and corresponds to the well-known Lifshits law[28]; the discrepancy at $\epsilon\to0$ is eliminated for finite cutoff parameter Λ. Oddly enough, for $\epsilon x\ll1$ Eqs. (51)–(53) and (62) have the same functional form as those for $d=4$ (Ref. 16), i.e., the behavior of all physical quantities in the vicinity of the mobility threshold turns out to be effectively four-dimensional. As in Refs. 15 and 16, the phase transition point is shifted into the complex plane, which ensures regularity of the density of states at all energies.

R_0^ϵ differs substantially from unity only when $\kappa^\epsilon/u\ll1/\epsilon$, the terms in braces in (63) cancelling almost exactly. Letting $R_0\to1$ in (60) is tantamount to completely neglecting $[\Sigma(p,\kappa)]_{\text{nonpert}}$, since the domain of integration in Eq. (118) of Ref. 15 is $p\gtrsim\kappa$. Thus, $[\Sigma(p,\kappa)]_{\text{nonpert}}$ is

significant only for large-magnitude negative E, and can be calculated using the Lipatov asymptotic form.

6. LIPATOV ASYMPTOTIC LIMIT

Calculation of the Lipatov asymptotic limit in $(4-\epsilon)$-theory closely follows the scheme for $d=4$ described in detail in Ref. 16. Therefore we discuss only the differences that arise, denoting by the numeral I references to equations from Ref. 16.

In massless four-dimensional theory there exists a specific zero mode—the dilatation mode, corresponding to variation of the radius R of the instanton.[16,18,29] As in the massive four-dimensional theory,[16] for $d=4-\epsilon$ this mode becomes soft and the integration over it bears a substantially non-Gaussian character. It is necessary to carry out this integration correctly to ensure that the correct limit is reached as $d\to 4$.

By analogy with (I.82), we introduce three expansions of unity inside the functional integral:

$$1=\left(\int d^dx|\varphi(x)|^4\right)^d\int d^dx_0 \times\prod_{\mu=1}^{d}\delta\left(-\int d^dx|\varphi(x)|^4(x-x_0)_\mu\right),$$

$$1=\int d^dx|\varphi(x)|^4\int_0^\infty d\,\ln R^2 \times\delta\left(-\int d^dx|\varphi(x)|^4\,\ln\left(\frac{x-x_0}{R}\right)^2\right),$$

$$1=\int d^nu\,\delta(\mathbf{u}-\mathbf{v}\{\varphi\}), \tag{65}$$

and in place of (I.82) we make the substitutions

$$x-x_0=R\tilde{x},\quad \varphi_\alpha(x_0+R\tilde{x})=R^{-(d-2)/2}\tilde{\varphi}_\alpha(\tilde{x}),\quad g=\tilde{g}R^{d-4}. \tag{66}$$

As a result, we have

$$[G_M]_{N-1}=\int_0^\infty d\,\ln R^2 Z_0(\kappa_R)^{-1} \times\int d^dx_0\int d^nuR^{-4-(d-2)M/2}\int\frac{dg}{2\pi i} \times\int D\varphi\prod_{\mu=1}^{d}\delta\left(-\int d^dx|\varphi(x)|^4x_\mu\right) \times\delta\left(-\int d^dx|\varphi(x)|^4\,\ln x^2\right)\delta(\mathbf{u}-\mathbf{v}) \times\left(\int d^dx|\varphi(x)|^4\right)^{d+1}\varphi_{\alpha_1}\left(\frac{x_1-x_0}{R}\right)\ldots\varphi_{\alpha_M} \times\left(\frac{x_M-x_0}{R}\right)\exp[-H\{\kappa_R,g,\varphi\}-N\,\ln g +N\epsilon\,\ln R], \tag{67}$$

the main difference of which from (I.83) consists in the appearance of the term $N\epsilon\,\ln R$ in the exponential. The choice of instanton, as before, is dictated by Eq. (I.94), which after transforming to the function $\phi_c(x)$ according to (I.72) takes the following form in spherical coordinates ($r\equiv|x|$):

$$\phi_c''(r)+\frac{3-\epsilon}{r}\,\phi_c'(r)-\kappa_R^2\phi_c(r)+\phi_c^3(x)-\mu_0\phi_c^3(r)\ln r^2=0. \tag{68}$$

In the region $r\ll\kappa_R^{-1}$, terms with ϵ, κ_R, and μ_0 are treated as a perturbation, and by analogy with (I.99) we obtain

$$\phi_c(r)=\frac{2\sqrt{2}}{z+1}\left[1+\frac{1-z}{1+z}\,v(z)\right]_{z=r^2},$$

$$v(z)=\int_0^z dz\,\frac{(1+z)^4}{(1-z)^2z^2}\left\{\epsilon\,\frac{z^2(z-3)}{12(1+z)^3} +\frac{\kappa_R^2}{4}\left[-\ln(1+z)+\frac{z+2z^2}{(1+z)^2}\right] +\mu_0\left[\frac{\ln z}{(z+1)^4}-\frac{z+3}{6(z+1)^3}\right]z^2\right\}. \tag{69}$$

Calculation of the asymptotic limit of $v(z)$ for $z\gg 1$ with allowance for only the growing terms in z gives for the region $1\ll r\ll\kappa_R^{-1}$

$$\phi_c(r)=\frac{2\sqrt{2}}{r^2}\left\{1+\frac{1}{2}\,\kappa_R^2r^2\,\ln r+\left[\frac{1}{6}\,\mu_0-\frac{3}{4}\,\kappa_R^2-\frac{1}{12}\,\epsilon\right]r^2 +3\kappa_R^2\,\ln^2 r+\left[2\mu_0-\frac{11}{2}\,\kappa_R^2\right]\ln r-\frac{1}{r^2}\right\}. \tag{70}$$

In the region $r\gg 1$, treating the nonlinear terms in (68) as a perturbation, we obtain after separating out the asymptotic limit for $r\ll\kappa_R^{-1}$

$$\phi_c(r)=\frac{2\sqrt{2}}{r^2}\left\{1+\frac{1}{2}\,\kappa_R^2r^2\,\ln r +\frac{2C-1+2\,\ln(\kappa_R/2)}{4}\,\kappa_R^2r^2+3\kappa_R^2\,\ln^2 r +\left[\epsilon+\kappa_R^2\left(6C+\frac{1}{2}+6\,\ln\frac{\kappa_R}{2}\right)\right]\ln r-\frac{1}{r^2}\right\}. \tag{71}$$

The matching condition for (70) and (71) has the form

$$2\mu_0=\epsilon+6\kappa_R^2\,(\ln\,\kappa_R+C+1-\ln 2) \tag{72}$$

Using Eq. (69) to calculate the integral in (I.70) (making the substitution $d^4x\to d^dx$), we obtain

$$N\,\ln\,g_c=N\,\ln\left(-\frac{\overline{I_4}}{4N}\right)+Nf(\kappa_R), \tag{73}$$

where $f(x)$ and $\overline{I_4}$ are defined by (14). In comparison with the case $d=4$, the function $f(x)$ differs by a constant $\sim\epsilon$.

Another modification arises when the divergences are separated out of the determinants defined by the sum rule [cf. (I.114)]

$$\sum_s \frac{1}{\mu_s^2} = 9\int_0^{\Lambda R} \frac{d^d k}{(2\pi)^d} \times \int_0^{\Lambda R} \frac{d^d q}{(2\pi)^d} \frac{\langle \phi_c^2\rangle_q \langle \phi_c^2\rangle_{-q}}{(k^2+\kappa_R^2)[(k+q)^2+\kappa_R^2]} \approx 9K_d I_4(\kappa_R)\frac{1-(\Lambda R)^{-\epsilon}}{\epsilon} + 12\left(\frac{1}{3}+C-\ln 2\right). \tag{74}$$

For the Nth coefficient of the expansion of the Green's function, instead of (I.113) we obtain

$$[G_M(x_1,\alpha_1,\ldots,x_M,\alpha_M)]_N = c(-1)^N\left(\frac{4}{\overline{I_4}}\right)^{N+(M+d+1)/2}\Gamma\left(N+\frac{M+n+d}{2}\right) \times \int d^n u\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M}\int_0^\infty d\,\ln R^2 \times \int d^d x_0 R^{-d-M(d-2)/2}\phi_c\left(\frac{x_1-x_0}{R}\right)\ldots\left(\frac{x_M-x_0}{R}\right) \times \exp\left\{-Nf(\kappa_R)+N\epsilon\,\ln R + \frac{n+8}{4}K_d I_4(\kappa_R)\frac{1-(\Lambda R)^{-\epsilon}}{\epsilon}\right\}, \tag{75}$$

where the constant c is calculated in the lowest order in ϵ and is given by formula (I.114). Going over to the vertex part, instead of (I.127) we obtain

$$[\Gamma^{(0,2M)}(p_1,\ldots,p_{2M})]_N = c(-1)^N\frac{2\pi^{n/2}}{2^M\Gamma(M+n/2)}\left(\frac{4}{\overline{I_4}}\right)^{N+M+5/2} \times \Gamma\left(N+\frac{2M+n+d}{2}\right)\int_0^\infty d\,\ln R^2 R^{-d+(d-2)M} \times \langle\phi_c^3\rangle_{Rp_1}\ldots\langle\phi_c^3\rangle_{Rp_{2M}}\exp\left\{-Nf(\kappa_R)+N\epsilon\,\ln R + \frac{n+8}{4}K_d I_4(\kappa_R)\frac{1-(\Lambda R)^{-\epsilon}}{\epsilon}\right\}, \tag{76}$$

where $\langle\phi_c^3\rangle_p$ is the Fourier component of the function $\phi_c^3(x)$. To lowest order in ϵ this Fourier component is given by (14). The vertex $\Gamma^{(0,2)}$ coincides with the eigenenergy and for $M=1$, $n=0$, (76) follows from (13).

7. INSTANTON RESULTS FOR $\epsilon\sim 1$

In order to compare with the results of other authors,[7,25–27] let us discuss instanton calculations for $d<4$ without assuming that d is close to 4. Such calculations closely follow the scheme for $d>4$ described in Ref. 15 (we denote references to the corresponding formulas by the numeral II), with the replacements $\Sigma_x\to\int d^d x$ and $\epsilon(p)\to p^2$. The difference has to do with the need to separate out the zero translational modes along with the rotational modes; the dilatation mode is considered here, in contrast to the previous section, on general grounds. Accordingly, of the three expansions of unity (65) we use only the first and the third, but the substitution of variables (66) is carried out with $R=1$. In addition to (II.65), a transformation of the determinant D_L is required:

$$\frac{D_L'}{D_0} = \overline{D(1)}\prod_{\mu=1}^d \frac{\int d^d x\left(\frac{\partial\phi_c(x)}{\partial x_\mu}\right)^2}{3\int d^d x\,\phi_c^2(x)\left(\frac{\partial\phi_c(x)}{\partial x_\mu}\right)^2},$$

$$\overline{D(1)} = \prod_s{}'\left(1-\frac{1}{\mu_s}\right). \tag{77}$$

The prime denotes omission of the contribution of the translational modes. The instanton equation reduces by this substitution of variables to the form

$$\Delta\phi_c(x)+\phi_c^3(x)-\overline{\kappa^2}\phi_c(x)=0, \tag{78}$$

where $\overline{\kappa}$ is an arbitrary parameter (see below). For the expansion coefficients of the Green's function we obtain

$$[G_M(x_1,\alpha_1,\ldots,x_M,\alpha_M)]_N = \frac{2^{n-1}}{(2\pi)^{(n+d+1)/2}}\left(\frac{I_6-\overline{\kappa^2}I_4}{d}\right)^{d/2}\left(\frac{4}{I_4}\right)^{(M+d)/2} \times\left(\frac{\kappa}{\overline{\kappa}}\right)^{(d-2)M/2}\left[-\overline{D(1)}\,\overline{D^{n-1}}(1/3)\right]^{-1/2} \times\left[-\frac{4}{I_4}\left(\frac{\kappa}{\overline{\kappa}}\right)^{d-4}\right]^N\Gamma\left(N+\frac{M+n+d-1}{2}\right) \times\int d^d x_0\,\phi_c\left(\frac{\kappa}{\overline{\kappa}}x_1-x_0\right)\ldots\phi_c\left(\frac{\kappa}{\overline{\kappa}}x_M-x_0\right) \times\int d^n u\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M}, \tag{79}$$

where

$$I_p=\int d^d x\,\phi_c^p(x). \tag{80}$$

For $d=4-\epsilon$, (75) and (79) are equivalent only for $N\epsilon\gg 1$, when the integration over R in (75), corresponding to the dilatation mode, can be carried out in the saddle-point approximation. The saddle point occurs for $\kappa_R=R_0$, where R_0 is a root of Eq. (58). In this case, by virtue of Eq. (72), we have $\mu_0=0$ and the instanton equation (68) reduces to Eq. (78) with $\overline{\kappa}=R_0$. Expression (79) with $\overline{\kappa}=R_0$, after estimating the pre-exponential in the zeroth order in ϵ, differs from the result for the saddle-point approximation in (75) by the constant factor

$$\left[\frac{\lambda_0^L+R_0^2}{3I_4R_0^2(-\ln R_0+\ln 2-C-3/2)}\int d^d x[e_0^L(x)]^2\right]^{1/2}$$

$$\approx\left[\frac{R_0^2-\rho^2}{R_0^2}\,\frac{\ln\rho}{\ln\epsilon}\right]^{1/2}, \tag{81}$$

where $\lambda_0^L\equiv-\rho^2$ and $e_0^L(x)$ are the eigenvalue and eigenfunction of the operator $-\Delta-3\phi_c^2(x)$ corresponding to the dilatation mode. The normalization of the function $e_0^L(x)$ is chosen so as to coincide with $[\partial\phi_c(x)/\partial R]_{R=1}$ in the region $|x|\lesssim 1$. The quantity (81) is equal to unity for $\rho\sim\epsilon$ or $-\lambda_0^L\sim\epsilon^2$; from perturbation theory it is easy to convince oneself that the contribution to λ_0^L that is first-order in ϵ vanishes, due to the divergence of the normalization integral for $e_0^L(x)$ for $d=4$.

For $2\leqslant d<4$, the determinants $\overline{D}(1)$ and $\overline{D}(1/3)$ contain divergences,[25] which can be eliminated by renormalization according to (II.75) with simultaneous transformation to the renormalized energy E (the Thomas–Fermi method yields $\mu_s\sim s^{2/d}$ for $s\gg 1$, and the first sum in (II.69) diverges). Setting $\overline{\kappa}=1$ and summing the non-leading terms of the perturbation-theory series for the two-point ($M=2$) Green's function according to (II.90), it is not hard to obtain an expression for the fluctuation tail of the density of states:

$$\nu(E)=\frac{(4-d)2^{d-1}}{(2\pi)^{(d+1)/2}}\left(\frac{I_6-I_4}{I_4 d}\right)^{d/2}\left|\frac{\overline{D_R}(1/3)}{\overline{D_R}(1)}\right|^{1/2}|E|^{(d-2)/2}$$
$$\times\left(\frac{I_4|E|^{(4-d)/2}}{2a_0^dW^2}\right)^{(d+1)/2}\exp\left(-\frac{I_4|E|^{(4-d)/2}}{2a_0^dW^2}\right) \tag{82}$$

(where $4I_2=(4-d)I_4$). The energy dependence of this expression for the fluctuation tail of the density of states coincides with that obtained by Cardy.[27] Normalization to the unperturbed density of states $\nu_0(E)$ and changing over from the renormalized energy E to the unrenormalized energy E_B with a simultaneous shift of the origin (see formula (12) in Ref. 26) gives the results of Brézin and Parisi (Ref. 26)[1)]

$$\frac{\nu(E_B)}{\nu_0(-E_B)}=\left(\frac{I_6-I_4}{3}\right)^{3/2}\left|I_4\,\frac{\overline{D_R}(1/3)}{\overline{D_R}(1)}\right|^{1/2}\frac{|E_B|}{(a_0^dW^2)^2}$$
$$\times\exp\left(-\frac{I_4}{16\pi}-\frac{I_4|E_B|^{1/2}}{2a_0^dW^2}\right),\quad d=3,$$
$$\frac{\nu(E_B)}{\nu_0(-E_B)}=\frac{I_6-I_4}{8\pi^2}\left|I_4\,\frac{\overline{D_R}(1/3)}{\overline{D_R}(1)}\right|^{1/2}\left(\frac{4\pi|E_B|}{a_0^dW^2}\right)^{3/2-I_4/8\pi}$$
$$\times\exp\left(-\frac{I_4}{8\pi}-\frac{I_4|E_B|}{2a_0^dW^2}\right),\quad d=2. \tag{83}$$

For $d<2$, there are no divergences in the determinants, and (82) holds in terms of the unrenormalized quantities (i.e., after the substitutions $E\to E_B$, $D_R(1)\to D(1)$, and $D_R(1/3)\to D(1/3)$). For $d=1$, Eq. (78) with $\tilde{\kappa}=1$ has the solution $\phi_c(x)=\sqrt{2}/\cosh x$, and Eq. (II.64)

$$y''-y+\frac{\mu_s}{\cosh^2 x}\,y=0 \tag{84}$$

has eigenvalues $\mu_s=s(s+1)$, $s=1,2,\ldots$ since by the substitution $y=\tilde{y}\cosh^{-s}x$ it reduces to a form analogous to (I.121). Calculation of the parameters entering into (82)

$$\overline{D}(1)=\prod_{\substack{s=1\\ s\neq 2}}^{\infty}\frac{(s+3)(s-2)}{s(s+1)}=-\frac{1}{5},$$
$$\overline{D}(1/3)=\prod_{s=2}^{\infty}\frac{(s+2)(s-1)}{s(s+1)}=\frac{1}{3},$$
$$I_4=\frac{16}{3},\quad I_6=\frac{128}{15} \tag{85}$$

yields the result

$$\nu(E_B)=\frac{4}{\pi}\,\frac{|E_B|}{a_0^dW^2}\exp\left\{-\frac{8|E_B|^{3/2}}{3a_0^dW^2}\right\}, \tag{86}$$

which agrees with the exact solution due to Halperin.[10,30]

ACKNOWLEDGMENTS

I am grateful to the participants of the seminars at the Institute of Semiconductor Physics (IFP) and the Physics Institute of the Academy of Sciences (FIAN) for their interest in this work.

This work was carried out with the financial support of the International Science Foundation and the Russian Government (Grants No. MON 000 and No. MON 300) and the Russian Fund for Fundamental Research (Project No. 96-02-19527).

[1)]The left-hand sides of the final formulas (16) of Ref. 26 contain obvious typographical errors; substitution in expression (83) of the numerical values of the parameters obtained in Ref. 25 yields the coefficients shown in Ref. 26 on the right-hand sides of (16).

[1] A. Aharony and Y. Imry, J. Phys. C **10**, L487 (1977).
[2] A. L. Éfros, Usp. Fiz. Nauk **126**, 41 (1978) [Sov. Phys. Usp. **21**, 746 (1978)].
[3] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[4] D. J. Thouless, Phys. Rep. **13**, 92 (1974).
[5] M. V. Sadovskiĭ, Usp. Fiz. Nauk **133**, 223 (1981) [Sov. Phys. Usp. **24**, 96 (1981)].
[6] D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).
[7] M. V. Sadovskii, Sov. Sci. Rev. A. Phys. **7**, 1 (1986).
[8] D. Vollhardt and P. Wölfle, Phys. Rev. B **22**, 4666 (1980).
[9] I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].
[10] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Wiley, New York (1988).
[11] S.-K. Ma, *Modern Theory of Critical Phenomena*, Benjamin, Reading, Mass. (1976).
[12] A. Nitzan, K. F. Freed, and M. N. Cohen, Phys. Rev. B **15**, 4476 (1977).
[13] K. G. Wilson and J. Kogut, Phys. Rep. **12C**, 75 (1974).
[14] I. M. Suslov, Zh. Éksp. Teor. Fiz. **102**, 1951 (1992) [Sov. Phys. JETP **75**, 1049 (1992)].
[15] I. M. Suslov, Zh. Éksp. Teor. Fiz. **106**, 560 (1994) [JETP **79**, 307 (1994)].
[16] I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 220 (1997) [JETP **84**, (1997)].
[17] I. M. Suslov, JETP Lett. **63**, (1996).
[18] L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
[19] A. I. Larkin and D. E. Khmelnitskiĭ, Zh. Éksp. Teor. Fiz. **56**, 2087 (1969) [Sov. Phys. JETP **29**, 1123 (1969)]; A. M. Polyakov, Zh. Éksp. Teor. Fiz. **57**, 271 (1969) [Sov. Phys. JETP **30**, 151 (1969)].
[20] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, transl. of 1st Russ. ed., Pergamon Press, Oxford (1979); 2nd Russ. ed., Nauka, Moscow (1982).
[21] E. Brézin, J. C. Le Guillou, and J. Zinn-Justin, in *Phase Transitions and Critical Phenomena*, Vol. VI, C. Domb and M. S. Green (eds.), Academic Press, New York (1976).

[22] A. O. Gel'fond, *Calculus of Finite Differences* [in Russian], Nauka, Moscow (1967).
[23] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw–Hill, New York (1961).
[24] S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **66**, 647 (1974) [Sov. Phys. JETP **39**, 312 (1974)].
[25] E. Brézin and G. Parisi, J. Stat. Phys. **19**, 269 (1978).
[26] E. Brézin and G. Parisi, J. Phys. C **13**, L307 (1980).
[27] J. L. Cardy, J. Phys. C **11**, L321 (1978).
[28] I. M. Lifshitz, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 549 (1964)].
[29] E. Brézin, J. C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **15**, 1544 (1977).
[30] B. I. Halperin, Phys. Rev. **139**, A104 (1965).

Translated by Paul F. Schippnick

The effective action of W_3-gravity

D. P. Karakhanyan*)

Erevan Institute of Physics, 375036 Erevan, Republic of Armenia
(Submitted 24 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1537–1553 (May 1997)

A new method for integrating anomalous Ward identities and finding the effective action is proposed. Two-dimensional supergravity and W_3-gravity are used as examples to demonstrate its potential. An operator is introduced that associates each physical quantity with a Ward identity, i.e., a quantity that is transformed without anomalous terms and can be nullified in a consistent manner. A covariant form of the action for matter fields interacting with a gravitational and W_3-gravitational background is proposed. © *1997 American Institute of Physics.* [S1063-7761(97)00105-4] © *1997 American Institute of Physics.*

1. INTRODUCTION

The tremendous upsurge of interest in W_N-algebras[1] that followed their discovery by Zamolodchikov can be explained by the fact that the basic relationships in W_N-algebras, in contrast to those of ordinary Lie algebras, are multilinear and that the mathematical aspects had not been systematically studied. A big achievement in this area of research was the use of the Drinfel'd–Sokolov reduction scheme,[2] which reduces W-algebras to Lie algebras and relates them to the second Hamiltonian structure of the generalized Korteweg–de Vries hierarchies; W_N-algebras contain the Virasoro algebra as a subalgebra. In the context of string theory, the appearance of the latter is a reflection of invariance under reparametrization of the string world surface. The extension of this symmetry to invariance under the W-gravity transformations leads to the theory of W-strings in the Polyakov approach, i.e., to the theory of the interaction of matter fields with gravitational (spin-2) and W-gravitational (spin-N) background fields. Thus, symmetry under transformations of W-gravity is the leading principle that makes it possible to write the interaction for fields with spins $\geqslant 2$, at least in two dimensions.

However, progress in this area of research was fraught with considerable difficulties. First the chiral theory of the interaction of matter and W-gravity was formulated by Hull.[3] Then Schoutens *et al.*[4] generalized the theory to the nonchiral case but encountered significant technical difficulties: the action in the theory proved to be infinitely nonlinear in the matter fields and nonlocal, so that any further analysis is extremely complicated. By calculating the functional integral over the matter fields with a central charge c interacting with W_3-gravity Schoutens *et al.*[5] also found the induced action of W_3-gravity in the form of a $1/c$-expansion. The same researchers (see Ref. 6) found the induced action of chiral W_3-gravity, a direct analog if the Polyakov's action[7] for ordinary gravity, by integrating the anomaly in the limit $c \to \infty$.

Clarification of the geometric meaning of W-transformations would help W-gravity studies considerably. This aspect was studied by Figueroa-O'Farrill *et al.* and Hull.[8]

At present it is generally hoped that W-gravity studies will help to overcome the strong-coupling barrier $c=1$ for a system consisting of conformal matter and two-dimensional gravitation, which will probably make it possible to avoid the fractional dimensionality established by Knizhnik *et al.*[9] for quantum gravity in the weak-coupling mode. Direct generalization of the results of Ref. 9 to W-gravity in the absence of matter fields was done by Matsuo.[10]

The present investigation develops a method for integrating two-dimensional anomalous Ward identities. Its application is illustrated by examples of two-dimensional gravity, supergravity, and W_3-gravity. The essence of the method consists in the following. By expressing anomalous currents in terms of free fields via bosonization formulas, we can lower the order of these differential equations and integrate them. The resulting effective action reproduces the anomaly correctly. When the regularization scheme changes, local counterterms are added to the nonlocal effective action, and the emergence of these counterterms changes the form and symmetry of the Ward identities. The bosonized fields, being free in one regularization scheme, in another scheme are related by the fact that they satisfy certain Ward identities. When the chiral Weyl-invariant regularization scheme is replaced by the diffeomorphism-invariant scheme, local counterterms are added in such a way that the kinetic part of the effective action becomes invariant both under diffeomorphisms and under Weyl transformations. The remaining (topological) part of the effective action is fixed by the requirement that the total action, being diffeomorphism-invariant under Weyl transformations, be symmetric in the quantum or projective sense, i.e., is transformed as a 1-cocycle.

In Secs. 2 and 3 the application of this method is demonstrated using the well-known examples of ordinary and $(N=1)$-supergravity, and a differential operator R is introduced, which with each physical quantity associates its Ward identity. The operator is actually a Slavnov operator, which was studied by Zucchini[11] in connection with two-dimensional gravity in conjunction with an auxiliary inhomogeneous term that destroys the anomalous contribution in the transformation law.

In Sec. 4 these calculations are generalized to the case of chiral W_3-gravity. It is found that the result is in full agreement with that of Ooguri *et al.*[6]

Finally, Sec. 5 deals with the covariant action of matter

interacting with nonchiral W_3-gravity. In addition to exhibiting parametrization symmetry and W-diffeomorphism symmetry, this action is W-Weyl invariant and can serve as the kinetic part of the effective action of W_3-gravity calculated in the diffeomorphism-invariant regularization scheme.

2. TWO-DIMENSIONAL GRAVITY

The Polyakov action, which was derived in Ref. 7 as the effective action induced by chiral matter interacting with two-dimensional gravity, is the determinant of the two-dimensional Laplacian calculated in a regularization scheme that conserves Weyl symmetry and half the reparametrization symmetry. The presence of a conformal anomaly manifests induces an explicit dependence of the Polyakov action on one of the reparametrization functions. In other words, this effective action can be calculated by integrating the appropriate variational equation, the Ward identity.

The Ward identity of two-dimensional gravitation theory in the lightlike gauge is well known:

$$R_T=(\overline{\partial}-h\partial-2\partial h)T-\partial^3 h=0. \tag{2.1}$$

It expresses the anomalous conservation of the system's energy–momentum tensor T. The field h in this expression denotes the nonvanishing metric component that remains after the lightlike gauge is specified. It is covariant under the transformations

$$\begin{aligned}&\delta h=(\overline{\partial}-h\partial+\partial h)\epsilon,\\ &\delta T=(\partial^3+2T\partial+\partial T)\epsilon,\end{aligned} \tag{2.2}$$

i.e.

$$\delta_\epsilon R_T=(\epsilon\partial+2\,\partial\epsilon)R_T. \tag{2.3}$$

Equation (2.3) expresses the Wess–Zumino self-consistency condition. If we use the bosonization formula and parametrize the energy–momentum tensor via a scalar field,

$$T=\partial^2\varphi-\tfrac{1}{2}(\partial\varphi)^2, \tag{2.4}$$

the order of the anomalous term in (2.1) can be reduced:

$$\begin{aligned}&R_\varphi=(\overline{\partial}-h\partial)\varphi-\partial h=0,\\ &\delta_\epsilon R_\varphi=\epsilon\partial R_\varphi.\end{aligned} \tag{2.5}$$

Comparing (2.1) with (2.4) and (2.5), we obtain

$$R_T=\partial^2 R_\varphi-\partial\varphi\partial R_\varphi. \tag{2.6}$$

The transformation law for the scalar field φ is also anomalous:

$$\delta\varphi=\partial\epsilon+\epsilon\partial\varphi. \tag{2.7}$$

If for the field φ we postulate the free-field Poisson bracket,

$$\{\partial\varphi(x);\partial\varphi(x')\}=\delta'(x-x'), \tag{2.8}$$

the ϵ-variation of any quantity A can be determined by its Poisson bracket with the energy–momentum tensor:

$$\delta_\epsilon A(x)=\int d^2x'\,\epsilon(x')\{T(x');A(x)\}. \tag{2.9}$$

Clearly, this definition for the field φ coincides with (2.7). The bracket of tensor T with itself is

$$-\{T(x);T(x')\}=\delta'''(x-x')+[T(x)+T(x')]\delta'(x-x'). \tag{2.10}$$

Although the energy–momentum tensor can be expressed in terms of φ, there is no way in which we can express the gauge field h in Eq. (2.5) in terms of φ in a local manner. To do this we must introduce a quantity that satisfies the regular Ward identity, i.e., a quantity that transforms as a scalar. The anomaly can be removed from the Ward identity by introducing a scalar field f in the following way:

$$\varphi=\ln\partial f. \tag{2.11}$$

The transformation law for f and the corresponding Ward identity have the form

$$\begin{aligned}&\delta_\epsilon f=\epsilon\partial f,\\ &R_f=(\overline{\partial}-h\partial)f=0,\\ &\delta_\epsilon R_f=\epsilon\partial R_f.\end{aligned} \tag{2.12}$$

Now, when all the quantities are expressed in terms of the function f locally, we can integrate the variational equation for the effective action of the theory, which can also be expressed in terms of f locally and is given by the Polyakov formula.[7] Detailed calculations are given in Sec. 3 for the more interesting case of supergravity.

Comparing Eqs. (2.3), (2.6) and (2.12), we see that the gauge variation δ and the Ward identities R are commutative operations on the fields T, φ, and f.

The relationship between R_f, R_φ, and R_T is specified by the following formulas:

$$\begin{aligned}&R_\varphi=\frac{\partial R_f}{\partial f},\\ &R_T=(\partial^3+2T\partial+\partial T)\frac{R_f}{\partial f}.\end{aligned} \tag{2.13}$$

We see that the operator R associates with each physical quantity X a covariant expression R_X, its Ward identity, which in view of its covariance under gauge transformations can be consistently made to vanish. But since the theory lacks quantities of the required dimensionality, this expression must be set to zero. Comparing (2.13) with (2.11) and (2.6) with (2.4), we see that R obeys the Newton–Leibniz rule. This property of R makes it possible to write the Ward identities for the correlation functions of the fields T, φ, etc. immediately.

If we apply the Legendre transformation

$$Z[h]=\Gamma[h]+\int d^2x\;Th, \tag{2.14}$$

Eq. (2.1) can be written as

$$(\partial^3+2T\partial+\partial T)\frac{\delta\Gamma}{\delta T}=-\overline{\partial T}, \tag{2.15}$$

where the Bol operator[12] on the left-hand side is the covariant form of ∂^3 on an arbitrary Riemann surface, and contains a projective connection, for which we may take T. This no-

tation expresses the covariance of the Ward identity (2.1), just as the Wess–Zumino self-consistency condition (2.3) does.

3. SIMPLE SUPERGRAVITY

In this section we generalize all the ideas of Sec. 2 to the case of simple supergravity and calculate the effective action of the theory.

Polyakov's result was generalized by Polyakov and Zamolodchikov[13] to the case of (1.0)-supergravity. The corresponding generalization of the Polyakov action represents the effective action obtained in a regularization scheme that conserves the Weyl and super-Weyl symmetries and half the supercoordinate symmetry.[14] The nontrivial dependence of the action on the other coordinate functions (odd- and even-parity) is determined by a superconformal anomaly.

The Ward identities of two-dimensional supergravity in the lightlike gauge can be written as[6]

$$\begin{aligned} R_T &= (\overline{\partial} - h\partial - 2\partial h)T - (\tfrac{1}{2}\chi\partial + \tfrac{3}{2}\partial\chi)S - \partial^3 h = 0, \\ R_S &= (\overline{\partial} - h\partial - \tfrac{3}{2}\partial h)S - \tfrac{1}{2}\chi T - \partial^2\chi = 0. \end{aligned} \tag{3.1}$$

They are covariant under the transformations

$$\begin{aligned} \delta h &= (\overline{\partial} - h\partial + \partial h)\epsilon + \tfrac{1}{2}\kappa\chi, \\ \delta\chi &= (\overline{\partial} - h\partial + \tfrac{1}{2}\partial h)\kappa + (\epsilon\partial - \tfrac{1}{2}\partial\epsilon)\chi, \\ \delta T &= (\partial^3 + 2T\partial + \partial T)\epsilon + (\tfrac{1}{2}\kappa\partial + \tfrac{3}{2}\partial\kappa)S, \\ \delta S &= (\epsilon\partial + \tfrac{3}{2}\partial\epsilon)S + (\partial^2 + \tfrac{1}{2}T)\kappa, \end{aligned} \tag{3.2}$$

i.e.,

$$\begin{aligned} \delta R_T &= (\epsilon\partial + 2\partial\epsilon)R_T + (\tfrac{1}{2}\kappa\partial + \tfrac{3}{2}\partial\kappa)R_S, \\ \delta R_S &= (\epsilon\partial + \tfrac{3}{2}\partial\epsilon)R_S + \tfrac{1}{2}\kappa R_T, \end{aligned} \tag{3.3}$$

which means that the Ward identity R_A transforms in the same way as the quantity A but without the anomalous terms.

Going over to the scalar multiplet of matter fields (φ,λ), with

$$\begin{aligned} \delta\varphi &= (\partial + \partial\varphi)\epsilon + \tfrac{1}{2}\kappa\lambda, \\ \delta\lambda &= (\epsilon\partial + \tfrac{1}{2}\partial\epsilon)\lambda + (\partial + \tfrac{1}{2}\partial\varphi)\kappa, \end{aligned} \tag{3.4}$$

which are related to the current fields by the rule

$$\begin{aligned} T &= \partial^2\varphi - \tfrac{1}{2}(\partial\varphi)^2 + \tfrac{1}{2}\lambda d\lambda, \\ S &= \partial\lambda - \tfrac{1}{2}\lambda\partial\varphi, \end{aligned} \tag{3.5}$$

we can reduce the order of the derivatives in (3.1). The operator R acts on the fields φ and λ in the following manner:

$$\begin{aligned} R_\varphi &= (\overline{\partial} - h\partial)\varphi - \tfrac{1}{2}\chi\lambda - \partial h, \\ R_\lambda &= (\overline{\partial} - h\partial - \tfrac{1}{2}\partial h)\lambda - \tfrac{1}{2}\chi\partial\varphi - \partial\chi. \end{aligned} \tag{3.6}$$

We see that the gauge fields h and χ cannot be expressed in terms of φ and λ locally, a situation resembling that of Sec. 2.

The fields φ and λ form an algebra of free fields in Poisson brackets:

$$\begin{aligned} \{\partial\varphi(x);\partial\varphi(x')\} &= \delta'(x-x'), \\ \{\lambda(x);\lambda(x')\} &= -\delta(x-x'). \end{aligned} \tag{3.7}$$

Then, with respect to this bracket, Eqs. (3.5) suggest the existence of the following algebra for the current fields:

$$\begin{aligned} -\{T(x);T(x')\} &= \delta'''(x-x') + (T(x)+T(x'))\delta'(x-x'), \\ -\{T(x);S(x')\} &= (S(x)+\tfrac{1}{2}S(x'))\delta'(x-x'), \\ -\{S(x);S(x')\} &= \delta''(x-x') - \tfrac{1}{2}T(x)\delta(x-x'). \end{aligned} \tag{3.8}$$

To parametrize the gauge fields in a convenient manner, we must introduce a scalar multiplet (f,ψ) without anomalous dimensionality,

$$\begin{aligned} \delta f &= \epsilon\partial f + \tfrac{1}{2}\kappa\psi, \\ \delta\psi &= \epsilon\partial\psi + \tfrac{1}{2}\partial\epsilon\psi + \tfrac{1}{2}\kappa\partial f, \\ R_f &= (\overline{\partial} - h\partial)f - \tfrac{1}{2}\chi\psi, \\ R_\psi &= (\overline{\partial} - h\partial - \tfrac{1}{2}\partial h)\psi - \tfrac{1}{2}\chi\partial f. \end{aligned} \tag{3.9}$$

This multiplet is related to the matter fields as follows:

$$\begin{aligned} \varphi &= \ln\,\partial f + \frac{\psi\partial\psi}{(\partial f)^2}, \\ \lambda &= 2\frac{\partial\psi}{\partial f} - \psi\frac{\partial^2 f}{(\partial f)^2}. \end{aligned} \tag{3.10}$$

There is no simple way in which we can deduce such a complicated relationship from the condition that the appropriate terms appear in the transformation laws. However, the problem can be simplified if superfields are introduced.

Since the superfield formulation of chiral supergravity contains no auxiliary fields, the meaning of all previous expressions is not altered when we go over to superfields:

$$\begin{aligned} R_U &= (\overline{\partial} - H\partial - \tfrac{1}{2}DHD - \tfrac{3}{2}\partial H)U - \partial^2 DH, \\ \delta U &= (D\partial^2 + \tfrac{3}{2}U\partial - \tfrac{1}{2}DUD + \partial U)E, \\ \delta H &= (\overline{\partial} - H\partial - \tfrac{1}{2}DHD + \partial H)E, \end{aligned} \tag{3.11}$$

where $U = S + \theta T$, $H = h + \theta\chi$, $D = \partial_\theta + \theta\partial$, and $E = \epsilon + \theta\kappa$, with θ the anticommuting coordinate. Then the current superfield U is related to the matter superfield $\Phi = \varphi + \theta\lambda$ as follows:

$$U = D\partial\Phi - \tfrac{1}{2}D\Phi\partial\Phi, \tag{3.12}$$

and accordingly,

$$\begin{aligned} \delta\Phi &= \partial E + E\partial\Phi + \tfrac{1}{2}DED\Phi, \\ R_U &= D\partial R_\Phi - \tfrac{1}{2}D\Phi\partial R_\Phi - \tfrac{1}{2}\partial\Phi DR_\Phi. \end{aligned} \tag{3.13}$$

The scalar multiplet $F = f + \theta\psi$, with

$$\delta F = E\partial F + \tfrac{1}{2}DEDF, \tag{3.14}$$

is related to the superfield Φ by

$$\Phi = \ln\,\partial F + \frac{DF}{\partial F}D\,\ln\,\partial F, \tag{3.15}$$

while the relationship between the corresponding Ward identities is

$$R_\Phi = \left[\frac{DF}{(\partial F)^2}\partial D + \left(\frac{1}{\partial F} - 2\frac{DFD\partial F}{(\partial F)^3} \right)\partial - \frac{D\partial F}{(\partial F)^2} D \right] R_F . \tag{3.16}$$

The formulas that link the current Ward identities with R_f and R_ψ are

$$\begin{aligned} R_T &= (\partial^3 + 2T\partial + \partial T)\left(\frac{R_f}{\partial f} - R_f \frac{\psi\partial\psi}{(\partial f)^3} + \frac{\psi R_\psi}{(\partial f)^2} \right) - \left(\frac{3}{2} S\partial \right. \\ &\quad \left. + \frac{1}{2}\partial S \right)\left(2\frac{R_\psi}{\partial f} - \frac{\psi}{\partial f}\partial\left(\frac{R_f}{\partial f} \right) - 2\frac{\partial\psi R_f}{(\partial f)^2} \right), \\ R_S &= \left(\partial^2 + \frac{1}{2}T \right)\left(2\frac{R_\psi}{\partial f} - \frac{\psi}{\partial f}\partial\left(\frac{R_f}{\partial f} \right) - 2\frac{\partial\psi R_f}{(\partial f)^2} \right) \\ &\quad - \left(\frac{3}{2} S\partial + \partial S \right)\left(\frac{R_f}{\partial f} - R_f \frac{\psi\partial\psi}{(\partial f)^3} + \frac{\psi R_\psi}{(\partial f)^2} \right). \end{aligned} \tag{3.17}$$

If we now use the Legendre transformations to proceed from the partition function to the effective action,

$$\Gamma[T,S] = Z[h,\chi] - \int d^2x (hT + \chi S), \tag{3.18}$$

the Ward identity becomes

$$(\partial^2 D + 3U\partial + DUD + 2\partial U)\frac{\delta\Gamma}{\delta U} = 0 \tag{3.19}$$

or, in components,

$$\begin{aligned} &(\partial^3 + 2T\partial + \partial T)\frac{\delta\Gamma}{\delta T} + \left(\frac{3}{2} S\partial + \frac{1}{2}\partial S \right)\frac{\delta\Gamma}{\delta S} = -\overline{\partial T}, \\ &\left(\partial^2 + \frac{1}{2}T \right)\frac{\delta\Gamma}{\delta T} + \left(\frac{3}{2} S\partial + \partial S \right)\frac{\delta\Gamma}{\delta S} = -\overline{\partial S}, \end{aligned} \tag{3.20}$$

i.e., there emerges a supersymmetric Bol operator, which has also been described in Ref. 12. In this form the covariance of Ward identities under the gauge transformations (3.2), which is equivalent to the Wess–Zumino conditions for an anomaly, becomes explicit.

Now let us turn to the problem of finding the partition function of the theory:

$$\begin{aligned} \delta Z &= \int d^2x (T\delta h + S\delta\chi) = \int d^2x d\theta U \delta H \\ &= -\int d^2x d\theta E\left(\overline{\partial} - H\partial - \frac{1}{2} DHD + \frac{1}{2}\partial H \right) U. \end{aligned} \tag{3.21}$$

We see that the integrand is the Ward identity R_U without the anomalous term. In the chiral scheme, i.e., a regularization scheme that conserves half the reparametrization symmetry and the Weyl symmetry as well as their superpartners, the fields φ and λ are related by (3.10), and the corresponding Ward identities R_φ and R_λ vanish. On the other hand, in a regularization scheme that preserves supercoordinate symmetry the group parameters f and ψ vanish and the fields φ and λ are unconstrained, but the Ward identities R_φ and R_λ still play an important role. Multiplying R_U by E and integrating by parts, we obtain

$$\begin{aligned} \int d^2x\, d\theta\, R_U E &= \int E\left(D\partial - \frac{1}{2} D\Phi\partial - \frac{1}{2}\partial\Phi D \right) R_\Phi \\ &= \int d^2x\, d\theta\, R_\Phi \left(D\partial E + \frac{1}{2}\partial(ED\Phi) \right. \\ &\quad \left. + \frac{1}{2} D(E\partial\Phi) \right) \\ &= \int d^2x\, d\theta\, R_\Phi D\delta\Phi. \end{aligned} \tag{3.22}$$

If we ''err'' twice, i.e., take (3.22) for δZ rather than (3.21) and ignore the relationship between H and Φ, expressed by the fact that R_Φ is zero, we can integrate this variational equation and arrive at the following expression for $Z[H]$:

$$Z[H] = -\frac{1}{2}\int d^2x d\theta (\overline{\partial}\Phi - H\partial\Phi - 2\partial H) D\Phi. \tag{3.23}$$

The fact that (3.17) reproduces the anomaly correctly can easily be verified. Thus, assuming that the superfield Φ is independent, we can reproduce the anomaly by directly adding the appropriate term to the action.

This conclusion agrees with our ideas about the ''transfer'' of the anomaly from one regularization scheme to another. A detailed description of the process in which a conformal anomaly is transformed into a gravitational anomaly in the two-dimensional gravitation theory can be found in Ref. 15.

4. W_3-GRAVITY

The difference between the theory of W_3-gravity and the above cases is that the chiral formulation of this theory is not only more convenient but is also the only one amenable to quantum analysis. The nonchiral version formulated in Ref. 4 contains an infinite number of arbitrary matter fields and is too complicated even at the classical level.

The Ward identities of chiral W_3-gravity are

$$\begin{aligned} R_T &= (\overline{\partial} - h\partial - 2\partial h)T - (2b\partial + 3\partial b)W - \partial^3 h, \\ R_W &= (\overline{\partial} - h\partial - 3\partial h)W + (2b\partial^3 + 9\partial B\partial^2 + 15\partial^2 b\partial \\ &\quad + 10\,\partial^3 b + 16bT\partial + 16\,\partial bT)T - \partial^5 h. \end{aligned} \tag{4.1}$$

Here b denotes the single nonvanishing component of a third-rank symmetric tensor—the gauge field of W-gravity, the partner of the metric in the multiplet—and W denotes the corresponding spin-3 current, the partner of the energy–momentum tensor. The chiral general-coordinate and W-transformations have the form

$$\begin{aligned} \delta T &= (\partial^3 + 2T\partial + \partial T)\epsilon + 3W\partial\lambda + 2\partial W\lambda, \\ \delta W &= \partial W\epsilon + 3W\partial\epsilon + (\partial^5 + 10T\partial^3 + 15\partial T\partial^2 + 9\partial^2 T\partial \\ &\quad + 2\partial^3 T + 16T^2\partial + 16T\partial T)\lambda, \\ \delta h &= (\overline{\partial} - h\partial + \partial h)\epsilon + 2\lambda\partial^3 b - 3\,\partial\lambda\partial^2 b + 3\partial^2\lambda\partial b \end{aligned}$$

$$-2\partial^3 b+16T(\lambda\partial b-b\partial\lambda),$$

$$\delta b=\epsilon\partial b-2\partial\epsilon b+(\overline{\partial}-h\partial+2\,\partial h)\lambda. \tag{4.2}$$

The quantities R_T and R_W are covariant under the transformations (4.2), i.e.,

$$\begin{aligned}\delta R_T&=(\epsilon\partial+2\partial\epsilon)R_T+(2\lambda\partial+3\partial\lambda)R_W,\\ \delta R_W&=(\epsilon\partial+3\partial\epsilon)R_W+(2\lambda\partial^3+9\partial\lambda\partial^2+15\partial^2\lambda\partial\\ &\quad+10\partial^3\lambda+32T\partial\lambda+16T\lambda\partial+16\lambda\partial T)R_T.\end{aligned} \tag{4.3}$$

Thus, taking R as the differential operator, we conclude that (4.3) yields a universal relation

$$[\delta,R]=0 \tag{4.4}$$

on the current fields T and W.

The transformations (4.2) of an arbitrary quantity A are generated by the currents T and W via Poisson brackets:

$$\delta A=\int d^2x[\epsilon(x)\{T(x);A\}+\lambda(x)\{W(x);A\}]. \tag{4.5}$$

In terms of these brackets, the transformations (4.2) themselves become

$$\begin{aligned}-\{T(x);T(x')\}&=\delta'''(x-x')+[T(x)+T(x')]\delta'(x-x'),\\ -\{T(x);W(x')\}&=[W(x)+2W(x')]\delta'(x-x'),\\ -\{W(x);T(x')\}&=[2W(x)+W(x')]\delta'(x-x'),\\ -\{W(x);W(x')\}&=\delta^{\mathrm{V}}(x-x')+5[T(x)+T(x')]\delta'''(x-x')\\ &\quad+8[T^2(x)+T^2(x')]\delta'(x-x')\\ &\quad-3[T''(x)+T''(x')]\delta'(x-x').\end{aligned} \tag{4.6}$$

This algebra can be reproduced by expressing the current fields in terms of the matter fields φ and ψ, which induce the algebra of free fields,

$$\begin{aligned}\{\varphi'(x);\varphi'(x')\}&=\delta'(x-x'),\\ \{\varphi'(x);\psi'(x')\}&=0,\\ \{\psi'(x);\psi'(x')\}&=-\delta'(x-x'),\end{aligned} \tag{4.7}$$

if we define $T(x)$ and $W(x)$ in the following manner:

$$\begin{aligned}T(x)&=\partial^2\varphi-\tfrac{1}{2}(\partial\varphi)^2+\tfrac{1}{2}(\partial\psi)^2,\\ W(x)&=\partial^3\psi+3\partial\varphi\partial^2\psi-\partial^2\varphi\partial\psi+2(\partial\varphi)^2\partial\psi+\tfrac{2}{3}(\partial\psi)^3.\end{aligned} \tag{4.8}$$

This corresponds to the following transformation law for the matter fields:

$$\begin{aligned}\delta\varphi&=\partial\epsilon+\partial\varphi\epsilon-4\lambda\partial\varphi\partial\psi+2\lambda\partial^2\psi-\partial\lambda\partial\psi,\\ \delta\psi&=\epsilon\partial\psi+\partial^2\lambda+3\partial\lambda\partial\varphi+2\lambda[\partial^2\varphi+(\partial\varphi)^2+(\partial\psi)^2].\end{aligned} \tag{4.9}$$

The anomalous equations of motion of the matter multiplet have the form

$$\begin{aligned}R_\varphi&=\overline{\partial}\varphi-h\partial\varphi+\partial b\partial\psi-2b(\partial^2\psi-2\partial\varphi\partial\psi)-\partial h,\\ R_\psi&=\overline{\partial}\psi-h\partial\psi-3\,\partial b\partial\varphi-2b[\partial^2\varphi+(\partial\varphi)^2+(\partial\psi)^2]\\ &\quad-\partial^2 b.\end{aligned} \tag{4.10}$$

With respect to ϵ- and λ-diffeomorphisms, these relationships are also covariant:

$$\begin{aligned}\delta R_\varphi&=(\epsilon\partial-4\lambda\partial\psi\partial)R_\varphi+(2\lambda\partial^2-\partial\lambda-4\lambda\partial\varphi\partial)R_\psi,\\ \delta R_\psi&=(\epsilon\partial+4\lambda\partial\psi)R_\psi+(3\partial\lambda\partial+2\lambda\partial^2+4\lambda\partial\varphi\partial)R_\varphi.\end{aligned} \tag{4.11}$$

This also establishes the validity of Eq. (4.4) when matter acts on the multiplet.

The Ward identities (4.1) can easily be transformed into

$$\begin{aligned}R_T&=(\partial^2-\partial\varphi)R_\varphi-\partial\psi\partial R_\psi,\\ R_W&=(4\,\partial\varphi\partial\partial\psi\partial-\partial\psi\partial^2-3\,\partial^2\psi\partial)R_\varphi+[\partial^3-3\,\partial\varphi\partial^2\\ &\quad+2(\partial\varphi)^2\partial+2(\partial\psi)^2\partial]R_\psi.\end{aligned} \tag{4.12}$$

The ϵ- and λ-transformations constitute a closed algebra on the multiplets of currents ($\{T,W\}$), matter fields ($\{\varphi,\psi\}$), and gauge fields ($\{h,b\}$):

$$\begin{aligned}&[\delta(\epsilon_1),\delta(\epsilon_2)]=\delta(\epsilon_3=\epsilon_2\partial\epsilon_1-\epsilon_1\partial\epsilon_2),\\ &[\delta(\epsilon_1),\delta(\lambda_2)]=\delta(\lambda_3=2\lambda_2\partial\epsilon_1-\epsilon_2\partial\lambda_1),\\ &[\delta(\lambda_1),\delta(\lambda_2)]=\delta(\epsilon_3=16T(\lambda_2\partial\lambda_1-\lambda_1\partial\lambda_2)+2\lambda_2\partial^3\lambda_1\\ &\qquad-3\,\partial\lambda_2\partial^2\lambda_1+3\,\partial^2\lambda_2\partial\lambda_1-2\,\partial^3\lambda_2\lambda_1).\end{aligned} \tag{4.13}$$

The partition function of the theory is calculated in the same way as in supergravitation theory: by multiplying Eqs. (4.12), respectively, by ϵ and λ, we obtain

$$\int d^2x(\epsilon R_T+\lambda R_W)=\int d^2x(R_\varphi\delta\partial\varphi-R_\psi\delta\partial\psi). \tag{4.14}$$

If the Ward identities (4.1) were to have no anomalous terms, the left-hand side of Eq. (4.14) would be the variation of the partition function multiplied by minus one. If, in addition, the fields φ and ψ were to be free and Eqs. (4.10) were not to link them with the gauge fields h and b, the right-hand side of Eq. (4.14) would be the total variation of the following expression:

$$\begin{aligned}Z[h,b]&=\int d^2x\Bigg(\frac{1}{2}\partial\varphi(\overline{\partial}\varphi-h\partial\varphi)-\frac{1}{2}\partial\psi(\overline{\partial}\psi-h\partial\psi)\\ &\quad-\partial h\partial\varphi+\partial^2 b\partial\psi+b\Bigg[2(\partial\varphi)^2\partial\psi-\partial^2\varphi\partial\psi\\ &\quad-3\,\partial\varphi\partial^2\psi+\frac{2}{3}(\partial\psi)^3\Bigg]\Bigg)\\ &=\int d^2x\left[\frac{1}{2}(\overline{\partial}\varphi\partial\varphi-\overline{\partial}\psi\partial\psi)+hT+bW\right].\end{aligned} \tag{4.15}$$

Equation (4.15) is the action of the matter fields interacting with two-dimensional chiral gravity and W_3-gravity. The variation of (4.15) with respect to (4.2) and (4.9) is

$$\delta Z=\int d^2x\ [h\partial^3\epsilon+b\partial^5\lambda+16\lambda(T^2\partial b+bT\partial T)]. \tag{4.16}$$

This expression differs from that for the quantum anomaly of the minimum type by the presence of terms quadratic in T. This discrepancy is due to the differences in defining the

transformation law for the field h under λ-diffeomorphisms. The transformation law (4.2) is motivated by the closure of the algebra (4.13) on the fields h and b and by the Wess–Zumino self-consistency condition, with Eqs. (4.3) being valid.

The following line of reasoning can motivate another definition of $\delta_\lambda h$: if we perform the Legendre transformation

$$Z[h,b]=\Gamma[T,W]+\int d^2x\ (hT+bW), \tag{4.17}$$

the expressions for R_T and R_W acquire the Bol operators L_3 and L_5:

$$R_T=\overline{\partial T}+(\partial^3+2T\partial+\partial T)\frac{\delta\Gamma}{\delta T}+(3W\partial+2\,\partial W)\frac{\delta\Gamma}{\delta W}, \tag{4.18}$$

$$R_W=\overline{\partial W}+(3W\partial+\partial W)\frac{\delta\Gamma}{\delta T}+(\partial^5+10T\partial^3+15\partial T\partial^2 +(9\partial^2T+16T^2)\partial+(2\partial^3T+16T\partial T))\frac{\delta\Gamma}{\delta W}.$$

Thus,

$$\delta Z=\int d^2x(T\delta h+W\delta b) =-\int d^2x[\epsilon\partial^3h+\lambda\partial^5b-16\lambda(T^2+bT\partial T)]. \tag{4.19}$$

To obtain an anomaly of the minimum type it appears reasonable to define $\delta_\lambda h$ with $\delta_{\text{extra}}h=-8T(\lambda\partial b-b\partial\lambda)$.

The anomaly can be completely removed from the transformation laws and the Ward identities if we transform to variables (f,g), which form a scalar multiplet:

$$\varphi=\ln\partial f+\tfrac{1}{2}\ln(1-\eth^2g), \qquad \psi=\gamma^{-1}\ln(1+\eth^2g), \tag{4.20}$$

where

$$\eth\equiv\frac{1}{\partial f}\partial \quad\text{and}\quad \gamma^2=-12.$$

The transformation law for the fields f and g is

$$\delta f=\epsilon\partial f-\gamma\left[\lambda\partial^2f+\frac{1}{2}\partial\lambda\partial f+\frac{2}{3}\lambda\partial f\partial\ln(1+\eth^2g)\right], \tag{4.21}$$

$$\delta g=\epsilon\partial g-\frac{1}{2}\gamma\partial\lambda\partial g-\gamma\lambda\left[(\partial f)^2-\partial^2g+\frac{2}{3}\partial g\partial\ln(1+\eth^2g)+2\partial g\frac{\partial^2f}{\partial f}\right].$$

Accordingly, the Ward identities are

$$R_f=\overline{\partial f}-h\partial f+\frac{\gamma}{2}\partial b\partial f+\gamma b\left[\partial^2f+\frac{2}{3}\partial f\partial\ln(1+\eth^2g)\right], \tag{4.22}$$

$$R_g=\overline{\partial g}-h\partial g+\frac{\gamma}{2}\partial b\partial f+\gamma b\left[2\partial g\frac{\partial^2f}{\partial f}+\frac{2}{3}\times\partial f\partial\ln(1+\eth^2g)-(\partial f)^2-\partial^2g\right].$$

Setting R_f and R_g to zero, we can express the gauge multiplet in terms of f and g:

$$b=\gamma^{-1}\frac{\overline{\partial g}-(\overline{\partial f}/\partial f)\partial g}{(\partial f)^2(1+\eth^2g)}, \tag{4.23}$$

$$h=\frac{\overline{\partial f}}{\partial f}+\frac{\gamma}{2}\partial b+\gamma b\left[\frac{\partial^2f}{\partial f}+\frac{2}{3}\partial\ln(1+\eth^2g)\right]. \tag{4.24}$$

These formulas coincide, to within renormalization of γ and λ, with the solution found by Ooguri *et al.*,[6] who interpreted W_3-gravity as a connected Wess–Zumino $SL(3,R)$-theory. Plugging (4.20) and (4.21) into (4.15), we reproduce the chiral action of Ooguri *et al.*,[6] which must be interpreted as the effective action induced by quantum fluctuations of the matter fields φ and ψ, which interact with the multiplet of chiral W_3-gravity via (4.15). The anomalous dependence of this action on the ''coordinate'' functions f and g is due to the W-gravity anomaly. Continuing the analogy with the cases of two-dimensional gravity and supergravity, we can assume that this action is a result of choosing a regularization scheme that conserves the Weyl and W-Weyl symmetries, as well as half the coordinate symmetries of the covariant action that describes the interaction of matter fields and ordinary gravity and W-gravity.

5. CONCLUSION

By generalizing some of the laws governing ordinary gravitation we were able to find the effective action of chiral W_3-gravity. It would be more interesting, however, to continue the analogy and find the W-analog of the Liouville action; namely, the covariant action that describes the interaction of matter fields with a gravitational and W-gravitational background.

To understand the nature of the symmetry properties of W-gravity theory it is advisable to first turn to classical theory.

As a classical gauge theory, W-gravity was first examined by Hull.[3] The nonchiral formulation of this theory was later performed by Schoutens *et al.*[4] They started with the action

$$S=\frac{1}{2}\int d^2x\ \partial_+\varphi\partial_-\varphi. \tag{5.1}$$

With respect to the ordinary diffeomorphisms $\delta\varphi=\epsilon^\alpha\partial_\alpha\varphi$, the variation of (5.1),

$$dS=\int d^2x\ [\partial_\alpha(\epsilon^\alpha\partial_+\varphi\partial_-\varphi)+\partial_+\epsilon^-(\partial_-\varphi)^2 +\partial_\epsilon^+(\partial_+\varphi)^2], \tag{5.2}$$

vanishes if $\partial_+\epsilon^-=\partial_-\epsilon^+=0$. To generalize this symmetry to a local one, we must, in accordance with Noether's theorem, add to (5.1) the currents

$$t_{++}=\tfrac{1}{2}(\partial_+\varphi)^2, \quad t_{--}=\tfrac{1}{2}(\partial_-\varphi)^2,$$

multiplied by the corresponding gauge fields kh_{++} and kh_{--}. Here k is an expansion parameter, which is set to unity in the final result. After an infinite number of steps the action can be summed as a geometric progression to produce

$$S=\frac{1}{2}\int d^2x\ \frac{(\partial_+\varphi-kh^{--}\partial_-\varphi)(\partial_-\varphi-kh^{++}\partial_+\varphi)}{1-k^2h^{++}h^{--}}. \tag{5.3}$$

Schoutens *et al.*[4] then stated that the action (5.1) for n real fields is invariant under holomorphic W-diffeomorphisms. Indeed, under the transformations

$$\delta\varphi^i=d^{ijk}(\lambda^{++}\partial_+\varphi^j\partial_+\varphi^k+\lambda^{--}\partial_-\varphi^j\partial_-\varphi^k) \tag{5.4}$$

the action transforms as

$$\delta S=\frac{1}{3}\int d^2x\ d_{ijk}\ (\partial_+\varphi^i\partial_+\varphi^j\partial_+\varphi^k\partial_-\lambda^{++} + \partial_-\varphi^i\partial_-\varphi^j\partial_-\varphi^k\partial_+\lambda^{--}). \tag{5.5}$$

The fact that the algebra of holomorphic ϵ- and λ-diffeomorphisms is closed imposes the following constraint on the symmetric constants d (see Ref. 3):

$$d^{k(ij}d^{l)mk}=\delta^{(ij}\delta^{l)m}. \tag{5.6}$$

The action (5.1) can be made invariant under local ϵ- and λ-transformations via the Noether procedure by introducing the appropriate gauge fields h^{++}, h^{--}, b^{+++}, and b^{---}. Unfortunately, in the given case the invariant action can be summed only by using auxiliary fields $F^i_\pm$:

$$S=\int d^2x\ e\Big[\partial_+\varphi^i\partial_-\varphi^i+F^i_+F^i_- + F^i_+\Big(\partial_-\varphi^i-\frac{1}{3}d_{ijk}b^{+++}F^j_+F^k_+\Big) + F^i_-\Big(\partial_+\varphi^i-\frac{1}{3}d_{ijk}b^{---}F^j_-F^k_-\Big)\Big]. \tag{5.7}$$

After the auxiliary fields are eliminated, they are replaced by ''nested'' covariant derivatives,

$$F^i_\pm\to\hat\partial\varphi^i_\pm=e^\alpha_\pm\partial_\alpha\varphi-b^{\mp\mp\mp}d_{ijk}\hat\partial_\pm\varphi^j\hat\partial_\pm\varphi^k, \tag{5.8}$$

and the action becomes infinitely nonlinear. To avoid difficulties and operate from the start in a covariant setting, we must introduce more gauge fields than required by the Noether approach. Specifically, in the case of pure gravitation, we must introduce the full tensor $h_{\alpha\beta}$ instead of the two components h^{++} and h^{--}. Then the Noether procedure terminates after the first step, and the invariant action has the form

$$S=\int d^2x\ h^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi.$$

Here we expect a new symmetry to appear, a symmetry that would balance the superfluous degree of freedom related to the h^{+-}-component of the metric. By requiring that the energy–momentum tensor of the theory be traceless,

$$T^\alpha_\alpha=T_{\alpha\beta}h^{\alpha\beta}=0,\quad T_{\alpha\beta}\equiv\frac{\delta S}{\delta h^{\alpha\beta}}, \tag{5.9}$$

we require that the theory be Weyl-invariant, so that the equation

$$h^{\alpha\beta}\frac{\delta S}{\delta h^{\alpha\beta}}=0$$

has a functional of the type $S=S(\sqrt{h}h^{\alpha\beta})$ as a solution, which is equivalent to invariance under $h^{\alpha\beta}\to e^\sigma h^{\alpha\beta}$ transformations. In this way the final expression for the invariant action is

$$S=\int\sqrt{h}d^2x\ h^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi. \tag{5.10}$$

In the case of W_3-gravity we propose introducing the h^{+-}, b^{++}, and b^{--+} components of the gauge fields, in order to produce the total tensors $h^{\alpha\beta}$ and $b^{\alpha\beta\gamma}$. The Noether procedure terminates after the first step, and the invariant action has the form

$$S=\int d^2x\ (h^{\alpha\beta}t_{\alpha\beta}+b^{\alpha\beta\gamma}\omega_{\alpha\beta\gamma}), \tag{5.11}$$

where

$$t_{\alpha\beta}=\tfrac{1}{2}(\partial_\alpha\varphi\partial_\beta\varphi-\partial_\alpha\psi\partial_\beta\varphi),$$
$$\omega_{\alpha\beta\gamma}=\tfrac{2}{3}(\partial_\alpha\varphi\partial_\beta\varphi\partial_\gamma\psi+\partial_\alpha\varphi\partial_\beta\psi\partial_\gamma\varphi+\partial_\alpha\psi\partial_\beta\varphi\partial_\gamma\varphi + \partial_\alpha\psi\partial_\beta\psi\partial_\gamma\psi).$$

The action (5.11) is assumed to be invariant under the w-diffeomorphisms

$$\delta_\lambda\varphi=-4\lambda^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\psi,$$
$$\delta_\lambda\psi=2\lambda^{\alpha\beta}(\partial_\alpha\varphi\partial_\beta\varphi+\partial_\alpha\psi\partial_\beta\psi), \tag{5.12}$$

defined with a traceless parameter λ, i.e., $\lambda^{\alpha\beta}h_{\alpha\beta}=0$. The variation of (5.11) under the transformations (5.12) can be written in the form

$$\delta S=\int d^2x[\,\delta_\lambda b^{\alpha\beta\gamma}\omega_{\alpha\beta\gamma}-\omega_{\alpha\beta\gamma}(h^{\alpha\mu}\nabla_\mu\lambda^{\beta\gamma} + h^{\beta\mu}\nabla_\mu\lambda^{\alpha\gamma}+h^{\gamma\mu}\nabla_\mu\lambda^{\alpha\beta})-(h^{\alpha\mu}\lambda^{\beta\gamma} + h^{\beta\gamma}\lambda^{\alpha\mu})\nabla_\mu\omega_{\alpha\beta\gamma}+\delta_\lambda h^{\alpha\beta}t_{\alpha\beta} + 16b^{\alpha\beta\gamma}\lambda^{\mu\nu}(2t_{\gamma\nu}\nabla_\mu t_{\alpha\beta}-t_{\beta\gamma}\nabla_\alpha t_{\mu\nu}) + 16b^{\alpha\beta\gamma}\nabla_\alpha\lambda^{\mu\nu}(2t_{\beta\mu}t_{\gamma\nu}-t_{\beta\gamma}t_{\mu\nu})]. \tag{5.13}$$

Defining the λ-variations of the gauge fields in such a way that the coefficients of the currents $t_{\alpha\beta}$ and $\omega_{\alpha\beta\gamma}$ vanish, we ensure that the action (5.11) is invariant. The transformations (5.12) represent a specific realization of the constants d^{ijk} in (5.4) for the case of two fields. Such a restriction is not accidental, the point being that when there are three or more fields, the condition that the covariant algebra of the W-transformations

$$\delta\varphi^i=d^{ijk}\lambda^{\alpha\beta}\partial_\alpha\varphi^j\partial_\beta\varphi^k \tag{5.14}$$

be closed imposes additional constraints on the d^{ijk}, restrictions which together with the conditions (5.6) have only a vanishing solution. In addition, the algebra becomes closed

on the gauge fields only if the equations of motion are included, as happens in the simpler case of two-dimensional supergravitation.[16]

Thus, to guarantee invariance under ordinary and λ-diffeomorphisms, we introduced seven gauge fields. Now we would like to impose constraints on the theory in such a way so as to obtain a three-parameter symmetry group that ''balances'' the three superfluous degrees of freedom.

By imposing the conditions that the energy–momentum tensor and W-current be traceless,

$$h^{\alpha\beta}\frac{\delta S}{\delta h^{\alpha\beta}}=0, \quad h^{\alpha\beta}\frac{\delta S}{\delta b^{\alpha\beta\gamma}}=0, \tag{5.15}$$

conditions with a solution of the form $S=S(\hat{h}^{\alpha\beta},\hat{b}^{\alpha\beta\gamma})$, with the quantities

$$\hat{h}^{\alpha\beta}\equiv\sqrt{h}h^{\alpha\beta},$$

$$\hat{b}^{\alpha\beta\gamma}\equiv b^{\alpha\beta\gamma}-\tfrac{1}{4}h_{\mu\nu}(h^{\alpha\beta}b^{\gamma\mu\nu}+h^{\beta\gamma}b^{\alpha\mu\nu}+h^{\alpha\gamma}b^{\beta\mu\nu})$$

invariant under Weyl and W-Weyl transformations, respectively, i.e.,

$$\begin{aligned} &h^{\alpha\beta}\to e^{\sigma}h^{\alpha\beta} \quad \text{(Weyl)},\\ &b^{\alpha\beta\gamma}\to b^{\alpha\beta\gamma}+(\zeta^{\alpha}h^{\beta\gamma}+\zeta^{\beta}h^{\alpha\gamma}+\zeta^{\gamma}h^{\alpha\beta}) \quad (W\text{-Weyl}), \end{aligned} \tag{5.16}$$

we reduce the number of degrees of freedom. However, comparison of the variations $\delta S(\hat{h},\hat{b})$ with (5.13) shows that the symmetry of the action under λ-diffeomorphisms is incompatible with Weyl invariance, since if the variation $\delta_{\lambda}t_{\alpha\beta}$ is traceless, the variation $\delta_{\lambda}\omega_{\alpha\beta\gamma}$ is not, with the result that the latter cannot be made equal to the traceless variation $\delta\hat{h}$. Thus, Weyl invariance is incompatible with W-symmetry even at the classical level, and to reduce the number of degrees of freedom we must replace the requirements that the energy–momentum tensor be traceless by a different one.

Schoutens *et al.*[17] proposed a covariant formulation of W-gravity. With each annihilation operator of the classical W_3-algebra they associated a gauge field and a local parameter, and with each annihilation operator they associated a field in the adjoint representation. Then they required that all corresponding curvatures vanish. The resulting theory has a finite number of degrees of freedom and, in addition to being ordinary-invariant and W-diffeomorphism invariant, it is locally Weyl- and Lorentz-symmetric and W-Weyl and W-Lorentz symmetric.

It would also be interesting to study this theory as a system with constraints.

Many thanks go to R. L. Mkrtchyan and O. M. Khudaverdyan for fruitful and stimulating discussions and especially to A. G. Sedrakyan for a thorough review of the manuscript and for the critical remarks. The present work has in part been supported by grants 211-5291 YPI (German Bundesministerium für Forschung und Technologie) and INTAS-2058.

*)e-mail: karakhan@vx2.yerphi.am

[1] A. B. Zamolodchikov, Teoret. Mat. Fiz. **65**, 347 (1985).
[2] I. M. Gel'fand and L. A. Dikiĭ, Preprint IFM of the USSR Academy of Sciences, Moscow (1978); V. G. Drinfel'd and V. V. Sokolov, *Current Problems of Mathematics* [in Russian], Vol. 24, VINITI, Moscow (1984).
[3] C. M. Hull, Phys. Lett. B **240**, 110 (1990).
[4] K. Schoutens, A. Servin, and P. van Nieuvenhuizen, Phys. Lett. B **243**, 248 (1990).
[5] K. Schoutens, A. Servin, and P. van Nieuvenhuizen, Nucl. Phys. B **364**, 584 (1991); **371**, 315 (1992).
[6] H. Ooguri, K. Schoutens, A. Servin, and P. van Nieuvenhuizen, Comm. Math. Phys. **145**, 515 (1992).
[7] A. M. Polyakov, Mod. Phys. Lett. A **2**, 893 (1987).
[8] J. M. Figueroa-O'Farrill, S. Stauciu, and E. Ramos, Phys. Lett. B **297**, 289 (1992); C. M. Hull, Phys. Lett. B **269**, 257 (1991).
[9] V. G. Knizhnik, A. M. Polyakov, and A. B. Zamolodchikov, Mod. Phys. Lett. A **3**, 819 (1988).
[10] Y. Matsuo, Phys. Lett. A **227**, 117 (1991).
[11] R. Zucchini, Phys. Lett. B **260**, 296 (1991).
[12] F. Gieres, Preprint CERN 366 (1991).
[13] A. M. Polyakov and A. B. Zamolodchikov, Mod. Phys. Lett. A **3**, 819 (1988).
[14] D. R. Karakhanyan, Phys. Lett. B **365**, 56 (1996).
[15] D. R. Karakhanyan, R. P. Manvelyan, and R. L. Mkrtchyan, Phys. Lett. B **329**, 185 (1994).
[16] S. Deser and B. Zumino, Phys. Lett. A **65**, 369 (1976).
[17] K. Schoutens, A. Servin, and P. van Nieuvenhuizen, Nucl. Phys. B **349**, 791 (1991).

Translated by Eugene Yankovsky

Scattering of relativistic electrons by high-power laser light tightly focused

Yu. N. Eremenko and L. S. Mkhitar'yan

Russian Federal Nuclear Center, National Research Institute of Experimental Physics, 607190 Sarov, Nizhniĭ Novgorod Region, Russia

(Submitted 26 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1554–1562 (May 1997)

We obtain an approximate solution for the drift and oscillatory components of the motion of relativistic electrons in the field of temporally extended high-power laser light under strong focusing of the light (the size of the focal region is of the order of the light wavelength). This makes it possible to start numerically integrating the equations of electron motion near the focal region. We estimate the impact parameters of the electrons when they are still efficiently accelerated in the focal region. © *1997 American Institute of Physics.* [S1063-7761(97)00205-9]

1. INTRODUCTION

Investigations of the acceleration of electrons by high-power laser radiation ($E \gg e/a^2$, where e is the electron charge, and a_0 is the Bohr radius) in the focal region of a lens have a long history.[1–3]

The mechanism of such acceleration is fairly simple. In a strong electromagnetic field the oscillatory motion of an electron is relativistic, and the energy of this motion can be as high as desired (in comparison to the initial energy of electron motion). If the amplitude of the electron oscillations becomes comparable to the beam waist radius w_0, the electron can leave the focal region with an energy comparable to the energy of the oscillatory motion,[4] which in turn is proportional to the light intensity in the focal spot. By measuring the spectrum of the scattered electrons one can estimate the parameters of the laser pulse in the waist region.

The feasibility of such measurements strongly depends on the number of accelerated electrons, i.e., in other words, on the volume of the region from which the electrons are drawn into the acceleration process.

The size of this region is usually estimated by numerically integrating the equations of electron motion in the field of the focused laser light.

Most studies of this problem are devoted to investigating ultrashort pulses (tens to hundreds of femtoseconds). Lately, however, it has become possible, thanks to the development of wavefront reversal methods,[5] to focus laser light with an energy ~ 500 J and a pulse length ~ 200 ps into a spot with a diameter of $(3–5)\lambda$ (see Ref. 6). This makes it possible to reach intensities $I_L \sim 10^{19}$ W/cm^{-2} and electric field strengths $E \approx 10^{11}$ V/cm ($E \sim 10e/a_0^2$) in the focal region. The number of field oscillations for a pulse of such a length reaches 10^4 and even higher values, with numerical integration encountering difficulties related to round-off errors and long computation times.

In the present paper we develop a method for calculating the electron spectrum and the size of the region from which the electrons are drawn into the acceleration process. The method is based on separating the electron motion into drift and oscillatory components. It makes it possible to derive analytic expressions for the state of electron motion near the beam waist, which can be used as initial data for numerical integration of the equations of motion. By applying this method, the number of calculated field oscillations is reduced by a factor of almost one hundred. For a short pulse ($\tau \sim 0.1$ ps) we compare the direct method of calculation with the approximate one, and demonstrate that the accuracy of the approximation is satisfactory. For a long pulse ($\tau \sim 200$ ps) we give an example of calculating the spectrum of accelerated electrons and estimate the size of the region from which the electrons are drawn into the acceleration process.

2. STATEMENT OF THE PROBLEM

We take a beam of relativistic electrons moving with initial velocity $\beta_z = v_z/c = \beta_0$ ($\beta_x = \beta_y = 0$) along the axis of a lens (the z axis) with impact parameter x_0, and interacting with a laser pulse whose field we describe (following the approach adopted in Ref. 7) in the paraxial approximation for a linearly polarized Gaussian beam:

$$E_x = Q(\phi,x,z)\sin\left(\phi - \frac{kx^2}{z+z_0^2/z} - \arctan\frac{z}{z_0}\right), \qquad H_z = E_x, \tag{1}$$

$$Q(\phi,x,z) = \frac{E_0 g(\phi)}{\sqrt{1+z^2/z_0^2}}\ \exp\left[-\frac{x^2}{w^2(z)}\right]. \tag{2}$$

Here $\phi = \omega[t - z(t)/c]$ is the phase of the field at the point occupied by the electron, k is the wave vector of the light, $w(z) = w_0\sqrt{1+z^2/z_0^2}$ is the beamwidth at distance z from the focal point, and $z_0 = \pi w_0^2/\lambda$ is the length of the beam waist, with w_0 the waist radius. The function $g(\phi)$ specifies the temporal pulse shape.

In terms of the dimensionless electron velocity $\boldsymbol{\beta} = \mathbf{v}/c$ and electron energy

$$\gamma = \frac{1}{\sqrt{1-\beta^2}},$$

the equation of electron motion (with radiative corrections ignored) have the following form:[8]

 1063-7761/97/050851-05$10.00

$$\frac{d}{dt}(\gamma\boldsymbol{\beta}) = -\frac{e}{mc}(\mathbf{E}+\boldsymbol{\beta}\times\mathbf{H}),$$
$$\frac{d}{dt}\gamma = -\frac{e}{mc}\boldsymbol{\beta}\cdot\mathbf{E}. \tag{3}$$

For short pulses, the scattered-electron spectrum can be obtained by numerically integrating the equations of electron motion (3) in the field of the light beam. For long pulses, where the number of field oscillations reaches 10^4 and may even be higher, numerical integration encounters difficulties related to round-off errors and long computation times.

However, the fact that the pulses are long makes it possible to approach this problem from another perspective. When the pulse length is large, the relative variation of the field amplitude over one period, $(1/Q)(dQ/d\phi)$, is small, and electron motion can be separated in oscillatory and drift components, for which one can easily set up approximate solutions in the form of series of increasing orders of the phase derivatives of the field amplitude, e.g.,

$$\gamma\beta_x = \sum_{n=0}^{\infty} \mathbf{a}_n \frac{d^n Q}{d\phi^n}.$$

Here the derivatives $d^nQ/d\phi^n$ depend on the temporal envelope (or profile) of the pulse, $g(\phi)$, and the profile of the light beam formed by the lens. Below we establish that for long pulses, terms in $d^ng/d\phi^n$ are negligible in comparison with terms in d^nQ/dz^n. The latter depend on z as $1/z^{n+1}$, so that the series rapidly converge. This makes it possible to start numerically integrating the electron paths near the focal region ($\sim$200–300 periods), taking the initial conditions from the approximate solution.

3. APPROXIMATE SOLUTIONS FOR THE OSCILLATORY AND DRIFT COMPONENTS OF ELECTRON MOTION

Replacing the time variable t by the phase ϕ in the equations of motion (3), we arrive at the following system of equations:

$$\frac{d}{d\phi}(\gamma\beta_x) = -\frac{eE_x}{mc\omega}, \tag{4}$$

$$\frac{d\xi}{d\phi} = \frac{\gamma\beta_x}{B}, \quad B = 2\pi\gamma_0(1-\beta_0), \tag{5}$$

$$\frac{d\eta}{d\phi} = a\left[1+\frac{(1+\beta_0)(\gamma\beta_x)^2}{2\beta_0}\right], \quad a = \frac{\beta_0}{2\pi^2 w_0^2(1-\beta_0)}. \tag{6}$$

Here $\xi=(x-x_0)/\lambda$ and $\eta=z/z_0$. In deriving (6) we used the constant of the motion $\gamma(1-\beta_z)=\gamma_0(1-\beta_0)$, from which it follows that

$$\frac{\gamma}{\gamma_0} = 1+\frac{1+\beta_0}{2}(\gamma\beta_x)^2.$$

The initial conditions for Eqs. (4)–(6) are $\gamma\beta_x|_{\phi=0}=0$, $\xi|_{\phi=0}=0$, and $\eta|_{\phi=0}=\eta_0$. Since the values of the impact parameter x_0 of practical interest are those not exceeding w_0, and since the solution will be set up for $\eta\gg 1$, for the function Q in (1) we take the first two terms in the expansion:

$$Q(\phi,x,z) = Q(\phi,x_0,z) + \left.\frac{\partial Q}{\partial x}\right|_{x=x_0}(x-x_0),$$

with the exponentials in $Q(x_0)$ and $\partial Q/\partial x|_{x=x_0}$ assumed equal to unity.

In Eq. (6) we select the term that is linear in ϕ by introducing a new variable, $\eta' = \eta - a\phi - \eta_0$, for which we have

$$\frac{d\eta'}{d\phi} = a\frac{1+\beta_0}{2\beta_0}(\gamma\beta_x)^2 \tag{7}$$

with initial condition $\eta'|_{\phi=0}=0$. Now we separate the motion into an oscillatory component $(\gamma\tilde{\beta}_x, \tilde{\xi}, \tilde{\eta}')$ and a drift component $(\langle\gamma\beta_x\rangle, \langle\xi\rangle, \langle\eta'\rangle)$. To this end in Eqs. (4), (5), and (7) we group terms that vanish under averaging over ϕ, and terms that do not. The result is

$$\frac{d}{d\phi}\gamma\tilde{\beta}_x = -G(\phi,\eta)\sin\left[\phi - \frac{(\xi_0+\xi)^2\eta}{w_0^2(1+\eta^2)} - \arctan\,\eta\right], \tag{8}$$

$$\frac{d}{d\phi}\tilde{\xi} = \frac{\gamma\tilde{\beta}_x}{B}, \quad G = \frac{\alpha g(\phi)}{\sqrt{1+\eta^2}}, \quad \alpha = \frac{eE_0}{mc\omega}, \tag{9}$$

where $\xi_0 = x_0/\lambda$.

Now we can easily obtain the approximate solution by successively integrating these equations by parts. Thus,

$$\gamma\tilde{\beta}_x \approx G\cos\phi'|_0^{\phi} - \sin\phi'\frac{d}{d\phi}[G(\phi)]|_0^{\phi} + \int_0^{\phi}\sin\phi'\frac{d^2}{d\phi^2}[G(\phi)]\,d\phi,$$

where

$$\phi' = \phi - \frac{(\xi_0-\xi)^2\eta}{w_0^2(1+\eta^2)} - \arctan\,\eta.$$

Clearly, the third term is $\sim 1/\eta^3$. Restricting the series to terms quadratic in $1/\eta$ and performing the elementary differentiation of Q (we limit ourselves to the case in which $Q|_{\phi=0} = dQ/d\phi|_{\phi=0} = 0$), we find that

$$\gamma\tilde{\beta}_x \approx G(\phi)\cos\phi' + G(\phi)\left(\frac{\eta a}{1+\eta^2} - \frac{1}{g}\frac{dg}{d\phi}\right)\times\sin\phi' + O\left(\frac{1}{\eta^3}\right), \tag{10}$$

where we have replaced $d\eta/d\phi$ by a, allowing for the fact that

$$\frac{d\eta'}{d\phi} \sim (\gamma\tilde{\beta}_x)^2 \sim \frac{1}{\eta^2}.$$

Similarly, integrating (9) by parts and restricting the series to terms quadratic in $1/\eta$, we obtain

$$\tilde{\xi} = \frac{G(\phi)}{B}\left[\sin\phi' + 2\cos\phi'\left(\frac{1}{g}\frac{dg}{d\phi} - \frac{a\eta}{1+\eta^2}\right)\right]. \tag{11}$$

The drift components of motion can easily be found by averaging Eqs. (4), (5), and (7) over one period. The first nonvanishing term on the right-hand side of Eq. (4) after averaging is

$$\left.\frac{\partial Q}{\partial x}\right|_{x=0}(x-x_0)\sin\phi',$$

since, as Eq. (11) shows, $x-x_0\sim\sin\phi'$. Changing the variable ϕ in (4) again to η_0, we find that

$$\frac{d}{d\eta}\langle\gamma\beta_x\rangle=\frac{2\xi_0}{aB\overline{w_0^2}}\frac{G^2}{1+\eta^2}\left\langle\frac{\sin^2\phi'}{1+0.5(1+\beta_0)(\gamma\beta_x)^2}\right\rangle,$$

where $\overline{w_0}=w_0/\lambda$. The angle brackets denote averaging over phase, and the function $g(\phi)$, which changes little in the 2π interval, is taken outside the averaging sign.

Allowing only for the first term in the denominator of the expansion for $\gamma\beta_x$ from (10), and averaging, we obtain

$$\langle\gamma\beta_x\rangle=\frac{2\xi_0}{aB\overline{w_0^2}}\frac{2\beta_0}{1+\beta_0}\int_{\eta_0}^{\eta}\left(\sqrt{1+\frac{1+\beta_0}{2\beta_0}G^2}-1\right)\frac{d\eta}{1+\eta^2}$$

$$\equiv\frac{2\xi_0}{aB\overline{w_0^2}}\frac{2\beta_0}{1+\beta_0}\,\mathrm{er}(\eta). \tag{12}$$

Similarly, for drift along the x and z axes we have

$$\langle\xi\rangle=\frac{1}{aB}\int_{\eta_0}^{\eta}d\eta\left\langle\frac{\langle\gamma\beta_x\rangle}{1+(\gamma\beta_x)^2(1+\beta_0)/2}\right\rangle$$

$$=\frac{2\xi_0}{a^2B^2\overline{w_0^2}}\frac{2\beta_0}{1+\beta_0}\int_{\eta_0}^{\eta}\frac{\mathrm{er}(\eta)}{\sqrt{1+G^2(1+\beta_0)/2\beta_0}}\,d\eta, \tag{13}$$

$$\langle\eta'\rangle=\frac{1+\beta_0}{2\beta_0}\int_{\eta_0}^{\eta}d\eta\left\langle\frac{(\gamma\beta_x)^2}{1+(\gamma\beta_x)^2(1+\beta_0)/2\beta_0}\right\rangle$$

$$=\int_{\eta_0}^{\eta}d\eta\left[1-\frac{1}{\sqrt{1+G^2(1+\beta_0)/2\beta_0}}\right]. \tag{14}$$

It is sufficient to replace the argument of $g(\phi)$ in the integrals in Eqs. (12)–(14) by the approximate expression $\phi=(\eta-\eta_0)/a$.

Thus, an electron that was initially deflected away from the axis by ξ_0 drifts in the transverse direction. Its displacement along the x axis is determined largely by the parameter[1)]

$$D=\frac{4\beta_0}{a^2B^2w_0^2(1+\beta_0)}=\frac{4\pi^2w_0^2}{\gamma_0^2\beta_0(1+\beta_0)},$$

which depends on the initial electron energy γ_0 and the beam waist radius w_0. The integral in (13) determines the dependence of the traverse electron drift on the characteristics of the laser pulse, α and $\Delta\phi$.

Now it is clear that for temporally extended laser pulses ($|g^{-1}dg/d\phi|\ll1$) there is no need to numerically integrate the equations of motion over the entire electron–field interaction range. Instead the integration can start from a certain value η_{in} ($|\eta_{\mathrm{in}}|\gg1$), with the initial state of the electron at this point determined by solving Eqs. (10)–(14), which are actually an expansion of the solution in inverse powers of η.

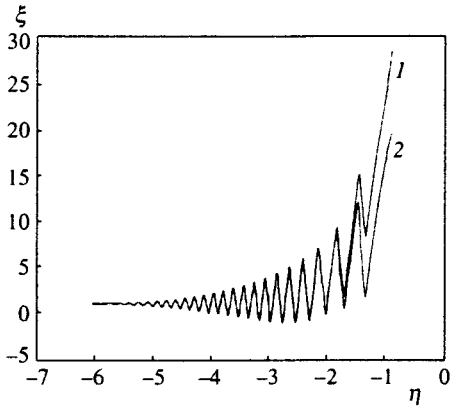

FIG. 1. The paths of electron motion with an impact parameter $\xi_0=1$ obtained by numerical calculations (curve *1*) and by an approximate method (curve *2*) for the case of a short laser pulse ($\alpha=3.41$, $\Delta\phi=60\pi$, $w_0=20$, and $\gamma_0=10$).

Figures 1 and 2 depict the paths of an electron and the electron drift along the z axis as functions of the distance to the focal point calculated by Eqs. (4)–(7) and (10)–(14). In all numerical examples we used

$$g(\phi)=\sin^2\left(\frac{\pi}{2}\frac{\phi}{\Delta\phi}\right).$$

Clearly, the electron path is described fairly well up to the waist region ($\eta\geqslant2$). At the same time, the accuracy of Eq. (14) is not sufficient for determining the phase of motion, $\phi=a^{-1}(\eta-\eta_0-\eta')$, at such distances from the focal point, which leads to a considerable uncertainty in the energy with which the electron is scattered by the focal region.

If, however, we are less interested in the result of an individual scattering event than in the spectrum of the electrons that begin to interact with the laser pulse at different points η_0 on the lens axis, there is no need to accurately determine the phase. Indeed, the relationship between the phase with which an electron arrives at the point η_{in} and the coordinate η_0 at which the electron begins to interact with the field is $\eta_{\mathrm{in}}=\eta_0+a\phi+\eta'$. This yields

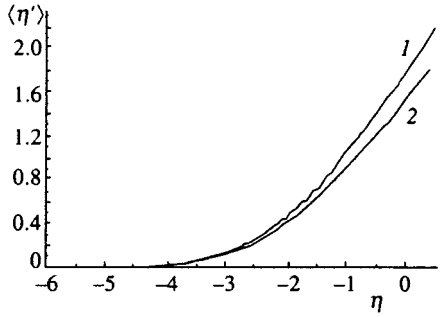

FIG. 2. The dependence of electron drift $\langle\eta'\rangle$ along the z axis on the distance from the focal points. The curves were obtained by numerical integration of Eqs. (3)–(6) (curve *1*) and by an approximate method of solving Eq. (13) (curve *2*) with the same parameters of the laser pulse as in Fig. 1.

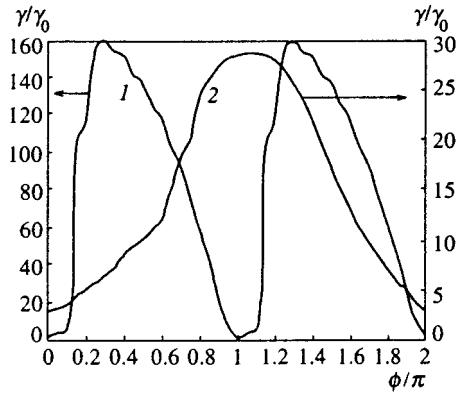

FIG. 3. The spectrum of an electron arriving at the focal point simultaneously with the peak of the laser pulse ($\alpha=5.97$, $\Delta\phi=46\,000\pi$, $w_0=2.55$, and $\gamma_0=1.8$) for the following values of the impact parameter: $\xi_0=0$ (curve *1*), and $\xi_0=0.01$ (curve *2*).

$$\frac{d\phi}{d\eta_0}=-\frac{1}{a}\left(1+\frac{d\eta'}{d\eta_0}\right).$$

Using Eq. (14), we can easily show that $d\langle\eta'\rangle/d\eta_0$ does not exceed $(\alpha/\eta_{\rm in})^2$, and at distances where this parameter is much smaller than unity there exists a linear relationship between the coordinate at which the electron begins to interact with the field and the phase with which the electron arrives at the point $\eta_{\rm in}$. By going through all the phases belonging to the interval $[0,2\pi]$ in specifying the initial data needed for numerical calculations, we reproduce the spectrum of scattered electrons that started their interaction with the beam in the interval $\Delta\eta_0=-2\pi a$.

Figure 3 depicts an example of the spectra calculated in this manner, for electrons that arrive at the beam waist region together with the pulse peak ($\eta_0\approx-a\Delta\phi$). Clearly, in the case at hand an electron moving along the axis can become accelerated to energies that exceed the initial electron energy by a factor of about 160. However, for an impact parameter $\xi_0=0.01$, the electron energy drops by a factor of about 5 after scattering.

4. ELECTRON SCATTERING CROSS SECTION AT THE FOCAL POINT OF A LASER PULSE

As noted in the Introduction, for an electron to be scattered in the beam waist region, its oscillation amplitude in the waist must exceed the waist radius:

$$\frac{\alpha}{2\pi\gamma_0(1-\beta_0)}>w_0.$$

It is clear, however, that if the electron oscillation amplitude is too large, the electron will scatter even before it gets to the region of peak field strength in the waist, and will not acquire the energy it could if it reached the waist region. Thus, having the highest possible energy imparted to the electron when it is scattered at the focal point imposes an additional constraint on the electron's maximum oscillation amplitude:

$$\frac{\alpha}{2\pi\gamma_0(1-\beta_0)}<w_0(\,1.5-1.7).$$

We see that for fixed parameters of the laser pulse there exists an optimum value of the initial electron energy at which the electron acquires the maximum possible energy when it is scattered in the field of the focused laser pulse.

Thus, in the case of the short laser pulse under discussion ($\Delta\phi=60\pi$, $w_0=20$, and $\alpha=3.41$), the optimum initial electron energy is $\gamma_0\approx10$, and for a long pulse ($\Delta\phi=46\,000\pi$, $w_0\approx2.55$, and $\alpha\approx6$) the optimum value is $\gamma_0\approx1.8$.

For the transverse drift not to have a significant effect on the electron–field interaction, $\langle\xi\rangle$ at the waist much be much smaller than the electron oscillation amplitude a/B (see Eq. (12)). Since the integral in (14) is of order unity, we can easily see that the value of the impact parameter at which the electron is still accelerated satisfies the inequality

$$\xi_0\ll\frac{\alpha}{8\pi^3\overline{w_0^2}}\gamma_0^3\beta_0(1+\beta_0)^2, \tag{15}$$

so that the effective electron cross section rapidly increases with initial electron energy.

For the long pulse under discussion, plugging numerical values into (15) yields $\xi_0\ll0.07$, while for the short pulse we obtain $\xi_0\ll0.33$.

Hartemann *et al.*[4] studied a picosecond laser pulse with $\alpha=5.33$, $\Delta\tau=0.8$ pc, and $w_0=4.95$ μm. According to (15), the electrons with $\gamma_0=10$ reach the waist of such a pulse from the region around the axis with $\xi_0\ll9$, which constitutes a considerable fraction of the waist size.

For the long pulse ($\Delta\phi=60\pi$, $w_0=20$, and $\alpha=3.41$), we now estimate the volume occupied by electrons that can be accelerated by the field of the laser light to energies higher than half the peak energy that an electron can acquire. Since the peak energy that an electron can acquire as a result of acceleration is proportional to the square of the field, the electrons of interest must arrive at the waist when the field there is no lower than $E_{\max}/\sqrt{2}$, which means that

$$\sin\left(\frac{\pi}{2}\frac{\phi}{\Delta\phi}\right)\geqslant\frac{1}{\sqrt{2}}.$$

Thus, the size of the region of interest is $\Delta\phi/2$, or (in terms of length) $\sim a\Delta\phi/2$, and the radius amounts to roughly 0.01λ. For the long pulse under discussion, this yields $V\approx\pi(0.01\times10^{-4})^2\times5.5=1.7\times10^{-11}$ cm^3.

Figure 3 shows that only roughly half the electrons belonging to the specified region will be accelerated to the required energies (for this to occur, the electrons must enter the waist region with specific phases). Thus, for the pulse under discussion, the number of electrons accelerated to significant energies in the waist region to significant energies is $N\approx0.5Vn$, or $N\approx0.85\times10^{-11}n$, where n is the electron beam density.

We see that studying the intensity of such a laser pulse at the focal point requires using an electron beam with a particle number density much greater than 10^{11}cm^{-3}.

5. CONCLUSION

We have studied the problem of electron scattering by a temporally extended laser pulse focused to a region several wavelengths long. We have found that for extended pulses the electron motion at large distances from the waist region can be represented to high accuracy by the superposition of two types of motion: drift perpendicular to the lens axis, and oscillatory motion.

For both types of motion we have derived analytic expressions in the form of series in inverse powers of the distance to the waist, which makes it possible to calculate the initial values of the electron coordinates and velocity for numerical integration of the equations of motion in the vicinity of the waist.

For various values of the impact parameter of electrons that arrive at the waist region together with the peak of the laser pulse, we give the electron spectra after scattering.

The expressions for electron drift perpendicular to the lens axis enables one to estimate the electron scattering cross section. For a pulse ~ 0.2 ns long and a waist diameter of about 3λ, the volume V occupied by electrons that can become accelerated at the focal point of the lens and can be used to estimate the light intensity at the focal point has been found to be about 10^{-11} cm^3.

Hence determining the peak light intensity in such pulses from the scattered electron spectrum becomes possible when beams with an electron number density $n \gg 10^{11}$ cm^{-3} are employed.

The authors are grateful to G. G. Kochemasov for useful remarks. The present work was sponsored by the International Science and Technology Center (Project 111).

[1)]The dependence of drift on D is clear from the start. Indeed, in the nonrelativistic limit the ponderomotive force acting on an electron in a plane wave is $F_{\text{pond}}=(e^2/4m_0\omega^2)\nabla E_0^2$. Since the ponderomotive force is significant only in the waist region, the displacement of the electron due to the force is $x_{\text{dr}}=(F_{\text{pond}}/m_0)(t^2/2)$, where $t=z_0/v_0$ is the time it takes the electron to travel through the beam waist region. Plugging in $z_0=\pi w_0^2$ and noting that $\nabla E_0^2=(4\xi_0/w_0^2)E_0^2$, we find that $\xi_{\text{dr}}\sim\xi_0 w_0^2\alpha^2/\beta_0^2$, which coincides with (13) calculated in the approximation in which $G^2\ll 1$. The factor γ_0^{-2} in the formula for D simply allows for the increase in the inertial mass of the electron.

[1] V. V. Apollonov, A. I. Artem'ev, Yu. L. Kalachev *et al.*, JETP Lett. **47**, 91 (1988).
[2] M. V. Fedorov, *Electrons in a Strong Light Field* [in Russian], Nauka, Moscow (1991).
[3] W. Scheid and H. Hora, Laser Part. Beams **7**, 315 (1989).
[4] F. V. Hartemann, J. Woodworth, M. D. Perry *et al.*, Phys. Rev. E **51**, 4833 (1955).
[5] Yu. V. Dolgopolov, A. M. Dudov, L. I. Zykov *et al.*, Izv. Ros. Akad. Nauk, Ser. Fiz. **58**, 35 (1994).
[6] G. G. A. Kirilov, G. G. Kochemasov, and S. M. Kulikov, in *Proc. 12th Int. Conf. on Laser Interaction Plasma Phenomena, Osaka, Japan, 1995*, AIP Conf. 369, II, p. 866.
[7] M. V. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Theory of Waves* [in Russian], Nauka, Moscow (1979).
[8] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford (1975).

Translated by Eugene Yankovsky

Evolution of the width of the wave packet of a charged particle interacting with a quantum electromagnetic field

S. V. Faleev*)

G. I. Budker Nuclear Physics Institute, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 17 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1563–1578 (May 1997)

The path integral method is used to study the width of the wave packet of a relativistic charged particle interacting with a quantum electromagnetic field. A general expression is derived for the density distribution of a particle moving in arbitrary external potentials. An electron synchrotron with weak focusing is studied as a specific example, and the width of the wave packet of an electron moving in this accelerator is found. © *1997 American Institute of Physics.* [S1063-7761(97)00305-3] © *1997 American Institute of Physics.*

1. INTRODUCTION

The width of the wave packet of an electron moving in a circular accelerator is known to be determined, after a long enough time, by two competing factors, classical radiation damping,[1,2] and an increase in packet width due to the quantum nature of radiation.[3] However, the method used by Sands[1] and Kolomenskiĭ and Lebedev[2] to calculate the wave-packet width was semiclassical and statistical. The electron in this approach is a classical particle that spontaneously emits photons with a given (classical) average radiation intensity. Strictly speaking, in such an approach one must use such a notion as the width of a particle beam instead of the width of the wave packet of a single particle.

The present investigation develops the general method for a rigorous quantum calculation of the wave packet density of a relativistic charged particle interacting with a quantum electromagnetic field. To solve this problem it is convenient to formulate relativistic quantum mechanics using the language of path integrals. To allow for the effect of a quantum electromagnetic field on the particle in this approach one can use the influence functional technique.[4,5]

In Sec. 2 we derive an expression for the influence functional for the case where initially the electromagnetic field is in thermal equilibrium and is characterized by a temperature T. In the semiclassical approximation, where the packet is much narrower than the characteristic scale on which the external potentials vary, the fact that the interaction with a quantum electromagnetic field is taken into account leads to the emergence of three additional forces in the equation for the particle's classical path: the radiative reaction force and two fluctuation forces, which allow for the quantum nature of radiation. Section 3 is devoted to finding a general expression (in the form of path integrals of the fluctuation forces) for the density distribution of the wave packet of a charged particle moving in arbitrary external potentials. Finally, in Sec. 4 we discuss the evolution of the wave packet of an electron moving in a specific system—a synchrotron with weak focusing. We find that for each fixed configuration of the fluctuation forces the wave packet has an ''inherent'' relative density distribution, which is determined by the initial wave function and is independent of the fluctuation forces and whose width falls off exponentially with the passage of time. One the other hand, the coordinates of the center of this distribution as a whole depend strongly on the fluctuation forces, and in averaging over these forces we arrive at what is known as the ''Brownian'' contribution to the packet width, the principal contribution a large times.

The system we consider here is an example of a dissipative quantum system. The simplest model of a dissipative quantum system, a particle linearly coupled to an ensemble of independent harmonic oscillators, was studied by Caldeira and Leggett.[6] Despite the formal differences between the present model and model discussed in Ref. 6 (a relativistic setting and a more complicated coupling scheme), the packet width in both cases has the same structure, i.e., it can be separated into ''Brownian'' and ''intrinsic'' parts, each of which has a characteristic time dependence.

2. THE INFLUENCE FUNCTIONAL

The action of the system under investigation is given by the following expression:

$$S[q,A]=S_0[q]+S_I[q,A]+S_A[A]$$
$$=\int_0^t d\tau\,(-M\sqrt{1-\dot{q}^2}+e\dot{q}A^{(\mathrm{ex})}(q,\tau)-eU^{(\mathrm{ex})}(q,\tau))$$
$$+e\int_0^t d\tau\;\dot{q}A(q,\tau)-\frac{1}{4}\int_0^t d\tau\int d^3x(F_{\mu\nu}(x,\tau))^2, \tag{1}$$

where q, e, and M are the position, electric charge and mass of the electron, $A^{(\mathrm{ex})}$ and $U^{(\mathrm{ex})}$ are external vector and scalar potentials, and $F_{\mu\nu}(x,\tau)=\partial_\mu A_\nu(x,\tau)-\partial_\nu A_\mu(x,\tau)$ is the quantum electromagnetic field tensor. The speed of light and the Planck constant in the present paper are set equal to unity: $\hbar=c=1$. Here and in what follows (where possible), vector indices for the electron position vector $\mathbf{q}$, electromagnetic field $\mathbf{A}$, etc. are dropped. We also select the most convenient (for our purposes) Coulomb gauge for the field:

$$A_0=0 \quad \text{and} \quad \nabla\cdot\mathbf{A}(x)=0. \tag{2}$$

 1063-7761/97/050856-08$10.00

We express the Lagrangian of the electromagnetic field, L_A, in terms of the field's Fourier transform $A(k,\tau)=\int d^3x A(x,\tau)\exp(-ikx)$:

$$L_A[A]=\sum_{k_z>0}\left((\mathrm{Re}[\dot{A}(k,\tau)])^2+(\mathrm{Im}[\dot{A}(k,\tau)])^2 -k^2(\mathrm{Re}[A(k,\tau)])^2-k^2(\mathrm{Im}[A(k,\tau)])^2\right), \quad (3)$$

where for the sake of computational convenience we make the momentum $\mathbf{k}$ discrete and replace the integral over $\mathbf{k}$ by a sum over k:

$$\int \frac{d^3k}{(2\pi)^3}\to\sum_k .$$

The notation for the summation range in $\Sigma_{k_z>0}$ in (3) means that for the independent variables we take the real and imaginary parts of the Fourier components $A(k)=\mathrm{Re}[A(k)]+i\ \mathrm{Im}[A(k)]$ in the half-space $k_z>0$ $(A(-k)=\mathrm{Re}[A(k)]-i\ \mathrm{Im}[A(k)]$, since the field $A(x)$ is a real quantity).

In terms of the variables $\mathrm{Re}[A(k)]$ and $\mathrm{Im}[A(k)]$, the interaction Lagrangian L_I can be written

$$L_I=e\int d^3x\ \dot{q}(\tau)\delta^3(x-q(\tau))A(x)$$

$$=e\sum_{k_z>0}\left(2\ \cos(kq)\mathrm{Re}[A(k)] -2\ \sin(kq)\mathrm{Im}[A(k)]\right)\left(\dot{q}-k\frac{(\dot{q}k)}{k^2}\right). \quad (4)$$

The choice of gauge (2) $(\mathbf{A}(k)\cdot\mathbf{k}=0)$ makes it possible to add a term proportional to $\mathbf{k}$ to the right-hand side of Eq. (4). After doing so, we can assume that we also have longitudinal polarization of field $\mathbf{A}(k)\propto\mathbf{k}$, which, however, does not interact with the electron.

Suppose that initially an electron is in a pure state with wave function $\psi(q)$, and that the electromagnetic field is in thermal equilibrium at temperature T and is described by the density matrix $\rho(A,A')$.

The electron density matrix at time t is given by the given by the convolution of the density matrix propagator and the initial particle density matrix:

$$\rho(q_f,q_f',t)=\int dq_i\ dq_i'\ J(q_f,q_f',t;q_i,q_i',0)\psi(q_i)\psi^*(q_i'). \quad (5)$$

The density matrix propagator can be expressed in terms of path integrals over the electron paths:

$$J(q_f,q_f',t;q_i,q_i',0)$$

$$=\int\int Dq\ Dq'\ \exp(iS_0[q]-iS_0[q'])F[q,q'], \quad (6)$$

where the influence functional $F[q,q']$ can be expressed in terms of integrals over the fields $A(k)$:[4,5]

$$F[q,q']=\int\prod_{k_z>0} dA_f(k)\ dA_i(k)\ dA_i'(k)\ \rho_A(A_iA_i') \times\int\int\prod_{k_z>0} DA(k)\ DA'(k)\ \exp(i(S_I[q,A] -S_I[q',A']+S_A[A]-S_A[A'])). \quad (7)$$

All path integrals in (5) and (7) are evaluated along the paths $q(\tau)$, $q'(\tau)$, $A(\tau)$, and $A'(\tau)$ with boundary points $q(t)=q_f$, $q'(t)=q_f'$, $q(0)=q_i$, $q'(0)=q_i'$, $A(t)=A'(t)=A_f$, $A(0)=A_i$, and $A'(0)=A_i'$.

If we set the influence functional to unity, $F[q,q']=1$, the path integral (6) corresponds to the ''positive-frequency'' part of the Klein–Gordon equation for the particle's wave function.

$$[(i\partial_t-eU^{(\mathrm{ex})})-\sqrt{(-i\nabla-e\mathbf{A}^{(\mathrm{ex})})^2+M^2}]\psi(q,t)=0. \quad (8)$$

Note that we can speak of the wave function of a single particle only if the particle energy is much higher than the uncertainty in energy, which is related to both photon emission and the finite size of the wave packet. Otherwise we must use field theory with an arbitrary number of electron–positron pairs. Thus, in a comoving reference frame the energy of the emitted photons and the reciprocal wave-packet width expressed in energy units must be much less than the particle's rest mass Mc^2. In a circular accelerator the first condition yields the following restriction on the electron energy:

$$E\ll E_{1/2}\equiv Mc^2\left(\frac{RMc}{\hbar}\right)^{1/2},$$

where R is the accelerator's radius.

Another restriction on the applicability of our approach is the fact that the time intervals considered cannot be too short. The function

$$\psi(q_f,t)=\int^{q_f} Dq\ \exp(iS_0[q]) \quad (9)$$

is the solution of Eq. (8) only in the limit $dtM\gg 1$, where dt is the length of the time intervals into which the entire time interval t in the path integral with respect to Dq in (9) is partitioned. Furthermore, at $dtM\sim 1$ the path integral (9) is generally ill-defined. Thus, within our approach we cannot derive a correct expression for the (infinite) correction to the particle mass, which builds up precisely on short time intervals (see the discussion following Eq. (17)).

As noted earlier, the electron does not interact with the longitudinal polarization of the electromagnetic field (see the Lagrangian (4)), so that in (7) we can integrate over all three polarizations of the field $A(k)$.

The electromagnetic-field Lagrangian (3) consists of a set of Lagrangians of independent harmonic oscillators. The problem of calculating the influence functional for a harmonic oscillator linearly coupled to another system (see Eq. (4)) was solved exactly in Refs. 4 and 5. Thus, we can obtain the following expression for the influence functional (7):

$$F[q,q']=\exp(iI)\ \exp(-R), \quad (10)$$

where

$$I=I_{qq}-I_{q'q'}+I_{qq'}-I_{q'q}$$
$$=\int\frac{d^3k}{(2\pi)^3}\int_0^t d\tau\int_0^\tau ds\,\frac{e^2}{2|k|}\sin(|k|(\tau-s))\{\cos(k(q_\tau-q_s))$$
$$\times(\dot q_\tau\dot q_s-(\dot q_\tau n)(\dot q_s n))-\cos(k(q'_\tau-q'_s))(\dot q'_\tau\dot q'_s-(\dot q'_\tau n)$$
$$\times(\dot q'_s n))+\cos(k(q_\tau-q'_s))(\dot q_\tau\dot q'_s-(\dot q_\tau n)(\dot q'_s n))$$
$$-\cos(k(q'_\tau-q_s))(\dot q'_\tau\dot q_s-(\dot q'_\tau n)(\dot q_s n))\}\qquad(11)$$

and

$$R=\int\frac{d^3k}{(2\pi)^3}\int_0^t d\tau\int_0^t ds\,\frac{e^2}{4|k|}\cos(|k|(\tau-s))$$
$$\times\coth\frac{|k|}{2k_BT}\{\cos(k(q_\tau-q_s))(\dot q_\tau\dot q_s-(\dot q_\tau n)(\dot q_s n))$$
$$+\cos(k(q'_\tau-q'_s))(\dot q'_\tau\dot q'_s-(\dot q'_\tau n)(\dot q'_s n))$$
$$-2\cos(k(q_\tau-q'_s))(\dot q_\tau\dot q'_s-(\dot q_\tau n)(\dot q'_s n))\}.\qquad(12)$$

Here $\mathbf{n}\equiv\mathbf{k}/|k|$, k_B is Boltzmann's constant, and for the sake of brevity we have put $q_\tau\equiv q(\tau)$.

We start with the contribution of the term $I_{qq'}$. In integrating by parts the expression

$$\sin(|k|(\tau-s))\cos(k(q_\tau-q'_s))(\dot q_\tau n)(\dot q'_s n)$$
$$=\sin(|k|(\tau-s))\frac{d^2}{d\tau ds}\cos(k(q_\tau-q'_s))k^{-2}$$

in (11) we can replace the term $(\dot q_\tau n)(\dot q'_s n)$ in the integrand of (11) by unity. Then the dependence on the vector $\mathbf{n}$ in (11) becomes trivial, and the integral with d^3k can easily be evaluated:

$$I_{qq'}-I_{q'q}=\frac{\alpha}{2}\int_0^t d\tau\int_0^t ds\,\frac{1}{|q_\tau-q'_s|}\{\delta(\tau-s-|q_\tau-q'_s|)$$
$$-\delta(\tau-s+|q_\tau-q'_s|)\}(\dot q_\tau\dot q_s-1).\qquad(13)$$

Here

$$\alpha\equiv\frac{e^2}{4\pi\hbar c}\simeq\frac{1}{137}$$

is the electromagnetic coupling constant. The contribution of $I_{q'q}$ enters into (13) by virtue of the fact that the domain of integration with respect to s extend over the entire time interval t.

The particle-path variation $\delta q_\tau\equiv q_\tau-q'_\tau$, which coincides in order of magnitude with the size of the wave packet, is much smaller than the characteristic scale over which the external potentials vary, L. This means that it suffices to keep only terms that are linear in δq in (13). Further calculations yield

$$I_{qq'}-I_{q'q}=\frac{2\alpha}{3}\int_0^t d\tau\;\delta q_\tau\left(\frac{\ddot v}{1-v^2}+\frac{3\dot v(v\dot v)}{(1-v^2)^2}\right.$$
$$\left.+\frac{v(v\ddot v)}{(1-v^2)^2}+\frac{3v(v\dot v)^2}{(1-v^2)^3}\right)\left(1+O\left(\frac{(\delta q)^2}{L^2}\right)\right),\qquad(14)$$

where $v=\dot{\overline{q}}$ is the velocity of the ''average'' particle path $\overline{q}_\tau\equiv(q_\tau+q'_\tau)/2$. Clearly, the contribution of (14) to the influence functional leads to the appearance of the well-known radiative reaction forces[7] in the equation for the classical path.

Let us write the term I_{qq} in (11) in Lorentz-invariant form:

$$I_{qq}=-\alpha\int_0^t d\tau\int_0^t ds\int d^3\mathbf{r}\int d^3\mathbf{r}'\,j_\mu(r,\tau)j^\mu(r',s)$$
$$\times\delta((\tau-s)^2-(r-r')^2),\qquad(15)$$

where $j_\mu=(\rho,\mathbf{j})$ is the ''particle'' density four-vector, with

$$\rho(r,\tau)=\delta^3(\mathbf{r}-\mathbf{q}_\tau)$$

and

$$\mathbf{j}(r,\tau)=\dot{\mathbf{q}}_\tau\delta(\mathbf{r}-\mathbf{q}_\tau).$$

We calculate the relativistic scalar I_{qq} in the comoving reference frame:

$$I_{qq}=\int_0^t d\tau\;\sqrt{1-\dot q^2}(-\delta M),\qquad(16)$$

where $d\tau\sqrt{1-\dot q^2}$ is the ''proper-time'' differential, and

$$\delta M=\int d^3\mathbf{r}\int d^3\mathbf{r}'\rho_0(r)\rho_0(r')\frac{\alpha}{2}\frac{1}{|r-r'|},\qquad(17)$$

where $\rho_0(r)$ is the ''particle'' density distribution in the particle's rest frame. We see that I_{qq} in (16) leads to the renormalization of the electron mass, $M+\delta M\to M$, with the expression for δM coinciding with the classical expression for the self-energy of a distributed charge. In our case the correction to the mass proves to be infinite, since the ''charge'' is a point charge: $\rho_0(r)=\delta(\mathbf{r})$.

In Eq. (17) we have the classical linear divergence of the electron mass, $\delta M\sim e^2/r_e$, where r_e represents the electron ''radius.'' This contradicts the quadratic divergence of the particle mass in scalar QED, $\delta(M^2)\sim\alpha\Lambda^2$, where Λ is the cutoff in integrals over momenta in QED. The reason for this inconsistency is that, as noted earlier, the path integral (6) does not describe very short time intervals $|\tau-s|\leqslant 1/M$, which, as Eq. (15) shows, contribute to the divergence of the electron self-energy.

The expression (12) for R in leading (quadratic in δq) order has the following form:

$$R=\int\frac{d^3k}{(2\pi)^3}\int_0^t d\tau\int_0^t ds\,\frac{e^2\cos(|k|(\tau-s))}{4|k|}\coth\left(\frac{|k|}{2k_BT}\right)$$
$$\times\cos(k(\overline{q_\tau}-\overline{q_s}))\{(\delta\dot q_\tau\delta\dot q_s-(\delta\dot q_s n)(\delta\dot q_\tau n))+(k\delta q_\tau)$$
$$\times(k\delta q_s)[\dot{\overline{q_\tau}}\dot{\overline{q_s}}-(\dot{\overline{q_s}}n)(\dot{\overline{q_\tau}}n)]\}\left(1+O\left(\frac{(\delta q)^2}{L^2}\right)\right).\qquad(18)$$

To leave only terms that are linear in δq in the effective action, we write the exponential e^{-R} in terms of several new path integrals of what is known as fluctuation forces, $F_\parallel(\tau)$ and $F_\perp(\tau)$, with weight functions $P_\parallel[F_\parallel]$ and $P_\perp[F_\perp]$ given by the identity

$$e^{-R}\equiv\int\int dF_\parallel\, DF_\perp\; P_\parallel[F_\parallel]P_\perp[F_\perp]$$
$$\times\exp\left\{ i\int_0^t d\tau\, F_\parallel(\tau)\,\delta q(\tau)\right.$$
$$\left.-i\int_0^t d\tau\, F_\perp(\tau)\,\delta\dot{q}(\tau)\right\}. \tag{19}$$

The correlation functions of $F_\parallel$ and $F_\perp$ at different times can easily be found by employing the second functional derivative of the exponential (19) with respect to δq and $\delta\dot{q}$:

$$\langle F_{\parallel i}(\tau)F_{\parallel j}(s)\rangle=\int\frac{d^3k}{(2\pi)^3}\,\frac{e^2\cos(|k|(\tau-s))}{2|k|}$$
$$\times\coth\left(\frac{|k|}{2k_BT}\right)\cos(k(\overline{q}_\tau-\overline{q}_s))$$
$$\times k_ik_j(\dot{\overline{q}}_\tau\dot{\overline{q}}_s-(\dot{\overline{q}}_\tau n)(\dot{\overline{q}}_s n)), \tag{20}$$

$$\langle F_{\perp i}(\tau)F_{\perp j}(s)\rangle=\int\frac{d^3k}{(2\pi)^3}\,\frac{e^2\cos(|k|(\tau-s))}{2|k|}$$
$$\times\coth\left(\frac{|k|}{2k_BT}\right)\cos(k(\overline{q}_\tau-\overline{q}_s))$$
$$\times(\delta_{ij}-n_in_j). \tag{21}$$

These correlation functions clearly show that in the ultrarelativistic case, with the photons emitted chiefly along the particle velocity vector, the force $F_\parallel$ acts parallel to the velocity and the force $F_\perp$ perpendicular to the velocity.

In the nonrelativistic limit we can ignore the contribution of $F_\parallel$, since it contains two extra powers of the velocity over and above the contribution of $F_\perp$. On the other hand, in the ultrarelativistic case, $F_\parallel$ provides the principal contribution, since it contains the frequency $|k|$ of the emitted photons as a factor, and this frequency is much higher than the characteristic frequencies of particle motion in external potentials.

In the semiclassical limit we are examining here, where the wave-packet width can be assumed small compared to the scale on which the external potentials vary, we can interpret the path $\overline{q}$ in (20) and (21) as a classical path obtained by solving classical equations of motion without allowing for corrections associated with fluctuation forces.

The physical meaning of fluctuation forces is clear. These forces allow for the quantized, fluctuating nature of the radiation. In connection with the problem of electron motion in a circular accelerator, such forces were introduced in Refs. 1 and 2 statistically on the basis of the fact that photons are emitted independently of one another. In the present work the fluctuation forces appeared in a natural manner from exact quantum calculations.

3. DENSITY DISTRIBUTION

In the path integrals in (6) we introduce new variables $\overline{q}$ and δq instead of $q(\tau)$ and $q'(\tau)$:

$$\overline{q}(\tau)=\frac{q(\tau)+q'(\tau)}{2},\quad \delta q(\tau)=q(\tau)-q'(\tau). \tag{22}$$

In terms of these new variables, the expression (6) for the particle density distribution is

$$\rho(\overline{q}_f,t)=\int d\overline{q}_i\, d(\delta q_i)\;\psi\left(\overline{q}_i+\frac{\delta q_i}{2}\right)\psi^*\left(\overline{q}_i-\frac{\delta q_i}{2}\right)$$
$$\times\int DF_\parallel\, DF_\perp\; P_\parallel[F_\parallel]P_\perp[F_\perp]$$
$$\times\int D\overline{q}\; D(\delta q)\;\exp(iS_{\rm eff}[\overline{q},\delta q]). \tag{23}$$

The path integrals in (23) are along the paths $\overline{q}(\tau)$ and $\delta q(\tau)$ the boundary points $\overline{q}(0)=\overline{q}_i$, $\overline{q}(t)=\overline{q}_f$, $\delta q(0)=\delta q_i$, and $\delta q(t)=0$. The effective action $S_{\rm eff}$ in an approximation linear in δq was obtained in Sec. 2 (see Eqs. (10), (14), and (19)):

$$S_{\rm eff}[\overline{q},\delta q]=\int_0^t d\tau\left\{\delta\dot{q}\,\frac{Mv}{\sqrt{1-v^2}}+\delta q(eE(\overline{q})\right.$$
$$+e[v,H(\overline{q})]+F_{\rm fr}+F_\parallel)-\delta\dot{q}F_\perp$$
$$\left.+e\frac{d}{dt}(\delta qA^{(\rm ex)}(\overline{q}))\right\}, \tag{24}$$

where $v(\tau)\equiv\dot{\overline{q}}(\tau)$, $E(\overline{q})$ and $H(\overline{q})$ are the external electric and magnetic fields, and $F_{\rm fr}$ is the radiative reaction force,

$$F_{\rm fr}=\frac{2\alpha}{3}\left(\frac{\ddot{v}}{1-v^2}+\frac{3\dot{v}(v\dot{v})}{(1-v^2)^2}+\frac{v(v\ddot{v})}{(1-v^2)^2}+\frac{3v(v\dot{v})^2}{(1-v^2)^3}\right). \tag{25}$$

The path integrals along the electron paths in (23) can be evaluated by the saddle-point method in the semiclassical approximation. The saddle point of the effective action $S_{\rm eff}$ can be found by solving the equations

$$\frac{\delta}{\delta(\delta q)}S_{\rm eff}[\overline{q}_{\rm cl},\delta q_{\rm cl}]=0,\quad \frac{\delta}{\delta q}S_{\rm eff}[\overline{q}_{\rm cl},\delta q_{\rm cl}]=0 \tag{26}$$

with appropriate boundary conditions. The first equation in (26) is independent of $\delta q_{\rm cl}$, and has the form of a classical equation of motion in external fields with a radiative reaction force and fluctuation forces on the right-hand side:

$$\frac{d}{dt}\,\frac{M\dot{\overline{q}}}{\sqrt{1-\dot{\overline{q}}^2}}=eE+e[\dot{\overline{q}},H]+eF_{\rm fr}+F_\parallel+\dot{F}_\perp. \tag{27}$$

The saddle-point value of the effective action (24) calculated along the classical path $\overline{q}_{\rm cl}$ is independent of $\delta q_{\rm cl}$ and is purely a surface term:

$$S_{\rm eff}[\overline{q}_{\rm cl}]=-\delta q_i(p_i+eA^{(\rm ex)}(\overline{q}_i)-F_\perp(0)), \tag{28}$$

where $p_i = p_i(\overline{q_i}, \overline{q_f}; t; F_\parallel; F_\perp)$ is the initial momentum of the classical path that satisfies Eq. (27) with boundary conditions $\overline{q}(0) = \overline{q_i}$ and $\overline{q}(t) = \overline{q_f}$.

Clearly, the path integrals along the particle paths in Eq. (23), written in terms of variables shifted by their classical values, $\overline{q} - \overline{q}_{\rm cl}$ and $\delta q - \delta q_{\rm cl}$, do not depend in the semiclassical approximation on the fluctuation forces and the boundary points $\overline{q_f}$, $\overline{q_i}$, and δq_i, and reduce to a function that depends only on time.

Thus, only path integrals of the fluctuation forces remain in the expression for the electron density distribution:

$$\rho(\overline{q_f}, t) = \widetilde{N}(t) \int DF_\parallel \, DF_\perp \, P_\parallel[F_\parallel] P_\perp[F_\perp] \times \int d\overline{q_i} \, f(\overline{q_i} p_i(\overline{q_i}, \overline{q_f}; t; F_\parallel, F_\perp)), \qquad (29)$$

where the function $f(q,p)$ signifies the initial distribution function in the phase space,

$$f(q,p) = \int d(\delta q) \; \psi^*\left(q - \frac{\delta q}{2}\right) \psi\left(q + \frac{\delta q}{2}\right) \times \exp\{-i \delta q (p + eA^{(\rm ex)}(q) - F_\perp(0))\}. \qquad (30)$$

The presence of a term in braces that is proportional to $F_\perp(0)$ can be explained by the fact that, as Eq. (27) clearly shows, the time derivative $\dot{F}_\perp$ is actually a force, and $-F_\perp/e$ is thereby an addition to the vector potential of the external field.

As Eq. (29) clearly shows, for any fixed configuration of the fluctuation forces, the wave function has a density distribution coinciding with that of a beam of classical particles, a distribution that was initially specified by the distribution function (30) in the phase space, and that then evolved with the passage of time in accordance with the classical equation of motion (27) with a radiative damping force and fluctuation forces on the right-hand side.

Thus, in this section we have derived a closed expression, consisting of (29) and (30), for the density distribution of the wave packet of a charged scalar particle moving in arbitrary external potentials with an arbitrary initial temperature of the quantum electromagnetic field (the temperature enters into the expressions for the correlation functions (20) and (21)).

Now let us recall the principal assumptions made in deriving Eq. (29) for the wave-packet density distribution. In the particle's rest frame, the characteristic energy of the emitted photons and the reciprocal wave-packet width expressed in energy units must be much less than the particle's rest mass Mc^2 (only then can we sensibly speak of a single particle). On the other hand, the packet width must be small compared to the scale on which the external potentials vary (only then can we speak of a wave packet).

It is not very difficult to calculate the width of a wave packet for a specific system of external fields using the adopted method, since the entire calculation reduces to solving the classical equation of motion (27). In the approximation of a small deviation δq from the path $q_{\rm cl}$ followed by the center of a beam of classical particles that is fixed initially by the distribution function (30) in the phase space, Eq. (27) becomes linear in this deviation and contains fluctuation forces on its right-hand side.

The solution of the resulting linear equation with the specified right-hand side can be represented by the sum of a particular solution δq_p, which depends on the fluctuation forces, and the general solution of the linear equation (with zero on the right-hand side) that satisfies the corresponding boundary conditions. In this way we arrive at an expression (similar to the expression (41) in Sec. 4) that relates the boundary points $\overline{q_i}$ and $\overline{q_f}$ to the initial momentum p_i (which is an argument of the distribution function in (29)), with the final point $\overline{q_f}$ and the particular solution δq_p entering into the expression for p_i only as the difference $\delta q_f - \delta q_p$, where $\delta q_f \equiv \overline{q_f} - q_{\rm cl}$ is the deviation of the final point from the classical path. Since the particular solution δq_p is the only quantity that depends on the fluctuation forces, for any fixed configuration of these forces the wave packet has an ''intrinsic'' relative density distribution that is independent of the fluctuation forces and is determined by the initial wave function. The role of the fluctuation forces reduces to simply shifting the derived (classical) distribution as a whole.

Thus, the square of the packet width (along the axes $j = x,y,z$)

$$\sigma_j^2 \equiv \left\langle \int d^3 q_f \; \rho(q_f, t) (\delta q_f^j)^2 \right\rangle$$

separates into what is known as the ''Brownian'' part and the intrinsic part. The Brownian contribution to the square of the packet width is given by the radiation fluctuations and is determined by the average over the fluctuation forces of the square of the particular solution obtained earlier: $(\sigma_j^{(\rm Br)})^2 \equiv \langle (\delta q_p^j)^2 \rangle$. The intrinsic width, on the other hand, is independent of the fluctuation forces and hence coincides with the width of a beam of classical particles that initially is specified by the distribution function (30) in the phase space and evolves with time in accordance with the classical equation of motion (27) with the radiative reaction force on the right-hand side but without the fluctuation forces.

4. WIDTH OF THE ELECTRON WAVE PACKET IN A CIRCULAR ACCELERATOR

In this section we attempt to use the general formulas obtained earlier to calculate the width of the wave packet of an ultrarelativistic electron moving in a circular accelerator, i.e., a synchrotron with weak focusing.

Suppose that an electron is moving in a uniform magnetic field $\mathbf{H} = (0,0,H)$ directed along the z axis. The radius R of an equilibrium path determines the electron velocity $v_0 = R\omega$ and the Larmor frequency $\omega = (eH/M)\sqrt{1 - v_0^2}$. Let us assume that within a small range of angles $\Delta\varphi$ about an angle β measured clockwise in the xy plane from the x semiaxis ($\varphi \in [\beta - \Delta\varphi/2, \beta + \Delta\varphi/2]$) there exists a nonvanishing accelerating electric field that is a harmonic function of time (i.e., varies with the Larmor frequency ω). Within this interval the electric field is described by the following potential:

$$U(q_x,q_y,\tau)=E_0(R+r)(\varphi-\beta)\cos\,\omega\tau,$$

where $R+r\equiv(q_x^2+q_y^2)^{1/2}$ is the radius and φ is the angle in cylindrical coordinates. The amplitude E_0 of the electric field is selected in such a way that, when moving along the equilibrium path from point $\mathbf{R}=(R,0,0)$, in the acceleration region the electron acquires an amount of energy that exactly balances the radiative losses:

$$|e|E_0R\Delta\varphi\,\cos\,\beta=\frac{2\pi}{\omega}W_0. \tag{31}$$

Here $W_0=(2/3)\alpha\omega^2v_0^2\gamma^4$ is the average power of radiative losses,[7] and $\gamma=(1-v_0^2)^{-1/2}$ is the (large) relativistic factor.

Electron motion along the z axis is regulated by a potential of the general type $U_z(q_z)$, with absolute minimum at $q_z=0$. Thus, the electron action $S_0[q]$ is

$$S_0[q]=\int_0^t d\tau\left\{-M\sqrt{1-\dot{q}^2}+\frac{1}{2}e\dot{\mathbf{q}}\cdot(\mathbf{H}\times\mathbf{q})\right.$$
$$\left.-eU(q_x,q_y,\tau)-eU_z(q_z)\right\}. \tag{32}$$

As noted at the end of Sec. 2, in the ultrarelativistic case being investigated, $\gamma\gg1$, we can ignore the force $F_\perp$. The contribution of this force is smaller than that of $F_\parallel$ at least by a factor of $k_{\rm ch}^2/\omega^2\sim\gamma^6$, where $k_{\rm ch}\sim\omega\gamma^3$ is the characteristic frequency of the photons emitted in the accelerator.

Suppose that initially the electron is in a pure state described by a Gaussian wave packet with its center at $\mathbf{R}$, width σ_0, and momentum $\mathbf{p}_0=(0,Mv_0\gamma,0)$:

$$\psi(q)=\frac{1}{(2\pi\sigma_0^2)^{3/2}}\exp\left\{i\dot{\mathbf{q}}\left(\mathbf{p}_0+\frac{1}{2}e\mathbf{H}\times\mathbf{R}\right)-\frac{(\mathbf{q}-\mathbf{R})^2}{4\sigma_0^2}\right\}. \tag{33}$$

The second term in braces allows for the fact that in the presence of a magnetic field the momentum operator must be ''extended'':

$$\hat{p}=-i\frac{\partial}{\partial q}-eA^{(\rm ex)}(q).$$

The wave function (33) specifies the distribution function in phase space (see Eq. (30)):

$$f(q.p_i)\propto\exp\left\{-\frac{(\mathbf{q}-\mathbf{R})^2}{2\sigma_0^2}-2\sigma_0^2\left(p_i-p_0\right.\right.$$
$$\left.\left.+\frac{1}{2}e\mathbf{H}\times(\mathbf{q}-\mathbf{R})\right)^2\right\}. \tag{34}$$

Now we must express the initial momentum of a classical path, p_i, in terms of the boundary points, i.e., we must solve the equation of motion (27). To this end we introduce new variables—the deviation from the equilibrium radius, r, and the deviation from the equilibrium phase, ϕ:

$$\overline{q_x}=(R+r)\cos(\omega\tau+\phi),\quad \overline{q_y}=(R+r)\sin(\omega\tau+\phi). \tag{35}$$

Since deviations from the equilibrium path are small, the electron moves along the z axis and in the xy plane independently. Hence the density distribution can be factorized. Below we are interested in the electron density distribution $\rho(r_f,\phi_f,t)$ only in the xy plane.

In an approximation that is linear in r and ϕ the classical equation of motion (27) is cumbersome:

$$\ddot{r}\gamma=\omega R\dot{\phi}\gamma^3+v_0^2\omega^2r\gamma^3+\frac{2\alpha}{3M}\{\gamma^2\dddot{r}-\gamma^4(3\dot{r}\omega^2$$
$$+3\omega R\ddot{\phi}+v_0^2\dot{r}\omega^2)\}+f_1,$$
$$R\ddot{\phi}\gamma^3=-\omega\dot{r}\gamma^3+\frac{2\alpha}{3M}\left\{\gamma^4(3\omega\ddot{r}-3R\dot{\phi}\omega^2-r\omega^3+R\dddot{\phi})\right.$$
$$-4v_0^2\omega^2\gamma^6(R\dot{\phi}+r\omega)+v_0\omega^2\gamma^4$$
$$\left.\times\left(-1+\frac{\cos(\phi-\beta)}{\cos\,\beta}\right)\right\}+f_2, \tag{36}$$

where

$$f_1=\frac{F_{\parallel x}\cos\,\omega\tau+F_{\parallel y}\sin\,\omega\tau}{M},$$
$$f_2=\frac{-F_{\parallel x}\sin\,\omega\tau+F_{\parallel y}\cos\,\omega\tau}{M}. \tag{37}$$

In Eq. (36) we left only the first harmonic from the Fourier expansion in the angle φ of the accelerating electric field. The contribution of other harmonics (say, $\propto\cos2\varphi\cos\omega\tau$) vanish in averaging over the rotation period.

We write Eq. (36) in matrix form:

$$\hat{M}_{ij}(\tau)\varphi_j(\tau)=f_i(\tau). \tag{38}$$

Here the indices i and j run through the values 1 and 2, and $\varphi_i\equiv(r(\tau),R\phi(\tau))$. The homogeneous equation ($f_i=0$) has four solutions:

$$\varphi_i^{(1)}(\tau)=(\cos\,\omega\tau,-\sin\,\omega\tau)e^{-\gamma_r\tau},$$
$$\varphi_i^{(2)}(\tau)=(\sin\,\omega\tau,\cos\,\omega\tau)e^{-\gamma_r\tau},$$
$$\varphi_i^{(3)}(\tau)=\left(\cos\,\Omega\tau,-\frac{v_0^2\omega}{\Omega}\sin\,\Omega\tau\right)e^{-\gamma_\phi\tau},$$
$$\varphi_i^{(4)}(\tau)=\left(\sin\,\Omega\tau,\frac{v_0^2\omega}{\Omega}\cos\,\Omega\tau\right)e^{-\gamma_\phi\tau}. \tag{39}$$

The first two are what is known as radial betatron oscillations with Larmor frequency ω and damping constant γ_r. These solutions correspond to translation of the orbit center in a uniform magnetic field and are damped because of the correct choice of the dependence of the amplitude of the accelerating voltage on the radial deviation r:

$$U\propto R+ar\quad\text{with}\quad a=1.$$

Clearly, if $a=0$ (the energy acquired by the electron in the acceleration interval does not depend on the radial deviation), the radial oscillations are undamped ($\gamma_r=0$), and if $a<0$, the become exponentially increasing ($\gamma_r<0$). The third and fourth solutions in (39) correspond to synchrotron radial-phase oscillations with frequency $\Omega\ll\omega$ and damping constant γ_ϕ. The following expressions can be obtained for

the radial damping constant γ_r, the phase damping constant γ_ϕ, and the frequency of synchrotron oscillations:

$$\gamma_r=\frac{\alpha}{3M}\omega^2\gamma^3, \quad \gamma_\phi=\frac{2\alpha}{3M}\omega^2\gamma^3,$$

$$\Omega^2=\frac{2\alpha}{3M}\omega^3\gamma^3\tan\beta. \tag{40}$$

The solution of the inhomogeneous equation (38) with given initial conditions in an approximation in which $\gamma_r,\gamma_\phi\ll\Omega\ll\omega$ has the form

$$\varphi_i(t)=\left(r(0)-\frac{\delta p_y}{\omega M\gamma}\right)\varphi_i^{(1)}+\left(R\phi(0)+\frac{\delta p_x}{\omega M\gamma}\right)\varphi_i^{(2)} + \frac{\delta p_y}{\omega M\gamma}\varphi_i^{(3)}-\frac{\Omega\,\delta p_x}{v_0^2\omega^2 M\gamma}\varphi_i^{(4)}+\varphi_{pi}, \tag{41}$$

where $\delta p=p_i-p_0$ is the deviation of the initial classical-path momentum p_i from the equilibrium momentum p_0:

$$\delta p_x=\gamma M(\dot r(0)-\omega R\phi(0)),$$

$$\delta p_y=\gamma^3 M(R\dot\phi(0)+r(0)\omega).$$

The last term on the right-hand side of Eq. (41), $\varphi_{pi}(t)=(r_p(t),R\phi_p(t))$, is the particular solution of Eq. (38) with initial conditions $\varphi_{pi}(0)=\dot\varphi_{pi}(0)=0$. This solution can be found by the Green's functions technique:

$$\varphi_{pi}(\tau)=\int_0^\tau d\tau'\ G_{ij}(\tau,\tau')f_j(\tau'). \tag{42}$$

The Green's function $G_{ij}(\tau,\tau')$ must satisfy the equation

$$\hat M_{ij}(\tau)G_{jk}(\tau,\tau')=\delta_{ik}\delta(\tau-\tau') \tag{43}$$

with initial conditions

$$G_{ij}(0,\tau')=0, \quad \left.\frac{dG_{ij}(\tau,\tau')}{d\tau}\right|_{\tau=0}=0.$$

Writing the Green's function in the form of an expansion in the solutions (39) of the homogeneous equation with $\Delta\tau\equiv\tau\tau'>0$, we obtain

$$G_{ij}(\tau,\tau')=\begin{pmatrix}\sin\omega\Delta\tau & -\cos\omega\Delta\tau\\ \cos\omega\Delta\tau & \sin\omega\Delta\tau\end{pmatrix}\frac{e^{-\gamma_r\Delta\tau}}{\omega\gamma} -\begin{pmatrix}0 & -\cos\Omega\Delta\tau\\ \cos\Omega\Delta\tau & \dfrac{v_0^2\omega}{\Omega}\sin\Omega\Delta\tau\end{pmatrix}\frac{e^{-\gamma_\phi\Delta\tau}}{\omega\gamma}. \tag{44}$$

To avoid cumbersome formulas in the final result, we assume that initially the packet is very narrow: $\sigma_0^2\omega M\ll1$. Then, in times of order of a single rotation period the width of the wave packet becomes much larger than the initial width σ_0, and we can ignore terms proportional to the initial deviations $r(0)$ and $\phi(0)$ in Eq. (41) and the term in (34) proportional to $e\mathbf{H}\times(\mathbf{q}-\mathbf{R})$. Here the integral with respect to $\overline{q_i}$ in (29) becomes trivial, and the final result for the density distribution in the electron wave packet in the xy plane assumes the form

$$\rho(r_f,\phi_f,t)=N(t)\int DF_\parallel(\tau)P_\parallel[F_\parallel] \times\exp\left\{-\frac{2(M\sigma_0\gamma\omega)^2}{(A^2+BC)^2}([A(r_f-r_p) +BR(\phi_f-\phi_p)]^2+[C(r_f-r_p) -AR(\phi_f-\phi_p)]^2)\right\}, \tag{45}$$

where

$$A=-\cos\omega te^{-\gamma_r t}+\cos\Omega te^{-\gamma_\phi t}, \quad B=\sin\omega te^{-\gamma_r t}, \tag{46}$$

$$C=\sin\omega te^{-\gamma_r t}-\frac{\omega v_0^2}{\Omega}\sin\Omega te^{-\gamma_\phi t}.$$

Equation (45) clearly shows that for any fixed configuration of $F_\parallel(\tau)$ the wave packet has an ''intrinsic'' relative density distribution, which is independent of $F_\parallel$. Only the coordinates (r_p,ϕ_p) of the center of this distribution as a whole depend on fluctuation forces. As noted at the end of Sec. 3, such behavior of the density distribution is general in nature, and does not depend on the shape of the external potentials.

For the radial and angular dispersions of the electron coordinate we have the following expressions:

$$\sigma_r^2(t)\equiv\frac{\int dr_f d(R\phi_f)\rho(r_f,\phi_f,t)r_f^2}{\int dr_f d(R\phi_f)\rho(r_f,\phi_f,t)} =(\sigma_r^{(\mathrm{int})})^2+\langle r_p^2(t)\rangle, \tag{47}$$

$$\sigma_\phi^2(t)\equiv\frac{\int dr_f d(R\phi_f)\rho(r_f,\phi_f,t)(R\phi_f)^2}{\int dr_f d(R\phi_f)\rho(r_f,\phi_f,t)} =(\sigma_\phi^{(\mathrm{int})})^2+\langle(R\phi_p(t))^2\rangle. \tag{48}$$

The intrinsic widths $\sigma_r^{(\mathrm{int})}$ and $\sigma_\phi^{(\mathrm{int})}$ are determined by the initial velocity spread; they decay exponentially with time because of classical radiative reaction:

$$(\sigma_r^{(\mathrm{int})})^2=\frac{1}{(2\sigma_0 M\gamma\omega)^2}\{(\cos\omega te^{-\gamma_r t}-\cos\Omega te^{-\gamma_\phi t})^2 +(\sin\omega te^{-\gamma_r t})^2\}, \tag{49}$$

$$(\sigma_\phi^{(\mathrm{int})})^2=\frac{1}{(2\sigma_0 M\gamma\omega)^2}\left\{(\cos\omega te^{-\gamma_r t}-\cos\Omega te^{-\gamma_\phi t})^2 +\left(\sin\omega te^{-\gamma_r t}-\frac{\omega v_0^2}{\Omega}\sin\Omega te^{-\gamma_\phi t}\right)^2\right\}.$$

The Brownian contributions to the radial and angular widths, $\sigma_r^{(\mathrm{Br})}$ and $\sigma_\phi^{(\mathrm{Br})}$, are determined by quantum radiation fluctuations:

$$(\sigma_r^{(\mathrm{Br})})^2 \equiv \langle r_p^2(t)\rangle = \frac{1}{2(M\gamma\omega)^2}\left(\frac{1-e^{-2\gamma_r t}}{2\gamma_r}+\frac{1-e^{-2\gamma_\phi t}}{2\gamma_\phi}\right) \times \int d^3k \coth\left(\frac{|k|}{2k_BT}\right)\frac{dI}{d^3k}|k|,$$

$$(\sigma_\phi^{(\mathrm{Br})})^2 \equiv \langle (R\phi_p(t))^2\rangle = \frac{1}{2(M\gamma\Omega)^2}\,\frac{1-e^{-2\gamma_\phi t}}{2\gamma_\phi} \times \int d^3k \coth\left(\frac{|k|}{2k_BT}\right)\frac{dI}{d^3k}|k|. \tag{50}$$

To average the squares of the particular solutions (42) over the forces $F_\parallel$ we use (44) for the Green's function and Eq. (20) for the correlation function of fluctuation forces.

The expression for the function dI/d^3k in (50) coincides with the well−known formula for the classical radiation intensity[8]

$$\frac{dI}{d^3k} = \frac{\alpha}{4\pi^2}\int_{-\infty}^{\infty} d\tau\,\{\dot q(\tau)\dot q(0)-(\dot q(\tau)n)(\dot q(0)n)\} \times \exp\{i|k|\tau - i\mathbf{k}\cdot(\mathbf{q}(\tau)-\mathbf{q}(0))\}, \tag{51}$$

where $q(\tau)$ is the classical electron path. At all reasonable temperatures we have the strong inequality $k_BT \ll k_{\mathrm{ch}} \sim \omega\gamma^3$. In this limit the expression for the integral in (50) is known:[1]

$$\int d^3k\,\frac{dI}{d^3k}|k| = \frac{55}{24\sqrt{3}}\alpha\omega^3\gamma^7. \tag{52}$$

In the problem under investigation and for times of the order of the photon emission time $|\tau - s| \sim (\omega\gamma)^{-1}$ we can ignore the structure of the correlation function (20), since in calculating the averages $\langle r_p^2(t)\rangle$ and $\langle (R\phi_p(t))^2\rangle$ the given function is present in the integral over τ and s, together with functions whose characteristic time scale is $\sim 1/\omega$ or $\sim 1/\Omega$. Bearing this in mind, we can simplify the expression (20) for the correlation function:

$$\langle F_\parallel(\tau)F_\parallel(s)\rangle = \delta(\tau - s)\int d^3k\,\frac{dI}{d^3k}|k|. \tag{53}$$

Clearly, the correlation function (53) is the same as the correlation function for the fluctuation forces introduced by Sands[1] and Kolomenskiĭ and Lebedev[2] on the basis of statistical ideas. Thus, the expressions (50) for the Brownian contributions to the packet width, which were obtained by rigorous quantum calculation, coincide with the corresponding expressions obtained by the semiclassical method.

Let us now assess the applicability of our results. The energy of the emitted photons, $\sim\omega\gamma^3$, must be much lower than the electron energy $E = M\gamma$. This leads to an upper bound on the electron energy,

$$E \ll E_{1/2} \equiv Mc^2\left(\frac{RMc}{\hbar}\right)^{1/2}.$$

For an estimate, we take the accelerator radius equal to ten meters. We see that all existing accelerators meet this condition: $E \ll E_{1/2} \sim 10^6$ MeV.

As for restrictions on the packet width, Eqs. (50) show that our results are valid only for times much shorter than $t \ll 1/\gamma_\phi$. For times $t \sim 1/\gamma_\phi \sim MR^2/\alpha\hbar\gamma^3 \sim 10^{-1}$s (for parameter values $R \sim 10^3$ cm and $\gamma \sim 10^3$), the Brownian width reaches a value comparable to the accelerator's radius, $\sigma_\phi^{(\mathrm{Br})} \simeq R/\sqrt{\alpha\gamma}$. This constitutes the main drawback of accelerators with weak focusing. To reduce the angular spread of the beam to values much smaller than the accelerator dimensions, accelerators with strong focusing are used.[9]

The intrinsic contribution to the packet's angular width (see the second equation in (49)) is inversely proportional to the initial packet width and rapidly falls off as the relativistic factor γ increases:

$$\sigma_\phi^{(\mathrm{int})} \sim \left(\frac{R^3\hbar}{\sigma_0^2 Mc\alpha\gamma^5}\right)^{1/2}.$$

For typical parameter values $R \sim 10^3$ cm, $\gamma \sim 10^3$, and $\sigma_0 \sim 10^{-4}$ cm, the intrinsic angular width $\sigma_\phi^{(\mathrm{int})}$ is of order 10^{-3} cm.

The characteristic values of the Brownian and intrinsic radial widths are smaller than the corresponding angular widths by a factor of $\omega/\Omega \sim (RMc/\hbar\alpha\gamma^3)^{1/2} \sim 10^3$ (at $R \sim 10^3$ cm and $\gamma \sim 10^3$).

The author is grateful to I. V. Kolokolov for suggesting the topic and for the useful remarks, and to I. B. Khriplovich for attending to this work and for valuable remarks. The present work supported in part by INTAS Grant No. 93-2492-ext.

*)e-mail: S. V. Faleev@INP.NSK.SU

[1] M. Sands, Phys. Rev. **97**, 470 (1955).
[2] A. A. Kolomenskiĭ and A. N. Lebedev, Zh. Éksp. Teor. Fiz. **30**, 205 (1956) [Sov. Phys. JETP **3**, 130 (1956)].
[3] A. A. Sokolov and I. M. Ternov, Zh. Éksp. Teor. Fiz. **25**, 698 (1953).
[4] R. P. Feynman and F. L. Vernon, Ann. Phys. (N.Y.) **24**, 118 (1963).
[5] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York (1965).
[6] A. O. Caldeira and A. J. Leggett, Physica A **121**, 587 (1983).
[7] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford (1975).
[8] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* [in Russian], 3rd ed., Nauka, Moscow (1989), p. 432 [English edition: Pergamon Press, Oxford (1991)].
[9] S. A. Kheĭfets, *Electron Synchrotron* [in Russia], Armenian SSR Academy of Sciences Publishing House, Erevan (1963).

Translated by Eugene Yankovsky

Possibility of suppressing quantum light fluctuations when excess photon fluctuations occur inside a cavity

Yu. M. Golubev and I. V. Sokolov

Physics Research Institute, St. Petersburg University, 198904 St. Petersburg, Russia

M. I. Kolobov

Physics Research Institute, St. Petersburg University, 198904 St. Petersburg, Russia; Universität Essen Gesamthochschule

(Submitted October 7, 1996)
Zh. Éksp. Teor. Fiz. **111**, 1579–1600 (May 1997)

Using the optical excitation of a high-Q cavity as an example, it is shown that when light is observed at the output of this cavity, effective suppression of the photocurrent shot noise below the quantum limit is in general independent of the parameters of the stationary state of the field oscillator (in particular, it is independent of the rms photon fluctuations) inside the cavity and can occur not only at any allowed negative value but even at a positive value of the Mandel parameter. It was assumed in solving the problem that the cavity is optically excited by superimposing the radiation of a sub-Poisson laser and a laser with excess photon noise. A formal solution was obtained in terms of the kinetic equation for the density matrix of the actual fields (inside the laser cavities and the empty cavity), which is derived here on the basis of the Heisenberg–Langevin quantum equations, taking into account directed propagation of the field from the laser cavities inside the empty cavity. The resulting kinetic equation can also be used to solve other physical problems, since it is applicable to optical systems that contain, in principle, an arbitrary number of coupled cavities and interference mixers. [S1063-7761(97)00405-8]

1. INTRODUCTION

A typical problem of quantum optics is to study the properties of the electromagnetic radiation coming out of a high-Q cavity. Depending on the specific physical conditions, an intracavity field oscillator is excited either by interacting with an active medium—for example, in lasers and micromasers—or by external electromagnetic action with or without the participation of a linear or nonlinear medium, as in narrow-band amplifiers, bistable systems, etc. If phenomena such as chaos are not involved, in many physical situations, and in particular in those of interest here, the field oscillator after some time enters a stationary state, while continuing to interact with the other subsystems. At first glance, it seems quite natural to compare the observable effects in the radiation field and this stationary state. In any case, this is precisely the tradition. For example, in an ordinary Poisson laser, it is said that a completely random intracavity flux of photons, in the final analysis, generates a completely random Poisson flux of photoelectrons in the photodetector circuit. However, in a sub-Poisson laser, a partially regulated photon flux inside the cavity generates a partially regulated flux of photoelectrons.

Following this logic, we would have to assume that the farther the state of an oscillator departs from the classical state, the more pronounced must be the observed quantum effect. This would undoubtedly be so if we were dealing with ensemble measurements of an isolated quantum oscillator. However, here we are considering quite a different situation, in which the field oscillator experiences conditions of steady-state action from other important subsystems (such as a nonlinear medium inside the cavity, external optical effects, interaction with the cavity, etc.). As can be understood from the most general considerations, the predictive role of the stationary state ceases to be absolute in this case, since the kinematic processes that bring the oscillator into a specific stationary state can now contribute to the averaged field characteristics. We shall give several specific examples that clearly illustrate this.

For qualitative treatments, it is convenient to introduce the so-called Mandel parameter ξ, which characterizes the total photon fluctuations inside a cavity:

$$\overline{\Delta n^2}=\overline{n}(1+\xi). \tag{1}$$

We now consider the limiting case $\xi=-1$, in which the stationary state of an oscillator is close to the Fock state in its properties (this occurs, for example, in a micromaser[1]). The fact that a field oscillator is in a Fock state (it is possible to treat the state with $\xi=-1$ in this way, since it means that there are no photon fluctuations) while undergoing continuous interaction with other subsystems can mean only one thing: any photon fluctuation inside the cavity damps out infinitely quickly. The emission of a photon from the cavity is perceived as a fluctuation inside the cavity, which is quickly compensated by relaxation processes. Therefore, any subsequent photon leaving the cavity can in no way be correlated with the previous one, and this means that the flux of photons leaving the cavity can only be a Poisson flux. Thus, even though the field oscillator is in a limiting quantum stationary state, no quantum effects can be expected to be observed.

It can already be seen from these considerations that an observable quantum effect is associated, at least, not only with a stationary state of the field oscillator but also with the relaxation properties of the system. Of course, the same conclusion can also be reached formally. In fact, it is well known that the photocurrent fluctuation spectrum during perfect photodetection of the radiation, for example, of the simplest laser or the simplest micromaser is represented as[2,3]

$$i_{\omega}^{(2)}=i_{\text{shot}}^{(2)}\left(1+2\xi\,\frac{\kappa\Gamma}{\Gamma^2+\omega^2}\right). \tag{2}$$

If ξ is negative, a nonclassical singularity occurs here at zero frequency: a dip, with a relative depth equal to

$$\delta=1-\frac{i_{\omega=0}^{(2)}}{i_{\text{shot}}^{(2)}}=2|\xi|\,\frac{\kappa}{\Gamma}. \tag{3}$$

This explicitly illustrates that an observable quantum effect is associated not only with the stationary state of a field oscillator (the Mandel parameter ξ) but also with the kinematic properties of the system, namely the rate κ at which photons leave the cavity in the direction of the photocathode and the damping rate Γ of the intracavity photon fluctuations. At the same time, it can also be seen that with widely varying ξ values, i.e., with widely varying stationary states of the field oscillator, one can in principle produce the same quantitative quantum effect (the same δ value) by suitably choosing the kinematic parameters.

There can be different relationships between the Γ and κ values, depending on the physical situation. For example, for the optimal sub-Poisson laser, $\Gamma=\kappa$. This equality can be treated in the same way as the damping of photon fluctuations inside a cavity except for processes associated with the emission of photons from the cavity. However, if not all the radiation in this case is incident on the photodetector, $\Gamma>\kappa$. Understandably, the stationary state of a field oscillator cannot be dependent on this, and thus $\xi=-1/2$ whether all the radiation leaving the cavity is incident on a photodetector or whether it is partially recorded. Also, as can be seen from Eq. (3), the depth of the quantum dip can differ for the same Mandel parameter. This example already compels us to think that the observation and the stationary state of a field oscillator are not associated with each other entirely unambiguously.

Another situation arises in a micromaser. Conditions in which $\xi=-1$ can be ensured there, but it turns out in this case that $\Gamma\gg\kappa$ even when the radiation is recorded completely. As a result, the depth of the observable quantum dip almost equals zero; i.e., the field inside the cavity is a totally quantum field, while it is actually a Poisson field outside it. The quantum properties of the radiation also break down when it leaves the cavity—not because the output mechanism itself is stochastic, but because the processes inside the cavity are too rigidly correlated.

We shall give one more example, which in some sense is intermediate between a sub-Poisson laser and a micromaser. The equality $\xi=-1/2$ is also true in a superradiant laser,[4] and $\Gamma=2\kappa$ for weak excitation of the active medium. Here, as in a micromaser, the depth of the quantum dip (there to zero, and here by a factor of two) is reduced because of the negative effect of intracavity relaxation processes. This effect can be reduced to zero if the active medium is significantly excited, and the depth of the quantum dip then reaches its maximum.

All the examples given here allow a physically transparent treatment: when light leaves a cavity, its quantum properties can break down for different reasons.

However, this is far from complete and is far from the most interesting conclusion that can actually be drawn. The very form of Eq. (3) compels us to think about situations in which the kinematic coefficients obey the inequality $\Gamma<\kappa$. If such cases are encountered in physics, it means that the shot noise can be completely suppressed even when the Mandel parameter is negative but close to zero (the stationary state of the intracavity field oscillator is essentially a Poisson state). Nothing prevents this formally, but of course it is important to have some specific physical examples here; in fact, there already are such examples. Thus, Ref. 5 treats the excitation of the active medium of an ordinary laser by the radiation of a sub-Poisson laser. Using the formulas obtained in Ref. 5, we show in Appendix B that the physical parameters can be chosen so that a situation is possible in which the Mandel parameters, although negative, are close to zero, and, at the same time, the observable quantum effect is complete, $\delta=1$. A completely analogous situation also occurs for a micromaser when its active medium is excited by the radiation of a sub-Poisson laser.[3]

All this compels us to conclude that, in general, there is no basis for associating the quantum optical effect observed earlier with the stationary state of an intracavity field oscillator. The example of optically exciting an empty cavity by superimposing the radiation from two lasers (a sub-Poisson laser and a laser with excess noise), which will be considered here, clearly illustrates this idea. We shall show that the shot noise in the photocurrent can be completely suppressed with the most varied (and even positive) values of the Mandel parameter. In particular, for example, large (super-Poisson) photon fluctuations inside a cavity can correspond to the smallest possible (sub-Poisson) fluctuations of the photoelectrons.

Section 2 of this paper describes the formulation of the problem, using the kinetic equation for the density matrix of the field in cavities associated with directed optical signals. The derivation of this equation is given in Appendix A. Sections 3 and 4 give calculations of the steady-state dispersion of the distribution of the number of photons and of the Mandel parameter of the internal field, as well as of a quantity that is observable during the photodetection of the output radiation—the spectrum of the photocurrent fluctuations. The simplest optical system, in which a passive cavity is illuminated by one laser source, is studied first (Sec. 3). A treatment is then given (in Sec. 4) of the illumination of a passive cavity by the light flux formed by optically mixing the radiation of an ordinary laser and a sub-Poisson laser. Section 5 of the paper is devoted to a discussion of the physical results. Appendix B shows the results of Ref. 5, rewritten from the viewpoint that interests us here.

2. COMBINED KINETIC EQUATION FOR OPTICAL CAVITIES EXCHANGING NONCLASSICAL LIGHT SIGNALS

The question of how a quantum kinetic equation for a light field could be used to describe the directed transfer of an optical signal from one cavity to another was first discussed by Kolobov and Sokolov[6] and later by Carmichael[7] and Gardiner.[8] These papers considered the situation in which there are two cavities. One of them (for example, a laser) forms the source, and the other forms a detector of the light emitted by the source, which in general is nonclassical (sub-Poisson, squeezed, etc.).

Here we shall discuss a somewhat more complex physical situation, in which a passive optical cavity is excited by radiation from two laser sources. Therefore, Appendix A generalizes the theory of Ref. 6 to the case of several resonant (or almost resonant) field oscillators concentrated in different cavities and exchanged in an arbitrary geometry by directed optical signals. The starting point of this treatment differs from that of Ref. 6. The theory of Ref. 6 was constructed in some sense from first principles. The quantization was carried out over the modes of the continuous spectrum of the entire space, taking into account the boundary conditions at the mirrors of the cavities, without isolating the earlier local oscillators of the modes in the cavities. The commutation relations of the light-field operators inside and outside the optical cavities (see Ref. 9 for the relations with the participation of the external fields) were obtained from this. The kinetic equation was also derived on the basis of these commutation relations.

In Appendix A, we use the simpler and more instructive method of the Heisenberg–Langevin quantum equations, which recently have been widely used for quantum optics problems. These equations are similar to the equations of classical electrodynamics. They can be used in quantum theory to give a transparent description of such wave phenomena as interference, diffraction, partial dispersion of a signal as it propagates, etc. The theory of the input and output of radiation from a cavity has been developed by Refs. 10 and 11. We shall show that the Heisenberg–Langevin equations can easily be combined with the approach used in Ref. 6 as the basis of the kinetic equation for the case of directed radiation transfer. This makes it easy to obtain the kinetic equation for several field oscillators concentrated in different cavities and exchanged by directed optical signals. When this is done, the intermediate optical elements used for beam splitting, signal mixing, etc. can be taken into account in a natural way.

When we deal with a system of cavities between which electromagnetic energy is exchanged along definite directions, this can be formally represented as a set of quantum electromagnetic field oscillators each of which is localized in its own cavity. To describe this situation, it is possible to introduce into the discussion a multi-oscillator density matrix and to attempt to construct a closed equation for it. Here the main problem of the theory is to make it adequately reflect the fact that the motion of the light wave is directed: this is not the simple standard interaction of any two localized oscillators, leading to the interchange of electromagnetic energy, but represents energy transfer from one oscillator to the other, and not the reverse.

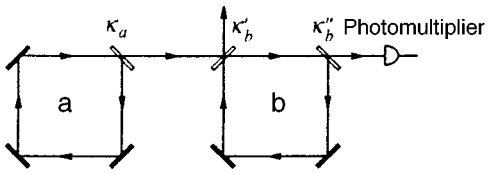

FIG. 1.

It will be shown that for a multi-oscillator density matrix, in the case of many optically coupled cavities, the basic kinetic equation can be written as

$$\dot{\rho}=\left(\sum_{m=1}^{M}\hat{L}_m+\sum_{m\neq m_0}^{M}\hat{L}_{m_0 m}\right)\rho. \tag{4}$$

Operator $\hat{L}_m$ describes the evolution of oscillator m (the field oscillator inside cavity m) in the absence of all other oscillators.

Operator $\hat{L}_{m_0 m}$ describes the action of oscillator m_0 on oscillator m:

$$\hat{L}_{m_0 m}\,\rho=g_{m m_0}[a_m^+,a_{m_0}\,\rho]+\text{H.c.} \tag{5}$$

Here a_i and a_i^+ $(i=m,m_0)$ are the photon operators of the corresponding localized oscillator, and $g_{m m_0}$ is a factor that describes the penetration of the signal from cavity m_0 into cavity m. It is given in the Appendix in the most general form, taking into account the possible presence of any optical elements in the spaces between the cavities. We shall now indicate what it equals for the two physical situations discussed below. For example, we are interested in a system consisting of just two cavities (Fig. 1): cavity a, in which sub-Poisson lasing occurs and whose radiation enters the other, passive cavity b. Then only one of all the constants $g_{m m_0}$ is nonzero:

$$g_{ba}=\sqrt{\kappa_b'\kappa_a}. \tag{6}$$

Here $\kappa_{a,b}^{-1}$ are the lifetimes of a photon inside cavities a and b. The value of κ_b results from transmission through the input and the output mirrors, which in the theory for cavity b we shall separate as $\kappa_b=\kappa_b'+\kappa_b''$.

A more complex system, which will also interest us here, consists of three cavities (Fig. 2). Two of them (1 and

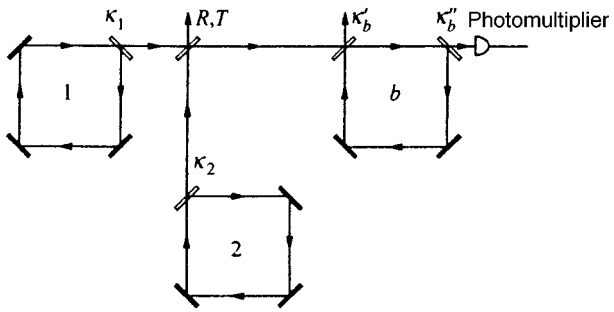

FIG. 2.

2) form independent laser sources, while the third (b), as in the preceding case, is passive and is excited by the electromagnetic fields from the sources. There are now two nonzero constants:

$$g_{b1}=T\sqrt{\kappa_b'\kappa_1},\quad g_{b2}=R\sqrt{\kappa_b'\kappa_2}. \tag{7}$$

Here $\dot{T}$ and R ($T^2+R^2=1$) are the transmission and reflection coefficients of the additional mirror at which the radiation from the two laser sources mixes before it enters cavity b (Fig. 2).

3. EXCITATION OF A PASSIVE CAVITY BY SUB-POISSON LASER RADIATION

3.1. The basic kinetic equation

We initially assume that the physical system consists of only two optical cavities in series (Fig. 1). Sub-Poisson lasing occurs in cavity a. The radiation of the sub-Poisson laser enters cavity b, is accumulated there, and, in the final analysis, being reflected from it or passing through it, is incident on a photodetector, by means of which it is analyzed. We are only interested in the part of the light that passes completely through the cavity. For simplicity, we assume that all the actual frequencies in this system coincide.

Accordingly, the following equation can be written for the two-oscillator field density matrix ρ that describes both the intracavity lasing field of the sub-Poisson laser and the field inside the second empty cavity:

$$\dot{\rho}=r_a\left(\hat{L}_a-\frac{1}{2}\hat{L}_a^2\right)\rho-\frac{\kappa_a}{2}\hat{R}_a\,\rho-\frac{\kappa_b}{2}\hat{R}_b\,\rho+g_{ba}\hat{L}_{a\to b}\,\rho. \tag{8}$$

Operator $\hat{L}_a$ determines the development of the sub-Poisson lasing inside cavity a produced by the active medium. It is derived in Ref. 2 and can be written as

$$\hat{L}_a=\left[2\underset{\to}{a^+}\underset{\leftarrow}{a}-\underset{\to}{a}\underset{\leftarrow}{a^+}-\underset{\leftarrow}{a}\underset{\to}{a^+}-\frac{1}{2}\beta_a(\underset{\to}{a}\underset{\leftarrow}{a^+}-\underset{\leftarrow}{a}\underset{\to}{a^+})^2\right]\times\left[\underset{\to}{a}\underset{\leftarrow}{a^+}+\underset{\leftarrow}{a}\underset{\to}{a^+}+\frac{1}{2}\beta_a(\underset{\to}{a}\underset{\leftarrow}{a^+}-\underset{\leftarrow}{a}\underset{\to}{a^+})^2\right]^{-1}. \tag{9}$$

The damping of the quantum oscillator is described by the well-known operator of the form

$$\hat{R}_a=\underset{\to}{a^+}\underset{\leftarrow}{a}+\underset{\leftarrow}{a^+}\underset{\to}{a}-2\underset{\to\leftarrow}{a\,a^+}. \tag{10}$$

Here r_a is the mean excitation rate of the upper laser level, a and a^+ are the photon operators for the laser mode $[a,a^+]=1$, and

$$\beta_a^{-1}=\frac{\gamma_b\gamma_{ab}}{2g^2}$$

is the number of photons that saturates the laser transition in the active medium of the sub-Poisson laser (γ_a and γ_b are the longitudinal relaxation constants relating to the upper and lower laser levels, $\gamma_a=0$, and γ_{ab} is the transverse relaxation constant). The arrows under the operators determine the direction of action with respect to the operator expressions, including the density matrix standing at the right. The field-damping operator in cavity b, $\hat{R}_b$, is obtained from operator $\hat{R}_a$ by replacing photon operators a with photon operators b, and κ_a^{-1} and κ_b^{-1} are the photon lifetimes in cavities a and b because of their finite Q values.

An explicit expression for operator $\hat{L}_{a\to b}$, which determines the unidirectional action of oscillator a on oscillator b, recorded in the previous section, can be written as[9]

$$\hat{L}_{a\to b}=\underset{\to\to}{a\,b^+}+\underset{\leftarrow\leftarrow}{a^+\,b}-\underset{\to\leftarrow}{a\,b^+}-\underset{\to\leftarrow}{b\,a^+}. \tag{11}$$

The concomitant force is determined by the constant g_{ba} of Eq. (6).

The total spectral width of cavity b depends on the transmission of the input and output mirrors:

$$\kappa_b=\kappa_b'+\kappa_b''. \tag{12}$$

We transform Eq. (8) to the diagonal Glauber representation, using the integral relation

$$\rho(t)=\int d^2\alpha\; d^2\beta\; P(\alpha,\beta,t)|\alpha\rangle|\beta\rangle\langle\beta|\langle\alpha|, \tag{13}$$

where

$$d^2\alpha=d(\mathrm{Re}\;\alpha)d(\mathrm{Im}\;\alpha),\quad d^2\beta=d(\mathrm{Re}\;\beta)d(\mathrm{Im}\;\beta),$$

$$a|\alpha\rangle=\alpha|\alpha\rangle,\quad b|\beta\rangle=\beta|\beta\rangle.$$

We use the approximation of small photon fluctuations:

$$\alpha=\sqrt{n_a+\varepsilon_a}\,\exp(i\varphi_a),\quad \varepsilon_a\ll n_a,$$

$$\beta=\sqrt{n_b+\varepsilon_b}\,\exp(i\varphi_b),\quad \varepsilon_b\ll n_b. \tag{14}$$

Here n_a and n_b are the steady-state solutions of the semiclassical laser problem:

$$\frac{d}{dt}n_b=-\kappa_b n_b+2g_{ab}\sqrt{n_a n_b}\,\cos(\overline{\varphi_a}-\overline{\varphi_b})=0, \tag{15}$$

$$\frac{d}{dt}\varphi_b=g_{ba}\sqrt{\frac{n_a}{n_b}}\sin(\overline{\varphi_a}-\overline{\varphi_b})=0, \tag{16}$$

$$\frac{d}{dt}n_a=r_a-\kappa_a n_a=0,\quad \frac{d}{dt}\varphi_a=0. \tag{17}$$

These equations can be directly obtained from Eq. (8) by multiplying it by operator a and taking the trace, and then by operator b and again taking the trace. Neglecting the fluctuations of the field variables in the expressions thus obtained, we get a closed system of equations for the complex amplitudes $\overline{\alpha}$ and $\overline{\beta}$. Transforming to amplitudes and phases, we get the required expression.

It is easy to see that the following stable steady-state solutions are valid:

$$\overline{\varphi_a}=\overline{\varphi_b}, \tag{18}$$

$$n_a=r_a/\kappa_a, \tag{19}$$

$$\frac{n_a}{n_b}=\frac{\kappa_b^2}{4g_{ab}^2}=\frac{\kappa_b^2}{4\kappa_b'\kappa_a}. \tag{20}$$

For the photon density matrix,

$$R(\varepsilon_a,\varepsilon_b,t)=\int d\varphi_a\, d\varphi_b\, P(\varepsilon_a,\varepsilon_b,\varphi_a,\varphi_b,t) \tag{21}$$

the following equation can be obtained in these approximations:

$$\frac{\partial R}{\partial t}=\frac{\kappa_b}{2}\frac{\partial}{\partial\varepsilon_b}\left(\varepsilon_b-\frac{n_b}{n_a}\varepsilon_a\right)R+\kappa_a\frac{\partial}{\partial\varepsilon_a}(\varepsilon_a R) -\frac{1}{2}\kappa_a n_a\frac{\partial^2 R}{\partial\varepsilon_a^2}+\{\ldots\}. \tag{22}$$

This is the basic kinetic equation, which makes it possible in what follows to obtain all the necessary information on the field in the second cavity. The notation $\{\ldots\}$ denotes the complete set of ε derivatives of third and higher orders that appear on the diagonal of the Glauber representation for quantum fields. These terms will make no contribution to the observable values that we shall consider below.

3.2. Mandel parameter

Starting with the basic kinetic equation, Eq. (22), the standard procedure can be used to multiply it by the appropriate quantity, after which it is integrated over the variables ε and φ to obtain the system of equations

$$\frac{d}{dt}\overline{\varepsilon_b^2}=-\kappa_b\overline{\varepsilon_b^2}+\kappa_b\frac{n_b}{n_a}\overline{\varepsilon_a\varepsilon_b}=0, \tag{23}$$

$$\frac{d}{dt}\overline{\varepsilon_a\varepsilon_b}=-\left(\kappa_a+\frac{\kappa_b}{2}\right)\overline{\varepsilon_a\varepsilon_b}+\frac{\kappa_b}{2}\frac{n_b}{n_a}\overline{\varepsilon_a^2}=0, \tag{24}$$

$$\frac{d}{dt}\overline{\varepsilon_a^2}=-2\kappa_a\overline{\varepsilon_a^2}-\kappa_a n_a=0. \tag{25}$$

The steady-state solution has the form

$$\overline{\varepsilon_a^2}=-\frac{1}{2}n_a=\xi_a n_a, \tag{26}$$

$$\overline{\varepsilon_a\varepsilon_b}=-\frac{1}{2}\frac{\kappa_b}{\kappa_b+2\kappa_a}n_b, \tag{27}$$

$$\overline{\varepsilon_b^2}=-\frac{1}{2}\frac{n_b}{n_a}\frac{\kappa_b}{\kappa_b+2\kappa_a}n_b=\xi_b n_b. \tag{28}$$

Using Eq. (20), we find that the statistical Mandel parameter in cavity b can be written as

$$\xi_b=-\frac{2\kappa_b'\kappa_a}{\kappa_b(\kappa_b+2\kappa_a)}. \tag{29}$$

Varying the ratio between the cavity widths κ_a and κ_b, we can get very different (negative) values of the Mandel parameter. In particular, in the two limiting cases,

$$\xi_b=-\frac{\kappa_b'}{\kappa_b}\to-\frac{1}{2} \tag{30}$$

when $\kappa_a\gg\kappa_b$ (which approaches $-1/2$ in the case of a symmetrical b cavity, with $\kappa_b'=\kappa_b''$) and

$$\xi_b=-\frac{2\kappa_b'\kappa_a}{\kappa_b^2},\quad |\xi|\ll 1, \tag{31}$$

when $\kappa_a\ll\kappa_b$.

3.3. Photocurrent spectrum

To understand the role played by the stationary state of the intracavity field oscillator in forming the quantum singularities in observation, it is sufficient to analyze any one of the actually observed singularities. For example, this can be the photocurrent noise spectrum, which, when the radiation from cavity b is observed with perfect photodetection, can be represented by[2]

$$i_\omega^{(2)}=i_{\text{shot}}^{(2)}\left[1+\frac{2\kappa_b''}{n_b}\,\text{Re}\int_0^\infty dt\; e^{i\omega t}\overline{\varepsilon_b(0)\varepsilon_b(t)}\right]. \tag{32}$$

The basic kinetic equation, Eq. (22), and the standard procedure can be used to obtain the system of equations

$$\frac{d}{dt}\overline{\varepsilon_b(0)\varepsilon_b(t)}=-\frac{\kappa_b}{2}\overline{\varepsilon_b(0)\varepsilon_b(t)}+\frac{\kappa_b}{2}\frac{n_b}{n_a}\overline{\varepsilon_b(0)\varepsilon_a(t)}, \tag{33}$$

$$\frac{d}{dt}\overline{\varepsilon_b(0)\varepsilon_a(t)}=-\kappa_a\overline{\varepsilon_b(0)\varepsilon_a(t)}. \tag{34}$$

From this we can obtain

$$\overline{\varepsilon_b(0)\varepsilon_b(t)}=\overline{\varepsilon_b^2}\exp\left(-\frac{\kappa_b}{2}t\right)+\overline{\varepsilon_a\varepsilon_b}\frac{\kappa_b}{\kappa_b-2\kappa_a}\frac{n_b}{n_a} \times\left[\exp(-\kappa_a t)-\exp\left(-\frac{\kappa_b}{2}t\right)\right]. \tag{35}$$

After substituting this expression into Eq. (32) and using Eqs. (27) and (28), we find

$$i_\omega^{(2)}=i_{\text{shot}}^{(2)}\left[1-k^2\frac{\kappa_a^2}{\omega^2+\kappa_a^2}\frac{\kappa_b^2}{4\omega^2+\kappa_b^2}\right],\quad k^2=\frac{4\kappa_b'\kappa_b''}{\kappa_b^2}. \tag{36}$$

When oscillator b is symmetric ($k^2=1$), it can be seen that the relative depth of the quantum dip at zero frequency is $\delta=1$ regardless of any other physical conditions, including the ratio of the cavity widths κ_a and κ_b. This is an important fact for our subsequent calculations.

4. EXCITING A PASSIVE CAVITY WITH THE LIGHT FROM TWO LASER SOURCES

4.1. Kinetic equation

We now consider the physical system schematically shown in Fig. 2. We discuss the case in which empty cavity b is excited by the light of two laser sources. We assume that regular (noiseless) pumping of the upper laser level occurs in the active medium of laser 1, which provides some Mandel parameter ξ_1 that under optimum conditions can become equal to $-1/2$. In the active medium of laser 2, ordinary, completely random pumping of the active atoms to the upper

laser level occurs. In this case, Mandel parameter ξ_2 is large in weak lasing fields and goes to zero during saturation. In the final formulas, we shall be interested in the case where $\xi_1 = -1/2$ and $\xi_2 \sim 1$.

The equation for the three-oscillator density matrix ρ that describes the intracavity fields of the two laser sources and the passive cavity can be written as

$$\dot{\rho} = r_1\left(\hat{L}_1 - \frac{1}{2}\,L_1^2\right)\rho - \frac{\kappa_1}{2}\,\hat{R}_1\rho + g_{b1}\hat{L}_{1\to b}\rho + r_2\hat{L}_2\rho - \frac{\kappa_2}{2}\,\hat{R}_2\rho - g_{b2}\hat{L}_{2\to b}\rho - \frac{\kappa_b}{2}\,\hat{R}_b\rho. \tag{37}$$

Operators $\hat{L}_{1,2}$ and $\hat{R}_{1,2}$ determine the evolution of the electromagnetic field inside the laser cavities:

$$\hat{L}_1 = \frac{1}{2}\,\beta_{1a}\left[2\underset{\to}{a_1}^{+}\underset{\leftarrow}{a_1} - \underset{\to}{a_1}\underset{\leftarrow}{a_1}^{+} - \underset{\leftarrow}{a_1}\underset{\to}{a_1}^{+} - \frac{1}{2}\,\beta_{1b}\left(\underset{\to}{a_1}\underset{\leftarrow}{a_1}^{+} - \underset{\leftarrow}{a_1}\underset{\to}{a_1}^{+}\right)^2\right]\frac{1}{\hat{Q}_1}, \tag{38}$$

$$\hat{Q}_1 = 1 + \frac{1}{2}\,\beta_{1+}\left(\underset{\to}{a_1}\underset{\leftarrow}{a_1}^{+} + \underset{\leftarrow}{a_1}\underset{\to}{a_1}^{+}\right) + \frac{1}{4}\,\beta_{1a}\beta_{1b}\left(\underset{\to}{a_1}\underset{\leftarrow}{a_1}^{+} - \underset{\leftarrow}{a_1}\underset{\to}{a_1}^{+}\right)^2,$$

$$\hat{L}_2 = \frac{1}{2}\,\beta_{2a}\left[2\underset{\to}{a_2}^{+}\underset{\leftarrow}{a_2} - \underset{\to}{a_2}\underset{\leftarrow}{a_2}^{+} - \underset{\leftarrow}{a_2}\underset{\to}{a_2}^{+} - \frac{1}{2}\,\beta_{2b}\left(\underset{\to}{a_2}\underset{\leftarrow}{a_2}^{+} - \underset{\leftarrow}{a_2}\underset{\to}{a_2}^{+}\right)^2\right]\frac{1}{\hat{Q}_2}, \tag{39}$$

$$\hat{Q}_2 = 1 + \frac{1}{2}\,\beta_{2+}\left(\underset{\to}{a_2}\underset{\leftarrow}{a_2}^{+} + \underset{\leftarrow}{a_2}\underset{\to}{a_2}^{+}\right) + \frac{1}{4}\,\beta_{2a}\beta_{2b}\left(\underset{\to}{a_2}\underset{\leftarrow}{a_2}^{+} - \underset{\leftarrow}{a_2}\underset{\to}{a_2}^{+}\right)^2,$$

$$\hat{R}_1 = \underset{\to}{a_1^{+}}\underset{\leftarrow}{a_1} + \underset{\leftarrow}{a_1^{+}}\underset{\to}{a_1} - 2\underset{\to\leftarrow}{a_1}\underset{\to\leftarrow}{a_1}^{+}, \tag{40}$$

$$\hat{R}_2 = \underset{\to}{a_2^{+}}\underset{\leftarrow}{a_2} + \underset{\leftarrow}{a_2^{+}}\underset{\to}{a_2} - 2\underset{\to\leftarrow}{a_2}\underset{\to\leftarrow}{a_2}^{+}. \tag{41}$$

The nonlinear parameters for the laser sources have the form

$$\beta_{ia} = \frac{2|g_{iab}|^2}{\gamma_{ia}\gamma_{iab}}, \quad \beta_{ib} = \frac{2|g_{iab}|^2}{\gamma_{ib}\gamma_{iab}},$$

$$\beta_{i\pm} = \beta_{ia} \pm \beta_{ib}, \quad i = 1,2, \tag{42}$$

where γ_{ia} and γ_{ib} are the widths of the upper and lower laser levels, γ_{iab} is the homogeneous width of the gain curve of the working transition, and g_{iab} is the interaction constant of an atom with a plane laser wave in the dipole approximation.

Here r_i is the mean excitation rate of the upper laser level, a_i and a_i^+ are the photon operators for the laser mode, and κ_i^{-1} are the photon lifetimes in the cavities resulting from their finite Q values.

Recall that the presence of a term with the operator $\hat{L}_1^2$ in the basic kinetic equation indicates that the pumping of the active atoms in the first laser source is regular (with no fluctuations). When $\gamma_{1a} = 0$, we have an ideal sub-Poisson laser, and operator $\hat{L}_1$ will coincide in form with operator $\hat{L}_a$ from the first case, in which it was precisely the radiation of an ideal sub-Poisson laser that excited the passive cavity.

An explicit expression for operator $\hat{L}_{1\to b}$, which determines the unidirectional action of the radiation from cavity 1 on cavity b described in the preceding section, can be written as

$$\hat{L}_{1\to b} = \underset{\to\to}{a_1 b^{+}} + \underset{\leftarrow\ \leftarrow}{a_1^{+} b} - \underset{\to\leftarrow}{a_1 b^{+}} - \underset{\to\leftarrow}{b\, a_1^{+}}. \tag{43}$$

Likewise,

$$\hat{L}_{2\to b} = \underset{\to\to}{a_2 b^{+}} + \underset{\leftarrow\ \leftarrow}{a_2^{+} b} - \underset{\to\leftarrow}{a_2 b^{+}} - \underset{\to\leftarrow}{b\, a_2^{+}}. \tag{44}$$

The resulting force is determined by the constants g_{bi} of Eqs. (7).

The semiclassical equations can be obtained in the same way as before from the basic kinetic equation, Eq. (37), neglecting all field fluctuations:

$$\frac{d}{dt}\,n_b = -\kappa_b n_b + 2g_{1b}\sqrt{n_1 n_b}\,\cos(\overline{\varphi_1} - \overline{\varphi_b}) + 2g_{2b}\sqrt{n_2 n_b}\,\cos(\overline{\varphi_2} - \overline{\varphi_b}) = 0, \tag{45}$$

$$\frac{d}{dt}\,\overline{\varphi_b} = g_{1b}\sqrt{\frac{n_1}{n_b}}\,\sin(\overline{\varphi_1} - \overline{\varphi_b}) + g_{2b}\sqrt{\frac{n_2}{n_b}}\,\sin(\overline{\varphi_2} - \overline{\varphi_b}) = 0, \tag{46}$$

$$\frac{d}{dt}\,n_1 = \left(-\kappa_1 + \frac{r_1\beta_{1a}}{1+\beta_{1+}n_1}\right)n_1 = 0, \quad \frac{d}{dt}\,\overline{\varphi_1} = 0, \tag{47}$$

$$\frac{d}{dt}\,n_2 = \left(-\kappa_2 + \frac{r_2\beta_{2a}}{1+\beta_{2+}n_2}\right)n_2 = 0, \quad \frac{d}{dt}\,\overline{\varphi_2} = 0, \tag{48}$$

$$\alpha_1 = |\alpha_1|\exp(i\varphi_1), \quad \alpha_2 = |\alpha_2|\exp(i\varphi_2). \tag{49}$$

4.2. Mandel parameter and photocurrent spectrum

Here, as above, we transform to the diagonal Glauber representation, introducing into the formal scheme, in place of photon operators a_i and a_i^+ the corresponding c-number field amplitudes α_i and α_i^* and the derivatives with respect to them. Neglecting small photon fluctuations,

$$|\alpha_i|^2 = n_i + \varepsilon_i, \quad \varepsilon_i \ll n_i \tag{50}$$

and using the semiclassical equations for the photon density matrix,

$$R(\varepsilon_1,\varepsilon_2,\varepsilon_b,t) = \int d\varphi_1\, d\varphi_2\, d\varphi_b\, P(\varepsilon_1,\varepsilon_2,\varepsilon_b,\varphi_1,\varphi_2,\varphi_b,t) \tag{51}$$

we get the following basic kinetic equation, provided that there is a phase shift between the laser waves that is a multiple of π (otherwise, we would have to write an equation that simultaneously describes both the amplitude fluctuations and the phase fluctuations):

$$\frac{\partial R}{\partial t}=\frac{\kappa_b}{2}\frac{\partial}{\partial \varepsilon_b}(\varepsilon_b-\delta_1\varepsilon_1-\delta_2\varepsilon_2)R+\Gamma_1\frac{\partial}{\partial \varepsilon_1}(\varepsilon_1 R)+\Gamma_1\xi_1 n_1\frac{\partial^2 R}{\partial \varepsilon_1^2}+\Gamma_2\frac{\partial}{\partial \varepsilon_2}(\varepsilon_2 R)+\Gamma_2\xi_2 n_2\frac{\partial^2 R}{\partial \varepsilon_2^2}+\{\ldots\}. \tag{52}$$

Here we have introduced the notation

$$\delta_1=\frac{2g_{1b}}{\kappa_b}\sqrt{\frac{n_b}{n_1}}, \tag{53}$$

$$\delta_2=\frac{2g_{2b}}{\kappa_b}\sqrt{\frac{n_b}{n_2}}, \tag{54}$$

$$\Gamma_1=\kappa_1\frac{I_1}{1+I_1},\quad \xi_1=\frac{1}{I_1}-\frac{1}{2}\frac{\gamma_{1b}}{\gamma_{1a}+\gamma_{1b}}, \tag{55}$$

$$\Gamma_2=\kappa_2\frac{I_2}{1+I_2},\quad \xi_2=\frac{1}{I_2}, \tag{56}$$

$$I_i=\beta_{i+}n_i. \tag{57}$$

We get from the basic kinetic equation that

$$\frac{d}{dt}\overline{\varepsilon_b^2}=-\kappa_b\overline{\varepsilon_b^2}+\delta_1\overline{\varepsilon_1\varepsilon_b}+\delta_2\overline{\varepsilon_2\varepsilon_b}=0, \tag{58}$$

$$\frac{d}{dt}\overline{\varepsilon_1\varepsilon_b}=-\left(\Gamma_1+\frac{\kappa_b}{2}\right)\overline{\varepsilon_1\varepsilon_b}+\frac{\kappa_b}{2}\delta_1\overline{\varepsilon_1^2}=0, \tag{59}$$

$$\frac{d}{dt}\overline{\varepsilon_2\varepsilon_b}=-\left(\Gamma_2+\frac{\kappa_b}{2}\right)\overline{\varepsilon_2\varepsilon_b}+\frac{\kappa_b}{2}\delta_2\overline{\varepsilon_2^2}=0, \tag{60}$$

$$\frac{d}{dt}\overline{\varepsilon_1^2}=-2\Gamma_1\overline{\varepsilon_1^2}+2\Gamma_1\xi_1 n_1=0, \tag{61}$$

$$\frac{d}{dt}\overline{\varepsilon_2^2}=-2\Gamma_2\overline{\varepsilon_2^2}+2\Gamma_2\xi_2 n_2=0, \tag{62}$$

$$\frac{d}{dt}\overline{\varepsilon_b(0)\varepsilon_b(t)}=-\frac{\kappa_b}{2}\overline{\varepsilon_b(0)\varepsilon_b(t)}+\frac{\kappa_b}{2}\delta_1\overline{\varepsilon_b(0)\varepsilon_1(t)}+\frac{\kappa_b}{2}\delta_2\overline{\varepsilon_b(0)\varepsilon_2(t)}, \tag{63}$$

$$\frac{d}{dt}\overline{\varepsilon_b(0)\varepsilon_1(t)}=-\Gamma_1\overline{\varepsilon_b(0)\varepsilon_1(t)}, \tag{64}$$

$$\frac{d}{dt}\overline{\varepsilon_b(0)\varepsilon_2(t)}=-\Gamma_2\overline{\varepsilon_b(0)\varepsilon_2(t)}. \tag{65}$$

Solving this system, we find

$$\xi_b=\delta_1^2\frac{\kappa_b}{\kappa_b+2\Gamma_1}\frac{n_1}{n_b}\xi_1+\delta_2^2\frac{\kappa_b}{\kappa_b+2\Gamma_2}\frac{n_2}{n_b}\xi_2, \tag{66}$$

$$\overline{\varepsilon_b(0)\varepsilon_b(t)}=\delta_1^2\frac{\kappa_b^2}{\kappa_b^2-2\Gamma_1^2}\left[\exp(-\Gamma_1 t)-\frac{2\Gamma_1}{\kappa_b}\times\exp\left(-\frac{\kappa_b}{2}t\right)\right]\overline{\varepsilon_1^2}+\delta_2^2\frac{\kappa_b^2}{\kappa_b^2-2\Gamma_2^2}\left[\exp(-\Gamma_2 t)-\frac{2\Gamma_2}{\kappa_b}\exp\left(-\frac{\kappa_b}{2}t\right)\right]\overline{\varepsilon_2^2}. \tag{67}$$

Carrying out all the necessary calculations, we get the following expressions for the photocurrent spectrum and for the Mandel parameter:

$$i_\omega^{(2)}=i_{\text{shot}}^{(2)}\left[1-k^2T^2\frac{\kappa_1^2}{\omega^2+\kappa_1^2}\frac{\kappa_b^2}{4\omega^2+\kappa_b^2}+2k^2\frac{1+I_2}{I_2^2}R^2\frac{\Gamma_2^2}{\omega^2+\Gamma_2^2}\frac{\kappa_b^2}{4\omega^2+\kappa_b^2}\right], \tag{68}$$

$$\xi_b=-\frac{2\kappa_b'}{\kappa_b}\frac{\kappa_1 T^2}{\kappa_b+2\kappa_1}+\frac{4\kappa_b'}{\kappa_b}\frac{\kappa_2 R^2}{\kappa_b(1+I_2)+2\kappa_2 I_2}\frac{1+I_2}{I_2}. \tag{69}$$

In the last two formulas, we have assumed that $\gamma_{1a}\ll\gamma_{1b}$, as a consequence of which the first source is converted into an ideal sub-Poisson laser: $I_1\to\infty$ and $\xi_1\to-1/2$.

5. DISCUSSION

We first consider a physical situation with two cavities: a cavity with a laser source and an empty cavity (Fig. 1). As is well known, if the frequency of the external field (in this case the laser field) coincides with the natural frequency of the cavity and the cavity is symmetrical ($\kappa_b'=\kappa_b''$ and $k^2=1$), as a consequence of interference phenomena, the radiation incident on the input mirror from outside is not reflected from this mirror but passes completely through the cavity via the output mirror. Since no uncontrolled losses occur in this case, and the empty cavity, as it were, does nothing, it can be assumed that the radiation from the sub-Poisson laser will not be distorted. We should see the photocurrent shot noise, which is completely suppressed at the lower frequencies, both when the radiation is observed immediately after the laser cavity and when it is observed after the additional empty cavity. This fact is probably independent of any other circumstances and, in particular, of the ratio between κ_a and κ_b. It is easy to convince oneself of this if we assume that $k^2=1$ in the formula for the photocurrent spectrum, Eq. (36). The depth of the dip in the shot noise at zero frequency is then a maximum and reaches the zero mark, regardless of κ_a and κ_b.

At the same time, it is clear even from qualitative considerations that the stationary state of a field oscillator (and, in particular, the Mandel parameter) must depend on what kinematic constants are actually chosen. In fact, let $\kappa_a\gg\kappa_b$, for example. The spectral profile of the empty cavity is then filled with Fourier components correlated in the same way as in the cavity of the sub-Poisson laser. The int-

racavity field situation is therefore the same as it is inside the sub-Poisson laser, and we can expect that $\xi_b=-1/2$ in this limiting case.

However, in the other limiting case of $\kappa_a \ll \kappa_b$, the ''correctly organized'' Fourier components occupy only a very narrow central part of the spectral profile of the empty cavity, and all the other components (which are far larger) are occupied by vacuum fluctuations. It can thus be expected that the Mandel parameter is close to zero. We now see that for a symmetrical passive cavity, shot noise is completely suppressed at the lower frequencies for any ratios between the cavity widths. At the same time, the Mandel parameter can vary from $-1/2$ to zero. It is easy to show that Eqs. (29) and (36) confirm these qualitative conclusions:

$$\xi_b=-\frac{2\kappa_b'}{\kappa_b}\frac{\kappa_a}{\kappa_b}, \tag{70}$$

i.e., $|\xi_b| \ll 1$.

We can now go even further by assuming, for example, that the ''free'' Fourier components in the limiting case $\kappa_a \ll \kappa_b$ are excited not by vacuum fluctuations but by fluctuations from an additional laser source. This can be accomplished in the system shown in Fig. 2 by mixing at an intermediate mirror the light from a sub-Poisson laser and from a laser with excess noise. Understandably, if the addition from the super-Poisson laser is negligible (which can be the case in which $T\sim 1$ and $R\ll 1$), it must not appreciably degrade the observable quantum effect. At the same time, ξ_b can be expected to become completely positive, since it is precisely the radiation of the super-Poisson laser that forms the main part of the spectral profile of the empty cavity. We now proceed to specific formulas and write them for a symmetric empty cavity in the limiting case of $\kappa_1 \ll \kappa_b \ll \kappa_2$, assuming that the dimensionless lasing power I_2 of the super-Poisson laser is intermediate and has the order of unity:

$$\xi_b=-\frac{\kappa_1}{\kappa_b}T^2+\frac{1+I_2}{I_2^2}R^2, \tag{71}$$

$$i_\omega^{(2)}=i_{\text{shot}}^{(2)}\left[1-T^2\frac{\kappa_1^2}{\omega^2+\kappa_1^2}+2\frac{1+I_2}{I_2^2}R^2\frac{\kappa_b^2}{4\omega^2+\kappa_b^2}\right]. \tag{72}$$

Writing the latter relationship for zero frequency, we get an expression for the depth of the quantum dip in the form

$$\delta=1-R^2\left(1+2\frac{1+I_2}{I_2^2}\right). \tag{73}$$

If $R^2 \ll 1$, then $\delta \approx 1$. At the same time, if this value is still not too small and if

$$\frac{\kappa_1}{\kappa_b} \ll R^2 \ll 1, \tag{74}$$

then the Mandel parameter ξ_b, as we assumed, becomes positive:

$$\xi_b=\frac{1+I_2}{I_2^2}R^2. \tag{75}$$

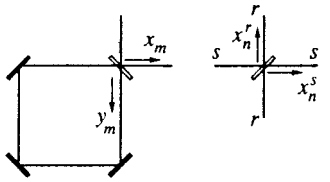

FIG. 3.

APPENDIX A: CONSTRUCTING THE COMBINED KINETIC EQUATION FOR THE DENSITY MATRIX OF THE ELECTROMAGNETIC FIELD IN THE CASE OF OPTICALLY COUPLED CAVITIES

Let there be local field oscillators in the cavities, and let the annihilation operators (positive-frequency slow operator amplitudes) be $a_m(t)$, $m=1,\ldots,M$. Since the cavity-coupling mirrors possess finite transmittance, damping of the mode oscillators arises in the Heisenberg–Langevin-equation method, and random quantum forces simultaneously appear that are responsible for coupling to the heat bath of the quantized light modes of the external space.

To describe the light field inside cavity m and on the outside in the neighborhood of its coupling mirror, it is convenient to introduce longitudinal coordinates y_m and x_m, as shown in Fig. 3. Here $0<y_m<l_m$ and $-\infty<x_m<\infty$, where l_m is the round-trip length of the cavity. We assume that the normal frequencies ω_m of the cavities are close to one another. Their difference must be much less than intermode splitting in each of the cavities.

Let us choose some carrier frequency ω_0, equal (or close in the nondegenerate case) to ω_m. The field of the light wave with $x_m<0$, i.e., incident on the coupling mirror, is related to the slow amplitude $e_m^{(\text{in})}(x_m,t)$ by

$$E(x_m,t)=-i\sqrt{\frac{2\pi\hbar\omega_0}{S}}\times\exp\{i(k_0x_m-\omega_0t)\}e_m^{(\text{in})}(x_m,t). \tag{A1}$$

For $x_m>0$, a similar definition gives the slow amplitude $e_m^{(\text{out})}(x_m,t)$ of the field leaving the cavity. Here S is the cross section of the light flux. In this normalization, the mean value of (e^+e) corresponds to the number of photons per unit length of the ray. For the field inside the cavity, we use

$$E(y_m,t)=-i\sqrt{\frac{2\pi\hbar\omega_0}{S}}\times\exp\{i(k_my_m-\omega_0t)\}\frac{1}{\sqrt{l_m}}a_m(t). \tag{A2}$$

The condition $E(y_m=l_m,t)=-E(y_m=0,t)$ is satisfied; i.e., when the phase shift by π accompanying reflection from the weakly transmissive coupling mirror is taken into account, the field is periodic after a round trip around the cavity. The reflection and transmission coefficients of the coupling mirror of cavity m are taken for definiteness in the form $\rho_m=-|\rho_m|$, $\tau_m=i|\tau_m|$.

The relationship between fields a_m, $e_m^{(\mathrm{in})}$, and $e_m^{(\mathrm{out})}$ is given[8,9] by

$$\dot{a}_m = -\left(i\Delta_m + \frac{\kappa_m}{2} \right) a_m + \frac{c\tau_m}{\sqrt{l_m}} e_m^{(\mathrm{in})}(0,t), \tag{A3}$$

$$e_m^{(\mathrm{out})}(0,t) = -e_m^{(\mathrm{in})}(0,t) - \frac{\tau_m}{\sqrt{l_m}} a_m(t). \tag{A4}$$

Here $\Delta_m = \omega_m - \omega_0$ is the frequency offset, and $\kappa_m = c|\tau_m|^2/l_m$ is the damping rate of the field energy. It is assumed that $1-|\rho_m| \ll 1$, which is necessary for systematic isolation of the mode oscillators. These relationships reflect the field-addition rules at the coupling mirror. On the other hand, Eqs. (A3) and (A4) can be thought of as the Heisenberg–Langevin equations in the absence of an active substance in the cavities.

Let the optical system contain $n = 1..., N$ nodes, where $N \geqslant M$. They are linear elements; i.e., cavities of the form described above, as well as lossless interference mixers. If there are any partially absorbing linear filters, we represent them as mixers that extract part of the light flux from the system. We do not consider optical configurations in which the light signals connecting the elements of the system form closed paths. Moreover, we exclude any complex cavities that explicitly or implicitly appear.

We number the inputs of an optical system on which independent light fluxes in the vacuum state are incident with the superscript $r = 1, \ldots, R$. Starting with each input of the system, we trace to its output the light rays with the same numbers. This can be done by various means, since each of the n elements of the system has an equal number of inputs and outputs: one each for cavities and two each for mixers.

For mixer n, we introduce longitudinal coordinates x_n^r and x_n^s connected with it, lying along rays r and s, which intersect in it (see Fig. 3). We define by an expression analogous to Eq. (A1) the input and output fields of the mixer in its normal coordinates. They are related by the transformation

$$e_n^{r(\mathrm{out})}(0,t) = \sum_s R_n^{rs} e_n^{s(\mathrm{in})}(0,t), \tag{A5}$$

where $\{R_n\}$ is a unitary matrix of the mixer. On the path from element n to the neighboring element m along ray r, the light signal is delayed only by distance d_{nm}^r, which can be expressed by

$$e_m^{r(\mathrm{in})}(0,t) = \exp(ik_0 d_{nm}^r) e_n^{r(\mathrm{out})}(0, t - d_{nm}^r/c). \tag{A6}$$

Below we assume that the delay time is small compared with the other times involved in the problem, and neglect it.

Equations (A3)–(A6) must be considered along with the commutation relations for the independent light fields that come from free space to the inputs of the optical system:

$$[\tilde{e}^{\,r(\mathrm{in})}(x,t), \tilde{e}^{\,r'(\mathrm{in})+}(x',t')] = \delta_{r,r'}\delta((x-x') - c(t-t')). \tag{A7}$$

Here $\tilde{e}^{\,r(\mathrm{in})}(x,t) = e_n^{r(\mathrm{in})}(x,t)$ are the input fields of the elements placed first in the ray paths.

We consider below the question of how consistent the description introduced above is from the viewpoint of quantum theory.

Let us transform Eqs. (A3), keeping in mind Eqs. (A4)–(A6). We represent the input field of each cavity as a sum of the fields of the cavities that act as the signal sources for that cavity and the fields at the inputs of the optical system:

$$\dot{a}_m = -\left(i\Delta_m + \frac{\kappa_m}{2} \right) a_m + \sum_{m_0 \neq m} g_{mm_0} a_{m_0} + \sum_r f_m^r \tilde{e}^{\,r(\mathrm{in})}(0,t). \tag{A8}$$

Here g_{mm_0} is the propagator that describes the penetration of the signal from cavity m_0 into cavity m:

$$g_{mm_0} = c\,\frac{|\tau_m \tau_{m_0}|}{\sqrt{l_m l_{m_0}}} \sum_{\mathrm{path}} \{\exp(ik_0 d_{\mathrm{path}}) S_{n_p}^{r_p r_{p-1}} \ldots S_{n_1}^{r_1 r_0}\}. \tag{A9}$$

The summation in this equation is over all signal paths from m_0 to m. If a path passes through intermediate elements $n_1, \ldots, n_p$, the summation includes the product of the transmission coefficients of these elements. According to (A4), we have $S_{n_i}^{r_i r_{i-1}} = -1$ for the reflection from the cavity. The path length is

$$d_{\mathrm{path}} = d_{m_0 n_1}^{r_0} + \ldots + d_{n_p m}^{r_p}. \tag{A10}$$

The propagator that describes the penetration of the vacuum field from input r of the optical system into cavity m has a similar form:

$$f_m^r = c\,\frac{i|\tau_m|}{\sqrt{l_m}} \sum_{\mathrm{path}} \{\exp(ik_0 d_{\mathrm{path}}) S_{n_p}^{r_p r_{p-1}} \ldots S_{n_1}^{r_1 r}\}, \tag{A11}$$

where

$$d_{\mathrm{path}} = d_{n_1 n_2}^{r_1} + \ldots + d_{n_p m}^{r_p}. \tag{A12}$$

We now proceed from these constructions to the kinetic equation. We assume that unless there is an active substance in the cavities, the evolution of the fields $a_m(t)$ according to the Heisenberg–Langevin equations is equivalent to their evolution in the form

$$a_m(t) \sim \exp(iH_0 t/\hbar) a_m \exp(-iH_0 t/\hbar). \tag{A13}$$

The unperturbed field-energy operator H_0 includes the energy of undamped oscillators a_m in cavities with perfect mirrors, the energy of the heat bath field, and also the effective interaction energy, which connects the internal and external fields when the mirrors have finite transmittance. We assume that before the interaction was turned on ($t=0$), the total light field of the isolated oscillators and of the heat bath was in the ground state $|\{0\}\rangle$.

We introduce perturbation V, which couples the field in the cavities to the substance located in the cavities. The density matrix of the system is

$$M(t) = v(t)|\{0\}\rangle\langle\{0\}| \otimes |\mu\rangle\langle\mu| v^+(t), \tag{A14}$$

where $|\mu\rangle\langle\mu|$ is the initial density matrix of the substance, and $v(t)$ is the evolution operator,

$$v(t)=\exp\left(-\frac{iH_0t}{\hbar}\right)v^{(i)}(t)=\exp\left(-\frac{iH_0t}{\hbar}\right)T$$
$$\times\exp\left\{-\frac{i}{\hbar}\int_0^t dt'V^{(i)}(t')\right\}. \quad \text{(A15)}$$

Superscript (i) indicates the interaction picture, in which V serves as the perturbation. We exclude the oscillators of the heat bath field, proceeding to the density matrix

$$\rho=\operatorname*{Tr}_{\{\text{term}\}} M. \quad \text{(A16)}$$

We transform the density matrix of the explicitly identified modes into the antinormal representation over coherent states. We introduce the coherent state $|\{z_m\}\rangle$ in the usual way by means of shift operators $D_m(z_m)$, $D(\{z_m\})=\Pi D_m(z_m)$. Instead of the density matrix, it is convenient to consider the somewhat more general quantity

$$\langle\{z_m\}|A\rho B|\{z_m\}\rangle=\langle\{0\}|D^+(\{z_m\})A\rho BD(\{z_m\})|\{0\}\rangle, \quad \text{(A17)}$$

where A and B are operators related to the identified modes. In order to consider the average (A17) as the trace over all degrees of freedom of the field, we use the equation

$$\langle\{0\}|..|\{0\}\rangle=\operatorname*{Tr}_{\{a_m\}}\left\{\exp\left(-\sum_m \underset{\rightarrow}{a_m}\underset{\leftarrow}{a_m^+}\right)\ldots\right\}. \quad \text{(A18)}$$

Here the operators $\underset{\rightarrow}{a_m}$ and $\underset{\leftarrow}{a_m^+}$ are placed to the left and right of the quantity to be averaged. Using simple transformations, Eq. (A17) becomes

$$\langle\{z_m\}|A\rho B|\{z_m\}\rangle=\langle\{0\}|v^{(i)^+}(t)\exp\left(\frac{iH_0t}{\hbar}\right)B:$$
$$\times\exp\left\{-\sum_m (a_m^+-z_m^*)(a_m-z_m)\right\}:A\exp\left(-\frac{iH_0t}{\hbar}\right)v^{(i)}(t)|\{0\}\rangle, \quad \text{(A19)}$$

where the symbol $:\ldots:$ denotes normal ordering. Let us consider the antinormal representation of the density matrix, i.e., $A=B=1$. As follows from Eq. (A13), the time-dependent field operators $a_m(t)$ and $a_m^+(t)$ appear in the enclosing functions $\exp(iH_0t/\hbar)$ and $\exp(-iH_0t/\hbar)$ in Eq. (A19). They satisfy the Heisenberg–Langevin equations for cavities coupled by optical signals in the absence of an active substance.

In order to obtain the kinetic equation, we take the derivative with respect to time t in Eq. (A19), expanding the resulting derivatives of the field amplitudes by means of the Heisenberg–Langevin equations, Eqs. (A8). On the right-hand side, in particular, we get

$$\frac{d}{dt}:\exp\{\ldots\}:=-\sum_m (\dot a_m^+(t):\exp\{\ldots\}:(a_m(t)-z_m)$$
$$+(a_m^+(t)-z_m^*):\exp\{\ldots\}:\dot a_m(t). \quad \text{(A20)}$$

Here it is taken into account that the quantities $\dot a_m(t)$ contain only positive-frequency contributions and can be permuted with the amplitudes $a_n(t)$, and the negative-frequency quantities can likewise be permuted.

Using Eqs. (A8), factors appear in Eqs. (A20) that equal the field amplitudes in the cavities, as well as the factors z_m and z_m^*, which can easily be converted into operators. It follows from Eq. (A17) that multiplying by z_m is equivalent to replacing B with Ba_m, and multiplying by z_m^* is equivalent to replacing A with a_m^+A. Moreover, factors proportional to the incident fields (random forces) $\tilde e^{\,r(\text{in})}(t)$ and $\tilde e^{\,r(\text{in})+}(t)$ appear in part of the contributions. All the contributions from the given factors equal zero. Actually, it follows from Eqs. (A8) that the solutions for the amplitudes $a_m(t')$ for $t'<t$ must have the form of a convolution of random forces over time, with these being taken at an instant $t''<t'<t$. It follows that, for $t'<t$,

$$[a_m(t'),\tilde e^{\,r(\text{in})+}(0,t)]=0 \quad \text{(A21)}$$

Therefore, the positive-frequency random forces can be carried to the right through the amplitudes of the isolated modes entering at times $t'<t$ into the evolution operator $v^{(i)}(t)$; see Eq. (A15). The negative-frequency random forces are likewise carried to the left. When they act on the initial vacuum state, the given factors annihilate the contributions in which they are contained.

Thus, by differentiating with respect to time, the amplitudes of the isolated modes that are attached to operators A and B in Eq. (A17) can be separated out. It is easy to show that the kinetic equation in the form of Eq. (4) (see Section II) follows from the described construction and the explicit form of the Heisenberg–Langevin equations, Eqs. (A8). The Liouville evolution operator of isolated mode m has the usual form:

$$\hat L_m\rho=\left\{-\left(i\Delta_m+\frac{\kappa_m}{2}\right)a_m^+a_m\rho+\frac{\kappa_m}{2}a_m\rho a_m^+ +\text{H.c.}\right\}$$
$$-\frac{i}{\hbar}[V_m,\rho]. \quad \text{(A22)}$$

Here V_m is the interaction-energy operator of mode m with the active substance in the cavity. The factor g_{mm_0} entering into operator $\hat L_{m_0m}$ of Eq. (5) is nonzero only when there are paths in the optical system from cavity m_0 to cavity m [see Eq. (A9)].

In the special case of two cavities (a source and a signal detector), the kinetic equation, Eq. (A22), transforms into that obtained earlier.[6–8]

Based on the approach used above, we assumed the Heisenberg–Langevin equations, Eq. (A3), the coupling equations, Eqs. (A4)–(A6), and the commutation relations of the input free fields, Eq. (A7). We now show that the given formulation of the problem is consistent in the case of optical systems in which the light signal does not close on itself. The elucidation of this question is of interest not only in justifying the kinetic equation (see above), but also for understanding the limits of applicability of the Heisenberg–Langevin-equation method.

We form a vector from the set of input amplitudes $\tilde{e}^{r(\mathrm{in})}(0,t)$, where $r=1,\ldots,R$. The transformation of the fields at the elements of the optical system, beginning with its inputs, can be thought of as a linear transformation of this vector. We replace the linear attenuating elements with partially transmitting mixer devices, which extract part of the light flux from the optical system and simultaneously introduce the vacuum field into the light flux. This is required to take into account in the calculation the number of inputs, internal rays, and outputs of the system. For elements acting on independent sets of rays, the order in which the corresponding transformations are applied is not fixed. It can be chosen arbitrarily, since such transformations obviously commute with each other.

The transformation of the fields in a mixer, Eq. (A5), is unitary. It is convenient to consider the transformation of the fields at the coupling mirror of the cavity, given by Eqs. (A3) and (A4), in a particular representation. For the Fourier transforms of the amplitudes, $t\to\Omega$, it is easy to obtain

$$e_m^{(\mathrm{out})}(0,\Omega)=e_m^{(\mathrm{in})}(0,\Omega)\,\frac{\kappa_m/2-i(\Delta_m-\Omega)}{\kappa_m/2+i(\Delta_m-\Omega)}, \tag{A23}$$

which is also a unitary transformation. The vector of the initial amplitudes thus undergoes a unitary transformation as the field passes through an arbitrary number of elements. It is easy to obtain from this that the commutation relations for the components of the transformed vector of the input amplitudes have the form of Eqs. (A7) for the fields of free space.

This in turn provides the necessary commutation relations for the amplitudes of the isolated modes. Amplitude a_m can easily be expressed by means of Eq. (A3) via a convolution of the input field $e_m^{(\mathrm{in})}(0,t)$ over time. As we have just explained, the input fields of any elements of the system (and not only the initial elements on the ray path) obey the commutation relations for the fields of free space. From this it is easy to find

$$[a_m(t),a_m^{+}(t')]=\exp\{-i\Delta_m(t-t')-(\kappa_m/2)|t-t'|\}. \tag{A24}$$

When $t=t'$, the usual commutation relation for an oscillator appears.

We should point out that it is important for these considerations that there be no light rays closed on themselves in the optical system. It can be assumed that in the case of completely or partially closed system, a description based on the coupled Heisenberg–Langevin equations and the free commutation relations for the intermediate light fields will not be valid.

APPENDIX B: EXCITING THE ACTIVE MEDIUM OF A LASER WITH RADIATION FROM AN AUXILIARY SUB-POISSON LASER

The following physical model is considered in Ref. 5. Inside a high-Q cavity with a resonant medium is placed another high-Q cavity with another resonant medium. Ideal sub-Poisson lasing is provided in the inner cavity, and this excites the resonant medium of the large cavity in the above-threshold state, as a consequence of which secondary lasing appears. Reference 5 discusses in detail the conditions under which the secondary lasing becomes ideally sub-Poisson with a Mandel parameter of $\xi=-1/2$, i.e., it becomes similar in its statistical properties to the primary radiation as it was before the secondary radiation arose. Using the formulas obtained in Ref. 5, we can find an explicit expression for the photocurrent spectrum during recording of the radiation of the secondary lasing in the form

$$i_\omega^{(2)}=i_{\mathrm{shot}}^{(2)}\left[1-\frac{\kappa}{\kappa+\kappa_1}\,\frac{(\kappa_1+\kappa)^2}{\omega^2+(\kappa_1+\kappa)^2}\,\frac{\kappa_2^2}{\omega^2+\kappa_2^2}\right]. \tag{B1}$$

Here κ_1 and κ_2 are the spectral widths of the primary (inner) and secondary (outer) laser cavity, and κ is the linear absorption coefficient for the primary laser light in the medium of the secondary laser (it determines the population rate of the upper active level of the medium of the secondary laser).

We use for our analysis the formula obtained in Ref. 5 for the Mandel parameter of the secondary lasing:

$$\xi_2=-\frac{1}{2}\,\frac{\kappa}{\kappa_1+\kappa_2+\kappa}. \tag{B2}$$

The most interesting case for us is that in which the cavity losses of the primary lasing are mainly associated with absorption in the medium of the secondary laser:

$$\kappa\gg\kappa_1. \tag{B3}$$

Then the quantum singularity in the form of a dip in the photocurrent spectrum, Eq. (B1), will be the most pronounced (the depth of the dip will be $\delta=1$). As for a passive cavity (see Section V), this fact is not at all dependent on the ratio between the cavity widths κ_1 and κ_2 of the primary and secondary lasers.

At the same time, the Mandel parameter depends on this ratio. In fact, if $\kappa\gg\kappa_a,\kappa_b$, $\xi=-1/2$. However, if $\kappa_2\gg\kappa\gg\kappa_1$, then $|\xi|\ll 1$.

This work was carried out with the partial support of INTAS (93-1914-EXT).

[1] P. Filipowicz, J. Javanainen, and P. Meystre, Phys. Rev. A **34**, 4547 (1986); L. Lugiato, M. O. Scully, and H. Walther, Phys. Rev. A **36**, 740 (1987); J. Krause, M. O. Scully, and H. Walther, Phys. Rev. A **34**, 2032 (1986).
[2] Yu. M. Golubev and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **87**, 408 (1984) [Sov. Phys. JETP **60**, 234 (1984)].
[3] Yu. M. Golubev, Zh. Éksp. Teor. Fiz. **107**, 401 (1995) [Phys. JETP **80**, 212 (1995)].
[4] F. Haake, M. Kolobov, C. Fabre, E. Giacobino, and S. Reynaud, Phys. Rev. Lett. **71**, 995 (1993).
[5] Yu. M. Golubev, Zh. Éksp. Teor. Fiz. **103**, 832 (1993) [JETP **76**, 408 (1993)].
[6] M. I. Kolobov and I. V. Sokolov, Opt. Spektrosk. **62**, 112 (1987) [Opt. Spectrosc. (USSR) **62**, 69 (1987)].
[7] H. J. Carmichael, Phys. Rev. Lett. **70**, 2273 (1993).
[8] C. W. Gardiner, Phys. Rev. Lett. **70**, 2269 (1993).
[9] M. I. Kolobov and I. V. Sokolov, Opt. Spektrosk. **67**, 122 (1989) [Opt. Spectrosc. (USSR) **67**, 68 (1989)].
[10] C. W. Gardiner and M. J. Collett, Phys. Rev. A **31**, 3761 (1985).
[11] C. W. Gardiner, *Quantum Noise*, Springer, Berlin (1992).

Translated by W. J. Manthey

Rydberg matter—a long-lived excited state of matter

L. Holmlid

Reaction Dynamics Group, Physical Chemistry University of Göteborg, S-412 96 Göteborg, Sweden

E. A. Manykin

Superconductivity and Solid State Physics Institute, Russian Research Center Kurchatov Institute, 123182 Moscow, Russia

(Submitted 11 June 1996)
Zh. Éksp. Teor. Fiz. **111**, 1601–1610 (May 1997)

The theory of condensed excited matter, the so-called Rydberg matter (RM), is examined briefly. Explicit results are given for several physical quantities, notably, the work function and the resistivity, for which experimental results exist. The most important aspects of the experiments, which are fully described elsewhere, are discussed. Large densities of Rydberg species are formed in the experiments with cesium vapor in contact with carbon (graphite) surfaces. The resistivity of the RM formed is found to be 10^{-2}–10^{-3} $\Omega\cdot$m under varying conditions, while theory gives the order of 10^{-3} $\Omega\cdot$m. The work function is experimentally found to be less than 0.7 eV, perhaps even less than 0.5 eV. Two different methods were used to extract this quantity from thermionic diode data. These work function values are much lower than reported for any known material, especially at the high temperatures, and they thus give strong support for the description of RM as a very dilute metal. Theory gives values ranging from 0.6 down to 0.2 eV, depending on the principal quantum number, which is estimated to be $n=12$–14 from the lifetime calculations and from the known pressure. Supporting evidence is found from spectroscopic studies of RM, from jellium calculations, and from recent confirming experiments. From the good agreement between theory and experiment we conclude that RM exists. [S1063-7761(97)00505-2]

1. INTRODUCTION

The theoretical treatment of a phase consisting of highly excited atoms of the so-called Rydberg type (hydrogenic atoms with one excited electron) indicates that a metal with a density as low as a gas can exist. Recent experimental results on the resistivity and work function indicate that such metals do exist, in the form of Rydberg matter (RM). We will here consider the new theoretical and experimental arguments for the existence of RM.

The theoretical reason for introducing RM is that local excitations cannot satisfactorily explain the behavior of electronically highly excited systems. Instead, the interactions between the excitations in the system must be taken into account. Such a collective or condensed state of matter, which is formed from excited atoms or other excited species was proposed in a series of theoretical studies.[1–4] When atoms are excited to high electronic states of the Rydberg type, their size increases rapidly. Since the excited electrons spend almost all their time far from the core ions in the atoms, the wave functions of highly excited atoms overlap strongly when such atoms are brought together. At large densities and moderately high temperatures, excited atoms can form a solid-like condensed phase with the ions in a regular lattice, which we call a Rydberg crystal, as shown in Fig. 1. At higher temperatures, a liquid state of excited matter, which is another form of RM, should exist.

A prerequisite for valid experimental tests of the existence of RM is that RM can be formed in macroscopic quantities, which in turn requires that large densities of Rydberg species can be obtained. The use of new methods of forming Rydberg species has now led to this achievement, and in our experiments, a volume of 30 mm^2 of RM of cesium atoms can be formed routinely in a flow system which continuously renews RM. We will show that the accumulated experimental evidence now shows good agreement with theoretical predictions, and that this, by itself, proves the existence of RM. However, it may be even more convincing if unique results can be provided by experiments, i.e., results which cannot be attributed to any known material. This is found, we believe, in the case of the work function of RM. Of course, low work functions of surfaces have been studied for a long time. In many technical applications of electron physics low work functions of electrodes are required, and the combination of low work function and metallic properties is extremely important in many situations. Examples are provided by thermionic energy converters and MHD generators, where current densities up to the order of 100 $A\cdot cm^{-2}$ should be obtained to make the devices work efficiently. The difficulties in forming or constructing such surfaces have for a long time been serious obstacles for the technical development. Thus, several approaches have been tested, but with very limited success. The alkali metals have the lowest work functions among the elements, at approximately 1.8 eV for Cs, and a layer of Cs on metallic surfaces can have an even lower work function, at 1.5 eV for nearly a monolayer. The standard theory for work functions of metals[5] states that the average electron density, or the corresponding positive ion density in the material, should be decreased if very low work functions are found. This explains, at least in part, why the low density monolayers of alkali metals deposited on other

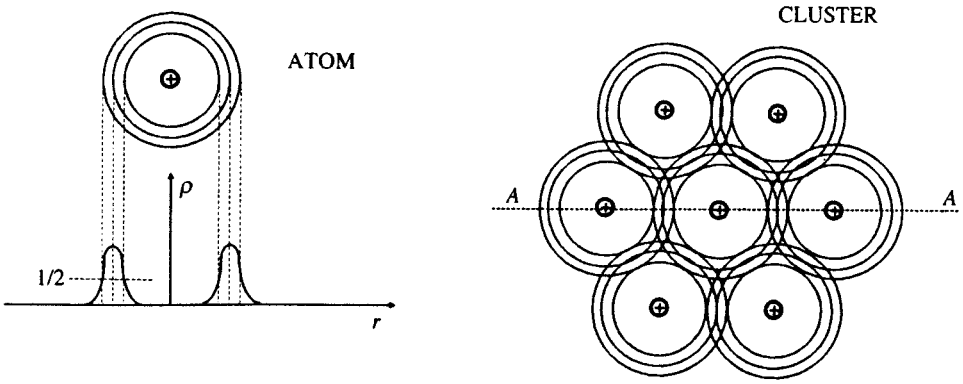

FIG. 1. The electron density ρ in a Rydberg atom as a function of distance (top panel) and a cluster of RM. The central positive core ions are indicated by plus signs. The interatomic distance in RM with atoms in state 13 S is 15 nm according to theory.[7]

metals have lower work function than the pure alkali metals themselves. Following this idea, various experiments have attempted to decrease the work function by diluting the alkali-metal atoms with other nonmetal atoms, and very low work functions, down to 1.2 eV, have also been obtained for various oxides and alkali metal-oxygen coadsorbed phases on metal surfaces.[6] Of course, the resistivity increases for such materials, but at high temperatures and for thin layers this drawback is not of great importance. The goal to reach even lower work functions is thus clear: to further increase the distance between the alkali atoms. This is where the theory of RM is important also from an applied point of view, since it states that a low density, and thus low work function, metallic material can be constructed. The prerequisite is that Rydberg states should be formed in large densities and be condensed on a surface which can remove the condensation energy.

2. THEORY OF RYDBERG MATTER

In Rydberg matter the excited valence electrons are shared between many atoms in the material and form a degenerate Fermi liquid. Due to delocalization, the kinetic energy of the electrons decreases considerably when the condensed phase is formed. As a result, the binding energies of the excited atoms become high enough to keep the atoms in place. The potential energy for the valence electrons in Fig. 2 shows that the probability of these electrons coming close to the core ions is very small.

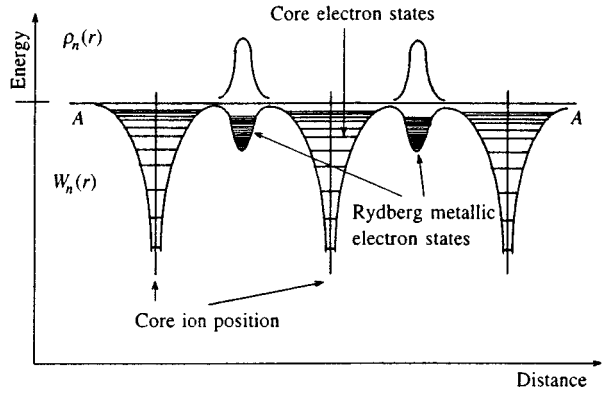

FIG. 2. Potential energy diagram for the electrons in RM, along the line $A-A$ in Fig. 1. The electron density ρ is indicated.

The different properties of RM are well described by pseudopotential theory and density functional theory similar to the theory of simple metals.[4] The Rydberg matter is therefore assumed to consist of identical Rydberg atoms. A substantial role should be played by effects due to the inhomogeneity of the electron density. This is a new aspect, which is characteristic of far-from-equilibrium matter like RM, while this effect is very small for ordinary condensed matter. The theory of RM shows that the exchange and correlation effect leads to self-trapping of the electrons where the electron density differs substantially from zero.[4] Figure 3 shows the result of a numerical calculation within the framework of the Wigner-Seitz cell potential for RM with the excitation level $n=10$ (Refs. 4 and 7), as well as the relevant energy parameters as a function of n. The question of applying the density functional theory to the RM problem is discussed in detail in Ref. 4. It is shown there, that the problems of describing excited states by density-functional methods can be overcome by combining this theory with the pseudopotential concept.

The lifetime of the Rydberg crystal is, of course, of great importance. The decay of RM is caused by electron transitions to low-lying unfilled energy levels. The Rydberg matter is therefore inherently unstable and has a finite lifetime. At first glance, the lifetime of RM seems to be shorter than that of the isolated atoms, which is attributable to the strong Coulomb interaction of the electrons. As the excitation level increases, however, the lifetime of RM is expected to increase rapidly due to the spatial separation of the initial and final states of the electrons, as in free Rydberg atoms, and due to specific effects on the local field in the strongly nonuniform electron liquid of RM. The higher the level of excitation, the stronger the electrons are drawn to the boundaries of the unit cells, while the final states for the decay transitions remain localized at the core ions. Recombination, therefore, cannot occur into all low-lying states corresponding to isolated atoms. The interaction between the electrons leads to the formation of an effective potential barrier to the penetration of valence electrons into the region close to the core ions. Thus, RM can be a long-lived excited state of matter. The high excitation energy in the condensed matter also makes possible decay processes involving several electrons, for example, the Auger processes. The lifetimes including such processes are calculated in Ref. 8 and are still found to be of

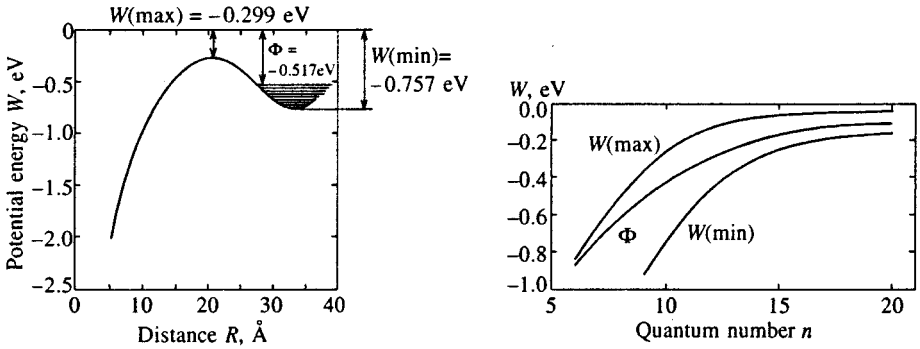

FIG. 3. One-electron potential energy for the excited electrons in RM with $n=10$. The energy parameters indicated are also shown separately as functions of n. The work function Φ is the absolute value of the Fermi energy E_F.

the order of several seconds or longer for some excitation levels.

3. CALCULATED QUANTITIES

Various parameters characterizing the Rydberg crystal state can be calculated by the methods mentioned above.[4] Some examples are represented in Table I. The Rydberg crystal is a good electrical conductor,[8] and its work function is thus a well-defined quantity of great interest. Due to the low electron density, the Rydberg crystal is transparent to visible light. It becomes opaque to electromagnetic radiation only at frequencies lower than its own plasma frequency or at wavelengths longer than some transparency boundary wavelength, which is far out in the IR for the excitation levels in Table I.

A basic problem with this type of calculation is that the excitation level, i.e., the principal quantum number for the combining atoms, must be assumed. This parameter is not known directly from the experiments. However, the vapor pressure before condensation in the experiments in Refs. 9 and 10 is of the order of 1 mbar. Assuming that the diameters of the Rydberg atoms are the same as the interatomic distances in a gas of this pressure, one finds n approximately equal to 16. It is likely that some contraction is taking place as a result of the condensation, and it is thus reasonable to perform the calculations for $n=12-14$. The calculated values for the density at $n=13$ in Table I, correspond to a pressure of 60 mbar or an equivalent n value of 9. Thus, $n=12-14$ appears to be a good choice of excitation level, lacking more detailed information from the experiments. A more detailed discussion is found in Ref. 8.

4. RESULTS AND DISCUSSION

The experimental results have been obtained using new techniques to create large densities of highly excited Rydberg species. Alkali atom Rydberg species are generally the simplest to form because of the low ionization energy of alkali atoms. The new techniques employ diffusion of alkali ions from the bulk of nonmetallic materials. For example, in the case of graphite surfaces it was shown that the emission of Cs^+ ions gives rise to a near-resonant process that forms Rydberg species from the emitted ions and thermal electrons.[10–12] By increasing the pressure of Cs vapor in contact with such surfaces and by ensuring that diffusion in the material takes place, large densities of Rydberg species of Cs and K have been formed. Small particles (clusters)[13] and macroscopic amounts of RM[9,10] can be formed by using Cs vapor. Also, alkali-doped metal-oxide surfaces (e.g., promoted catalyst surfaces) have been shown to give large densities of Rydberg species of K.[14–16] The behavior observed for RM formed by such methods is qualitatively the same as the predicted behavior: it is, for example, transparent to visible light, and it emits much less light than an ordinary plasma when it carries a large current. That the observed matter is very energetic is easy to observe through explosions of the matter under simultaneous emission of visible light or charged particles, both for small particles of RM[17] and for large layers of RM on surfaces.[18] In one experiment clusters of RM were collected on a liquid-nitrogen-cooled surface.[17] They were de-excited by ion impact, which gave small microflashes (small white explosions) with an energy content approximately as expected from RM theory. Collection times longer than a few minutes gave no increase in the number of flashes during de-excitation, and the intrinsic lifetime of RM at that temperature was estimated to be a few minutes. This lifetime is of the same order of magnitude as that given in Table I, with values ranging between 5 s and 80 h. The main recombination channels for the kind of RM, which consists of highly excited Cs atoms, have been recently investigated. A detailed description of this study is found in Ref. 8. The decay proceeds, as a rule, by the Auger recombination mechanism.

TABLE I. Parameters of the cesium RM crystal.

State of atoms	Density ρ, cm^{-3}	Binding energy B, eV/atom	Melting point T_m, K	Work function Φ, eV	Transparency boundary λ, μm	Lifetime τ, s
12 S	1.1×10^{18}	0.14	460	0.23	32	25
13 S	5.3×10^{17}	0.11	460	0.2	46	5
14 S	2.8×10^{17}	0.1	540	0.18	63	80 h

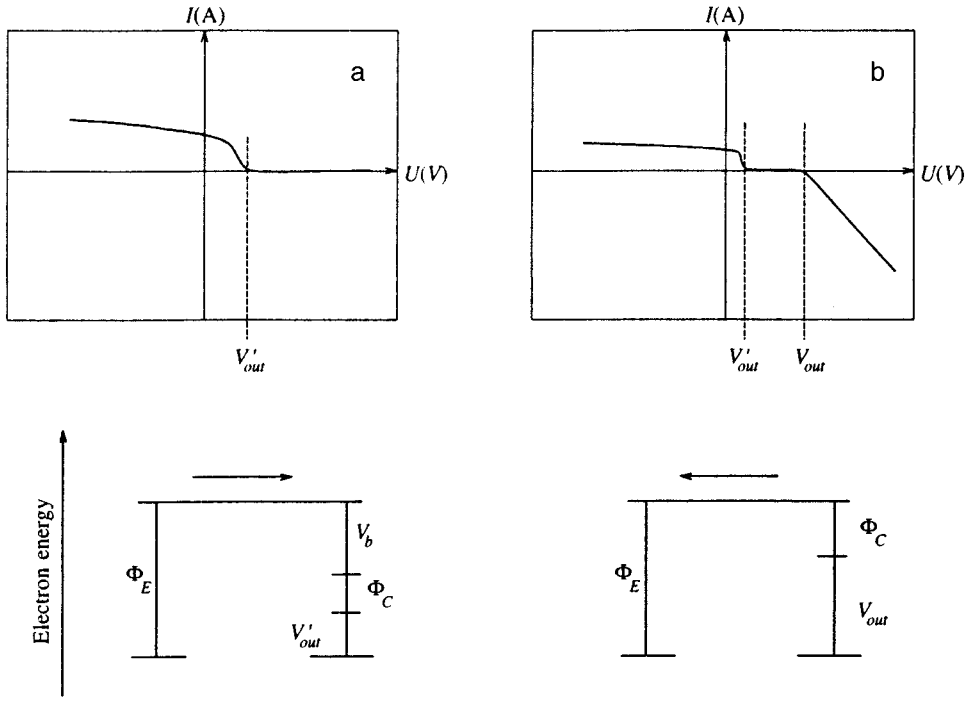

FIG. 4. Current–voltage (I–V) characteristics of the cesium-vapor-filled thermionic diode. The ordinary behavior, together with a description of the energetics for the electrons in the diode, is shown in (a). The signs used are conventional; the positive current visible in the second quadrant is emitted from the hot electrode. The special behavior caused by RM formation, with two break points in the curves and a very large electron current from the cold electrode in the fourth quadrant, is shown in (b). The energetics shown in (b) corresponds to the break point farthest to the right.

Of special importance are the values of the resistivity, which were determined for RM of cesium at a temperature of approximately 800 K. These measurements were made in a flow system which renews the RM continuously between two partially graphite-covered electrodes in a vacuum chamber. This chamber is a thermionic diode, which is characterized as a thermionic energy converter. The complete apparatus has recently been described elsewhere.[19] The electron current flows from the cold to the hot electrode. The interelectrode space can be observed visually during the measurements; this space is not glowing. (When a plasma is formed between the electrodes, this region is glowing and the maximum possible conducted current is much lower.) The experiment is designed in such a way that possible artifacts, like leakage paths on isolators, can be rejected directly from the experimental results. Further, the hot electrode is heated by the impinging electrons, which shows that the current passes through the interelectrode space. A resistivity of 10^{-3}–10^{-2} $\Omega\cdot$m is found from the linear resistive behavior of the current-voltage characteristics.[9] The range in values is due to the varying conditions for a large number of experiments done at different electrode temperatures and cesium pressures. The resistive behavior is found only with graphite layers on the electrodes. Calculations give the resistivity of RM for the levels $n=12$–14 of the order of 10^{-3} $\Omega\cdot$m. The agreement between theory and experiment is therefore very good.

The surface work function of RM was determined in similar experiments in the same kind of apparatus,[19] which forms RM from Cs vapor in contact with a pair of electrodes that are partially covered with graphite. It is worth noting that the colder electrode of the two plane-parallel electrodes supports the RM, and that the electron current which is observed is emitted from the cold surface to the hot surface (the reverse current is usually much smaller). The maximum current density is very large, up to 500 A$\cdot$cm^{-2} at an applied voltage of 30 V. As stated in the introduction, the extremely low values of the work function found in the experiments are conclusive evidence that RM exists. We must therefore examine the measurement methods in detail.

Two different methods were used to extract the work function from the current-voltage behavior observed: 1) maximum current density method, and 2) analysis of the diode characteristics of the plasma-RM setup. From the Richardson formula for the electron emission current density, we clearly see that a current density, $i_{\max}=500$ A$\cdot$cm^{-2}, corresponds to a work function of 0.82 eV at 800 K, which is the temperature of the cold emitting surface. This value of $i_{\max}$ is the highest current density measured in the setup because of the experimental limitations like power supply availability and melting of the electrodes. The current density at this point still increases linearly with the applied current. We thus can state that the work function of RM in this experiment is <0.8 eV. The measurement of the current density depends on the correct determination of the electron emitting area. Through a window of the apparatus we can observe the faint glow from the current-carrying part of the interelectrode space, and thus ascertain that there is no large current emission from adjacent parts of the apparatus. If the emitting area is 50% larger in reality, the work function would be 0.85 eV. In fact, the area used in the calculation is not emitting homogeneously, so the central part, which has a larger emission, also has a somewhat lower work function.

The diode analysis procedure is carried out in the following way. The typical current-voltage characteristics of the plasma diode is shown in Fig. 4a. In the first and second quadrants, the electron current from the hot electrode is measured as a positive current by definition. With an increase in the voltage to the right, i.e., with a more negative voltage on

the cold electrode, the thermal current from the hot electrode will decrease, showing a Boltzmann tail. The steepest descent of this decay curve indicates the approximate output voltage, at which the work function Φ_E of the hot electrode is equal to the sum of the so-called barrier index (which contains the work function Φ_C of the cold electrode and the plasma drop) and the output voltage V'_{out}:

$$\Phi_E = V_b + V'_{\text{out}}. \tag{1}$$

The decay curve must be limited by the density of the thermally emitted electron current from the hot electrode, as expressed by the Richardson equation:

$$i_e = AT_E^2 \exp(-V_b/k_B T_E), \tag{2}$$

which gives the barrier index V_b from the measured current density and the hot electrode temperature T_E. This can be understood more easily from the lower part of Fig. 4a, where the energy of the electrons is shown. This analysis is a standard procedure for the study of thermionic converter performance.

In the case of RM formation in the thermionic diode, there exist two break points or knees on the curve, as seen in Fig. 4b. The break point at the lowest voltage V'_{out} (to the left in the figure) corresponds to the point at which the electron emission from the hot electrode is small, i.e., to a case similar to that analyzed with the help of Fig. 4a. The right-hand break point at V_{out} indicates the onset of electron emission from the cold electrode, which increases approximately linearly with the applied voltage above this voltage point. Here another relation can be found from the lower part in Fig. 4b:

$$\Phi_E = \Phi_C + V_{\text{out}}. \tag{3}$$

Combining Eqs. (1) and (3), we obtain

$$V_b = (V_{\text{out}} - V'_{\text{out}}) + \Phi_C. \tag{4}$$

This relation indicates that the work function of the cold electrode covered with RM should be regarded a constant (axis cutoff) in a linear relationship between the barrier index and the voltage difference between the break points. The data from a large number of runs with different temperatures of the hot electrode is plotted in Fig. 5, with the barrier index and the break-point voltage difference on the x and y axes, respectively. As in Ref. 10, an approximate linear relationship is observed. In the figure, the temperature of the hot electrode is now plotted in the z direction. We see that most of the results fall within a band with the work function Φ_C (barrier index value at zero difference voltage) between zero and 0.7 eV. The few data points at low difference voltage and high barrier index are found for higher temperatures of the hot electrode than the other data points, which might mean that the plasma voltage drop has a somewhat different character in those cases. It is important to realize that no data point with $V_b < (V_{\text{out}} - V'_{\text{out}})$ should exist. This condition is satisfied (excluding one data point); this strongly supports the analysis of the data. The work function of the RM was thus found to be less than 0.7 eV.[10] No known material has such a low work function, especially at such a high temperature. The calculated value of the work function for RM is extremely small, ranging from 0.6 eV down to 0.2 eV for a

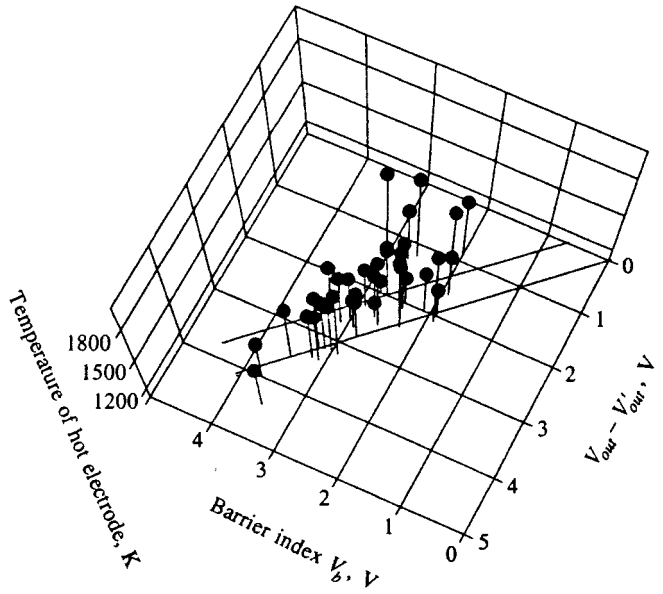

FIG. 5. A plot of the barrier index V_b versus the voltage difference ($V_{\text{out}} - V'_{\text{out}}$) for the two break points in the I–V characteristics for the thermionic diode with RM formation. The two lines demonstrate the linear relationship between the two parameters at moderately high temperatures of the hot electrode.

value of n approaching 20, as shown in Table I and Fig. 3. This value is just below the upper limit found experimentally. Similar diode experiments, performed recently by a group in the Kurchatov Institute using graphite surfaces, were shown to give extremely large current densities.[20] This confirms the results of our experiments.

Very low values of work function have also been found recently in simple jellium calculations.[21,22] Such calculations are preformed for a dilute alkali metal, replacing the inhomogeneous electron density in RM with a constant, smeared-out, electron density in the jellium approximation. In Ref. 22, the same densities of RM were used for the jellium calculations, as in the more accurate calculations in Refs. 1–4, 7, and 8. The work function from the jellium calculation follows the same trend with the excitation level (value of n) as the RM calculations. However, the work function values are even lower, which is expected since the volume over which the electrons can move is much larger in the jellium model, where there is no excluded volume around the ion cores. This point is discussed in greater detail in Ref. 22. It is encouraging that the two widely different theoretical methods give similar values of the work function, and the results in Ref. 22 strongly support the correctness of the RM theoretical treatment. The real work function for RM is likely to be between 0.5 and 0.1 eV at $10 < n < 20$ (Ref. 22).

A direct spectroscopic study of RM in the same apparatus as used for the other experiments discussed here[19] is quite instructive.[22] In this test, the spectra of the free Cs atoms were recorded under conditions of an ordinary plasma and also during conditions of RM formation with no glowing plasma. With RM, the current was generally higher and the voltage drop lower than for the plasma. It was barely possible to run comparison experiments where the current in the presence of RM was as low as that during plasma conditions.

Under these conditions, the spectral lines were a factor of 10–20 weaker in RM mode than in the plasma mode. Since the emitting Cs atoms must exist outside the RM within the viewing angle of the optical fiber, the fraction of free Cs atoms in the diode interelectrode space becomes very low, i.e., probably lower than 0.05 of all atoms there. This clearly indicates that the other atoms there are not free, i.e., they are bound in the RM, which is a much better conductor than the plasma.

5. CONCLUSIONS

The good agreement between theory and experiment makes it possible for us to conclude confidently that RM exists. Even if the discussion here has mainly described RM of cesium, it is now clear that RM can also be formed from other atoms and molecules. The Rydberg matter of hydrocarbons as clusters[24] and of hydrogen as surface layers has been reported.[18] It is likely that RM appears naturally in many situations, e.g., RM seems to be the likely explanation for the phenomena known as ball lightning.[25] This explanation was proposed by Manykin *et al.*[3]

We thank Robert Svensson for performing the experiments discussed in this article.

This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 95-02-06384-a).

[1] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Pis'ma Zh. Tekh. Fiz. **6**(2) 218 (1980) [Sov. Phys. Tech. Phys. Lett. **6**, 95 (1980)].
[2] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Dokl. Akad. Nauk SSSR **260** 1096 (1981) [Sov. Phys. Dokl. **26**, 974 (1981)].
[3] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Tekh. Fiz. **52** 1474 (1982) [Sov. Phys. Tech. Phys. **27**, 905 (1982)].
[4] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Éksp. Teor. Fiz. **84**, 442 (1983) [Sov. Phys. JETP **57**, 256 (1983)].
[5] N. D. Lang and W. Kohn, Phys. Rev. B **3**, 1215 (1971).
[6] J.-L. Desplat, J. Appl. Phys. **54**, 5494 (1983.)
[7] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Éksp. Teor. Fiz. **102** 804 (1992) [Sov. Phys. JETP **75**, 440 (1992)].
[8] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, Zh. Éksp. Teor. Fiz. **102** 1109 (1992) [Sov. Phys. JETP **75**, 602 (1992)].
[9] R. Svensson, L. Holmlid and L. Lundgren, J. Appl. Phys. **70**, 1489 (1991).
[10] R. Svensson and L. Holmlid, Surface Sci. **269/270**, 695 (1992).
[11] J. B. C. Pettersson, L. Holmlid, and K. Möller, Appl. Surface Sci. **40**, 151 (1989).
[12] K. Möller and L. Holmlid, Surface Sci. **204**, 98 (1988).
[13] C. Åman, J. B. C. Pettersson, and L. Holmlid, Chem. Phys. **147**, 189 (1990).
[14] J. Lundin, K. Engvall, L. Holmlid, and P. G. Menon, Catal. Lett. **6**, 85 (1990).
[15] C. Åman and L. Holmlid, Appl. Surface. Sci. **62**, 201 (1992).
[16] C. Åman and L. Holmlid, Appl. Surface. Sci. **64**, 71 (1993).
[17] C. Åman, J. B. C. Pettersson, H. Lindroth, and L. Holmlid, J. Mat. Research **7**, 100 (1992).
[18] E. Wallin, T. Hansson, and L. Holmlid, J. Phys. Condensed Matter **4**, 9803 (1992).
[19] R. Svensson, B. Lönn, and L. Holmlid, Rev. Sci. Instrum. **66**, 3244 (1995).
[20] V. Kaibyshev and E. Kennel, private communication.
[21] A. Nyberg and L. Holmlid, Surface Sci. **292**, L801 (1993).
[22] M. Svanberg and L. Holmlid, Surface Sci. **315**, L1003 (1994).
[23] B. E. R. Olsson, R. Svensson, and J. Davidsson, J. Phys. D: Appl. Phys. **28**, 479 (1995).
[24] C. Åman and L. Holmlid, J. Cluster Sci. **3**, 247 (1992).
[25] É. A. Manykin, M. I. Ozhovan, and P. P. Poluéktov, *Proc. 9th Int. Conf. on Atm. Electricity, St. Petersburg;* A. I. Voeikov, Main Geophysical Observatory, **3**, 838 (1992).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Transverse instability threshold in counterpropagating light beams for a nonlinear medium with local photorefractive response

B. I. Sturman and A. I. Chernykh

Siberian Branch of the International Institute of Nonlinear Studies, 630090 Novosibirsk, Russia
(Submitted 3 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1611–1623 (May 1977)

We derive the threshold conditions for the instability of counterpropagating waves in a nonlinear medium with local photorefractive response against the excitation of transverse small-angle structures. These conditions allow for all the important types of diffraction from refractive-index reflection gratings and are not limited to the case of strict frequency degeneracy of the waves. We study the dependence of the crystal-thickness threshold and the secondary wave emission angle on the crystal parameters and the pump conditions. We show that when the pump wave intensities differ considerably, excitation of standing light structures is replaced by excitation of traveling structures. Finally, we discuss the applications of the theory to experiments with the photorefractive crystals $LiNbO_3$ and $LiTaO_3$. © *1997 American Institute of Physics.* [S1063-7761(97)00605-7]

1. INTRODUCTION

It is well known that counterpropagating waves in a cubic nonlinear medium can be unstable against the spontaneous generation of small-angle structures.[1–4] This transverse instability is absolute and has a threshold in crystal thickness; positive distributed feedback lies at the basis of this phenomenon. By its nature, transverse instability is similar to what is known as cavityless lasing due to four-wave mixing.[5]

Experiments have shown that the result of the development of transverse instability is sensitive to the experimental conditions and the type of nonlinearity. The formation of ring, hexagonal, and more complicated small-angle light structure near the pump beams has been observed under various conditions.[3,4,6,7]

Theoretical studies of transverse instability have focused on finding the threshold conditions for the formation of light structures. These conditions were first established and studied in relation to Kerr and Brillouin media and to gases near absorption lines.[1–4] Lately there has been an upsurge of interest in transverse instability in media with photorefractive nonlinearity.[8–11]

The interest in photorefractive nonlinear media is not accidental. Such media are important in various applications and present exceptionally favorable conditions for studying secondary light structures. The required light intensities are within the intensity range of continuous-wave lasers, and typical nonlinear lengths amount to several millimeters. The mechanisms of photorefractive nonlinearity are well-studied and controllable.[12]

Nonlinear variations of the refractive index in a photorefractive medium are due to the formation of a space-charge field by light and the linear electrooptical effect.[12] Two limiting types of photorefractive linear response are usually distinguished: the local and the nonlocal.[1)] In the case of local response, a standing sinusoidal light pattern induces a (sinusoidal) spatial modulation of the refractive index. In the nonlocal photorefractive response, the induced refractive-index pattern is shifted by a quarter of the period relative to the light pattern. What determines the specific type of photorefractive response is the electron (hole) transport mechanism. For instance, photoelectron diffusion results in nonlocal photorefractive response. This is characteristic of experiments involving $KNbO_3$ and $BaTiO_3$ crystals. Local response usually occurs when the dominant effect is electron drift in an external field or the photovoltaic effect.[13] A typical example of a photorefractive medium with local photorefractive response is represented by $LiNbO_3$ crystals, which exhibit a strong photovoltaic effect.

Here is a brief history of studies of transverse instability in photorefractive media. The formation of traveling ring structures around counterpropagating pump beams was discovered in 1985 in $LiNbO_3$ crystals.[6] A similar effect in the same material was described later in Refs. 14 and 15. Unfortunately, no meaningful interpretation of these results was given at the time, nor was there any further development. The conditions under which the observations were conducted were not recorded. In 1993 transverse instability was observed in $KNbO_3$ crystals, which differ considerably in their photorefractive properties from $LiNbO_3$ crystals.[7] The result of such instability is the formation of hexagonal standing light patterns (hexagons). Both in Ref. 7 and in the work that followed,[16,17] a detailed study was made of emergent light structures. In particular, the possibility of controlled rotation and drift of hexagons was established. Honda and Matsumoto[18] reported on the formation of hexagons in $BaTiO_3$ crystals, which are similar to $KNbO_3$ in the type of photorefractive response.

Note that in all the experimental conditions described above, the formation of secondary light structures was related to the diffraction of light by refractive-index reflection gratings with a period close to half the wavelength of light in the medium. Transmission gratings with large periods, formed by accompanying waves, are extremely weak here.

Honda[7] was the first to study photorefractive transverse instability theoretically. The case of dominant rerfractive-index transmission gratings was studied, which had little relation to experiments. Next, in Refs. 9 and 10 the threshold

 1063-7761/97/050881-07$10.00

conditions for transverse instability were derived independently for dominant reflection gratings. Local photorefractive response was studied in Ref. 10, while the studies in Ref. 9 were of a more general nature. The application of the results of Saffman *et al.*[9] to the case of nonlocal response revealed a total lack of instability, which was in sharp contrast to the results of the experiments with $KNbO_3$ and $BaTiO_3$. Recently it was shown[11] that the previous work in Refs. 9 and 10 contains an elementary error, i.e., a number of important contributions related to the transverse modulation of the light intensity were omitted from the initial system of equations for the weak secondary waves. Honda and Banerjee[11] derived a general system of equations for the amplitudes of weak waves that allow for all important contributions. They applied it successfully in analyzing the transverse instability threshold in $KNbO_3$ with nonlocal photorefractive response.

The threshold conditions for instability were obtained in Ref. 11 with two restrictions: (a) the reflection of light from the feedback mirror positioned behind the crystal is complete, and (b) the waves are strictly frequency degenerate. Condition (a) is important for media with nonlocal photorefractive response. The point here is that the pump beams in such media exchange energy due to diffraction by the refractive-index reflection grating shifted by $\pi/2$ in relation to the light intensity distribution.[12] In this way the contrast of the initial interference light pattern varies in space. Studies of the stability of such a spatially inhomogeneous initial state are extremely difficult, however. For media with nonlocal response, condition (a) realizes the only known special case where the contrast of the interference light pattern is constant. Condition (b) involves not the experimcental conditions but the type of solution corresponding to instability. We believe that this condition is justified for media with nonlocal photorefractive response, since the secondary light structures observed with the experiment are immobile.

The goal of this paper is to correct the error we made in Ref. 10 and to study threshold conditions for the emergence of transverse instability in a medium with local photorefractive response. This is most important for $LiNbO_3$ and $LiTaO_3$ crystals, where the photovoltaic effect is the dominant transport mechanism.[13] The case of local response is also interesting from the theoretical viewpoint. Energy exchange between pump beams is absent here, so that stability of the initial state of the light field can be studied without restrictions on the experimental conditions. In particular, the pump beam intensity ratio can be assumed arbitrary. On the other hand, there is no reason to assume that the secondary light patterns are of the standing type in the case of local response. The frequency detunings between the side waves and the pump waves at the instability threshold are obtained through calculations.

2. THEORETICAL PREREQUISITES

The geometry of the problem is given in Fig. 1. There are two counterpropagating pump waves, α and β, and two symmetric pairs of weak side waves, γ, γ' and δ, δ'. All waves are assumed to have the same polarization. We also assume that the pump waves have equal frequencies ω, i.e., the interference pattern formed by the waves is of the standing type. Generally, the side wave frequencies differ from ω. The corresponding frequency detunings, however, cannot considerably exceed the reciprocal relaxation time of the space-charge field, t_d^{-1}, since otherwise, as shown in Ref. 12, photorefractive nonlinearity becomes ineffective. In the experiment the important detuning values do not exceed $(10^2-10^3)\ s^{-1}$. Such small detunings have little effect on the wavelengths of the light.

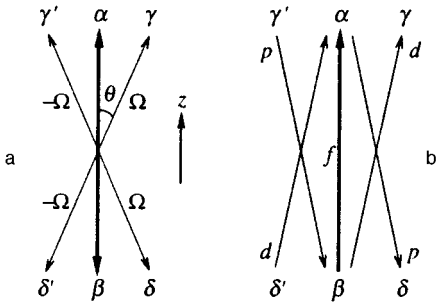

FIG. 1. Geometry of the problem: a—wave vectors of the pump waves (α, β) and the side waves (γ, γ' and δ, δ'); b—vectors of the spatial reflection gratings f, p, and d. The frequency detunings for the side waves are $\pm\Omega$.

The waves α, β, γ, γ', δ, and δ' form three reflection gratings of the space-charge field (see Fig. 1b); the fundamental grating f, with spatial frequency $\mathbf{K}_f=\mathbf{k}_\alpha-\mathbf{k}_\beta$, written by the pump beams, and gratings p and d, with vectors $\mathbf{K}_p=\mathbf{k}_\delta-\mathbf{k}_\alpha=\mathbf{k}_\beta-\mathbf{k}_{\gamma'}$ and $\mathbf{K}_d=\mathbf{k}_\gamma-\mathbf{k}_\beta=\mathbf{k}_\alpha-\mathbf{k}_{\delta'}$, formed by the weak side waves. Here we ignore reflection gratings written by pairs of weak waves and transmission gratings. The weakness of the latter in experiments with $LiNbO_3$ is due to the special features of electron transport.[12,13]

The slow wave amplitudes $A_j (j=\alpha,\beta,\gamma,\gamma',\delta,\delta')$ considered as functions of position z and time t vary because of the diffraction of light by the spatial gratings. If by $E_{f,p,d}$ we denote the complex-valued amplitudes of the space-charge field at frequencies $\mathbf{K}_{f,p,d}$, then according to the results of Ref. 10 we have the following expressions for the side wave amplitudes in the paraxial approximation:

$$\begin{aligned}
&\left(\frac{d}{dz}+i\Delta\right)A_\gamma=-is(E_dA_\beta+E_fA_\delta),\\
&\left(\frac{d}{dz}+i\Delta\right)A_{\gamma'}=-is(E_p^*A_\beta+E_fA_{\delta'}),\\
&\left(\frac{d}{dz}-i\Delta\right)A_\delta=is(E_pA_\alpha+E_f^*A_\gamma),\\
&\left(\frac{d}{dz}-i\Delta\right)A_{\delta'}=is(E_d^*A_\alpha+E_f^*A_{\gamma'}),
\end{aligned}\tag{1}$$

where $s=\pi n^3 r/\lambda$, and $\Delta=\pi n\theta^2/\lambda$, with n the refractive index, λ the vacuum wavelength of the light, r the corresponding component of the electrooptical tensor, and θ the propagation angle for weak waves in the crystals (Fig. 1a). As applied to lithium niobate and lithium tantalate crystals, $n=n_0$ is the ordinary refractive index, and $r=r_{13}$ is the tabu-

lated electrooptical constant. The parameter $\Delta(\theta)$ is simply the difference $k-|k_z|$ for the wave vectors of the side waves (see also Ref. 10).

The pump wave amplitudes are not constants. They change because of diffraction by the grating f,

$$\frac{dA_\alpha}{dz}=-isE_fA_\beta, \quad \frac{dA_\beta}{dz}=isE_f^*A_\alpha. \tag{2}$$

Equations (1) and (2) for the wave amplitudes do not contain time derivatives, since in the case of photorefractive nonlinearity they are small because of the large inertia of the space-charge field.[12,13]

Equations (1) and (2) for the wave amplitudes are not a complete set. They must be augmented by relationships expressing the grating amplitude $E_{f,p,d}$ in terms of wave amplitudes A_j. To obtain these expressions we use the following starting relationship for the space-charge field E:

$$\frac{\partial E}{\partial \tau}+\widetilde{IE}=E_{ph}\widetilde{I}, \tag{3}$$

where $\tau=t/t_d$ is dimensionless time, I is the light intensity normalized to the total intensity of the pump beams, E_{ph} is the characteristic photoinduced field, and the tilde stands for the spatially oscillating part of the particular quantity. For lithium tantalate and lithium niobate, E_{ph} is the photovoltaic field (an important characteristic of the crystal), and t_d is the dielectric relaxation time, which is inversely proportional to the pump intensity. The fact that the field E_{ph} in (3) is real guarantees that the photorefractive response is local, i.e., that there is no phase shift between the standing intensity grating and the distribution of the space-charge field.

Equation (3) implies that the amplitude of the standing grating f induced by the pump waves α and β is

$$E_f=E_{ph}\frac{A_\alpha A_\beta^*}{|A_\alpha|^2+|A_\beta|^2}. \tag{4}$$

In deriving (4) we ignored higher spatial harmonics with wave vectors that are multiples of $\mathbf{K}_f$. As is known, even if $|A_\alpha|^2=|A_\beta|^2$ and we allow for such harmonics, they provide no significant contribution to E_f. In accordance with Eqs. (2) and (4), the pump wave intensities $|A_{\alpha,\beta}|^2$ are independent of position z.

Generally, the reflection gratings p and d are not of the standing type. Their amplitudes can be found in the leading (linear) approximation from the side wave amplitudes. As Eq. (3) implies, there are two types of contribution to the amplitudes E_p and E_d. First, to obtain the p and d harmonics we must allow for modulation of intensity I at the spatial frequencies $\mathbf{K}_p$ and $\mathbf{K}_d$ on the right-hand side of Eq. (3) and ignore intensity modulation in the second term on the left-hand side. Contributions of this type were taken into account in Refs. 9 and 10. Second, we can take into account the fundamental harmonic of E in the second term and at the same time allow for transverse intensity modulation at the spatial frequency $\mathbf{K}_t=\mathbf{K}_d-\mathbf{K}_f\equiv\mathbf{K}_f+\mathbf{K}_p$ (Fig. 1b). Contributions of this type were ignored in Refs. 9 and 10 but were taken into account (for the case of strict frequency degeneracy) in Ref. 11. Clearly, the amplitudes of the intensity gratings at the spatial frequencies $\mathbf{K}_{p,d,t}$ are

$$I_p=\frac{A_\alpha^*A_\delta+A_\beta A_{\gamma'}^*}{|A|_\Sigma^2}, \quad I_d=\frac{A_\alpha A_{\delta'}^*+A_\beta^*A_\gamma}{|A|_\Sigma^2},$$
$$I_t=\frac{A_\alpha^*A_\gamma+A_\alpha A_{\gamma'}^*+A_\beta^*A_\delta+A_\beta A_{\delta'}^*}{|A|_\Sigma^2}, \tag{5}$$

where $|A|_\Sigma^2=|A_\alpha|^2+|A_\beta|^2$. Since the amplitude E_f is independent of t, the time dependence of $I_{p,d,t}$ fully determines the time dependence of the amplitudes $E_{p,d}$. Equations (5) readily suggest that only when the frequency detunings of the side waves, $\Omega_j=\omega-\omega_j$, satisfy the conditions

$$\Omega_\gamma=\Omega_\delta=-\Omega_{\gamma'}=-\Omega_{\delta'}\equiv\Omega, \tag{6}$$

each of the gratings p and d consists of a single (traveling) component, and each side wave contains only one temporal harmonic (see also Fig. 1). Assuming that these conditions are met and that $E_{p,d}\propto\exp(-i\Omega t)$, we see that Eq. (3) yields

$$E_p=\frac{E_{ph}I_p-E_f^*I_t}{1+i\nu}, \quad E_d=\frac{E_{ph}I_d-E_fI_t}{1+i\nu}, \tag{7}$$

where $\nu=\Omega t_d$. The dimensionless detuning ν can be interpreted as an internal degree of freedom for the side waves.

Plugging (4) and (7) into (1) and allowing for (5), we arrive at a closed system of linear equations for the side wave amplitudes. This system can be additionally simplified by going over to normalized wave amplitudes, $a_{\gamma,\gamma'}=A_{\gamma,\gamma'}A_\alpha^{-1}$ and $a_{\delta,\delta'}=A_{\delta,\delta'}A_\beta^{-1}$. Allowing for phase modulation of the amplitudes $A_{\alpha,\beta}$ via Eqs. (2) and doing simple calculations, we arrive at the following system of homogeneous linear equations with constant coefficients:

$$\left(\frac{d}{dz}+i\Delta\right)a_\gamma=ig(V_\beta a_\gamma+Va_{\gamma'}^*-V_\beta a_\delta-Va_{\delta'}^*),$$
$$\left(\frac{d}{dz}-i\Delta\right)a_{\gamma'}^*=-ig(Va_\gamma+V_\beta a_{\gamma'}^*-Va_\delta-V_\beta a_{\delta'}^*),$$
$$\left(\frac{d}{dz}-i\Delta\right)a_\delta=ig(V_\alpha a_\gamma+Va_{\gamma'}^*-V_\alpha a_\delta-Va_{\delta'}^*), \tag{8}$$
$$\left(\frac{d}{dz}+i\Delta\right)a_{\delta'}^*=-ig(Va_\gamma+V_\alpha a_{\gamma'}^*-Va_\delta-V_\alpha a_{\delta'}^*).$$

Here $g=\pi n^3rE_{ph}/\lambda$ is the couping constant, and V, V_α, V_β are dimensionless parameters:

$$V=\frac{m^2}{4(1+i\nu)}, \quad V_{\alpha,\beta}=\frac{m^2}{4}+\frac{i\nu}{1+i\nu}I_{\alpha,\beta}^2, \tag{9}$$

where $I_{\alpha,\beta}$ are the normalized pump wave intensities $(I_\alpha+I_\beta=1)$, and $m=2\sqrt{I_\alpha,I_\beta}$ is the original interference pattern contrast.

The system of equations (8) couples four amplitudes: a_γ, $a_{\gamma'}^*$, a_δ, and $a_{\delta'}^*$. The presence of at least one side wave means that the other three are also present. In the case of strict frequency degeneracy $(\nu=0)$ we have $V=V_{\alpha,\beta}=m^2/4$; the system of equations (8) coincides with the one

obtained in Ref. 11 for a real coupling constant. When the pump wave intensities are equal, $I_\alpha = I_\beta = 1/2$, with $m = 1$, we have

$$V = \frac{1}{4(1+i\nu)}, \quad V_\alpha = V_\beta = \frac{1+2i\nu}{4(1+i\nu)}. \tag{10}$$

The sign of the coupling constant g in (8) depends on the type of crystal and experimental conditions. The case $g>0$ corresponds to a defocusing nonlinearity, and the case $g<0$ to a focusing nonlinearity. In $LiNbO_3$ and $LiTaO_3$ with dominant photovoltaic transport, the nonlinearity is of the defocusing type, $g>0$.

Let us examine some properties of (8). At $\nu=0$ and $m=1$ the directions $+z$ and $-z$ prove physically equivalent. As a result of this spatial symmetry, the system (8) allows for solutions in the form of symmetric (S) and antisymmetric (A) modes, $a_{\gamma,\gamma'}(z) = \pm a_{\delta,\delta'}(-z)$. As shown later, this symmetry property simplifies the structure of the threshold equations for transverse instability considerably. For $\nu \neq 0$ and $m \neq 1$ the property of spatial symmetry is lost.

Note that in the case of strict frequency degeneracy ($\nu=0$) each of the parameters V, V_α, and V_β is proportional to m^2. However, when $\nu \neq 0$, only two of these parameters tend to zero as m decreases. This means that for $m^2 \ll 1$ the mutual coupling of the waves γ, γ', δ, and δ' strengthens because a frequency detuning is introduced. As shown later, this leads to the formation of traveling light patterns near the transverse instability threshold.

3. THRESHOLD CONDITIONS

3.1. General considerations

A threshold equation for transverse instability can be obtained from the condition that there can be nonzero output amplitudes of the side waves for zero input amplitudes. In other words, the equation corresponds to a situation in which the transmission coefficient for an infinitely weak signal (say, the wave γ) at the input becomes infinitely strtong.

The possibility for transverse instability to manifest itself is related solely to the fact that the boundary conditions for the waves γ, γ' and δ, δ' are fixed on the opposite faces of the crystal. To be definite, let us suppose that the entrance and exit faces are at $z=-l/2$ and $z=l/2$, so that l is the thickness of the nonlinear layer. Then, by solving the system of homogeneous linear differential equations (8) we can express $a_j(l/2)$ with $j=\gamma,\gamma',\delta,\delta'$ in terms of $a_j(-l/2)$ via a fourth-order matrix $\hat{T}(l)$. Next we must express algebraically the exit values $a_{\delta,\delta'}(-l/2)$ in terms of the entrance values $a_{\delta,\delta'}(l/2)$. The transmission coefficient for weak waves becomes infinite if the second-order determinant $T_{\delta\delta}T_{\delta'\delta'} - T_{\delta\delta'}T_{\delta'\delta}$ vanishes. A similar procedure for finding the threshold condition of cavityless lasing has been used earlier for a number of six-beam configurations (see, e.g., Refs. 19 and 20).

The type of threshold equation depends on the type of boundary condition. Below we discuss only vanishing boundary conditions:

$$a_{\gamma,\gamma'}(-l/2) = 0, \quad a_{\delta,\delta'}(l/2) = 0. \tag{11}$$

This type is most easily realized in experiments.

Generally the threshold condition is complex-valued and includes external parameters (the crystal thickness l and the pump beam intensity ratio $\rho = I_\beta/I_\alpha$) and internal parameters, which allow for automatic adjustment of the secondary waves (the propagation angle θ and the dimensionless detuning ν). Solution of the threshold equation yields the functions $l = l(\rho,\theta)$ and $\nu = \nu(\rho,\theta)$. Minimization of l as a function of θ yields the threshold value $l_{th}(\rho)$ and the threshold values of the emission angle and detuning, $\theta_{th}(\rho)$ and $\nu_{th}(\rho)$. When finding the threshold values one must bear in mind that the solution of the threshold equation may consist of a sequence of branches for $l(\theta,\rho)$ and $\nu(\theta,\rho)$.

3.2. Standing light patterns

What is important is that for strict frequency degeneracy the threshold equation becomes real. This means that it allows for the solution $\nu=0$ and $l=l(\theta,\rho)$. As noted in Sec. 2, at $\nu=0$ the system of equations (8) is symmetric under the $z \to -z$ transformation and a simultaneous permutation of indices $\gamma,\gamma' \leftrightarrow \delta,\delta'$. As a result the system is invariant under the substitution of ρ^{-1} for ρ. The vanishing boundary conditions (11) are also invariant under reflection in the $z=0$ plane. For this reason the threshold condition can be written as

$$S(l,\theta,m)A(l,\theta,m) = 0, \tag{12}$$

i.e., it splits into two equations, $S(l,\theta,m)=0$ and $A(l,\theta,m)=0$, where the first corresponds to the symmetric mode and the second to the antisymmetric. Calculations lead to the following expressions for the factors S and A:

$$\begin{aligned} S &= \Delta \sin\frac{\Delta l}{2}\sinh\frac{\Gamma l}{2} + \Gamma \cos\frac{\Delta l}{2}\cosh\frac{\Gamma l}{2}, \\ A &= \Delta \cos\frac{\Delta l}{2}\cosh\frac{\Gamma l}{2} - \Gamma \sin\frac{\Delta l}{2}\sinh\frac{\Gamma l}{2}, \end{aligned} \tag{13}$$

where

$$\Gamma = \sqrt{\Delta g m^2 - \Delta^2}. \tag{14}$$

The factors S and A are real for all values of the angular parameter Δ and any sign of the coupling constant g. At the same time, their structure depends on the sign of the difference $gm^2 - \Delta$. If the difference is negative, the hyperbolic functions of $\Gamma l/2$ in (13) become the corresponding trigonometric functions. Note that the pump wave intensity ratio ρ enters into (13) through the interference pattern contrast $m = 2/(\rho^{1/2} + \rho^{-1/2})$. This fact is a direct consequence of the spatial symmetry of the problem mentioned earlier.

Clearly, the parameters g, l, and m enter into (13) as two dimensionless combinations:

$$x = \frac{\Delta}{gm^2}, \quad y = lgm^2. \tag{15}$$

This simplifies the analysis of threshold conditions. If we know the values x_{th} and y_{th} corresponding to the absolute minimum of the function $y(x)$, we use (15) and immediately arrive at the dependence of l_{th} and θ_{th} on m and g. We note,

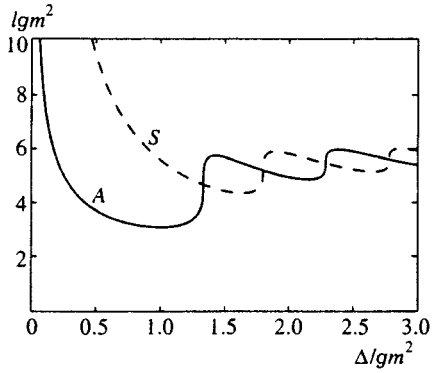

FIG. 2. The parameter $y=lgm^2$ as a function of $x=\Delta/gm^2$ for the antisymmetric (solid curve) and symmetric (dashed curve) modes.

however, that the $l_{th} \propto m^2$ and $\theta \propto m$ dependences that follow from (15) are sharper than those found in Ref. 10.

Each of the equations $S(x,y)=0$ and $A(x,y)=0$ produces a sequence of branches for the functions $y(x)$. Branches belonging to one symmetry (S or A) never intersect, but those belonging to different symmetries (S and A) do. The two lowest branches $y(x)$ corresponding to S- and A-modes are the most interesting. Their points of intersection can be found from the condition that $A(x,y)=S(x,y)=0$. The answer is

$$x_j=\frac{(j+1)^2}{2j+1}, \quad y_j=\pi\frac{2j+1}{j+1}, \tag{16}$$

where j is a positive integer.

We can now easily see that for $g>0$ the equation $A(x,y)=0$ has a solution $y=\pi$ at $x=1$, with this point corresponding to a local minimum of the function $y(x)$. Moreover, numerical calculations show that the particular solution established above corresponds to the absolute minimum of $y(x)$. Hence for the threshold values l_{th} and θ_{th}, we have

$$l_{th}=\frac{\pi}{gm^2}, \quad \theta_{th}=mn\sqrt{rE_{ph}}. \tag{17}$$

Note that at $m=1$, the angle θ_{th} coincides with the value found in Ref. 10, and the thickness l_{th} exceeds the previous value by a factor of $\sqrt{2}$.

Figure 2 depicts the two lower branches of $y(x)$ corresponding to the A- and S-modes. They fully agree with the above properties and are qualitatively similar in structure to the branches obtained in Ref. 10. Note that each branch of $y(x)$ exhibits a sequence of minima, with the values of y at adjacent minima being close. This fact is important for understanding the nonlinear stage in transverse instability, i.e., the structure of the emergent secondary light patterns.

3.3. Traveling light patterns

Let us assume that the dimensionless detuning ν is nonzero. We begin with the case of equal pump wave intensities, $\rho=1$. Here the system (8) still allows for a solution in the form of A- and S-modes, while the threshold condition may be written in the form (12). Calculations lead to new expressions for the factors A and S:

$$\begin{aligned} S&=\Delta\,\sin\frac{\kappa_+ l}{2}\sin\frac{\kappa_- l}{2}+\frac{\kappa_+\kappa_-}{\Delta}\cos\frac{\kappa_+ l}{2}\cos\frac{\kappa_- l}{2},\\ A&=\Delta\,\cos\frac{\kappa_+ l}{2}\cos\frac{\kappa_- l}{2}+\frac{\kappa_+\kappa_-}{\Delta}\sin\frac{\kappa_+ l}{2}\sin\frac{\kappa_- l}{2}, \end{aligned} \tag{18}$$

where

$$\kappa_+=\sqrt{\Delta(\Delta-\epsilon g)}, \quad \kappa_-=\sqrt{\Delta(\Delta-g)}, \tag{19}$$

with $\epsilon=i\nu/(1+i\nu)$. At $\nu\neq 0$ the factors A and S are complex-valued, while at $\nu=0$ they are given by the previous formulas (13) if we set m to unity. The first problem that emerges in studying the threshold equation $A=0$ and $S=0$ is whether these equations have at least one solution $l=l(\Delta)$ and $\nu=\nu(\Delta)$ with $\nu\neq 0$. Numerical calculations reveal a lack of such solutions. In other words, at $m=1$ the light patterns emerging near the instability threshold are of the standing type.

Investigation of the instability threshold in the general case of $\nu\neq 0$ and $\rho\neq 1$ entails enormous difficulties in view of cumbersome calculations resulting from the lack of spatial symmetry. Such an investigation can probably only be numerical. But when the pump wave intensities differ considerably, the problem can be simplified. Furthermore, from general considerations it follows that in this limit traveling light patterns correspond to a lower threshold than standing patterns. This becomes clear if we again turn to the original system (8) and assume, for the sake of definiteness, that $\rho=I_\beta/I_\alpha\ll 1$. From (9) it follows that the dimensionless parameters V, V_α, and V_β are

$$V\simeq\frac{\rho}{1+i\nu}, \quad V_\alpha\simeq\frac{i\nu}{1+i\nu}, \quad V_\beta\simeq\rho. \tag{20}$$

Only the parameter V_α remains finite as $\rho\to 0$. On the other hand, the structure of the system of equations (8) implies that for $V_\alpha\gg V$, V_β the coupling between pairs of counterpropagating waves, which is needed for instability, becomes much stronger that for $V_\alpha=0$ (i.e., at $\nu=0$).

To find the threshold condition we must calculate each of the matrix elements $T_{\delta\delta}$, $T_{\delta'\delta'}$, $T_{\delta\delta'}$, and $T_{\delta'\delta}$, which couple the amplitudes a_δ and $a_{\delta'}$, in the leading approximation in ρ. This can easily be done by employing the Laplace transformation in the coordinate z. As a result we arrive at the threshold equation

$$\cos[2l(\Delta-g\epsilon)]=\frac{2\Delta-g\epsilon}{2g^2\rho\Delta}\left(\frac{\Delta-g\epsilon}{1-\epsilon}\right)^2, \tag{21}$$

where $\epsilon=i\nu(1+i\nu)^{-1}$, in accordance with the adopted notation.

Equation (21) is complex-valued. Its solution consists of a sequence of branches for $l(\Delta)$ and $\nu(\Delta)$. Figure 3 depicts the two lowest branches for $l(\Delta)$ and the corresponding branches for $\nu(\Delta)$. We see that the function $l(\Delta)$ for each branch is characterized by a single smooth minimum and the dimensionless detuning ν corresponding to this minimum is of order unity. The minimum values of l for the branches *1* and *2* are close to each other, but correspond to substantially different values of the angular parameter Δ. Calculations show that for each value of ρ, the threshold, i.e., minimum,

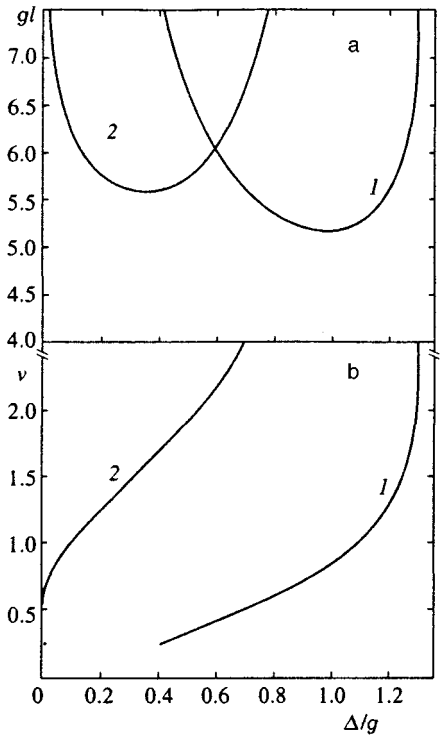

FIG. 3. (a) Two lowest branches (curves *1* and *2*) of the function $l(\Delta)$ at $\rho=0.1$. (b) The corresponding branches of the function $\nu(\Delta)$.

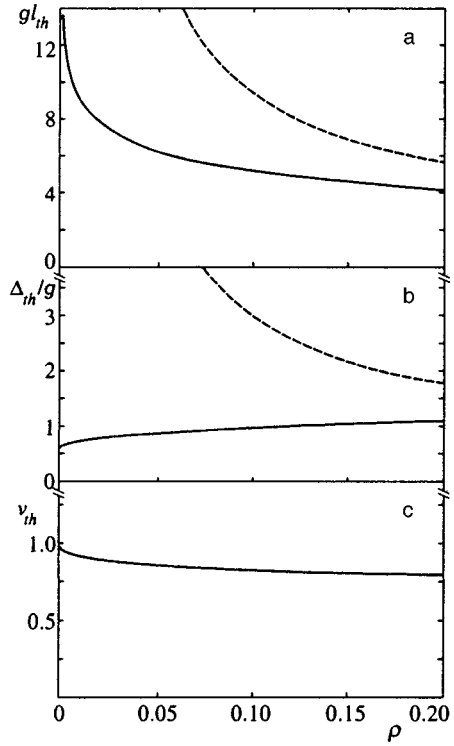

FIG. 4. The threshold values of (a) the thickness l_{th}, (b) the angular parameter $\Delta_{th}=\pi n\theta_{th}^2/\lambda$, and (c) the dimensionless detuning $\nu_{th}=t_d\Omega_{th}$ as functions of the pump wave intensity ratio ρ. The solid curves correspond to traveling light patterns and the dashed curves to standing light patterns.

value of the thickness l_{th} corresponds to the right-hand branch of $l(\Delta)$ (branch *1* in Fig. 3a). Figure 4 depicts l_{th}, Δ_{th}, and ν_{th} as functions of ρ. As ρ decreases, the threshold thickness l_{th} grows monotonically, tending to infinity logarithmically, while Δ_{th} and ν_{th} remain finite, with the former slowly increasing and the latter slowly decreasing.

The dashed curves in Figs. 4a and 4b depict the functions $l_{th}(\rho)$ and $\Delta_{th}(\rho)$ at $\nu=0$ corresponding to Eqs. (17). Clearly, over the entire range of ρ considered here, the threshold values l_{th} and Δ_{th} for traveling light patterns are smaller than for standing light patterns. The intersections between the solid and dashed curves are outside this range, where the adopted approximation of small values of ρ is probably inapplicable.

The velocity of the light pattern on a screen near the exit surface of the crystal is $v=\Omega/K_t$, where $K_t=2\pi n\theta/\lambda$. Using the definitions of g and Δ adopted earlier, we arrive at the following expression for this velocity at the instability threshold

$$v_{th}=\frac{1}{2\pi n^2\sqrt{rE_{ph}}}\frac{\lambda}{t_d}\frac{\nu_{th}}{\sqrt{\Delta_{th}/g}}. \tag{22}$$

The dimensionless combination $\nu_{th}/\sqrt{\Delta_{th}/g}$ slowly decreases as ρ grows, and remains close to unity for $\rho<0.2$.

4. DISCUSSION

Let us first estimate the main parameters of the theory as applied to experiments with $LiNbO_3$ crystals. Here the values of the photovoltaic field E_{ph} reach 10^2 kV/cm (see Ref. 13). Setting E_{ph} at 25 kV/cm, assuming that $n=n_0$ and $r=r_{13}$, and using the tabulated values $n_0\simeq 2.3$ and $r_{13}\simeq 8.6\times 10^{-10}$ cm/V, we arrive at the following estimates for the thickness l_{th} and the emission angle in air, $\theta_{th}^a=n_0\theta_{th}$, at equal pump beam intensities:

$$l_{th}=2\text{ mm},\quad \theta_{th}^a=1.5^\circ. \tag{23}$$

For the adopted values of the parameters, the coupling constant g is approximately 16 cm^{-1}. The estimates do not depend on the intensity of the light, and the low value of the threshold thickness suggests that the conditions for the excitation of standing and traveling light patterns can easily be met in the experiment.

To estimate the velocity of a light pattern, v_{th}, we must, additionally, know the space-charge field relaxation time t_d. In $LiNbO_3$ crystals t_d usually coincides with the dielectric relaxation time and is inversely proportional to the pump wave intensity. For $LiNbO_3$ samples with high E_{ph}, t_d at an intensity of 1 W/cm can roughly be estimated at 1 s. This, in accordance with (22), corresponds to a velocity v_{th} of order 10^{-3}cm/s. The accurate detection of such slow motion is improbable. This estimate means that laser beams in the experiments must be focused to a power density of order 10^2 W/cm^2.

As noted in the Introduction, we know of three experimental papers, Refs. 6, 14, and 15, which with high probability can be classified as reports on observations of transverse instability in $LiNbO_3$ crystals. The most detailed description of the observations can be found in Ref. 6. In this experiment one focused pump beam was directed at the crystal, while the counterpropagating beam emerged because of

the relatively weak effect of nonlinear reflection.[21,22] The intensity of the counterpropagating beam was not monitored, but it was certainly much lower than that of the incident beam. The ring light structures observed in the experiment were of the traveling type, and their velocity increased with pump wave intensity. The characteristic emission angle was 1.5–2°. Thus, we can claim qualitative agreement between the observed quantities and those calculated by the above theory.

We believe that it would be extremely interesting to conduct well-planned experiments on transverse instability in $LiNbO_3$ crystals, experiments similar in quality to those done by Banerjee *et al.*,[16] and Honda and Matsumoto[17,18] with $KNbO_3$ and $BaTiO_3$. The goal of such experiments would be to detect secondary light structures emerging near counterpropagating pump beams and to study the properties of these structures as functions of the intensity ratio ρ, the total light intensity, the photovoltaic field E_{ph}, and the sample thickness l. The existing preliminary observations and theoretical calculations suggest that the transverse instability threshold can easily be exceeded, and that above-threshold light structures are many and varied.

5. CONCLUSIONS

We have shown that when propagating in $LiNbO_3$ and $LiTaO_3$ crystals with dominant photovoltaic transport, counterpropagating laser beams become unstable against the formation of small-angle ($\theta^a \simeq 1°-2°$) light structures, starting at interaction lengths $l \simeq 1-2$ mm. As the original interference pattern contrast decreases, the excitation of standing light structures is replaced by the excitation of traveling structures. Published data suggest that it is easy to excite the discussed transverse instability in $LiNbO_3$:Fe and $LiNbO_3$:Cu crystals, and they agree qualitatively with the results of the theory. Thorough research on transverse instability (similar to the experiments involving $KNbO_3$ and $BaTiO_3$ crystals with dominant diffusive transport) in these nonlinear materials has yet to be conducted.

The authors are grateful to M. G. Stepanov for expert help with computer graphics. This work was sponsored by the Russian Fund for Fundamental Research (Grant No. 96-02-19126).

[1)]These are also known as the cases of real and imaginary coupling constants.

[1] W. J. Firth, A. Fitzgerald, and C. Pare, J. Opt. Soc. Am. B **7**, 1087 (1990).
[2] G. G. Luther and C. J. McKinstrie, J. Opt. Soc. Am. B **7**, 1125 (1990).
[3] G. Grynberg, Opt. Commun. **66**, 3231 (1988).
[4] J. Pender and L. Hesselink, J. Opt. Soc. Am. B **7**, 1361 (1990).
[5] A. Yariv and D. M. Pepper, Opt. Lett. **1**, 16 (1977).
[6] V. V. Lemeshko and V. V. Obukhovskiĭ, Pis'ma Zh. Tekh. Fiz. **11**, 1389 (1985) [Sov. Phys. Tech. Phys. Lett. **11**, 573 (1985)].
[7] T. Honda, Opt. Lett. **18**, 598 (1993).
[8] M. Saffman, D. Montgomery, A. A. Zozulja *et al.*, Phys. Rev. A **48**, 3209 (1993).
[9] M. Saffman, A. A. Zozulya, and D. A. Anderson, J. Opt. Soc. Am. B **11**, 1409 (1994).
[10] B. Sturman and A. Chernykh, J. Opt. Soc. Am. B **12**, 1384 (1995).
[11] T. Honda and P. P. Banerjee, Opt. Lett. **21**, 779 (1996).
[12] *Photorefractive Materials and Their Applications*, Vol.1: *Fundamental Phenomena*, P. Günter and J.-P. Huignard (eds.), Springer, Berlin (1988).
[13] B. I. Sturman and V. M. Fridkin, *The Photovoltaic Effect in Media Without a Symmetry Center and Related Phenomena* [in Russian], Nauka, Moscow (1992).
[14] V. Bazenov, S. Lyuksutov, R. Jungen *et al.*, Proc. SPIE **1273**, 48 (1990).
[15] S. M. Liu, G. Y. Zhang, J. L. Wang *et al.*, Opt. Commun. **70**, 185 (1989).
[16] P. P. Banerjee, H.-L. Yu, D. A. Gregory *et al.*, Opt. Lett. **20**, 10 (1995).
[17] T. Honda, Opt. Lett. **20**, 851 (1995).
[18] T. Honda and H. Matsumoto, Opt. Lett. **20**, 1755 (1995).
[19] A. D. Novikov, V. V. Obukhovskiĭ, S. G. Odulov, and B. I. Sturman, JETP Lett. **44**, 536 (1986).
[20] B. Sturman, S. Odulov, U. van Olfen *et al.*, J. Opt. Soc. Am. B **11**, 1700 (1994).
[21] I. F. Kanaev, V. K. Malinovskiĭ, and B. I. Sturman, Zh. Éksp. Teor. Fiz. **74**, 1599 (1978) [Sov. Phys. JETP **47**, 834 (1978)].
[22] K. R. MacDonald, J. Feinberg, M. Z. Zha, and P. Günter, Opt. Commun. **50**, 146 (1984).

Translated by Eugene Yankovsky

Multilevel rotational transitions in the intermediate stage of three-photon ionization of molecules

G. K. Ivanov, G. V. Golubkov, S. V. Drygin, and I. E. Cherlina

N. N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia
(Submitted 29 October 1996)
Zh. Éksp. Teor. Fiz. **111**, 1624–1632 (May 1997)

Ionization and dissociation of diatomic molecules induced by a weak field (after preliminarily populating an intermediate level) and by intense, linearly polarized monochromatic radiation have been studied. Field-induced mixing of rotational components of various electronic–vibrational states of molecules (such as CO, NO, etc.) at field strength $f \sim 10^{-4}$–10^{-5} atomic units can lead to migration among states with different angular momenta J. Therefore, ions with rotational momenta J^+ much higher than those prescribed by selection rules for three-photon absorption can be formed from molecules in the ground state. The possibility of selective formation of ions with $J^+ \gg 1$ and zero projection of the angular momentum on the polarization vector of the external electromagnetic radiation has been investigated. © *1997 American Institute of Physics.* [S1063-7761(97)00705-1]

1. INTRODUCTION

In this paper, we discuss a new class of phenomena related to the emergence of rotational structure in diatomic molecules subject to resonant multiphoton ionization and dissociation. The behavior of diatomic molecules in strong electromagnetic fields has attracted a lot of researchers' attention, primarily because these effects have numerous applications in various fields of modern physics.[1] Special attention has been focused on analyzing the feasibility of control over elementary chemical reactions through external radiation, and the generation of reaction products with specified properties.[2–4]

In recent years, several effects which show up most clearly in the ionization continuum have been detected. These include, above all, the above-threshold dissociation (ATD) reaction,[5–7] which is an analog of above-threshold ionization in atoms,[8] which has been studied in detail. Note that dissociation rate is appreciable at an external electromagnetic field intensity of order $I \sim 10^{13}$ W/cm^2. Furthermore, a process termed molecular bond softening (BS)[7,9] is dominant at higher intensities, $I \sim 10^{14}$ W/cm^2.

In addition to these phenomena, the effects of molecular stabilization[3,9–11] and dissociation of molecules aligned with the electromagnetic field polarization vector (molecule alignment) have been observed.[2,12] The former process occurs at high radiation intensities, $I \sim 10^{14}$ W/cm^2, and results from the formation of laser-induced bound states. The latter process is characterized by a maximum in the angular distribution of dissociation products in the direction of the polarization vector.

Theoretical methods of investigation of molecules in an electromagnetic field and recent experimental results were discussed in detail in the comprehensive review[13] and in the recent series of publications.[12,14,15]

Note that the role of molecular rotation, whose effect on elementary chemical reactions cannot be predicted *a priori*, has been widely discussed in the literature. In most cases, rotation can lead to a higher reaction rate because new reaction channels can be involved, owing to nonadiabatic coupling between electronic and rotational degrees of freedom. A similar situation occurs when molecules in excited electronic states are produced in an intermediate stage of a reaction. In some cases, rotation can lead to effective inhibition of a reaction. This occurs in resonant photoionization of diatomic molecules XY when laser-induced bound states overlapping with the dissociative continuum of X+Y states are produced.[16] In most publications on multiphoton ionization and dissociation, this issue has essentially been ignored,[17] since molecular spectra can be interpreted in terms of fixed molecular orientation owing to the small rotational constants of most molecules. Molecular rotation was taken into account by the authors of some recent publications,[12,14,15] but it is not important for the two-photon absorption discussed by them because it leads formally only to a modification of transition matrix elements in corresponding equations.

The situation is radically different in processes of three-photon (1+2) absorption described by the following chain of reactions:

$$\mathrm{XY}(i) \xrightarrow{\hbar\Omega} \mathrm{XY}^*(\rho_0 v_0 J_0) \xrightarrow{\hbar\omega_f} \mathrm{XY}^*(\rho_1 v_1 J_1) \xrightarrow{\hbar\omega_f} \begin{cases} \mathrm{XY}^+ + e^- \\ \mathrm{X+Y} \end{cases}, \tag{1}$$

where $|i\rangle$ is the initial state of the molecule XY. The states $|\rho_0 v_0 J_0\rangle$ can either be highly excited Rydberg states or low-lying electronic states. Rovibrational levels of the same group $(\rho_0 v_0 J_0)$ populated after absorption of a photon $\hbar\Omega$ of the weak (probing) field are coupled to the higher-lying group of levels $(\rho_1 v_1 J_1)$ by the intense electromagnetic field (here ρ_0 and v_0 are the electronic and vibrational quantum numbers, and J_0 is the total angular momentum of the molecule). If the coupling between vibrational components of these two groups (labeled 0 and 1) due to the field is comparable to the energy difference ΔE_J between rotational sublevels within these groups, many states will be involved in the process. The corresponding criterion is as follows:

 1063-7761/97/050888-05$10.00

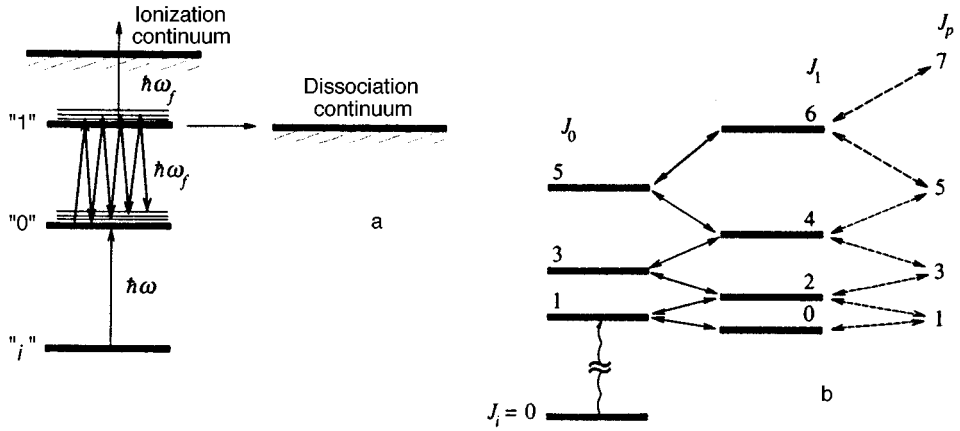

FIG. 1. (a) Diagram of intramolecular transitions; (b) groups of rotational levels of vibronic states ''0'' ($J_0=1, 3, 5$) and ''1'' ($J_1=0, 2, 4, 6$) resonant with the intense optical field. The level $J_0=1$ is populated by a photon of the probe optical field $\hbar\Omega$ from a state of the XY molecule. The ionization continuum corresponds to the states with total angular momentum $J_p=1, 3, 5, 7$ of the molecule.

$$V^{01}_{J_0J_1} \geqslant \Delta E_J,$$

where $V^{01}_{J_0J_1}$ is the matrix element of the field coupling between states ''0'' and ''1.'' If ρ_0 and ρ_1 are highly excited (Rydberg) states, the matrix element can be estimated by the simple analytic formula[18]

$$V^{01}_{J_0J_1} \sim \frac{f}{n_0^{3/2} n_1^{3/2} \omega_f^{5/3}}, \tag{2}$$

where f is the external field strength, and n_0 and n_1 are the principal quantum numbers of states ''0'' and ''1'' ($\hbar = m_e = e = 1$). For example, in the case of the atmospheric molecules N_2, NO, CO, and O_2, the parameter $\Delta E_J \sim 10^{-5}$ and 10^{-4} at $J=1$ and $J=10$, respectively. Therefore, starting with the values $n_0 \approx 5$ and $n_1 \approx 10$ and the external field frequency $\omega_f \sim 10^{-2}$, the condition (2) is satisfied at field strengths as low as $f \sim 10^{-5}$–10^{-4}. Under these conditions, the two-level resonant approximation does not apply, and more complex schemes of multilevel transitions must be used. An important point is that the intermediate states of the process are close to the boundary of the ionization continuum of the states $e^- + XY^+$ and can predissociate; the interacting resonant states have common (correlated) continua of final states. Therefore, introduction of phenomenological parameters for description of level broadening widely used in multilevel resonant approximation cannot be used in this case. The solution of the formulated problem is described below in terms of the stationary method of radiative-collision matrix,[19–21] which uses information about coupling between bound states and amplitudes of direct transitions between bound and free states. Since the process is described in terms of transition amplitudes, the contributions of correlated and uncorrelated continua of final states to effective interactions between unstable states can be consistently taken into account. This allows us to investigate in detail features of photoionization spectra and the shape of the distribution of produced ions over the angular momentum J^+. The application of the theory is illustrated by taking CO molecule as an example.

2. MULTILEVEL ROTATIONAL TRANSITIONS IN THREE-PHOTON IONIZATION OF MOLECULES

Let us assume that the initial molecule XY in the J_iM state is simultaneously exposed to two monochromatic sources of linearly polarized light, namely, a weak source emitting light with a frequency Ω and a powerful one at a frequency ω_f. As a result of multiple absorption with resonant excitation of two groups of intermediate rovibrational states ''0'' and ''1,'' the species $e^- + XY^+$ or $X + Y$ are produced. A diagram of the studied reaction described by Eq. (1) is given in Fig. 1.

Let us calculate the amplitude of the transition from the initial state of the molecule XY with the quantum numbers $\Lambda v_i J_i M$ to the final state $e^- + XY^+$ characterized by the quantum numbers $\Lambda_p v_p J_p M$ with the quantization axis aligned with the vector $\mathbf{f}$ (here v_i and v_p identify the vibrational states of XY and XY^+, respectively, and Λ is the projection of the electron angular momentum on the molecular axis). Supposing that LS-coupling takes place in the molecule, let us apply, as usual, the approximation adiabatic with respect to nuclear rotation. In accordance with the sum rules, the total angular momentum J_p and its projection M correlate in the usual manner with the angular momenta of the ion (J^+, M^+) and of the emitted photoelectron (l, m). Since the radiation of the frequency Ω is weak by definition ($f_\Omega \ll f$), the transition amplitude M_{ip} can be expressed as

$$M_{ip} = \langle i | \mathbf{D} | \Psi_p \rangle, \tag{3}$$

where $\mathbf{D}$ is the dipole-moment operator. The wave function Ψ_p should be constructed with account of the field-induced coupling between the groups of states ''0'' and ''1.'' Therefore, we use the technique[19,20] in which the wave functions of the continuum are expressed in terms of elements of the $\mathbf{T}$-matrix of radiative collisions, which describes the transitions in the system $e^- + XY^+$ taking into account the change in the number of photons of the external field. The transition rate is high when the distance between the electron and ion is

small, since our model presumes that separated species e^- and XY^+ do not interact with the external field. The latter condition holds provided that

$$\xi = f\omega_f^{-5/3} \ll 1. \tag{4}$$

This condition allows us to construct a wave function taking into account the coupled states ''0'' and ''1'' without using the perturbation theory.

The ionization amplitude can be expressed as[20]

$$M_{ip} = D_{i0}\frac{1}{E-E_0}T_{0p}, \tag{5}$$

where $D_{i0} = \langle i|\mathbf{D}|0\rangle$ is the dipole matrix element of the transition $(i \to 0)$, E is the system energy after absorption of the photon Ω, and E_0 is the unperturbed energy of the ''0'' state measured with respect to the ground state of the XY^+ ion. The elements T_{0p} are derived from the equation for the radiative collision matrix

$$\mathbf{T} = \mathbf{T}' + \mathbf{T}'\sum_{s,k}\frac{|s_k\rangle\langle s_k|}{E-E_s-k\omega_f}\mathbf{T}, \tag{6}$$

where the resonant discrete (or quasi-discrete) states are explicitly separated (in this specific case, these are ''0'' and ''1''). Nonresonant terms are included in the weakly energy-dependent matrix $\mathbf{T}'$, whose elements, if the condition (4) holds, can be expressed as

$$T'_{s_k s_{k'}} = t_{s_k s_{k'}} - i\sum_{p,k''} t_{s_k p_{k''}} t_{p_{k''} s_{k'}}. \tag{7}$$

The parameters $t_{s_k s_{k'}}$ with $k' = k \pm 1$ characterize the direct field-induced coupling between the states s_k and $s_{k'}$, which have different numbers of photons in the system ($k<0$ corresponds to a decrease in the number of photons, i.e., absorption). In order to characterize the partial amplitudes $t_{s_k,p}$, which describe the transition from the intermediate $|s_k\rangle$ state to the p-continuum, the equations include, in addition to the quantum numbers J_p, Λ_p and M, the index k, which identifies the change in the number of photons in the system. These amplitudes are related, as usual, to the total width of the level:

$$\Gamma_{s_k} = 2\gamma_{s_k} = 2\sum_p |t_{s_k,p}|^2, \tag{8}$$

which includes, accordingly, natural linewidths due to predissociation of the molecule XY.

3. FIELD-INDUCED POPULATION OF HIGH ROTATIONAL STATES OF IONS IN MULTIPHOTON IONIZATION

Let us investigate the possibility of populating high rotational states in the process of multiphoton ionization. It is sufficient to limit our discussion to a simple seven-level scheme of three-photon absorption, $\hbar\Omega + 2\hbar\omega_f$, whose diagram is given in Fig. 1b. Then the transition of the molecule XY from the lowest rotational state with $J_i = 0$ occurs under the conditions when groups of rotational sublevels with $J_0 = 1, 3, 5$ and $J_1 = 0, 2, 4, 6$ of the states ''0'' and ''1'' identified with the two respective electronic–vibrational (vibronic) levels $E_{\rho_0 v_0}$ and $E_{\rho_1 v_1}$ ($E_{\rho_0 v_0} \approx E_{\rho_1 v_1} - \omega_f$) are formed during the intermediate stage, owing to the field-induced coupling. An important point is that at a field strength $f \sim 10^{-4}$ the contribution of states with higher J is negligible because the matrix elements V_{01} are much smaller than the distance to the nearest level not included in our scheme. In our model, the working level (i.e., the one previously populated by the weaker radiation at the frequency Ω from the initial state of the molecule XY) is the vibronic level ($\rho_0 v_0 J_0 = 1$). The field-induced coupling between the groups of vibrational states ''0'' and ''1'' is shown in Fig. 1b by solid arrows, and the exchange with photons between the state ''1'' and continuum is shown by dashed arrows. As noted above, rotational sublevels of the groups ''0'' and ''1'' corresponding to different J can be, generally speaking, predissociative levels with natural widths $\Gamma_e = 2|t_{s_k,d}|^2$.

Above all, we wish to study the feasibility of formation and separation of ions in states with high angular momentum J^+. With this end in view, we ignore nonadiabatic vibronic coupling, which has little effect on the essential features of the process and only shifts the resonant levels in the resulting ionization spectrum. In our calculation, we take as an example the CO molecule, whose optical and electron spectra have been studied in detail.[22–24]

In selecting the intermediate states ''0'' and ''1,'' we consider two simple and illustrative physical situations. In the first case, these states are the lower levels of p and s Rydberg series with the effective principal quantum numbers $\nu_0 = 3.321$ and $\nu_1 = 5.123$. In the second case, which is quite different, multiphoton absorption populates s and p series with quantum numbers $\nu_0 = 4.099$ ($s\sigma$ series) and $\nu_1 = 6.337$ ($p\sigma$ series), respectively. For definiteness, let us analyze the following two schemes of optical transitions, assuming that the vectors of the weaker and stronger fields f_Ω and f are aligned with each other:

$$X^1\Sigma^+(J_i=0, v_i=0) \xrightarrow{\Omega} {}^1\Sigma^+(4p\sigma, J_0, v_0=1)$$
$$\overset{\omega_f}{\rightleftarrows} {}^1\Sigma^+(6s\sigma, J_1, v_1=1) \xrightarrow{\omega_f} \text{continuum}(J_p, v_p=1), \tag{9}$$

$$X^1\Sigma^+(J_i=0, v_i=0) \xrightarrow{\Omega} {}^1\Sigma^+(5s\sigma, J_0, v_0=1)$$
$$\overset{\omega_f}{\rightleftarrows} {}^1\Sigma^+(7p\sigma, J_1, v_1=1) \xrightarrow{\omega_f} \text{continuum}(J_p, v_p=1), \tag{10}$$

where the initial state ''i'' is the ground state $X^1\Sigma$ of the CO molecule, with quantum numbers $J_i = M = 0$. Then the state ''0'' is the set of rotational sublevels with odd $J_0 = 1, 3, 5$, and the state ''1'' is the set with $J_1 = 0, 2, 4, 6$. In accordance with the selection rules, dipole transitions for $M = 0$ (when $\Delta J = \pm 1$, $\Delta\Lambda = 0, \pm 1$) will populate continuum states with $J_p = 1, 3, 5, 7$. The predissociation states in this case belong to the s series, in which the natural widths of the levels with $n_1 = 6$, $v = 1$ (scheme (9)) and $n_0 = 5$, $v = 1$ (scheme (10)) are $\Gamma_e = 2.4\ \text{cm}^{-1}$ and $\Gamma_e = 1.9\ \text{cm}^{-1}$.[23] The transition probabilities to the states with various J_p as functions of optical frequency are shown in Fig. 2.

Figure 2a illustrates the first case, when the s series of the intermediate group ''1'' interacts with the ionized con-

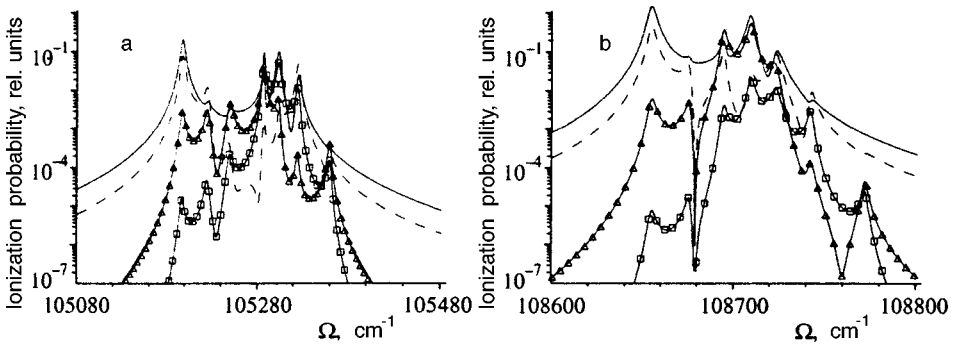

FIG. 2. Photoelectron spectra of CO molecule for continuum states with $J_p=1, 3, 5, 7$. The solid line corresponds to the states with $J_p=1$, the dashed line to $J_p=3$, triangles to $J_p=5$, and squares to $J_p=7$. The ionized states belong to (a) $s\sigma$ series (scheme (9)) and (b) $p\sigma$ series (scheme (10)).

tinuum. Similar curves for the second case (with the population of the p series) are given in Fig. 2b. The spectra of multiphoton ionization have been calculated with the initial data taken from Ref. 23 and frequency ω_f of the strong field close to the energy difference between unperturbed levels $\rho_0 v_0$ and $\rho_1 v_1$ with a detuning $\Delta=\omega_f-(E_{\rho_1 v_1}-E_{\rho_0 v_0})$. These conditions are optimal for populating states with large J^+. For example, in the first case $l=1$ (scheme (9)), and these conditions are satisfied at $\Omega=105287.1$ cm^{-1}, $\omega_f=5793.0$ cm^{-1}, which correspond to a detuning $\Delta=12.0\cdot B$ (B is the rotational constant of the ion CO^+) and $f=7.2\cdot 10^{-5}$. In the second case $l=0, 2$ (scheme (10)), and these conditions are satisfied at $\Omega=108696.1$ cm^{-1}, $\omega_f=3797.2$ cm^{-1} ($\Delta=-0.4\cdot B$), and $f=1.1\cdot 10^{-4}$.

Now let us discuss features of the resulting ionization spectra. Note that these spectra consist of typical resonant peaks, whose positions correspond to groups of rotational sublevels of the intermediate states ''0'' and ''1'' perturbed by the applied field. The peaks corresponding to larger J_p due to the multiphoton absorption are localized near these groups of states. The sidebands of the spectra are due to conventional three-photon absorption ($\Omega+2\omega_f$) with final states $J_p=1, 3$. We are mostly interested in bands where transitions to states with higher J_p are dominant. In the first case (Fig. 2a) these bands are near the probe frequencies $\Omega=105287.0$ cm^{-1} and $\Omega=105357.5$ cm^{-1}; in the second case, illustrated by Fig. 2b, they are near $\Omega=108719.0$ cm^{-1}.

One can see that at low applied field strength, $f\approx 10^{-5}$, resonances with intermediate states $J_0=5$ and $J_1=6$, and final states $J_p=1, 3, 5$ are essentially indistinguishable. When the final state has $J_p=7$, the resonances in the spectrum are well resolved. The ratio between the transition probability to the state with $J_p=5$ and the sum of transition probabilities to $J_p=1$ and 3 is as high as 4. An additional characteristic of the multiphoton ionization spectrum is the distribution function of produced ions over rotational states with angular momenta J^+ described by the expression

$$F(E)\sim\sum_{l,m,\Lambda,J_p}|(lJ^+mM^+|J_pM)(lJ^+\Lambda 0|J_p\Lambda)M_{ip}|^2, \tag{11}$$

where $(lJ^+mM^+|J_pM)$ are the coefficients of vector addition. The resulting histograms of ion distribution over J^+ are shown in Fig. 3.[1)] The function normalization to unity means that the sum of the function values at different J^+ equals unity. In the first case, when the angular momentum of the emitted electron $l=1$, the ion rotational momentum J^+ ranges between 0 and 6. In the second case, the electron angular momentum l can be 0 or 2, and, in accordance with the sum rules, the ion rotational momenta, $|J_p-l|\leq J^+\leq|J_p+l|$, ranges between 0 and 7. If $l=1$, ions in excited rotational states with $J^+=4, 6$ are generated with higher probabilities in the discussed spectral ranges, where

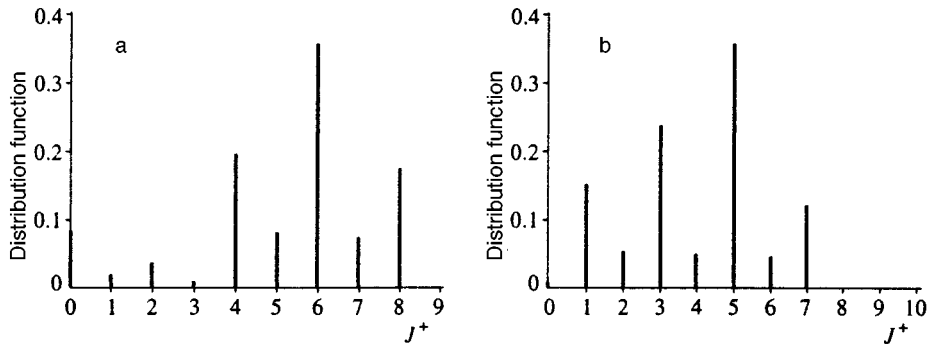

FIG. 3. Distribution of ions over angular momentum J^+ normalized to unity (the angular momentum of the ejected electron is (a) $l=1$, (b) $l=0, 2$).

the final states with total momentum $J_p=5,7$ dominate because $J^+=J_p+1$. In the case $l=0,2$ it takes place for $J^+=5,7$ (here $J^+=J_p, J_p\pm 2$).

It is important that the discussed scheme of calculation is very sensitive to the basic parameters of the theory, namely the applied field strength f and detuning Δ. For example, at $f\sim 10^{-3}$ the levels are strongly broadened by the field, and the shape of the ionization spectrum is affected by spectral diffusion, the relation between the parameters $|M_{ip}|^2$ being the same as in the spectrum wings. A similar effect is caused by an increase in the detuning. Nonetheless, the suggested scheme of calculation allows one to determine the optimal combination of frequencies Ω, ω_f, and field strength f for each specific molecule.

4. CONCLUSIONS

Owing to the interaction between excited molecules and intense monochromatic light, laser-induced states, which are coherent superpositions of rotational states coupled by the electromagnetic field, are formed when the separation between energy levels Δ_r is comparable to the field-induced splitting, $\Delta_r\sim V_{if}^{J_0J_1}$. Their number equals the total number of rovibronic states $\Psi_n=\Sigma_s C_s(f,\omega_f)\Phi_s$ involved in the process, which are autoionization or predissociation states and have lifetimes due to radiative decay longer than the lifetime due to the most efficient optical transition. Owing to the migration of population among states with different angular momenta due to reemission of photons of the stronger field, it is possible to bring about the optimal conditions required for generating molecular ions with high enough angular momenta $J^+\sim 5-7$. This allows one to design a new technique for generating slow polarized ion beams with predetermined distribution over rotational states through the multiphoton ionization of molecules in rotationally nonexcited initial states (in states with zero projection of the angular momentum M on the polarization vector). Generation of such beams, certainly, would be interesting from the viewpoint of studying experimentally the role of rotation in elementary reactions with molecular ions (reactions of dissociative recombination of electrons and molecular ions,[25] near-threshold photodissociation of molecules, etc.)

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-03-34113).

[1)]The function normalization to unity means that the sum of the function values at different J^+ equals unity.

[1] G. V. Golubkov and G. K. Ivanov, Izv. Akad. Nauk SSSR, Ser. Khim. **3**, 367 (1994).
[2] E. Charron, A. Giusti-Suzor, and F. H. Mies, Phys. Rev. A **49**, R641 (1994).
[3] O. Atabek, M. Chrysos, and R. Lefevre, Phys. Rev. A **49**, R8 (1994).
[4] P. Brumer and H. Shapiro, Sci. Amer. **272**(3), 34 (1995).
[5] A. Giusti-Suzor, X. He, O. Atabek, and F. H. Mies, Phys. Rev. Lett. **64**, 515 (1990).
[6] G. Jolicard and O. Atabek, Phys. Rev. A **46**, 5845 (1992).
[7] A. Zavriyev, P. H. Bucksbaum, G. H. Muller, and D. W. Schumacher, Phys. Rev. A **42**, 5500 (1990).
[8] P. Agostini, F. Fabre, G. Mainfray, G. Petite, and N. K. Rahman, Phys. Rev. Lett. **42**, 1127 (1979).
[9] G. Yao and S.-I. Chu, Phys. Rev. A **48**, 485 (1993).
[10] G. Yao and S.-I. Chu, Chem. Phys. Lett. **197**, 413 (1992).
[11] A. Zavriyev, P. H. Bucksbaum, J. Scuier, and F. Saline, Phys. Rev. Lett. **70**, 1077 (1993).
[12] J. F. McCann and A. D. Bandrauk, J. Chem. Phys. **96**, 903 (1992).
[13] A. Giusti-Suzor, F. H. Mies, L. F. DiMauro, E. Charron, and B. Yang, J. Phys. B **28**, 309 (1995).
[14] S. Banerjee, S. S. Bhattacharyya, and S. Saha, Phys. Rev. A **49**, 1836 (1994).
[15] E. Charron, A. Giusti-Suzor, and F. H. Mies, Phys. Rev. Lett. **75**, 2815 (1995).
[16] E. E. Aubanel, J. M. Gauthier, and A. D. Bandrauk, Phys. Rev. A **48**, 2145 (1993).
[17] V. S. Letokhov, *Laser Photoionization Spectroscopy* [in Russian], Nauka, Moscow (1987).
[18] B. N. Delone and V. P. Krainov, *Atoms in Strong Optical Fields* [in Russian], Énergoizdat, Moscow (1984).
[19] G. K. Ivanov, Khim. Vys. Énerg. **23**, 172 (1989).
[20] G. K. Ivanov and G. V. Golubkov, Zh. Éksp. Teor. Fiz. **99**, 1404 (1991) [Sov. Phys. JETP **72**, 783 (1991)].
[21] G. K. Ivanov, G. V. Golubkov, and S. V. Drygin, Zh. Éksp. Teor. Fiz. **107**, 1503 (1995) [JETP **80**, 840 (1995)].
[22] C. Letzelter, M. Eidelsberg, F. Rostus, J. Breton, and B. Thieblemont, Chem. Phys. **114**, 273 (1987).
[23] M. Komatsu, T. Ebata, T. Maeyama, and N. Mikami, J. Chem. Phys. **103**, 2420 (1995).
[24] G. Sha, D. Proch, Ch. Rose, and K. L. Kompa, J. Chem. Phys. **99**, 4334 (1993).
[25] M. Larsson, Phys. Scr. **59**, 270 (1995).

Translation provided by the Russian Editorial office.

Soliton relaxation in antiferromagnets

V. G. Bar'yakhtar

Institute of Magnetism, Ukrainian National Academy of Sciences, 340114 Kiev, Ukraine

A. L. Sukstanskiĭ and E. Yu. Melikhov

Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340144 Donetsk, Ukraine
(Submitted 11 July 1996; resubmitted 3 October 1996)
Zh. Éksp. Teor. Fiz. **111**, 1633–1650 (May 1997)

We study the relaxation two-parameter one-dimensional solitons in antiferromagnets using the phenomenological theory. Allowing for relaxation terms of a relativistic and exchange nature, we set up a system of evolution equations for the constants of the motion of a soliton and calculate the corresponding integral curves, which describe the variation of the soliton parameters in the relaxation process. © *1997 American Institute of Physics.* [S1063-7761(97)00805-6]

1. INTRODUCTION

One of the most important problems of the physics of magnetically ordered crystals, from both the theoretical and experimental viewpoints, is the group of questions associated with the analysis of relaxation processes in magnetic materials, which determine such characteristics of magnetic materials as ferromagnetic and antiferromagnetic resonance linewidths, spin-wave parametric excitation thresholds, and the width of the neutron scattering intensity peak. Lately there has been an upsurge of interest in studies of the relaxation characteristics of various nonlinear excitations, such as the mobility of domain walls and the soliton diffusion coefficient.

There are two basic approaches to investigating relaxation processes theoretically: the microscopic and the phenomenological. The first is based on a detailed microscopic study of the interactions of various excitations of the magnetic material (linear or nonlinear) with one another and with other subsystems of the crystal (say, the elastic subsystem). The advantage of the microscopic approach is that it makes it possible to find the dependence of the relaxation characteristics under investigation on the temperature, the defect concentration, and other parameters of the magnetic material, which in turn can be found from independent measurements.

However, when applied to nonlinear waves, the microscopic approach is much more complicated and can actually be used only to study the simplest solitons of the kink or domain-wall type. Describing the relaxation of more complicated solitons (say, of the two-parameter bion type) and generalizing to multidimensional excitations are nontrivial problems in the microscopic setting, since this requires knowing the exact spectrum and wave functions of the magnons against the soliton background, and these are known only for a relatively small number of one-dimensional problems.

An alternative approach is the phenomenological theory proposed in the classic work of Landau and Lifshitz[1] long before the microscopic approach was developed. Their theory did not yield the detailed characteristics of relaxation processes of the microscopic theory, but it made it possible to give a general picture of the relaxation of a nonlinear excitation. Within the phenomenological macroscopic approach, energy dissipation processes were taken into account by introducing what became known as relaxation terms into the dynamical equations of motion for magnetization (the Landau–Lifshitz equations).

They proposed the following equation for the magnetization vector $\mathbf{M}$ in a one-sublattice ferromagnet:

$$\dot{\mathbf{M}} = -g\mathbf{M}\times\mathbf{H} + \mathbf{R}, \tag{1}$$

where $\mathbf{H} = -\delta W/\delta\mathbf{M}$ is the effective field, W is the energy functional of the ferromagnet, g is the gyromagnetic ratio, and the dot stands for the time derivative. In this equation the first term on the right-hand side describes the dynamics of vector $\mathbf{M}$ and the second is the dissipative term, which describes how magnetization approaches its equilibrium value. This dissipative term was written in Ref. 1 as

$$\mathbf{R} = \frac{\lambda g}{M}\mathbf{M}\times(\mathbf{H}\times\mathbf{M}), \tag{2}$$

where λ is the only constant of the theory, the dimensionless relaxation constant.

Gilbert[2] suggested a slightly different form for the dissipative term, which, however (and this can easily be verified), is equivalent to (2):

$$\mathbf{R} = \frac{\lambda'}{M}\mathbf{M}\times\dot{\mathbf{M}}, \quad \lambda' = \frac{\lambda}{1+\lambda^2}. \tag{3}$$

An important property of the equation of motion with a dissipative term of the Landau–Lifshitz type (2) or of the Gilbert type (3) is that the equation conserves the length of the magnetization vector, $|\mathbf{M}| = \text{const}$. Hence, as noted by Landau and Lifshitz,[1] Eq. (1) is actually an equation for the unit vector $\mathbf{m} = \mathbf{M}/M$, $M = |\mathbf{M}|$, which describes the state of the spin system. Moreover, they also noted[1] that the dissipative term only describes relaxation that emerges because of relativistic interactions. Indeed, calculating the energy dissipation rate $W = -2Q$, where Q is what is known as the dissipative function, via (2) and (3), we obtain

$$Q=-\frac{1}{2}\int d\mathbf{r}\ \mathbf{H}\cdot\dot{\mathbf{M}}=-\frac{1}{2}\int d\mathbf{r}\ \mathbf{H}\cdot\mathbf{R}$$
$$=-\frac{\lambda}{2gM}\int d\mathbf{r}\ \dot{\mathbf{M}}^2. \tag{4}$$

Hence energy dissipation occurs in uniform precession of magnetization, too. Since only relativistic interactions can lead to the relaxation of uniform magnetization motion, the dissipative term **R** in form (2) or (3) is of a relativistic nature.

Starting out with Eq. (1) and the expression (2) or (3) for **R**, we can easily derive such important relaxation characteristics of magnetic materials as the ferromagnetic resonance linewidth $\Delta\omega=\lambda\omega_0$ (here ω_0 is the ferromagnetic resonance frequency), the dynamic braking coefficient η for a domain wall,[1] and the damping constant for spin waves.[3] However, a detailed comparison of the results with the experimental data and the results of microscopic calculations have revealed a number of striking contradictions. The most significant among these is the incorrect dependence of the damping constant of spin waves on the wave vector ($\gamma(k)\sim k^2$), while the microscopic calculations conducted by Dyson[4] and Koshcheev and Krivoglaz[5] (see also Refs. 1 and 6) yield the result $\gamma(k)\sim\omega^2(k)\sim k^4$ for short-wave magnons ($kx_0\gg 1$). In ferromagnets of the easy-plane type the result is really absurd: in the long-wave limit ($k\to 0$), when $\omega(k)\sim|\mathbf{k}|\to 0$, the calculation of $\gamma(k)$ on the basis of (2) yields $\gamma(k)\to\text{const}\neq 0$, i.e., $[\gamma(k)/\omega(k)]\to\infty$ as $k\to 0$ (on the contrary, microscopic calculations lead to the ''hydrodynamic'' result $\gamma(k)\sim\omega^2(k)\sim k^2$; see Ref. 7).

We also note that the relaxation constant λ obtained from data on the ferromagnetic resonance linewidth and the mobility of domain walls in high-quality ferrite films may differ considerably (by a factor of ten or even more; see Ref. 8). Moreover, microscopic calculations of the coefficient η by Abyzov and Ivanov[9] and Ivanov *et al.*[10] have revealed that exchange processes also contribute to the braking of domain walls (in addition to relativistic processes).

Thus, the above contradictions suggest that a phenomenological description of many relaxation phenomena in magnetic materials based on a dissipative term of the type (2) or (3) is inadequate.

Considerable progress in developing the phenomenological approach was achieved in Refs. 11–14. There a new type of relaxation terms consistently describing dissipation processes of both relativistic and exchange nature was proposed. Moreover, the studies revealed how the symmetry of the crystal and the hierarchy of various interactions affect the structure of the dissipative terns and the hierarchy of the corresponding relativistic constants.

In Refs. 11 and 12, Onsager equations with the components of the vector **M** taken as the independent generalized coordinates were used to find the relaxation terms. In this case the components of the effective field **H** are, as can easily be shown, the generalized forces. It was found that when there is weak spatial dispersion (the effective field $\mathbf{H}(\mathbf{r},t)$ slowly varies with the coordinates), the equation for the components of the magnetization vector has the form (for crystals with a symmetry center)

$$\dot{\mathbf{M}}=-g\mathbf{M}\times\mathbf{H}+gM\{\lambda_{ik}H_i\mathbf{e}_k-\lambda_e a^2\Delta\mathbf{H}\}. \tag{5}$$

The last term on the right-hand side, proportional to $\Delta\mathbf{H}$, describes relaxation processes caused by exchange interactions in the magnetic material[11] (''minus'' is chosen for the sake of convenience, so that $\lambda_e>0$), and the symmetric tensor λ_{ik} ($\lambda_{ik}=\lambda_{ki}$) describes the contribution of various interactions of a relativistic nature to the dissipation processes. The form of the tensor λ_{ik} is determined by the symmetry of the magnetic material: in a rhombic magnetic material the tensor λ_{ik} is diagonal along the principal axes: $\lambda_{ik}=\text{diag}(\lambda_1,\lambda_2,\lambda_3)$.

Dynamical symmetry plays an important role in selecting the constants λ_i. In particular, in the exchange approximation we have conservation of the total magnetic moment $\mathbf{M}_{\text{total}}$ of the body:

$$\mathbf{M}_{\text{total}}=\int d\mathbf{r}\ \mathbf{M}(\mathbf{r}), \tag{6}$$

which means that $\lambda_i=0$ (this suggests that the corresponding term in Eq. (5) is of a relativistic nature). In the model of a purely uniaxial magnetic material (C_∞ symmetry), one of the components of the vector $\mathbf{M}_{\text{total}}$, precisely $M_{\text{total},z}$ (here z is the selected axis), is a constant of the motion. This implies that in a uniaxial magnetic material $\lambda_{ik}=\text{diag}(\lambda_1,\lambda_1,0)$ (the fact that λ_x and λ_y are equal follows from the equivalence of the axes).

Allowing for anisotropy in the basal plane changes both the dynamical and relaxation terms in the equation of motion. If the uniaxial anisotropy energy is much higher than the energy of the interactions violating the above invariance, the corresponding relaxation constants must exhibit a similar hierarchy. Here the tensor λ_{ik} has the form

$$\lambda_{ik}=\text{diag}(\lambda_1,\lambda_1,\lambda_3),\quad \lambda_3\ll\lambda_1. \tag{7}$$

The ferromagnet's dissipative function corresponding to (5) can be written as[11,13]

$$Q=\frac{1}{2}\int d\mathbf{r}\ gM\{\lambda_{ik}H_iH_k+\lambda_e a^2(\nabla\cdot\mathbf{H})^2\}, \tag{8}$$

where the factors gM and a^2 (a is the lattice constant) have been introduced so that the relaxation constants λ_{ik} and λ_e are dimensionless.

Clearly, using Eq. (5) in the exchange approximation ($\lambda_{ik}=0$) and in the case of a uniaxial ferromagnet leads to the proper dependence of the spin-wave damping constant on the wave vector in the long-wave approximation, consistant with the one obtained by microscopic calculations. Moreover, the presence of at least two relaxation constants in the theory makes it possible to match experimental data on the ferromagnetic resonance linewidth and the braking of a domain wall.[11]

It is important to note that the equation of motion (5), in contrast to equations with a relaxation term of the type (2) or (3), does not conserve the length of the magnetization vector, $|\mathbf{M}|\neq\text{const}$, which complicates the study of relaxation in the system considerably.

For two-sublattice antiferromagnets, objects that we are studying in this paper, similar relaxation terms and a dissipative function allowing for the symmetry of the magnetic material and exchange relaxation were obtained in Refs. 12 and 14. There, on the basis of Onsager equations in which the generalized coordinates were the components of the vectors $\mathbf{M}$ and $\mathbf{L}$, where $\mathbf{M}=(\mathbf{M}_1+\mathbf{M}_2)/2$ and $\mathbf{L}=(\mathbf{M}_1-\mathbf{M}_2)/2$ are, respectively, the vectors of weak ferromagnetism and antiferromagnetism, with $\mathbf{M}_{1,2}$ the sublattice magnetization vectors, and the generalized forces were the effective fields $\mathbf{H}=-\delta W/\delta\mathbf{M}$ and $\mathbf{F}=-\delta W/\delta\mathbf{L}$ (W is the antiferromagnet's energy functional), it was established that the dissipative function of the antiferromagnet has the form

$$Q=\frac{1}{2}\int d\mathbf{r}\; g\;|\mathbf{L}|\{\lambda_{ik}H_iH_k+\lambda_e a^2(\nabla\cdot\mathbf{H})^2+\lambda_0\mathbf{F}^2\}, \quad (9)$$

where the tensor λ_{ik} is of a relativistic nature, and the constants λ_e and λ_0 of an exchange nature. The dissipative terms in the equations of motion for the vectors $\mathbf{M}$ and $\mathbf{L}$ are determined by the relationships $\mathbf{R}_M=\delta Q/\delta\mathbf{H}$ and $\mathbf{R}_L=\delta Q/\delta\mathbf{F}$ (see below).

In the present study we use a dissipative function of type (9) to develop the phenomenological approach for describing the relaxation of nonlinear excitations in two-sublattice antiferromagnets. As an example we study the relaxation of two-parameter solitons. A similar problem for two-parameter solitons in one-sublattice ferromagnets was studied in Ref. 15.

As in Ref. 15, to describe the evolution of the parameters of the excitations being considered we use the fact that the corresponding relaxation constants are small, which makes it possible to invoke perturbation theory techniques. For solitons in exactly integrable systems, there exists a specific form of perturbation theory based on the inverse scattering problem (see, e.g., Refs. 16 and 17). Here we use a simpler version of that theory, which amounts to setting up evolution equations for the constants of the motion of the unperturbed system. The equations describe the slow evolution of the parameters of the initial excitation as a result of dissipation. The simplest variant of this approach was used by McLaughlin and Scott[18] to study fluxon damping in Josephson junctions by applying the perturbed sine–Gordon equation. The advantage of such an approach is that it can be used even when the initial equation is not exactly integrable, say, in analyzing three-dimensional magnetic solitons.[19]

2. THE GENERAL EQUATIONS

If we use the two-sublattice antiferromagnet model, the equations of motion for the ferromagnetism and antiferromagnetism vectors with allowance for dissipative terms have the form

$$\begin{aligned}-\frac{2}{g}\dot{\mathbf{M}}&=\mathbf{M}\times\mathbf{H}+\mathbf{L}\times\mathbf{F}+\mathbf{R}_M,\\ -\frac{2}{g}\dot{\mathbf{L}}&=\mathbf{M}\times\mathbf{F}+\mathbf{L}\times\mathbf{H}+\mathbf{R}_L.\end{aligned} \quad (10)$$

Here $\mathbf{H}=-\delta W/\delta\mathbf{M}$ and $\mathbf{F}=-\delta W/\delta\mathbf{L}$ are the effective fields; $\mathbf{R}_M$ and $\mathbf{R}_L$ are relaxation terms determined by a dissipative function ($\mathbf{R}_M=\delta Q/\delta\mathbf{H}$ and $\mathbf{R}_L=\delta Q/\delta\mathbf{F}$). The structure of the dissipative function Q, which allows for both exchange and relativistic relaxation, was determined in Refs. 12 and 14 and can be written in the form (9).

As in ferromagnets, the symmetry of the tensor of the relativistic relaxation constants, λ_{ik}, is determined by the symmetry and hierarchy of relativistic interactions. In particular, in the case of the uniaxial antiferromagnet under consideration, the tensor λ_{ik} has the form (7), just as it does in the case of ferromagnets. Here the dissipative terms $\mathbf{R}_M$ and $\mathbf{R}_L$ are expressed as follows:

$$\begin{aligned}\mathbf{R}_M&=g|\mathbf{L}|(\lambda_1\mathbf{H}_\perp-\lambda_e a^2\Delta\mathbf{H}),\\ \mathbf{R}_L&=g|\mathbf{L}|\lambda_0\mathbf{F},\end{aligned} \quad (11)$$

where $\mathbf{H}_\perp$ is the component of the effective field $\mathbf{H}$ perpendicular to the Z axis.

In the nondissipative approximation, the magnetization vectors in an antiferromagnet whose temperature is much lower than the Néel temperature T_N have a constant length, $|\mathbf{M}_{1,2}|=M_0$, with the result that there are two identities linking the vectors $\mathbf{M}$ and $\mathbf{L}$:

$$p\equiv\mathbf{M}\cdot\mathbf{L}=0, \quad s\equiv\mathbf{M}^2+\mathbf{L}^2-M_0^2=0. \quad (12)$$

An important aspect of Eqs. (10) with the dissipative term (11) is the fact that these equations do not conserve the length of the magnetization vectors of the sublattices (in the same way as Eq. (5) does not conserve the length of vector $\mathbf{M}$ in a ferromagnet):

$$\begin{aligned}&\dot{M}_1=\mathbf{m}_1\cdot(\mathbf{R}_M+\mathbf{R}_L), \quad \dot{M}_2=\mathbf{m}_2\cdot(\mathbf{R}_M-\mathbf{R}_L),\\ &\mathbf{m}_i=\frac{\mathbf{M}_i}{|\mathbf{M}_i|}, \quad i=1,2,\end{aligned} \quad (13)$$

with $dM_i/dt\neq 0$ for vectors $\mathbf{R}_M$ and $\mathbf{R}_L$ of the form (11).

Consequently, the quantities p and s are not conserved either:

$$\dot{p}=\mathbf{R}_M\cdot\mathbf{L}+\mathbf{R}_L\cdot\mathbf{M}, \quad \dot{s}=2(\mathbf{R}_M\cdot\mathbf{M}+\mathbf{R}_L\cdot\mathbf{L}). \quad (14)$$

Describing the relaxation of a magnetic excitation requires solving Eqs. (11) with dissipative terms. In most cases the problem cannot be solved (except in the case of simple linear spin waves). Hence, allowing for the smallness of the relaxation constants, we can employ perturbative techniques. In this paper we analyze the relaxation of nonlinear waves in an antiferromagnet in an approximation that is linear in the relaxation constants and uses a simple perturbation scheme based on building evolution equations for the constants of the motion of the unperturbed system. The essence of this approach is as follows.

Let the magnetization distribution in a nonlinear wave be determined by a set of parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$ that in the nondissipative approximation are constants. When the relaxation terms are taken into account, these parameters become time-dependent. The corresponding evolution equations for α_j ($j=1,\ldots,n$) can be obtained from the constants of the

motion for the unperturbed system, $I_1, I_2, \ldots, I_n$ (if the system has a solution with n parameters, there are at least n constants of the motion).

One such constant of the motion is the magnetic excitation energy $E(\alpha_j)$. The rate at which this energy varies is determined by the dissipative function Q,

$$\frac{dE}{dt} = -2Q = \int d\mathbf{r}\ (\mathbf{H}\cdot\dot{\mathbf{M}} + \mathbf{F}\cdot\dot{\mathbf{L}})$$
$$= \int d\mathbf{r}\ (\mathbf{H}\cdot\mathbf{R}_M + \mathbf{F}\cdot\mathbf{R}_L). \tag{15}$$

On the one hand, we find dE/dt as a linear combination of the rates of variation of the nonlinear wave parameters, $d\alpha_j/dt$; on the other, we calculate the value of Q as a function of these parameters. Then we equate the two and arrive at an energy balance equation, which is one of the desired evolution equations, describing the variation of the parameters α_j caused by relaxation processes. Similar equations can be derived by calculating the rates of variation of other constants of the motion. As a result we arrive at a system of n first-order differential equations for the parameters α_j of the nonlinear wave.

The simplest variant of such an approach that uses only one constant of the motion (the energy) was used in Ref. 20 to analyze the dissipation of one-parameter nonlinear waves, domain walls (the parameter here is the velocity of the domain walls) with allowance for relaxation and a driving force. In Ref. 15 this approach was used to study the relaxation of two-parameter solitons in one-sublattice ferromagnets, and in Ref. 19 to analyze the relaxation of multidimensional precession solitons.

An important remark is in order. As noted earlier, the structure of the dissipative terms (11) is such that they do not conserve the lengths of the magnetization vectors of the sublattices. Since M_1 and M_2 are temperature-dependent, the equations for the vectors $\mathbf{M}$ and $\mathbf{L}$ with allowance for the relaxation terms (11) cannot, generally speaking, be studied independently. The system of equations must incorporate Eqs. (10), heat equations (for the lattice and spin specific heats), and the entropy balance equation. However, when we study the relaxation of a magnetic excitation in a temperature range far from the Néel temperature, we implicitly assume that relaxation leads only to changes in the parameters characterizing the corresponding solution of the dynamical equations (''slow'' relaxation). Here both the initial excitation and the final state of the magnetic system are found by solving the dynamical equations of motion in which the length of the magnetization vector is fixed. In this way we tacitly assume that the magnetic system is in contact with a heat bath that instantly balances all variations in $M_{1,2}$ by supplying or removing a certain amount of heat to or from the magnetic system.

To effectively allow for this implicit assumption, it is sufficient, in calculating the rate of variation of a constant of the motion, to take into account only relaxation not related to changes in $|\mathbf{M}_{1,2}|$.

Here we must set $d\mathbf{M}_i/dt = M_i d\mathbf{m}_i/dt + \mathbf{m}_i dM_i/dt$ in the expressions for the rates of variation of the constants of the motion, obtained from the equations of motion with allows for relaxation, and leave only the first term.

For instance, instead of Eq. (9) for the dissipative function Q, we obtain the following after performing some simple calculations:

$$\overline{Q} = -\frac{1}{2}\left(\frac{dE}{dt}\right)_{M_i=\text{const}} = \frac{1}{2}\int d\mathbf{r}[M_1\dot{\mathbf{m}}_1\cdot\mathbf{H}_1 + M_2\dot{\mathbf{m}}_2\cdot\mathbf{H}_2]$$
$$= \frac{1}{2}\int d\mathbf{r}\Big\{\mathbf{H}\cdot\mathbf{R}_M + \mathbf{F}\cdot\mathbf{R}_L - \frac{1}{M_0^2}[(H_M + F_L)(\mathbf{M}\cdot\mathbf{R}_M$$
$$+ \mathbf{L}\cdot\mathbf{R}_L) + (H_L + F_M)(\mathbf{L}\cdot\mathbf{R}_M + \mathbf{M}\cdot\mathbf{R}_L)]\Big\}, \tag{16}$$

where $H_L = (\mathbf{H}\cdot\mathbf{L})$, $F_L = (\mathbf{F}\cdot\mathbf{L})$, $H_M = (\mathbf{H}\cdot\mathbf{M})$, and $F_M = (\mathbf{F}\cdot\mathbf{M})$ (the first two terms in the integrand in (16) correspond to the total dissipative function).

In calculating $\overline{Q}$ in an approximation linear in the relaxation constants, the effective fields $\mathbf{H}$ and $\mathbf{F}$ in Eqs. (11) and (16) must be calculated in the leading (zeroth) approximation in these constants (this fact has already been employed in deriving Eq. (16) by assuming that $M_1 = M_2 = M_0$). In the nondissipative approximation, both $\mathbf{H}$ and $\mathbf{F}$ can be expressed in terms of the two scalar quantities H_L and F_L as follows:

$$\mathbf{H} = \frac{1}{L^2}\left\{\frac{2}{g}[\mathbf{L}\cdot\dot{\mathbf{L}}] + \mathbf{L}H_L + \mathbf{M}F_L\right\},$$
$$\mathbf{F} = \frac{1}{L^2}\left\{\frac{2}{g}[\mathbf{L}\cdot\dot{\mathbf{M}}] + \mathbf{M}H_L + \mathbf{L}F_L\right\}, \tag{17}$$

where $L = |\mathbf{L}|$.

To calculate H_L and F_L, which we call the collinear fields, we turn to an explicit expression for the energy of a ferromagnet:

$$W = \int d\mathbf{r}[f(M_1^2) + f(M_2^2) + w_0(\mathbf{M}\cdot\mathbf{L})]. \tag{18}$$

Here the function $f(M_i^2)$ determines the energy density of the uniform exchange interaction inside the sublattices, and is the main factor forming the length of the magnetization vectors $M_{1,2} = |\mathbf{M}_{1,2}|$. The term $w_0(\mathbf{M}\cdot\mathbf{L})$ describes the energy of nonuniform exchange interaction, the anisotropy energy, the energy of the interactions with an external magnetic field, etc.:

$$w_0(\mathbf{M}\cdot\mathbf{L}) = \frac{\alpha}{2}(\nabla\cdot\mathbf{L})^2 + \frac{\delta}{2}\mathbf{M}^2 + w_a - 2\mathbf{M}\cdot\mathbf{H}_e, \tag{19}$$

where w_a is the anisotropy energy, and $\mathbf{H}_e$ is the external magnetic field.

At temperatures far from the Néel temperature the function $f(x)$ in (18) has a deep minimum at $x = M_0^2(T)$, where $M_0(T)$ is the equilibrium value of the length of the sublattice magnetization vectors. In this case we can assume that only values of M_1 and M_2 close to M_0 play an important role. Hence, approximating $f(x)$ by the expression

$$f(x)=\frac{(x-M_0^2)^2}{4\chi_\parallel M_0^2}, \tag{20}$$

where $\chi_\parallel \ll 1$ has the meaning of the longitudinal susceptibility of the antiferromagnet, we use Eq. (18) and readily obtain, in the linear approximation in the parameters p and s, expressions for the collinear fields H_L and F_L:

$$\begin{aligned} H_L&=\frac{4}{\chi_\parallel}p+H_0, \quad H_0=-\left(\frac{\delta w_0}{\delta \mathbf{M}}\cdot\mathbf{L}\right),\\ F_L&=\frac{2}{\chi_\parallel}s+F_0, \quad F_0=-\left(\frac{\delta w_0}{\delta \mathbf{L}}\cdot\mathbf{L}\right). \end{aligned} \tag{21}$$

Naturally, in the final expressions we must let $\chi_\parallel$ go to zero. Nevertheless, the contribution of the first terms in (21) is finite, since $p\sim\chi_\parallel$ and $s\sim\chi_\parallel$.

In the static limit $H_L=F_L=0$, and the quantities p and s are finite:

$$p=-\frac{\chi_\parallel}{4}H_0, \quad s=-\frac{\chi_\parallel}{2}F_0.$$

In the presence of a dynamical magnetization wave, the quantities H_L and F_L are usually finite. We can easily derive an equation for these quantities by finding the time derivative of Eqs. (21) and using the explicit form of the dissipative terms (11) and Eq. (14):

$$\begin{aligned} &\frac{\chi_\parallel L}{4g}\dot{H}_L-\lambda_e a^2\mathbf{L}^2\Delta H_L-2\lambda_e a^2[(\mathbf{L}\nabla\cdot\mathbf{L})\\ &\times\nabla H_L+(\mathbf{L}\nabla\cdot\mathbf{M})\nabla F_L]+H_L[\lambda_1\mathbf{L}_\perp^2+\lambda_0\mathbf{M}^2\\ &-\lambda_e a^2(\mathbf{L}\Delta\mathbf{L})]+F_L[\lambda_1\mathbf{L}_\perp\cdot\mathbf{M}_\perp-\lambda_e a^2(\mathbf{L}\Delta\mathbf{M})]\\ &=\frac{\chi_\parallel L}{4g}\dot{H}_0+\frac{2}{g}\{\lambda_1 L_z(\mathbf{L}\cdot\dot{\mathbf{L}})_z+\lambda_0\mathbf{M}\cdot(\mathbf{L}\times\dot{\mathbf{M}})\\ &+\lambda_e a^2\mathbf{L}\cdot\Delta(\mathbf{L}\times\dot{\mathbf{L}})\}, \end{aligned} \tag{22}$$

$$\begin{aligned} &\frac{\chi_\parallel L}{4g}\dot{F}_L-\lambda_e a^2\mathbf{M}^2\Delta F_\perp-2\lambda_e a^2[(\mathbf{M}\nabla\cdot\mathbf{M})\\ &\times\nabla F_L+(\mathbf{M}\nabla\cdot\mathbf{L})\nabla H_L]+F_L[\lambda_1\mathbf{M}_\perp^2+\lambda_0\mathbf{L}^2\\ &-\lambda_e a^2(\mathbf{M}\Delta\mathbf{M})]+H_L[\lambda_1\mathbf{M}_\perp\cdot\mathbf{L}_\perp-\lambda_e a^2(\mathbf{M}\Delta\mathbf{L})]\\ &=\frac{\chi_\parallel L}{4g}\dot{F}_0+\frac{2}{g}\{\lambda_1 M_\perp(\dot{\mathbf{L}}\cdot\mathbf{L})_\perp+\lambda_e a^2\mathbf{M}\cdot\Delta(\mathbf{L}\cdot\dot{\mathbf{L}})\}. \end{aligned} \tag{23}$$

The solution of the system of equations (22) and (23) with zero right-hand sides describes the relaxation of H_L and F_L to the equilibrium values. The inhomogeneous solution of this system of equations is finite only in the presence of a dynamical magnetization wave.

We note once more that the system of equations (22) and (23) was derived to the lowest orders of the small relaxation constant λ and the quantity $\chi_\parallel$. Hence all coefficients and right-hand sides of these equations are determined by the distribution of magnetization in the particular excitation, calculated in the nondissipative approximation.

3. RELAXATION OF A TWO-PARAMETER SOLITON

For an example of how the above relationships can be used to analyze the relaxation of magnetic excitations in a uniaxial antiferromagnet, we examine the relaxation of a two-parameter soliton.

As is known, the relativistic interaction energy in antiferromagnets is low compared to that of exchange interaction, β, $b\ll\delta$, with the result that over the entire range of applicability of the phenomenological theory of magnetism ($x_0\gg a$, where x_0 is the characteristic relaxation size, and a is the lattice constant) we have $|\mathbf{M}|\ll|\mathbf{L}|\approx M_0$. Using this fact, we can show, following the results of Refs. 20, 22, and 23, that nonlinear dynamical excitations in two-sublattice antiferromagnets (and weak ferromagnets) can be analyzed by using the effective Lagrangian $\mathscr{L}$

$$\begin{aligned} \mathscr{L}=\int d\mathbf{r}\Big\{&\frac{\alpha}{2c^2}\dot{\mathbf{L}}^2-\frac{\alpha}{2}(\nabla\cdot\mathbf{L})^2-w_a(\mathbf{L})+\frac{4}{g\delta M_0}\\ &\times(h_e(\dot{\mathbf{L}}\cdot\mathbf{L}))-\frac{4}{\delta}(\mathbf{h}_e\cdot\mathbf{L})^2\Big\}, \end{aligned} \tag{24}$$

where $c=gM_0(\alpha\delta)^{1/2}/2$ is a characteristic velocity coinciding with the minimum phase velocity of linear spin waves, and $\mathbf{h}_e=\mathbf{H}_e/M_0$. Here, in the leading approximation in the small parameter $(\beta/\delta)\ll 1$, the length of vector $\mathbf{L}$ remains constant, $|\mathbf{L}|\approx M_0$, and the ferromagnetism vector $\mathbf{M}$ is linked to the unit vector $\mathbf{l}=\mathbf{L}/|\mathbf{L}|$ by the following relationship:

$$\mathbf{M}=\frac{2}{g}\left\{\frac{1}{g}(\dot{\mathbf{l}}\times\mathbf{l})+(\mathbf{l}\cdot\mathbf{H}_e)\times\mathbf{l}\right\}. \tag{25}$$

We write the magnetic anisotropy energy of a magnetic material, $w_a(\mathbf{l})$, in the following way:

$$w_a(\mathbf{L})=\frac{1}{2}\beta\mathbf{L}_\perp^2-\frac{1}{4M_0^2}b(\mathbf{L}_\perp^2)^2, \quad \beta>0, \quad b>0, \tag{26}$$

where β and b are the second- and fourth-order anisotropy constants, respectively.

Using the angular variables θ and φ that parametrize the vector $\mathbf{l}$,

$$\mathbf{l}=(\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta), \tag{27}$$

we can write the equations of the dynamics of the antiferromagnet under discussion in the following form:

$$\begin{aligned} &c^2\Delta\theta-\ddot{\theta}+\sin\theta\cos\theta\Big[(\dot{\varphi}-\Omega_e)^2-c^2(\nabla\varphi)^2\\ &-\omega_0^2\left(1-\frac{b}{\beta}\sin^2\theta\right)\Big]=0, \end{aligned} \tag{28}$$

$$\alpha\nabla(\sin^2\theta\nabla\varphi)-\frac{\alpha}{c^2}\frac{\partial}{\partial t}(\sin^2\theta\dot{\varphi})+\frac{4h_e}{\delta g M_0}\frac{\partial}{\partial t}(\sin^2\theta)=0,$$

$\omega_0=c/x_0$, $x_0=(\alpha/\beta)^{1/2}$, and $\Omega_e=gH_e$. The external magnetic field is assumed directed along the easy-magnetization axis (the Z axis).

Next we consider only one-dimensional soliton solutions of Eqs. (28), corresponding to the boundary conditions

$$\theta \to 0 \quad \text{or} \quad \pm\pi; \quad \frac{\partial\varphi}{\partial x} < \infty \quad \text{as} \quad |x| \to \infty. \tag{29}$$

We start with the case where $b=0$. Then the solution of the equations of motion (28) that satisfies the boundary conditions (29) has the following form:[22]

$$\varphi = kx - \Omega t, \quad \cos\theta = -\tanh[\tilde{\kappa}(x - Vt)],$$
$$k = \frac{(\Omega + \Omega_e)V}{c^2}, \quad \tilde{\kappa}^2 = \frac{\beta}{\alpha(1 - V^2/c^2)} - \frac{(\Omega + \Omega_e)^2}{c^2}. \tag{30}$$

This solution describes the distribution of magnetization in a two-parameter topological soliton of the kink type ($\theta(-\infty)=0$ and $\theta(+\infty)=\pm\pi$), which sets this solution apart from a dynamical two-parameter soliton in a ferromagnet, whose relaxation was studied in Ref. 15.

The parameters determining the soliton's structure are its translational velocity V and the precession frequency Ω.

The soliton solution (30) exists if $\tilde{\kappa}^2 > 0$, or

$$(\Omega + \Omega_e)^2 < \frac{\omega_0^2}{1 - V^2/c^2}. \tag{31}$$

The strength of the external magnetic field, H_e, is limited by the condition that there is an immobile soliton, $H_e < M_0(\beta\delta)^{1/2}$, or $\Omega_e < \Omega_0$, i.e., by the spin-flop transition field.

Since the solution is two-parameter, to analyze its relaxation we must examine two constants of the motion of the unperturbed system. For the constants, which we use to set up the evolution equations for the soliton parameters V and Ω, we take the soliton energy E and a quantity N proportional to the total z-projection of magnetization, which in the nondissipative case is conserved because the antiferromagnet is uniaxial:

$$N = \frac{1}{2\mu_0}\int d\mathbf{r} M_z, \tag{32}$$

where $\mu_0 = \hbar g/2$ is the Bohr magneton. Note that just as in the case of ferromagnets, the constant of the motion N can be interpreted as the number of magnons bound in the soliton.[24]

Note that using another constant of the motion, the momentum $\mathbf{P}$, results in no new equations because $dE = \hbar\omega dN + \mathbf{V}d\mathbf{P}$, which is true of any ferromagnet or antiferromagnet that allows for the existence of two-parameter solitons.[24] To analyze two-parameter nonlinear waves in a biaxial magnetic material, where I_z is not a constant of the motion, one can use the equations for dE/dt and $d\mathbf{P}/dt$.

The values of the constants of the motion E and N corresponding to the soliton solution (30) can easily be calculated:

$$E = \frac{2E_0}{\kappa}\left\{\frac{1}{1-\mu} - \omega_e(\omega + \omega_e)\right\}, \quad N = \frac{2n_0}{\kappa}(\omega + \omega_e),$$
$$E_0 = \beta M_0^2 x_0, \quad n_0 = \frac{E_0}{2\hbar\omega_0}, \tag{33}$$
$$\kappa^2 = \frac{1}{1-u} - (\omega + \omega_e)^2, \quad \omega_e = \frac{\Omega_e}{\omega_0},$$

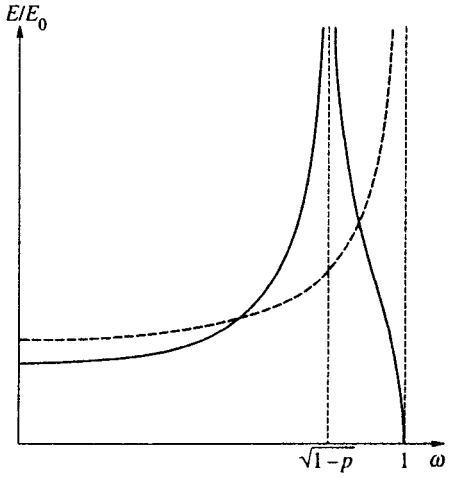

FIG. 1. Soliton energy E as a function of precession frequency at $V=0$.

where $\omega = \Omega/\omega_0$ and $u = (V/c)^2$ are the dimensionless soliton parameters, which prove useful in the calculations below.

The dependence of the soliton energy E on precession frequency ω given by (33) with $V=0$ and $H_e=0$ is depicted in Fig. 1 (the dashed curve).

In the leading approximation in the small parameter β/δ, we can drop all terms in the expression for the reduced dissipative function (16) that are proportional to $\mathbf{M}$. Allowing for (11) and (17), we can then write this function in the form

$$\overline{Q} \cong \frac{1}{2gM_0^2}\int d\mathbf{r}\mathbf{R}_M(\mathbf{L}\times\dot{\mathbf{L}}) = \frac{\delta g M_0}{2}E_0 q(u,\omega). \tag{34}$$

The function $q(u,\omega)$ consists of two terms, $q(u,\omega) = q_r(u,\omega) + q_e(u,\omega)$, the relativistic term and the exchange term:

$$q_r = \lambda_1\langle((\mathbf{l}\times\dot{\mathbf{l}})_\perp)^2 - h_L l_z(\mathbf{l}\times\dot{\mathbf{l}})_z\rangle,$$
$$q_e = \lambda_e'\langle((\mathbf{l}\times\dot{\mathbf{l}})')^2 + h_L(\dot{\mathbf{l}}'(\mathbf{l}\times\mathbf{l}') - \mathbf{l}''(\mathbf{l}\times\dot{\mathbf{l}}))\rangle, \tag{35}$$

where $\lambda_e' = \lambda_e(a/x_0)^2$, and $h_L = gH_L/2M_0\omega_0$; the prime and the angle brackets stand for, respectively, differentiation and integration with respect to the dimensionless spatial variable $\xi = x/x_0$, and the dot indicates differentiation with respect to the dimensionless time variable $\tau = \omega_0 t$.

Reasoning along similar lines, for the reduced (i.e., calculated with $M_{1,2} = \text{const}$) rate of variation of the second constant of the motion, $\dot{N}$, we obtain

$$\overline{\dot{N}} = -\frac{1}{2\mu_0 M_0^2}\int d\mathbf{r} L_z(\mathbf{R}_M\mathbf{L}) = \delta g M_0 n_0 \eta,$$
$$\eta = \eta_r + \eta_e, \tag{36}$$
$$\eta_r = \lambda_1\langle l_z^2(\mathbf{l}\times\dot{\mathbf{l}})_z - h_L l_z \mathbf{l}_\perp^2\rangle, \tag{37}$$
$$\eta_e = \lambda_e'\langle l_z\cdot\mathbf{l}(\mathbf{l}\times\dot{\mathbf{l}})'') + h_L(l_z'' - l_z\mathbf{l}'^2)\rangle.$$

On the other hand, differentiating the expressions (33) for E and N with respect to time, we can write the derivatives dE/dt and dN/dt as linear combinations of the derivatives du/dt and $d\omega/dt$. Equating the resulting expressions to

(34) and (36), respectively, we arrive at the desired equations, which describe the evolution of the soliton parameters ω and u:

$$\dot{u} = -\delta\kappa(1-u)^2(q+\omega\eta), \tag{38}$$

$$\dot{\omega} = -\frac{\delta\kappa}{2}\{(\omega+\omega_e)(1-u)q - [1-(1-u)(\omega+\omega_e)(2\omega+\eta_e)]\eta\}. \tag{39}$$

To calculate the expressions on the right-hand sides of Eqs. (38) and (39), we must find the collinear field h_L. Equations (22) and (23), which yield the quantities H_L and F_L, become independent in the leading approximation in the small parameter β/δ, with the result that the dimensionless quantity h_L in Eqs. (35) and (37) obeys the equation

$$\tilde{\chi}\dot{h}_L - \lambda_e' h_L'' + h_L\{\lambda_1 \mathbf{l}_\perp^2 + \lambda_e' \mathbf{l}'^2\} = \lambda_1 l_z(\mathbf{l}\times\dot{\mathbf{l}})_z + \lambda_e'(\mathbf{l}\cdot(\mathbf{l}\times\dot{\mathbf{l}})'') + \tilde{\chi}\omega_e \dot{l}_z, \tag{40}$$

where $\tilde{\chi} = \chi_\parallel \omega_0 / 4gM_0$.

Equation (40) is a homogeneous linear diffusion equation with a right-hand side. The general solution of this equation without a right-hand side describes the relaxation of h_L to the equilibrium value $h_L = 0$ with a characteristic relaxation time τ_e of order $\chi_\parallel/\lambda_1 gM_0$, which becomes very short as $\chi_\parallel \to 0$ (see Ref. 11).

The inhomogeneous solution of Eq. (40) differs from zero only when there are dynamical magnetization waves. Note that we are interested in the leading approximation for h_L in the relaxation constants, so that in the functions in the coefficients and the right-hand side of this equation we must use the magnetization distribution $\mathbf{l}(x,t)$ in the soliton calculated in the nondissipative approximation. We also note that in addition to having small relaxation constants, Eq. (40) contains the small parameter $\chi_\parallel$. Hence the structure of the solution strongly depends on the relationship between $\chi_\parallel$ and λ and the nature of the excitation of the magnetic material.

Apart from a difference in notation, Eq. (40) coincides with the equation for the collinear component of the effective field in a uniaxial ferromagnet studied in Ref. 15, so that we do not analyze this equation here in detail. The results of such an analysis in connection with the problem under consideration can be reduced to the following.

If the soliton's translational velocity is higher than a certain characteristic value $V_* \sim \lambda\omega_e x_0/\chi_\parallel$, then h_L is given by the simple formula:

$$h_L = \omega_e \cos\theta. \tag{41}$$

Obviously, such a situation occurs only if $\tilde{\chi} \gg \lambda$ and in an external magnetic field. But if $\tilde{\chi} \ll \lambda$, we can set h_L to zero in Eqs. (35) and (37) irrespective of the presence of an external field and the soliton velocity.

The evolution equations for the soliton parameters are simplest when $H_e = 0$. Plugging the explicit form of the magnetization distribution in the soliton (Eq. (38)) first into (35)–(37) and then into (38) and (39) and setting h_L to zero, we obtain

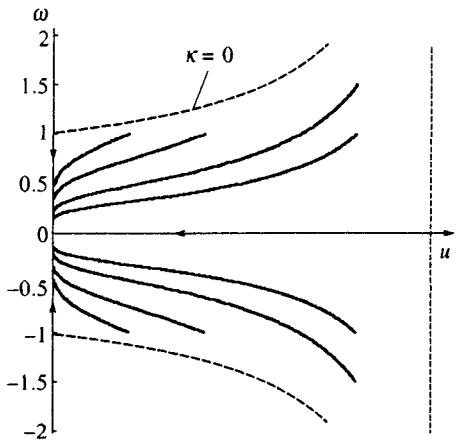

FIG. 2. Evolution of soliton parameters due to relativistic relaxation ($b = 0$ and $\lambda_e = 0$).

$$\dot{u} = f_r(u,\omega) + f_e(u,\omega), \quad \dot{\omega} = g_r(u,\omega) + g_e(u,\omega),$$
$$f_r = -2\lambda_1\delta u\kappa^2(1-u)^2, \quad g_r = -\lambda_1\delta\kappa^2\omega(1-u)(\tfrac{1}{3}+u), \tag{42}$$
$$f_e = -2\lambda_e'\delta u\kappa^2(1-u)[\tfrac{1}{3}+\omega^2(1-u)^2],$$
$$g_e = -2\lambda_e'\delta\kappa^2\omega\left[1-\frac{u}{3}-\omega^2(1+u)(1-u)^2\right].$$

Each of the right-hand sides of the evolution equations (42) consists of two terms (just as the functions $q(u,\omega)$ and $\eta(u,\omega)$ in (34)–(37) do), one related to relaxation processes caused by relativistic interaction and the other to the relaxation processes of an exchange nature.

An analysis of the evolution equations shows that the two types of dissipative terms, one due to relativistic interaction and the other to exchange interaction, have quite different effects on the evolution of the soliton parameters. Over the entire range of soliton parameters (see (31)), the relaxation caused by relativistic interaction and described by the functions $f_r(u,\omega)$ and $g_r(u,\omega)$ leads to a monotonic decrease in both the soliton velocity ($f_r < 0$) and the absolute value of the precession frequency ($\mathrm{sgn}(g_r) = -\mathrm{sgn}(\omega)$).

The nature of relativistic relaxation manifests itself most vividly in the integral curves of the equations $\dot{u} = f_r(u,\omega)$ and $\dot{\omega} = g_r(u,\omega)$, which can be obtained by direct integration of the equations

$$\omega = \omega(0)\left[\frac{u}{u(0)}\right]^{1/6}\left[\frac{1-u(0)}{1-u}\right]^{1/3}, \tag{43}$$

where $u(0)$ and $\omega(0)$ are the initial values of the soliton parameters. Schematically these curves are depicted in Fig. 2.

The contribution of exchange relaxation is described in Eqs. (42) by the alternating functions $f_e(u,\omega)$ and $g_e(u,\omega)$, with the result that the relaxation parameters u and ω vary nonmonotonically during the exchange relaxation process. The corresponding integral curves of the equations $\dot{u} = f_e(u,\omega)$ and $\dot{\omega} = g_e(u,\omega)$ obtained through numerical integration are depicted in Fig. 3.

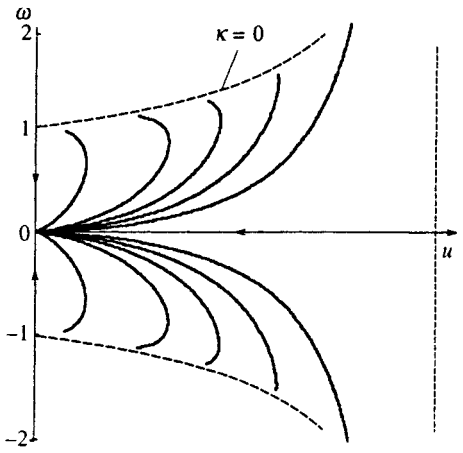

FIG. 3. Evolution of soliton parameters due to exchange relaxation ($b=0$ and $\lambda_r=0$).

We also note that if the soliton was initially at rest, i.e., $u(0)=0$, then $f_r(u,\omega)=f_e(u,\omega)=0$, which means that the soliton remains at rest at all future moments in time. Here the corresponding integral curves in Figs. 2 and 3 are sections of the vertical axis. Similarly, if the precession frequency was initially zero, $\omega(0)=0$, then $g_r(u,\omega)=g_e(u,\omega)=0$, and a soliton that initially does not precess will not do so during relaxation; in this case the integral curves are sections of the horizontal axis.

If the external magnetic field is finite and sufficiently high for V_* to be smaller than c, the soliton relaxation pattern is more complicated. For $V\geqslant V_*$, when h_L is given by (41), the functions $g(u,\omega)$ and $f(u,\omega)$ have the form

$$
\begin{aligned}
f_r&=-2\lambda_1\delta u\kappa^2(1-u)^2,\\
g_r&=-\lambda_1\delta\kappa^2(\omega+\omega_e)(1-u)(\tfrac{1}{3}+u),\\
f_e&=-2\lambda_e'\delta u\kappa^2(1-u)[\tfrac{1}{3}+(\omega+\omega_e)^2(1-u)^2],\\
g_e&=-2\lambda_e'\delta\kappa^2(\omega+\omega_e)\\
&\quad\times\left[1-\frac{u}{3}-(\omega+\omega_e)^2(1+u)(1-u)^2\right].
\end{aligned}
\tag{44}
$$

Clearly, the right-hand sides of (44) can be obtained from the corresponding right-hand sides of (42) by simply replacing ω with $\omega+\omega_e$. Since the range (31) in which a soliton can exist with $\omega_e\neq 0$ is actually obtained from the corresponding range with $\omega_e=0$ through a shift along the frequency axis by $-\omega_e$, the integral curves of Eqs. (44) are the same as those of Eqs. (42), the only difference being that they are shifted by $(-\omega_e)$.

Here the limit in the soliton state to which relaxation leads is a state with precession ($u\to 0$ and $\omega\to-\omega_e$ as $t\to\infty$). We must bear in mind, however, that for $V<V_*$ the value of h_L must be set to zero. Then instead of (44) we have

$$
\begin{aligned}
f_r&=-2\lambda_1\delta u\kappa^2(1-u)^2,\\
g_r&=-\lambda_1\delta\kappa^2(1-u)[u(\omega+\omega_e)+\tfrac{1}{3}\omega],\\
f_e&=-2\lambda_e'\delta u\kappa^2(1-u)\\
&\quad[\tfrac{1}{3}-\omega(\omega+\omega_e)^2(1-u)^2-\omega_e(\omega+\omega_e)(\tfrac{1}{3}-u)],
\end{aligned}
\tag{45}
$$

$$
g_e=-2\lambda_e'\delta\kappa^2(1-u)\left[\kappa^2\left(\omega-\frac{u}{3}(\omega+\omega_e)\right)+u(\omega+\omega_e)^2(u(\omega+\omega_e))-\frac{\omega}{3}\right].
$$

Clearly, the evolution of the soliton parameters described by Eqs. (45) leads to the equilibrium state $u=\omega=0$, as expected.

If the characteristic velocity V_* is higher than c, which is possible if $\tilde{\chi}\ll\lambda$ or if the external field is really weak, Eqs. (45) describe soliton relaxation for the entire range (31) of possible soliton parameters.

But if $V_*<c$, then for $V>V_*$ relaxation is described by Eqs. (44) and for $V<V_*$ by Eqs. (45). It is natural, then, for a transient region to exist in which the one relaxation mode changes to the other. To analyze this region we must use Eq. (40) to calculate h_L in general form, which constitutes a problem beyond the scope of the present investigation.

Let us now examine a more general model of an antiferromagnet, a model that allows for fourth-order terms in the anisotropy energy, i.e., a model with $b\neq 0$ (see Eq. (26)). In this case the solution of the dynamic equations of motion (28) can be written in the following form (at $H_e=0$; see Ref. 22):

$$
\tan\theta=\begin{cases}\dfrac{\kappa}{A^{1/2}\sinh\left(\dfrac{\kappa}{x_0}(x-Vt)\right)}, & A>0,\\[2ex] \dfrac{\kappa}{|A|^{1/2}\cosh\left(\dfrac{\kappa}{x_0}(x-Vy)\right)}, & A_{\min}<A<0,\end{cases}
\tag{46}
$$

where $A=(1-p)/(1-u)-\omega^2$, with $p=b/2\beta$, and the dimensionless parameters ω, u, and κ have been defined earlier.

The solution (46) exists if $\kappa^2>0$, which imposes a lower bound on A, i.e. $A\geqslant A_{\min}=-p/(1-u)$.

If $A>0$, the solution (46) still describes a soliton of the kink type with a topological charge (a domain wall, $\theta(-\infty)=0$ and $\theta(+\infty)=\pi$). But if $A<0$, we have $\theta(\pm\infty)=0$, and the solution (46) describes a dynamic soliton without a topological charge.

The values of the soliton energy E and the constant of the motion N corresponding to the solution (46) are

$$
E=E_0\left\{\frac{1}{2}\left(\frac{1-p}{1-u}+\omega^2\right)D_s+\kappa\right\},\quad N=n_0\omega D_s,
\tag{47}
$$

where $s=1,\ 2$; the case $A>0$ (a soliton with a topological charge) corresponds to $s=1$, and the case $A<0$ (a soliton without a topological charge) to $s=2$; and the functions D_1 and D_2 are given by

$$D_s = 2\left(\frac{1-u}{p}\right)^{1/2}\begin{cases}\sinh^{-1}\left[\dfrac{p}{(1-u)\kappa^2-p}\right]^{1/2}, & s=1,\\ \cosh^{-1}\left[\dfrac{p}{p-(1-u)\kappa^2}\right]^{1/2}, & s=2.\end{cases} \tag{48}$$

The dependence of the soliton energy E on the precession frequency ω described by (48) at $V=0$ is depicted in Fig. 1 by the solid curves. Note the important difference between this dependence and the one for $b=0$ (see Eqs. (33)). In contrast to (33) and the corresponding dashed curve in Fig. 1, the ω-dependence of E for $b \neq 0$ not only has a minimum at $\omega=0$ corresponding to an antiferromagnet state with a domain wall at rest, but also an absolute minimum at $\omega=1$ corresponding to the ground state of the magnetic material. We also note that the parameter ranges corresponding to the two types of soliton (with a topological charge and without) are separated by an infinitely high energy barrier, $E(A\to\pm 0)\to\infty$. Hence a dynamical soliton cannot become a topological soliton as a result of a relaxation process; nor can a topological soliton become a dynamic soliton.

Soliton relaxation for $b \neq 0$ can be analyzed in the same way as we did for $b=0$. For this reason we list only the results. The evolution equations for the soliton parameters u and ω for the case of relativistic relaxation are

$$\dot{u} = -2\lambda_1\delta u(1-u)\frac{\left(\dfrac{1-p}{1-u}-\omega^2\right)D_s+2\kappa}{\left(\dfrac{1-p}{1-u}+\omega^2\right)D_s+2\kappa},$$

$$\dot{\omega} = -\lambda_1\delta\omega\left\{\frac{u\left[\left(\dfrac{1-p}{1-u}-\omega^2\right)D_s+2\kappa\right]}{\left(\dfrac{1-p}{1-u}+\omega^2\right)D_s+2\kappa}\right.$$
$$\left.-\frac{(1-p-\omega^2-u\omega^2)\left[\left(\dfrac{1-p}{1-u}-\omega^2\right)D_s-2\kappa\right]}{2p\left[\left(\dfrac{1-p}{1-u}-\omega^2\right)D_s+\dfrac{2\omega^2}{\kappa}\right]}\right\}. \tag{49}$$

The exchange terms have a similar but more cumbersome form, and we do not give them here.

The integral curves of Eqs. (49) are depicted in Fig. 4. Clearly, topological solitons relax, just as they did at $b=0$, to a state with an immobile domain wall, while dynamical solitons relax to the homogeneous ground state of the antiferromagnet. Such relaxation of a soliton is quite natural: the energy of a soliton with a topological charge drops to the local minimum at $\omega=0$ and $V=0$, corresponding to the antiferromagnetic state with an immobile domain wall (see Fig. 1), while the evolution of the parameters of a soliton without a topological charge is terminated at the curve $\kappa=0$, on which the soliton decays, i.e., its amplitude and energy (and the number N of magnons bound to it) vanish, which corresponds to a transition to the ground state of the magnetic material (a similar situation occurs in the relaxation of a one-dimensional dynamical soliton in a ferromagnet;[15] in the process of relaxation of multidimensional solitons at the edge of their existence range, the solitons decay into a finite number ($N \neq 0$)) of independent linear spin waves[19]).

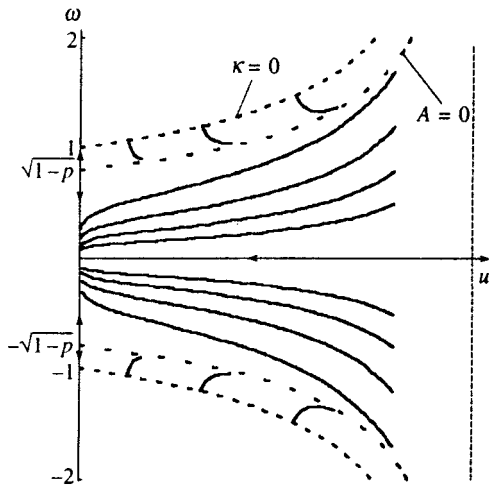

FIG. 4. Evolution of soliton parameters due to relativistic relaxation ($b \neq 0$).

The authors are sincerely grateful to B. I. Ivanov for numerous discussions and useful remarks. The present work was sponsored by the International Science Foundation (Grant UB2300 (V.G.B.)) and the International Soros Science Education Program (ISSEP, Grant No. 452 (E.Yu.M.)).

[1] L. D. Landau and E. M. Lifshitz, Phys. Zs. Sowjet. **8**, 153 (1935).
[2] T. L. Gilbert, Phys. Rev. **100**, 1243 (1955).
[3] E. M. Lifshitz, Zh. Éksp. Teor. Fiz. **15**, 1 (1945).
[4] F. Dyson, Phys. Rev. **102**, 1217 (1956).
[5] V. N. Koshcheev and M. A. Krivoglaz, Fiz. Tverd. Tela (Leningrad) **3**, 1541 (1961).
[6] V. G. Bar'yakhtar, V. N. Krivoruchko, and D. A. Yablonskiĭ, *Green's Functions in the Theory of Magnetism* [in Russian], Naukova Dumka, Kiev (1984).
[7] A. B. Harris, D. Kumar, B. I. Halperin, and P. C. Hohenberg, Phys. Rev. B **13**, 961 (1971).
[8] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Academic Press, New York (1979).
[9] A. S. Abyzov and B. A. Ivanov, Zh. Éksp. Teor. Fiz. **76**, 1700 (1979) [Sov. Phys. JETP **49**, 865 (1979)].
[10] B. A. Ivanov, Yu. N. Mitsaĭ, and N. V. Shakhova, Zh. Éksp. Teor. Fiz. **87**, 289 (1984) [Sov. Phys. JETP **60**, 168 (1984)].
[11] V. G. Bar'yakhtar, Zh. Éksp. Teor. Fiz. **87**, 1501 (1984) [Sov. Phys. JETP **60**, 863 (1984)].
[12] V. G. Bar'yakhtar, Fiz. Nizk. Temp. **11**, 1198 (1985) [Sov. J. Low Temp. Phys. **11**, 662 (1985)].
[13] V. G. Bar'yakhtar, Fiz. Tverd. Tela (Leningrad) **29**, 1317 (1987) [Sov. Phys. Solid State **29**, 754 (1987)].
[14] V. G. Bar'yakhtar, Zh. Éksp. Teor. Fiz. **94**, No. 4, 196 (1988) [Sov. Phys. JETP **67**, 757 (1988)].
[15] V. G. Bar'yakhtar, B. A. Ivanov, T. K. Soboleva, and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **91**, 1454 (1986) [Sov. Phys. JETP **64**, 857 (1986)].
[16] D. J. Kaup and A. C. Newell, Proc. Roy. Soc. London, Ser. A **361**, 413 (1978).
[17] V. I. Karpman and E. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1977) [Sov. Phys. JETP **46**, 281 (1977)].
[18] D. W. McLaughlin and A. C. Scott, ''A multisoliton perturbation theory,'' in *Solitons in Action* (Proc. Workshop Sponsored by the Mathematics Division, Army Research Office, Redstone Arsenal, October, 26–27, 1977), K. Lonngren and A. C. Scott (Eds.), Academic Press, New York (1978).

[19] V. G. Bar'yakhtar, B. A. Ivanov, and A. L. Sukstanskii, Phys. Lett. A **119**, 191 (1986).
[20] V. G. Bar'yakhtar, B. A. Ivanov, and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **78**, 1509 (1980) [Sov. Phys. JETP **51**, 757 (1980)].
[21] B. A. Ivanov and A. L. Sukstanskii, J. Magn. Magn. Mater. **117**, 102 (1992).
[22] I. V. Bar'yakhtar and B. A. Ivanov, Zh. Éksp. Teor. Fiz. **85**, 328 (1983) [Sov. Phys. JETP **58**, 190 (1983)].
[23] A. F. Andreev and V. I. Marchenko, Usp. Fiz. Nauk **130**, 39 (1980) [Sov. Phys. Usp. **23**, 21 (1980)].
[24] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, *Nonlinear Magnetization Waves. Dynamical and Topological Solitons* [in Russian], Naukova Dumka, Kiev (1983).

Translated by Eugene Yankovsky

Theory of magnetic-breakdown oscillations of the galvanomagnetic properties of aluminum taking account of the spin of the conduction electrons

N. Kh. Useinov

Ul'yanovsk State University, 432700 Ul'yanovsk, Russia
(Submitted 16 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1651–1666 (May 1997)

The theory of magnetic breakdown, taking account of the spin of the conduction electrons, is used to calculate the galvanomagnetic properties of aluminum, where the system of electron trajectories in a magnetic field contains small β orbits. Expressions are obtained for the magnetoresistance and Hall resistance in the case of a two-dimensional magnetic-breakdown network of trajectories on the basis of stochastic electron motion on large orbits and coherent electron motion on the β orbits. Qualitative agreement is obtained with the existing experimental data. © *1997 American Institute of Physics.* [S1063-7761(97)00905-0]

1. INTRODUCTION

It is well known that magnetic breakdown changes the character of the field dependences of the magnetoresistance of metals.[1,2] The study made in Refs. 3–8 of magnetic breakdown in aluminum ($T=4.2$ K) showed that for $\mathbf{H} \parallel [001]$ a narrow magnetic-breakdown layer of open trajectories in $\mathbf{k}$ space with $\delta k \approx 4\times 10^{-2}$ a. u. is formed in fields exceeding 20 kG and coherence with a breakdown field $H_0 \approx 4$ kG is present. However, in explaining the magnetic breakdown oscillations and in the calculation of the galvanomagnetic characteristics of Al the spin of the conduction electrons and the spin–orbit interaction were neglected in these papers.

A detailed review of the theory of magnetic breakdown, taking account of the spin degrees of freedom of the conduction electrons, has been given recently.[9] It was shown, on the basis of an analysis of the dispersion relation in regions of anomalous convergence of the trajectories in different zones, that the spin–orbit interaction leads to the possibility of breakdown with spin flip, the main dynamical characteristic of magnetic breakdown was obtained — the s matrix is of rank 4, the basic principles of the theory of coherent magnetic breakdown were generalized, and the magnetic-breakdown spectrum of the conduction electrons was investigated. Applications of the theory to galvanomagnetic phenomena, the de Haas–van Alphen effect, and paramagnetic resonance in conduction electrons were examined.

In the case of aluminum, the transverse cross-sectional areas of the β orbits and the corresponding magnetic-breakdown fields as a function of the wave vector k_z were calculated in Ref. 8, which is devoted to explaining the anomalies of the magnetic-breakdown oscillations of the magnetization, from a calculation of the band structure by the pseudopotential method, including the spin–orbit interaction. It was noted that the energy gap through which electrons tunnel in a magnetic field appears on account of the weak spin–orbit interaction. In Ref. 10 the g-factor of the conduction electrons was calculated on the entire Fermi surface of aluminum, using a scheme with four orthogonalized plane waves, taking the spin–orbit interaction into account as a perturbation. In Refs. 8 and 10 it was concluded that the quantities calculated taking account of the spin–orbit interaction agree better with the experimental data on the de Haas–Alphen effect and on the spin resonance of the conduction electrons.

A more complete list of the literature devoted to the effect of the spin of the conduction electrons and the spin–orbit interaction on the energy spectrum and the galvanomagnetic and other properties of metals under magnetic-breakdown conditions is given in Ref. 9.

Thus there arises the problem of re-examining the theoretical interpretation of the field-dependences of the magnetoresistance of Al taking account of the electronic spin degrees of freedom. For this, it is necessary to calculate the conductivity tensor for the two-dimensional system of open electronic trajectories (see Fig. 1) that arise in Al during magnetic breakdown, taking into account in a natural manner the orientation of electron spin in a magnetic field and incorporating the spin–orbit interaction. This can be done on the basis of the results in Refs. 9 and 11, where the particular oscillatory characteristics of a real metal (zinc) were calculated from first principles on the basis of the theory of magnetic breakdown, taking account of the spin degrees of freedom of the conduction electrons by the ''effective path'' method (simple examples of this method are given in Refs. 1 and 12).

But the application of the effective-path method is complicated by the fact that for each particular real metal, depending of the geometry of the Fermi surface, a separate procedure must be developed for calculating the conductivity tensor. For this reason, this paper develops a more general method that makes it possible to unify this procedure for an arbitrary two-dimensional system of electronic trajectories arising under magnetic-breakdown conditions that possesses definite symmetry properties (and that is faithful to the topology of the Fermi surface). We underscore that the calculations of the components of the conductivity tensor in this case can be performed independently for each spin orientation. This is especially important in studying transport phenomena in the case of transition and ferromagnetic metals, where as a result of exchange and the spin–orbit interaction, the sheets of the Fermi surface with oppositely-directed spins are shifted.

We note that in Ref. 9 the method proposed here is il-

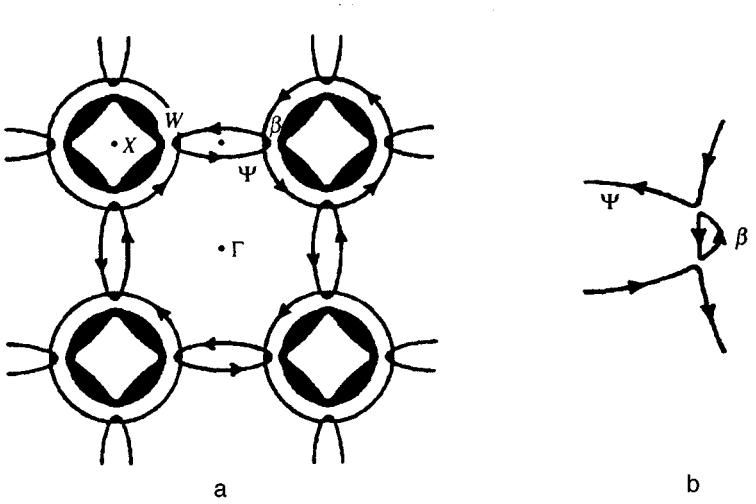

FIG. 1. Transverse sections of the Fermi surface in Al: a — from the second hole and third electron zones, b — enlarged views of the ψ and β orbits. The sections from the third zone are filled. The arrows show the direction of motion of the electrons.

lustrated in detail for a simple model of a metal with a periodic one-dimensional network of electronic trajectories associated with magnetic breakdown, taking account of the spin degrees of freedom of the conduction electrons.

The present paper is essentially a continuation and elaboration of the review in Ref. 9 for the case of a two-dimensional system of trajectories in Al. In Sec. 2 the required definitions and quantities, aiding in understanding the calculation of the conductivity tensor, are introduced on the basis of the theory of magnetic breakdown. Small β orbits are replaced by effective magnetic-breakdown sites (points). The effective breakdown probabilities calculated in Ref. 13 taking account of the electron spin and the spin–orbit interaction are used to describe the sites mathematically. The system of electronic trajectories in Al with $\mathbf{H} \parallel [001]$ is replaced by a two-dimensional reduced magnetic-breakdown network. In Sec. 3 a more general expression is obtained for the components of the conductivity tensor on the basis of the Boltzmann equation and the theory of stochastic magnetic breakdown. In Sec. 4 these components of the tensor are systematically calculated under the assumption that the electron motion on the two-dimensional reduced magnetic-breakdown network is stochastic and the motion on the β orbits is coherent. In Sec. 5 the required computational parameters are introduced, the components of the resistance tensor are calculated, and the results are compared with theoretical and experimental data.

2. MAGNETIC-BREAKDOWN NETWORK AND EFFECTIVE PROBABILITIES

It is well known[14,15] that many properties of metals in a magnetic field can be described by means of semiclassical equations. In this approximation the electrons move in the reciprocal space on trajectories which arise from the intersections of the Fermi surface and a surface perpendicular to the magnetic field $\mathbf{H}=(0,0,H)$

$$\varepsilon_{ms}(\mathbf{k})=\varepsilon_F, \quad k_z=k_{z0}=\text{const}, \tag{1}$$

where $\varepsilon_{ms}(\mathbf{k})$ is the dispersion relation, m is the zone number, s is the spin index, ε_F is the Fermi energy, and $\mathbf{k}$ is the wave vector. In Refs. 1 and 2 it was shown that the topology of the trajectories (1) changes under magnetic-breakdown conditions: a planar network consisting of sections of semiclassical motion, which belong to different zones and are associated with magnetic-breakdown sites, is formed. The probability of electron tunneling into a neighboring zone at a magnetic-breakdown site is $w=\exp(-H_0/H)$, where $H_0=H_0(\varepsilon_F, k_z)$ is the breakdown field. The network of trajectories which appears upon breakdown is called a magnetic-breakdown configuration or a magnetic-breakdown network.

When the spin–orbit interaction is included in the theory of magnetic breakdown,[9] sections with different projections of the electronic spins are united into a single magnetic-breakdown network. In this case the tunneling probability equals the sum of the probability w^0 of breakdown without spin flip and the probability w^s of breakdown with spin flip

$$w^0=\frac{w}{1+\alpha^2}, \quad w^s=\frac{\alpha^2 w}{1+\alpha^2}, \tag{2}$$

where α is the spin–orbit interaction parameter, which is determined by the ratio of the off-diagonal (in the zone number and spin index) matrix elements of the electron velocity operator near the maximum convergence of the trajectories.[9] An estimate of α shows that for a weak spin–orbit interaction $0 \leqslant \alpha \leqslant 1$.

The Fermi surface of Al is well known. It consists of a closed hole surface in the second zone and three electronic surfaces in the third zone, centered, respectively, at the points Γ and X of the Brillouin zone. The third zone forms closed ''quarter rings'' or ''sleeves'' which are arranged along the edges and connect near the points W (see Fig. 1a).

Electron tunneling during magnetic breakdown occurs between the ψ orbit of the second zone and the β orbit of the third zone, the splitting between which, shown in Fig. 1b, is due to the weak spin–orbit interaction.[8] This results in the formation of diverse closed and open orbits with two, four, or more magnetic-breakdown sites.[5] The enlarged and open orbits in turn change the period of the oscillations, decrease the amplitude of the de Haas–van Alphen effect, and produce large changes in the galvanomagnetic properties, mani-

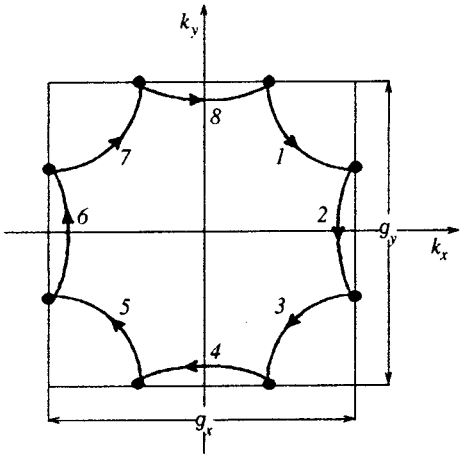

FIG. 2. Reduced two-dimensional magnetic-breakdown network within the first Brillouin zone, corresponding to Fig. 1a. The dots represent the effective magnetic-breakdown sites. The numbers denote sections of the magnetic-breakdown network.

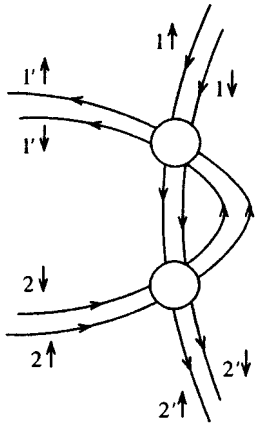

FIG. 3. Effective magnetic-breakdown site from a small β orbit, taking account of the electron spin, which connects the large ψ orbits (see Fig. 1). The arrows show the direction of electron motion and spin orientation.

fested in oscillations as a result of the quantum interference of the conduction electrons.

Figure 1 gives a schematic representation of the magnetic-breakdown network in Al in a magnetic field $\mathbf{H} \parallel [001]$. The network consists of a large closed hole ψ orbit and small electronic β orbits. One can see from Fig. 1 that the Larmor period $T_{c\beta}$ of electron motion on a β orbit is much shorter than the electron residence time T_c on large sections of the ψ orbit. A large difference of the periods is also observed in other metals (see, for example, Refs. 1, 2, and 9).

In such a situation it is possible to simplify the calculation of the conductivity tensor by applying the theory of stochastic magnetic breakdown.[1,2] In this theory it is assumed that small-angle scattering has no effect on the phase of the wave function of an electron in a small orbit, and on large sections of the trajectory it transforms the electron motion into a random walk. In other words, two types of electron motion exist in metals: coherent motion within small orbits and stochastic motion on large sections.

In this case a reduced magnetic-breakdown network, where small orbits are replaced by effective magnetic-breakdown sites, is used. Such a reduced two-dimensional magnetic-breakdown network for Al within the first Brillouin zone is shown in Fig. 2. Each effective magnetic-breakdown site ''switches'' the electron motion on the reduced magnetic-breakdown network. The corresponding effective breakdown probabilities are periodic functions of the phase acquired by an electron during coherent motion on a small orbit; this results in oscillations of the transport coefficients with a period corresponding to the area of the small orbit.[1,2,9]

In studying the electron dynamics in Al, taking account of the spin degrees of freedom and the spin–orbit interaction, we assume that there exist two independent reduced magnetic-breakdown networks (one embedded in the other), on which an electron moves with spin up on one and spin down on the other. In this case, for each reduced magnetic breakdown network we must know the effective probabilities of traversing small orbits or reflecting from them with their own direction of spin. These probabilities were obtained in Ref. 13 for a symmetric double-angle orbit with equivalent magnetic-breakdown sites. In the case of β orbits with equivalent magnetic-breakdown sites they have the same form

$$
\begin{aligned}
P^{\uparrow} &= \frac{w^2}{1+\alpha^2}\left(\frac{1}{|1-(1-w)\exp(i\gamma_\beta^{\uparrow})|^2}\right. \\
&\quad \left. + \frac{\alpha^2}{|1-(1-w)\exp(i\gamma_\beta^{\downarrow})|^2}\right), \\
P^{\downarrow} &= \frac{w^2}{1+\alpha^2}\left(\frac{\alpha^2}{|1-(1-w)\exp(i\gamma_\beta^{\uparrow})|^2}\right. \\
&\quad \left. + \frac{1}{|1-(1-w)\exp(i\gamma_\beta^{\downarrow})|^2}\right), \\
Q^{\uparrow,\downarrow} &= 1-P^{\uparrow,\downarrow},
\end{aligned}
\tag{3}
$$

where the arrows $\uparrow$, $\downarrow$ show the direction of the spin of an electron approaching an effective magnetic-breakdown site (see Fig. 3). The phases acquired by an electron on a β orbit have the form

$$
\gamma_\beta^{\uparrow\downarrow} = \frac{c\hbar}{eH} A_\beta(k_z) \pm \pi g^{\mathrm{eff}} + 2\Lambda, \tag{4}
$$

where $A_\beta(k_z)$ is the ''spinless'' area of the orbit and g^{eff} is the spin contribution to the phase and is called the effective g factor. The effective g factor is given by

$$
g^{\mathrm{eff}} = \frac{1}{2} g_\beta \frac{m_{c\beta}}{m_0}, \tag{5}
$$

where $m_{c\beta}$ and g_β are the electron cyclotron mass and the electron g factor associated with the β orbit, and m_0 is the free-electron mass. The plus sign in Eq. (4) corresponds to electron motion along a small orbit with spin up, and the minus sign corresponds to motion with spin down (see Fig. 3). The third term in Eq. (4) determines the jump in the phase of the wave function when the electron passes through

a magnetic-breakdown site, and Λ is a function of H and H_0 that is not important for us here; the exact form of this function is presented in Ref. 9.

In closing this section, we note that the introduction of effective breakdown probabilities (3) and the reduction of the magnetic-breakdown network to a reduced network with stochastic electron motion make it possible to study electron motion with spin up and down independently. For this reason, in what follows we calculate the conductivity tensor for a single reduced magnetic-breakdown network with a definite direction of spin. In so doing, we do not indicate the spin index $s=\uparrow,\ \downarrow$, but we average the final result over spin. Furthermore, in addition to the term *reduced magnetic-breakdown network*, we employ the term *magnetic-breakdown network*.

3. STATIC CONDUCTIVITY TENSOR UNDER STOCHASTIC MAGNETIC BREAKDOWN CONDITIONS

The stochastic character of the electron motion examined above makes it possible to calculate the conductivity tensor in the semiclassical approximation.[2] In this case magnetic breakdown starts to play the role of a stochastic factor, mixing the electrons over all sections of the magnetic-breakdown network. The time between two successive magnetic-breakdown scatterings plays the role of the free-flight time. It is determined by the characteristic cyclotron frequency ω_c and the corresponding magnetic-breakdown probabilities. The components $\sigma_{\alpha\beta}$ of the conductivity tensor depend on the topology of the magnetic-breakdown network. In the limiting cases $H \ll H_0$ or $H \gg H_0$ the sections of the magnetic-breakdown network transform into ordinary open or closed trajectories. Then the field dependence of $\sigma_{\alpha\beta}$ starts to determine the true momentum relaxation time τ.

The Boltzmann equation with a linearized collision integral corresponds to this picture. The general solution of this equation under the condition $\omega_c\tau \gg 1$ can be sought in the form

$$f_m(\mathbf{k}) = f_F(\mathbf{k}) - \frac{\partial f_F}{\partial \varepsilon} e\mathbf{E}\cdot\boldsymbol{\Psi}_m(\mathbf{k}), \tag{6}$$

where $f_m(\mathbf{k})$ is the nonequilibrium distribution function, $f_F(\mathbf{k})$ is the Fermi–Dirac distribution function, $\mathbf{E}$ is the electric field, $\boldsymbol{\Psi}_m(\mathbf{k})$is the vector distribution function of the electrons over the magnetic-breakdown network (analog of the density matrix). It is convenient to write the Boltzmann equation in the electric and magnetic fields in the variables ε, k_z, and t_m. Then to lowest order in $\boldsymbol{\Psi}_m(\mathbf{k})$ the equations acquire the form[14,15]

$$\frac{\partial \boldsymbol{\Psi}_i}{\partial t_i} + \hat{I}_k\{\boldsymbol{\Psi}_i\} = \mathbf{v}_i, \quad \hat{I}_k\{\dots\} = \left(\frac{\partial f_F}{\partial \varepsilon}\right)^{-1} \hat{I}\left\{\frac{\partial f_F}{\partial \varepsilon}\dots\right\}, \tag{7}$$

where $\boldsymbol{\Psi}_i$ are the values of the functions $\boldsymbol{\Psi}_m(\mathbf{k})$ on the sections of the magnetic-breakdown network, $i=i(m)$ is the number of the section, $\hat{I}$ is the linear collision integral, and $\mathbf{v}_i$ is the velocity vector.

Boundary conditions are required to solve Eqs. (7). If a trajectory for given values of ε and k_z is closed, then $\boldsymbol{\Psi}_i$ must obviously be a periodic function of t_i. If a trajectory is open, however, then the boundary condition is that $\boldsymbol{\Psi}_i$ remains finite at $t_i = \pm\infty$. When the magnetic field $\mathbf{H}$ is parallel to the z axis, the function $\boldsymbol{\Psi}_i$ should be periodic with the period of the reciprocal lattice. The difference from the semiclassical situation (see Ref. 14) lies in the fact that under stochastic magnetic-breakdown conditions the functions $\boldsymbol{\Psi}_i$ on different sections are related by the equations[2]

$$\boldsymbol{\Psi}_{i'}(0) = \sum_i w_{i'i}\boldsymbol{\Psi}_i(T_i), \tag{8}$$

where $\boldsymbol{\Psi}_{i'}(0)$ are the values of the functions $\boldsymbol{\Psi}_{i'}(\mathbf{k})$ on the initial sections leaving a magnetic-breakdown site; $w_{i'i}$ is the probability of magnetic breakdown; $\boldsymbol{\Psi}_i(T_i)$ are the values of $\boldsymbol{\Psi}_i(\mathbf{k}(t_i))$ at the ends of the sections entering a magnetic-breakdown site; and, T_i is the time during which an electron moves along the i-th section. The boundary conditions (8) show that the particle flux leaving a magnetic-breakdown site, for example, on the section $i'=1'$, consists of particles moving along incoming sections $i=1,\ 2$ with weights $1-w$ and w, respectively.

The solution of Eqs. (7) with the boundary conditions (8) makes it possible to obtain the components $\sigma_{\alpha\beta}$ of the conductivity tensor, which can be expressed in terms of $\boldsymbol{\Psi}_i$:

$$\sigma_{\alpha\beta} = \frac{e^2}{(2\pi)^3}\frac{|h|}{\hbar}\sum_i \int_0^\infty \left(-\frac{\partial f_F}{\partial \varepsilon}\right) d\varepsilon \times \int_{-k_0}^{k_0} dk_z \int_0^{T_i} v_i^\alpha(t_i)\Psi_i^\beta(t_i)\,dt, \tag{9}$$

where $h = eH/c\hbar$ and the summation over i extends up to N, the total number of nonequivalent sections of the magnetic-breakdown network. Here it must be noted that to obtain the total conductivity it is also necessary to sum over spin. At low temperatures ($T \approx 4.2$ K) it is assumed that $(-\partial f_F/\partial\varepsilon)$ in Eq. (9) behaves like a delta function with argument $\varepsilon - \varepsilon_F$; we assume that this is the case in what follows. However, this substitution is generally inadmissible, since the finiteness of the temperature T affects the characteristic electron energy k_BT, whose ratio to $\hbar\omega_\beta$, where ω_β is the cyclotron frequency of an electron in a β orbit, can be arbitrary. We take this circumstance into account at the end of the next section.

4. CALCULATION OF THE CONDUCTIVITY TENSOR OF A TWO-DIMENSIONAL MAGNETIC BREAKDOWN-NETWORK IN ALUMINUM

We apply the above theory to the reduced two-dimensional magnetic-breakdown network in Al, portrayed in Fig. 2. First, we obtain the dissipative component σ_{xx} of the conductivity tensor. For this, we must find the components $\boldsymbol{\Psi}_i$ as functions of the effective probabilities P and Q and the reciprocal-lattice vectors.

Neglecting the collision integral in Eqs. (7), we obtain

$$\Psi_i^x(t_i)=\Psi_i^x(0)-\frac{1}{h}[k_y^i(t_i)-k_y^i(0)],$$

$$\Psi_i^y(t_i)=\Psi_i^y(0)+\frac{1}{h}[k_x^i(t_i)-k_x^i(0)]. \tag{10}$$

Here $\Psi_i^x(0)$ and $\Psi_i^y(0)$ are the values of the distribution function at the start of the ith section; the second terms follow from the semiclassical equations of motion.[14]

At $t_i=T_i$ the expressions (10) assume the form

$$\Psi_i^x(T_i)=\Psi_i^x(0)+\frac{\Delta_i^y}{h}, \quad \Psi_i^y(T_i)=\Psi_i^y(0)-\frac{\Delta_i^x}{h}, \tag{11}$$

where Δ_i^x and Δ_i^y are the increments in the coordinates k_x and k_y as a result of the passage of an electron through the ith section.

We note that for the magnetic-breakdown networks of some metals one of the conditions

$$\sum_i^N \Delta_i^x=g_x\sum_i^N n_i=0, \quad \sum_i^N \Delta_i^y=g_y\sum_i^N n_i=0,$$

is often satisfied, or these two conditions are satisfied simultaneously, as in our case. Here g_x and g_y are components of a reciprocal lattice vector, and $n_i=0$ if i is the number of an interior section and $n_i=\pm 1$ if the section i intersects the boundary of the unit cell. For Al, obviously, $g_x=g_y=g$.

In the case at hand (see Fig. 2) the increments Δ_i^x and Δ_i^y in the coordinates are easy to find, provided that the coordinates of the start and end of each of the eight sections are known. For example, the x components (11) of four sections have the form

$$\Psi_1^x(T_1)=\Psi_1^x(0)+\delta, \quad \Psi_2^x(T_2)=\Psi_2^x(0)+2\delta,$$

$$\Psi_3^x(T_3)=\Psi_3^x(0)+\delta, \quad \Psi_4^x(T_4)=\Psi_4^x(0), \tag{12}$$

where the notation $\delta=g/4h$ is introduced to simplify the equations. The remaining x and y components $\mathbf{\Psi}_i(T_i)$ can be easily found by an elementary calculation or from the symmetry properties of the magnetic-breakdown network (see Fig. 2)

$$\mathbf{\Psi}_1+\mathbf{\Psi}_5=0, \quad \mathbf{\Psi}_3+\mathbf{\Psi}_7=0,$$

$$\mathbf{\Psi}_2+\mathbf{\Psi}_6=0, \quad \mathbf{\Psi}_4+\mathbf{\Psi}_8=0. \tag{13}$$

Using Eqs. (9) and (7) with $\hat{I}=0$ for a $2k_{zm}$ thick layer of the magnetic-breakdown network, taking account of the fact that the magnetic-breakdown parameters and the cross-sectional area A_β of a small β orbit are functions of k_z, we obtain

$$\sigma_{xx}=\frac{1}{2}\frac{e^2}{(2\pi)^3}\frac{|h|}{\hbar}\sum_{i=1}^{8}\int_{-k_{zm}}^{k_{zm}}dk_z\{[\Psi_i^x(T_i)]^2-[\Psi_i^x(0)]^2\}. \tag{14}$$

Now we write the boundary conditions (8) for the reduced magnetic-breakdown network with stochastic electron motion over large sections:

$$\mathbf{\Psi}_1(0)=P\mathbf{\Psi}_3(T_3)-Q\mathbf{\Psi}_4(T_4),$$

$$\mathbf{\Psi}_2(0)=Q\mathbf{\Psi}_1(T_1)-P\mathbf{\Psi}_2(T_2),$$

$$\mathbf{\Psi}_3(0)=Q\mathbf{\Psi}_2(T_2)-P\mathbf{\Psi}_1(T_1),$$

$$\mathbf{\Psi}_4(0)=Q\mathbf{\Psi}_3(T_3)-P\mathbf{\Psi}_4(T_4), \tag{15}$$

To obtain these relations we employed the periodic equivalence of the sections of the magnetic-breakdown network and the symmetry properties (13). Furthermore, these properties make it possible to decrease the number of terms in the sum over i in the expression (14). Substituting the expressions (15) and then the expressions (12) into Eq. (14) and using the fact that $P+Q=1$, we obtain an expression for σ_{xx} in terms of $\Psi_i^x(0)\equiv\Psi_i^x$:

$$\sigma_{xx}=2\frac{e^2}{(2\pi)^3}\frac{|h|}{\hbar}\int_{-k_{zm}}^{k_{zm}}dk_zQP[(\Psi_1^x+\Psi_2^x+3\delta)^2+(\Psi_3^x+\Psi_4^x+3\delta)^2]. \tag{16}$$

To determine the unknown quantities Ψ_i^x $(i=1, 2, 3, 4)$, the values of the functions (12) must be substituted for $\Psi_i^x(T_i)$ in the boundary conditions (15) and the system of four equations must be solved. The determinant of this system depends on the effective probabilities as $\Delta=Q^2+4P^2$, and the distribution functions at the start of the sections 1, 2, 3, and 4 have the form

$$\Psi_1^x=2\frac{\delta}{\Delta}Q(3P-1), \quad \Psi_2^x=4\frac{\delta}{\Delta}PQ-\delta,$$

$$\Psi_3^x=4\frac{\delta}{\Delta}P(1-3P)+\delta, \quad \Psi_4^x=2\frac{\delta}{\Delta}Q^2. \tag{17}$$

Substituting (17) into Eq. (16) we arrive at the expression

$$\sigma_{xx}=\frac{ec}{H}\frac{g^2}{(2\pi)^3}\int_{-k_{zm}}^{k_{zm}}dk_z\frac{2QP}{Q^2+4P^2}. \tag{18}$$

We now obtain the Hall component of the conductivity tensor. In terms of the unknown distribution functions $\Psi_i^x(t_i)$ and $\Psi_i^y(t_i)$ it has the form

$$\sigma_{yx}=\frac{e^2}{(2\pi)^3}\frac{|h|}{\hbar}\sum_{i=1}^{8}\int_{-k_{zm}}^{k_{zm}}dk_z\int_0^{T_i}\frac{\partial\Psi_i^y(t_i)}{\partial t_i}\Psi_i^x(t_i)\,dt_i. \tag{19}$$

Substituting (10) into Eq. (19) we obtain

$$\sigma_{yx}=\frac{e^2}{(2\pi)^3}\frac{|h|}{h}\sum_{i=1}^{8}\int_{-k_{zm}}^{k_{zm}}dk_z\left[(\Psi_i^y(T_i)-\Psi_i^y)\left(\Psi_i^x+\frac{k_y^i(0)}{h}\right)-\frac{1}{h^2}\int_{(\varepsilon,k_z)}k_y^i dk_x^i\right], \tag{20}$$

where $k_y^i(0)$ is the coordinate of the wave vector at the start of the ith section, and the second term in brackets determines the transverse cross-sectional area of the Fermi surface (see Eq. (1)).

To calculate the first term in Eq. (20), it is necessary to know the increment Δ_i^x in the coordinates k_x when an electron traverses the ith section. In a manner similar to the derivation of the equations (12), we obtain for the y component of the distribution functions

$$\Psi_1^y(T_1)=\Psi_1^y(0)+\delta, \quad \Psi_2^y(T_2)=\Psi_2^y(0),$$
$$\Psi_3^y(T_3)=\Psi_3^y(0)-\delta, \quad \Psi_4^y(T_4)=\Psi_4^y(0)-2\delta. \tag{21}$$

Using Eqs. (13), substituting (21) and (17) into Eq. (20), and expressing $k_y^i(0)/h$ in terms of δ (see Fig. 2), we obtain for the Hall component

$$\sigma_{yx}=\frac{ec}{H}\frac{1}{(2\pi)^3}\int_{-k_{zm}}^{k_{zm}}dk_z\left[\frac{4g^2P^2}{Q^2+4P^2}+A(\varepsilon,k_z)\right]. \tag{22}$$

Here $A(\varepsilon, k_z)$ is the cross-sectional area of the ψ orbit, taking account of the sign determined by the sign of the effective mass. For a ψ orbit in Al the effective mass is negative, $m_{c\psi}<0$; therefore a plus sign appears in the brackets in Eq. (22).

We note that if the calculation starts with the Hall component σ_{xy}, then it would be necessary to calculate the y component of the distribution functions Ψ_i^y at the start of the sections $i=1, 2, 3, 4$, substituting (21) into the boundary conditions (15). However, this is not necessary, since $\sigma_{xy}=-\sigma_{yx}$. This assertion can be proved by a direct calculation, and it also follows from the Onsager relations.[15] It follows from the calculation of Ψ_i^y and from the symmetry of the magnetic-breakdown network (see Fig. 2) that

$$\Psi_1^x=-\Psi_3^y, \quad \Psi_3^x=\Psi_1^y,$$
$$\Psi_2^x=-\Psi_4^y, \quad \Psi_4^x=\Psi_2^y. \tag{23}$$

To compare this result with the data for the real situation in Al, the expressions (18) and (22) must be modified somewhat. First, we take account of the contribution to σ_{xx} and σ_{yx} due to the corresponding conductivities arising from the electrons that do not participate in magnetic breakdown, i.e. electrons moving along closed trajectories. As is well known,[14] for uncompensated metals the components of the conductivity tensor for $\omega_c\tau\gg 1$ are determined by $\sigma_{xx}\sim a/H^2$ and $\sigma_{yx}\sim 1/RH$, where a is a constant and R is the Hall coefficient. Second, we take account of the temperature factor $A(T)=X/\sinh X$, where $X=147.0m_{c\beta}T/m_0H$ is a known expression (see, for example, Ref. 12). The factor $A(T)$ (see the remark after Eq. (9)) is always present in the general expressions for σ_{xx} and σ_{yx}, and it follows from the systematic quantum theory in the derivation of the oscillatory parts of the transport coefficients.[2] Third, we multiply (18) and (22) by 2 (the factor of 2 is taken into account in the standard derivation of σ_{xx} and σ_{yx}) in order to study the theory of magnetic breakdown in Al while neglecting the spin degrees of freedom of the conduction electrons. Finally, assuming that the density of electrons that move on a ψ orbit within a narrow magnetic-breakdown layer does not depend on k_z (the second term in Eq. (22)), we obtain finally

$$\sigma_{xx}=\sigma_{yy}=\frac{A(T)b}{Hk_{zm}}\int_0^{k_{zm}}\frac{2QP}{Q^2+4P^2}dk_z+\frac{a}{H^2}, \tag{24}$$

$$\sigma_{yx}=-\sigma_{xy}=\frac{A(T)b}{Hk_{zm}}\int_0^{k_{zm}}\frac{4P^2}{Q^2+4P^2}dk_z+\frac{1}{RH}, \tag{25}$$

where $b=4g^2k_{zm}ec/(2\pi)^3$ characterizes the unbalance between electrons and holes in the $2k_{zm}$ layer as a result of magnetic breakdown. We stress that the second term in Eq. (25) includes not only the conductivity arising from the closed ψ orbit in the magnetic breakdown layer but also the conductivity from closed orbits on the rest of the Fermi surface.

In summary, neglecting the electron spin and the spin–orbit interaction, for $g_\beta=0$ and $\alpha=0$ the formulas (24) and (25) can be used directly to calculate the field dependences of the components of the conductivity tensor of Al under magnetic breakdown conditions. For $g_\beta\neq 0$ and $\alpha\neq 0$ the entire calculation of σ_{xx} and σ_{yz} can be performed from the outset by equipping Q and P with a spin index. Then the magnetic breakdown terms in Eqs. (24) and (25) must be two times smaller.

5. MAGNETIC-BREAKDOWN OSCILLATIONS OF THE GALVANOMAGNETIC PROPERTIES OF ALUMINUM

In this section we shall obtain, using the theoretical and experimental data obtained in Refs. 3–8 and 16–18, theoretical relations for the galvanomagnetic characteristics of Al in order to substantiate the correctness of (24) and (25). We compare the theoretical relation for the magnetoresistance with the experimental curve,[4] and we also discuss the effect of the spin degrees of freedom of the conduction electrons on the period and amplitude of the magnetic-breakdown oscillations.

Substituting into Eqs. (24) and (25) the effective probabilities (3) with each direction of spin and averaging, we obtain the total conductivities, taking account of the electron spin and the spin–orbit interaction:

$$\sigma_{xx}=\frac{1}{2}(\sigma_{xx}^{\uparrow}+\sigma_{xx}^{\downarrow}), \quad \sigma_{yx}=\frac{1}{2}(\sigma_{yx}^{\uparrow}+\sigma_{yx}^{\downarrow}). \tag{26}$$

To calculate the integrals in Eqs. (24) and (25) within the magnetic-breakdown layer, we find the functions $H_0(k_z)$ and $A_\beta(k_z)$, using the results of Refs. 3 and 8. In Ref. 3, the cross sections of the Fermi surface of Al along k_z near the points W were calculated (see Fig. 1) neglecting the spin–orbit interaction, and plots of the cross-sectional areas of the β orbits and the width of the band gap as functions of k_z were presented. The latter makes it possible to determine the breakdown field in different sections in the direction k_z.

As follows from the general theory of magnetic breakdown with spin flip,[9] taking account of the spin–orbit interaction at the locations of breakdown results in a renormal-

ization of the breakdown field: $H_0=H_0^0/(1+\alpha^2)^{1/2}$, where H_0^0 is the breakdown field neglecting the spin–orbit interaction ($\alpha=0$). However, we do not employ this renormalization for H_0 in the calculations of Eq. (26), but instead we employ the results of Ref. 8, where the functions $H_0(k_z)$ and $A_\beta(k_z)$ with k_z oriented in the direction [001] were obtained from a calculation of the band structure of Al taking account of the spin–orbit interaction.

We expand the breakdown field $H_0(k_z)$ and the transverse cross-sectional area $A_\beta(k_z)$ to lowest order in k_z:

$$H_0(k_z)=H_0(0)+\frac{1}{2}H_0''k_z^2, \tag{27}$$

$$A_\beta(k_z)=A_\beta(0)+\frac{1}{2}A_0''k_z^2, \tag{28}$$

where $H_0(0)$ and $A_\beta(0)$ are the extremal values for $k_z=0$; H_0'' and A'' are the second derivatives with respect to k_z, calculated at $k_z=0$. We call H_0'' and A'' the variation parameter of the breakdown field and the curvature parameter of the β orbit in the direction k_z. Fitting the functions (27) and (28) with a half-width of the magnetic-breakdown layer $k_{zm}\approx(1.3-2.0)\times10^{-2}$ a.u. to the data of Refs. 3 and 8 gives $H_0''\approx4.64\times10^6$ kG/(a.u.)^2 and $A''\approx-(0.37-0.17)$.

We now determine the parameters a, b, and R in Eqs. (24) and (25). We start with the Hall coefficient R. The value of R is taken from Ref. 16, where it is shown that R depends on the magnetic field at $T=4.2$ K. For $H>5$ kG, R approaches 10.2×10^{-10} $\Omega\cdot$cm/kG, which agrees with the results obtained by other investigators (see, for example, Ref. 3). The parameters b and a can be determined in terms of the Hall coefficient. It follows from Ref. 17 that in strong fields with $\omega_c\tau\gg1$ the Hall coefficient in Al asymptotically approaches $3/ecn_{\text{eff}}$, where n_{eff} is the effective number of carriers per unit cell. We assume that this is the number of electrons participating in breakdown in the layer $2k_{zm}$. On the other hand, it follows from the definition of b (see above) that $b=ecn_{\text{eff}}$. Equating the effective electron densities, we obtain $b=3/R$.

To determine a and check the validity of (24) and (25), we find the components of the resistance tensor, which, as is well known,[14] are given by

$$\rho_{xx}=\frac{\sigma_{xx}}{\sigma_{xx}^2+\sigma_{yx}^2},\quad \rho_{yx}=\frac{\sigma_{yx}}{\sigma_{xx}^2+\sigma_{yx}^2}. \tag{29}$$

Substituting the conductivity from closed orbits (the second terms in Eqs. (24) and (25)) into ρ_{xx} and neglecting the infinitesimal $(aR/H)^2$ for $H>H_0$, we obtain $\rho_{xx}\sim aR^2$. The field dependence of the magnetoresistance corresponding to closed orbits is shown in Ref. 4. In fields up to 60 kG it remains constant with $\rho_{xx}\sim3.5\rho_0$, where ρ_0 is the resistance at $H=0$. At $T=4.2$ K $\rho_0\approx1\times10^{-10}$ $\Omega\cdot$cm, which, as indicated in Refs. 3 and 4, agrees well with existing calculations. Comparing the results for ρ_{xx}, we obtain $a\approx3.5\rho_0/R^2$.

In summary, the technology for calculating the galvanomagnetic characteristics of Al under magnetic-breakdown conditions, taking account of the spin degrees of freedom of the conduction electrons, has been reduced to numerical modeling of the theoretical magnetic-field dependences of the quantities (29) and finding the microscopic parameters of the model. The experimental curve from Ref. 4 (see Fig. 4) was used to improve the values of the parameters.

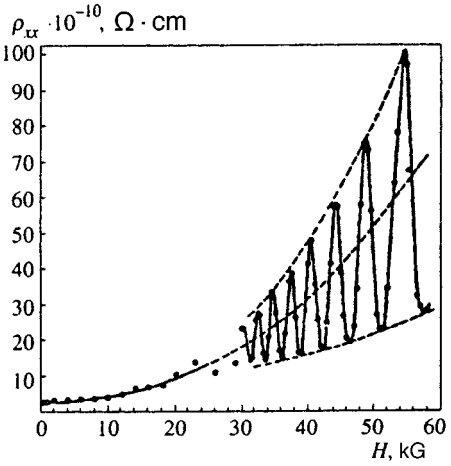

FIG. 4. Experimental oscillatory curve of the magnetoresistance of Al at $T=4.2$ K with current $\mathbf{j}\perp\mathbf{H}$, $\mathbf{H}\parallel[001]$.[4]

Such a calculation was performed at $T=4.2$ K, with cyclotron mass $m_{c\beta}=0.102m_0$ for an electron in a β orbit,[2,6] using the known values of the parameters (see Table I).

A least-squares fit of ρ_{xx} (see Fig. 5a) to the experimental magnetoresistance (Fig. 4) was made for a wide range of values of the microscopic parameters. In so doing, the simplex method was used to find the optimal values of the parameters. Equations (24) and (25) were integrated numerically by Simpson's method. The relative error did not exceed 1%, and the number of layers in the interval $0<k_z<k_{zm}$ reached 400.

Curves of the magnetic-breakdown oscillations of the magnetoresistance ρ_{xx} and Hall resistance ρ_{yx} with $g^{eff}=0$ and $\alpha=0$ are displayed in Fig. 5. One can see from Fig. 5a that the field dependence of ρ_{xx} is qualitatively identical to the experimental curve (see Fig. 4), and the parameters agree quantitatively with the published data (see Table I), which supports the model adopted. The parameters H_0'' and A'' have the greatest effect on the amplitude and form of the oscillations of ρ_{xx}. For example, exponential growth of the amplitude of the oscillations does not occur for $H_0''=0$; the curve

TABLE I.

Computational parameters	Published data	Theoretical curves	
		Fig. 5	Fig. 6
$a\times10^{10}$, kG$^2/\Omega\cdot$cm	0.034	0.034	0.034
$b\times10^{10}$, kG$^2/\Omega\cdot$cm	0.29	0.14	0.16
$R\times10^{-10}$, $\Omega\cdot$cm/kG	10.2 (Refs. 3, 16, and 17)	10.2	10.2
$H_0(0)$, G	3.6–4.0 (Refs. 2 and 8)	3.6	3.6
$H_0''\times10^6$, kG/(a.u.)^2	≈4.8 (Ref. 8)	2.16	4.88
$A_\beta(0)\times10^{-3}$, (a.u.)^2	1.24 (Refs. 3, 7, and 8)	1.24	0.65
A''	≈−3.3 (Ref. 8)	−1.74	−1.23
$k_{zm}\times10^{-2}$, a.u.	2.0 (Refs. 4 and 8)	2.0	2.0
g^{eff}	–	0	0.49
α	–	0	≈0

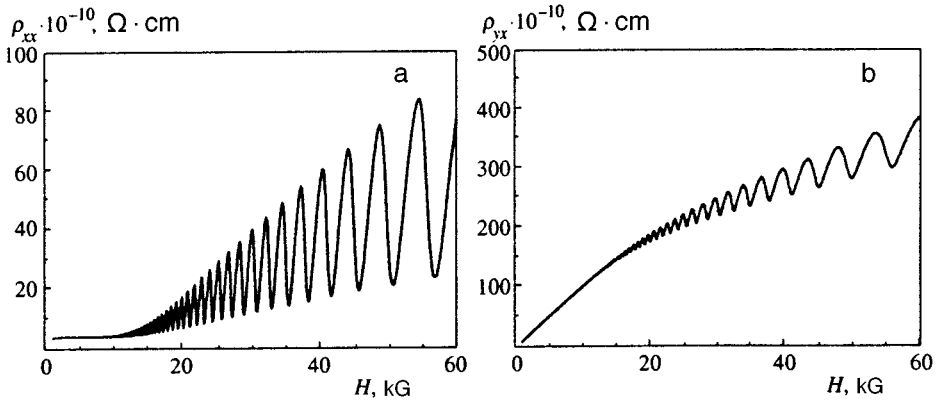

FIG. 5. Theoretical curves of magnetic-breakdown oscillations in Al at $T=4.2$ K, $\mathbf{H} \parallel [001]$: a — magnetoresistance $\rho_{xx}=\rho_{yy}$, b — Hall resistance $\rho_{yx}=-\rho_{xy}$. The parameters for which the curves were obtained are presented in Table I.

of ρ_{xx}, oscillating, saturates below 20×10^{-10} Ω · cm. For $A''=0$ the peaks of the oscillations are asymmetric.

The oscillatory curves of ρ_{xx} and ρ_{yx}, corresponding to Fig. 5, with $g^{\text{eff}}=0.49$ and $\alpha\approx 0$ are shown in Fig. 6. One can see from Fig. 6a that qualitative agreement with the experimental curve obtains in this case as well, but the amplitudes of the oscillations are somewhat smaller than in Fig. 5a. If α falls in the range $0<\alpha\leqslant 1$, then the oscillations of ρ_{xx} are less symmetric, but the average values of the oscillations are shifted upwards, approaching the average values of the experimental curve. In order for the periods of the oscillations of ρ_{xx} in Figs. 6a and 4 to be the same, the experimental transverse cross-sectional area $A_\beta(0)$ must be halved. If for $g^{\text{eff}}=0.49$ the area $A_\beta(0)=1.24\times 10^{-3}$ (a.u.)^2 is retained, which corresponds to the ordinarily observed period $\Delta(1/H)=2.15\times 10^{-6}$ G^{-1}, then the number of peaks in ρ_{xx} and ρ_{yx} doubles. Therefore, taking account of the g factor of conduction electrons in a β orbit changes the extremal area $A_\beta(0)$, provided that the expansion (28) is valid.

It can be assumed that a doubled structure of the oscillation peaks, which arises on account of the lifting of the double spin degeneracy of the Landau levels by the magnetic field, is observed in experiments. This can be verified by studying the experimental curve (Fig. 4) more carefully. The amplitudes of two successive oscillation peaks differ by a small amount. Furthermore, we note that a situation where the spin splitting exactly equals the splitting between the Landau levels is possible. This is approximately the case for energy bands where $m_c/m_0\ll 1$.[12]

Knowing the cyclotron mass $m_{c\beta}=0.102m_0$, the value of the g factor of an electron in a β orbit can be estimated from the value obtained for g^{eff} (see Table I): $g_\beta\approx 9.8$. However, there is still some uncertainty here, since the spin contribution to the phase (4) can only be found to within $2\pi n$, where n is an integer.

The effect of temperature is observed in the monotonic part and in the character of the oscillations in ρ_{xx} below 30 kG. At temperatures $T<4.2$ K the intensity of the oscillations increases and the average field dependence of ρ_{xx} becomes linear. At temperatures above 4.2 K the oscillatory part becomes small, in agreement with the results of the temperature investigations.[4] Therefore a comparison of the theoretical field-dependences of ρ_{xx} in Figs. 5a and 6a with the experimental curve shows that the introduction of the temperature factor $A(T)$ in our model is justified.

Qualitative agreement with experiment also obtains for the Hall effect, shown in Figs. 5b and 6b, computed with $\mathbf{H} \parallel [001]$. In both cases the Hall resistance ρ_{yx} oscillates weakly about a value proportional to the field, but since $g^{\text{eff}} \neq 0$ for $0<\alpha\leqslant 1$, as one can see from Fig. 6b, the average magnetic field dependence of ρ_{yx} is more linear, which agrees better with the results of Refs. 3 and 18.

In summary, the plots presented show that a systematic theory of magnetic breakdown that takes account of the spin of the conduction electrons and the spin–orbit interaction, as well as the more general method presented above for calculating the conductivity tensor for a two-dimensional system of electronic trajectories in Al, also explain well not only the

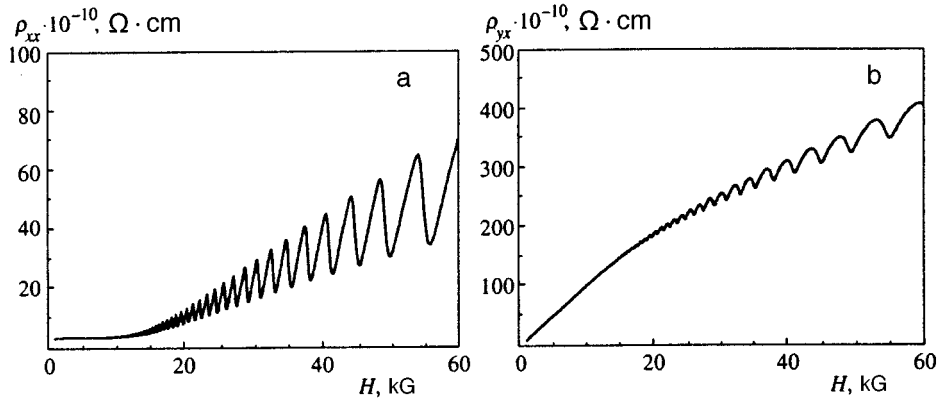

FIG. 6. Theoretical curves of magnetic-breakdown oscillations in Al at $T=4.2$ K, $\mathbf{H} \parallel [001]$. The curves correspond to Fig. 5 but were obtained taking account of the electron spin; the computational parameters are given in Table I.

behavior of the experimental curves,[3–5,7] but they also yield surprising (for such complicated oscillating functions (25) and (26)) qualitative agreement with experiment. The facts presented show that in observations of the magnetic-breakdown oscillations of the transport and thermodynamic properties of Al, the electron spin and the spin–orbit interaction strongly influence the character of the magnetic-field dependences.

It is my pleasant duty to thank B. I. Kochelaev and Yu. N. Proshin for their interest in this work and for providing the opportunity to serve a scientific apprenticeship in the Department of Theoretical Physics at the Kazan' University.

[1] R. W. Stark and L. M. Falicov, Progr. Low Temp. Phys. **5**, 235 (1967).
[2] M. I. Kaganov and A. A. Slutskin in *Conduction Electrons* [in Russian], Nauka, Moscow (1985), p. 101; N. E. Alekseevskiĭ and V. I. Nizhankovskiĭ, ibid., p. 197.
[3] R. J. Balcombe and R. A. Parker, Phil. Mag. **21**, 533 (1970).
[4] V. N. Morgun, V. I. Khorkevich, N. N. Chebotaev, and V. A. Bondar', Fiz. Nizk. Temp. **7**, 1301 (1976) [Sov. J. Low Temp. Phys. **7**, 634 (1976)].
[5] W. Kesternich and C. Papastaikoudis, J. Phys. F: Metal Phys. **7**, 837 (1977).
[6] V. I. Gostishchev, M. A. Glin'skiĭ, A. A. Drozd, and S. E. Dem'yanov, Zh. Éksp. Teor. Fiz. **74**, 1102 (1978) [Sov. Phys. JETP **47**, 579 (1978)].
[7] H. Hosoda, Y. Ueda, amd T. Kino, J. Phys. Soc. Jpn. **56**, 2858 (1987).
[8] G. G. Lonzarich and P. M. Holtham, Proc. R. Soc. Lond. A **400**, 145 (1985).
[9] Yu. N. Proshin and N. Kh. Useinov, Usp. Fiz. Nauk **165**, 41 (1995).
[10] F. Beuneu, J. Phys. F: Metal Phys. **10**, 2875 (1980).
[11] Yu. N. Proshin and H. Kh. Useinov, Zh. Éksp. Teor. Fiz. **100**, 1088 (1991) [Sov. Phys. JETP **73**, 602 (1991)].
[12] D. Shoenberg, *Magnetic Oscillations in Metals*, Cambridge University Press, N. Y. (1984)
[13] Yu. N.Proshin and N. H. Useinov, Phys. Stat. Sol. B **166**, 173 (1991).
[14] I. M. Lifshitz, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals*, Consultants Bureau, N. Y. (1973).
[15] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, North-Holland, Amsterdam (1988).
[16] R. D. Barnard and A. E. E. Abdel Rahiem, J. Phys. F: Metal Phys. **10**, 2739 (1980).
[17] R. Lück, Phys. Stat. Sol. **18**, 49 (1966).
[18] J. Feder and J. Lothe, Phil. Mag. **12**, 107 (1965).

Translated by M. E. Alferieff

Modification of the spin-wave dispersion law in multilayer films by changes in the symmetry of the boundary conditions

A. M. Zyuzin and A. G. Bazhanov

N. P. Ogaryov Mordovian State University, 430000 Saransk, Russia
(Submitted 28 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1667–1673 (May 1997)

The present paper is the first attempt to study the transformation of spin-wave resonance spectra when symmetric boundary conditions are smoothly replaced by asymmetric. The transition is done by gradually reducing the thickness of one of the layers in a three-layer film. Spin deexcitation is caused by a dissipation mechanism. We find that in the transition region between symmetric and asymmetric boundary conditions the dispersion curve experiences a break, whose position depends on the degree of deexcitation (the thickness of the upper layer). The break is caused by the appearance of asymmetric transitional spin-wave modes, which cannot be excited under symmetric boundary conditions. © *1997 American Institute of Physics.*
[S1063-7761(97)01005-6]

The boundary conditions constitute the most important factor determining the characteristics of spin-wave resonance spectra. This problem has been studied by many researchers (see, e.g., Refs. 1–3), who, in particular, investigated spin-wave resonance spectra for the following boundary conditions: symmetric, asymmetric (spins deexcited only at one boundary of the excitation layer), and antisymmetric. The analysis of the spectra in the majority of the papers is based on a surface anisotropy model, in which the degree of deexcitation is described by a phenomenological parameter, the surface anisotropy,[4,5] which, however, is difficult to measure and monitor. Such analysis does not allow for such physical parameters as the resonant field, the magnetization, the damping constant, and the film thickness (which is assumed to be negligible).

Schlömann[6] studied the dynamical mechanism of spin deexcitation, which is related to the nonuniformity of the distribution of the resonant field over the film thickness. The most detailed analysis of spin-wave resonance spectra for multilayer films has been done by Wiltz and Prasad[7] and Hoecstra *et al.*,[8] who found that under the dynamical mechanism there is a strong dependence of the spectrum on the orientation of the external magnetic field **H** relative to the film, and a deviation of the dispersion law from quadratic. The reason is the dependence of the spin-wave localization region on H and on the orientation of **H**. In Refs. 4, 9 and 10 and in a number of other papers, spin-wave resonance was studied in films with fluctuations of the magnetization and the exchange parameter, and in multilayer films with one-dimensional modulation of the magnetic parameters. The presence of fluctuations leads to a break in the dispersion curve, with the wave vector at the break related to the correlation radius of the fluctuations or the spatial modulation parameter.

Despite the large number of papers on spin-wave resonance, the features of the spectra in the transitional region between symmetric and asymmetric boundary conditions have yet to be studied. Such a situation is quite common, however. We also believe that it is extremely important, when analyzing spin-wave resonance spectra, to allow for the dependence of the degree of deexcitation on the values of the wave number.

The spin-wave resonance spectra for symmetric boundary conditions differ from those for antisymmetric boundary conditions, in that in the former case a uniform microwave field excites standing harmonic modes with an odd number of half-waves fitting into the thickness of the excitation layer, while in the latter case such a field excites modes with an odd number of quarter-waves. Hence, other things being equal, twice the number of spin-wave modes are excited in the same interval of wave numbers k in the latter case as in the former.

The decrease in the degree of deexcitation at one of the boundaries of an excitation layer with initially symmetric boundary conditions should lead to a transitional situation and to the emergence of previously ''forbidden'' asymmetric modes, with the configuration transformed in such a way that the total variable magnetic moment becomes nonzero. It is evident, then, that the earlier excited, symmetric, modes must also undergo a transformation in this case.

The aim of our research was to study the transformation of spin-wave resonance spectra and the modification of the dispersion law for spin waves when symmetric boundary conditions gradually become asymmetric. The transition was done by lowering the degree of deexcitation at one of the boundaries of the excitation layer via gradual reduction of the thickness of the upper layer with strong damping.

The studies involved three-layer single-crystal films of ferrite garnets, where spin deexcitation is achieved by a dissipation mechanism.[11] The films were grown by the liquid-phase epitaxy method on a (111) gadolinium–gallium garnet substrate by subsequent immersion into two differing melts. To measure and monitor the parameters, we grew one-layer analogs of each layer of the three-layer film on pure substrates. The lower and upper layers (the deexcitation layers), which had a large damping parameter $\alpha=\Delta H\gamma/\omega=0.2$ (here ΔH is the absorption-line halfwidth, γ is the gyromagnetic ratio, and ω is the circular frequency of the microwave field), had the composition $(SmEr)_3Fe_5O_{12}$, thickness $h=0.74$–1.2μm, saturation magnetization $4\pi M=1328$ G, effective uniaxial anisotropy field $H_k^{\text{eff}}=96$ Oe, and

 1063-7761/97/050912-04$10.00

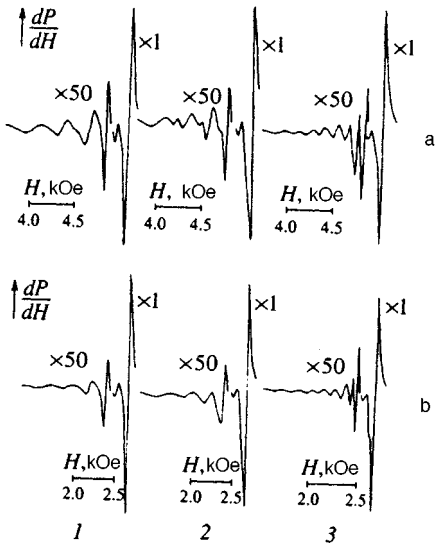

FIG. 1. Changes in the spin-wave resonance spectrum induced by a decrease in the thickness of the upper deexcitation layer: *1*—$h=0.74$ μm, *2*—$h=0.042$ μm, and *3*—the layer has been etched away completely; (a) perpendicular orientation of **H** relative to the plane of the film, and (b) parallel orientation.

$\gamma=1.38\times10^7$ Oe^{-1}s^{-1}. The middle layer (where standing harmonic modes are excited) had the composition $Y_{2.98}Sm_{0.02}Fe_5O_{12}$, thickness ranging from 0.6 to 0.95 μm for different samples, $\alpha=0.003$, $4\pi M=1730$ G, $\gamma=1.76\times10^7$ Oe^{-1}s^{-1}, and $H_k^{\text{eff}}=-1715$ Oe. The thickness of the various layers was measured interferometrically using the one-layer analogs. The etching rate was estimated by the time it took to etch away the layer completely. The etching step amounted to 0.014 μm. The spectra were recorded at room temperature at a frequency of 9.34×10^9 Hz. Since the spin-wave resonance spectra were recorded at constant ω, and the difference in the resonant fields of the zeroth and nth modes, H_0-H_n, is quadratic in k (like the frequency of spin waves for constant H), the dispersion curves were plotted in the $\{H_0-H_n,(2n+1)^2\}$ plane, where n is the mode number.[8,12,13]

Six spin-wave modes were confidently registered in the spectrum of the initial three-layer field with symmetric boundary conditions, in both perpendicular and parallel orientations (Fig. 1, *1*). The dispersion relations for the two orientations were found to be linear (Fig. 2, *1*). In the process of etching the upper level down to 0.07 μm, no appreciable variations were observed. When the layer was made even thinner, the spin-wave resonance spectrum for the case of perpendicular orientation was found to acquire peaks between those that existed earlier (Fig. 1, *2*).

These new peaks emerged in the following manner. The first peaks to appear were those with higher numbers n. They grew in height as the upper layer became thinner, and then new modes with lower numbers n appeared. All this was accompanied by a slight increase in the resonant field strength and a decrease in the height of the peaks that had appeared earlier. As a result, the monotonic dependence of the amplitude of spin-wave modes on the mode number disappeared.

There is a dramatic effect accompanying this process. Within the transitional region between symmetric and asymmetric boundary conditions, a break appears in the dispersion curve (Fig. 2, *2*), and it moves toward smaller values of n as the thickness of the upper layer and hence the degree of deexcitation become smaller. The break (and this can be verified) is due to the emergence of previously ''forbidden'' intermediate modes. When the upper layer is etched away completely, all intermediate modes are present (Fig. 1, *3*), the distribution of the amplitude becomes monotonic, and the break in the dispersion curve disappears (Fig. 2, *3*). The slope of the entire dispersion curve built on the basis of the functional dependence corresponding to symmetric boundary conditions,

$$H_0-H_n=\frac{2A}{M}(2n+1)^2\frac{\pi^2}{h^2},$$

changes. Clearly, the modification of the dispersion law is due to the transition from symmetric boundary conditions to asymmetric, rather than to a change in $(2A/M)(\pi^2/h^2)$ or some other influence. The dispersion curves set up for asymmetric boundary conditions, for which the wave number k assumes values $(n+\frac{1}{2})\pi/h$ rather than $(2n+1)\pi/h$ ($n=0,1,2,\ldots$ is the mode number), coincide to high accuracy with the original curves set up for symmetric boundary conditions.

Note that the number of modes was found to be the same for both perpendicular and parallel orientations, but certain

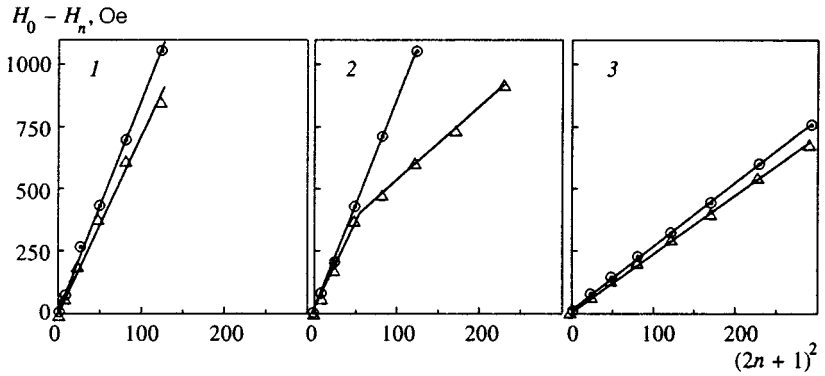

FIG. 2. Changes in the H_0-H_n vs $(2n+1)^2$ dependence induced by a decrease in the thickness of the upper deexcitation layer. The dispersion curves *1*, *2*, and *3* correspond to the spectra *1*, *2*, and *3* in Fig. 1. △—perpendicular orientation, and ⊙—parallel orientation.

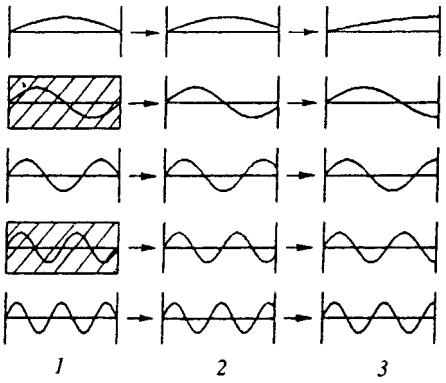

FIG. 3. Transformation of the first five spin-wave modes as a result of transition from symmetric boundary conditions to asymmetric. The hatched areas depict modes that are not excited when the boundary conditions are symmetric.

differences in the slopes of the dispersion curves were present. Establishing the reason for these differences, related to the additional effect of the fact that the deexcitation layer is in one case a dispersive medium and in the other a reactive medium, requires further analysis.

Another interesting result of our investigation is the qualitatively different way in which the intermediate modes for one orientation appear as compared with those for the other. While for perpendicular orientation the modes, as described earlier, emerged sequentially starting with those with higher numbers, for parallel orientation no such pattern was observed, up to approximately the middle of the process of mode emergence for perpendicular orientation. Only then did all the intermediate modes suddenly appear, with a monotonic intensity distribution.

In analyzing the results we approximated spin waves by harmonic waves in the excitation layer and exponentially damped waves in layers with large α. The intensity of excitation of spin-wave modes in a linearly polarized microwave field can be described by[1,7]

$$I_n \sim \frac{\left(\int m_x\, dz\right)^2}{\int (\alpha/2\gamma M)(m_x^2+m_y^2)\, dz}, \tag{1}$$

where m_i are the components of the variable magnetization in a system of coordinates in which the direction of **M** coincides with that of the z axis and the direction of **h** with that of the x axis.

Using Eq. (1), we can easily show that in the case of symmetric boundary conditions, only modes with an odd number of half-waves in the thickness of the excitation layer (Fig. 3, *1*) have nonzero intensity when the total variable magnetic moment is nonzero. Modes with an even number of half-waves (asymmetric modes, depicted in Fig. 3 by hatched rectangles) are not excited.

The damped spin wave in the deexcitation layer is characterized by a certain penetration depth l, which generally depends on the wave number in the excitation layer, $k_e \sim n$. A drop in the thickness of the upper layer to values comparable to, or less than, l leads to a drop in the degree of deexcitation at the corresponding boundary of the excitation layer. This in turn changes the wave numbers of the harmonic spin waves and their phases at the given boundary (Fig. 3, *2*). As a result, the total variable magnetic moment of the spin-wave modes that earlier were asymmetric becomes nonzero, with the corresponding peaks appearing in the spin-wave resonance spectra. This is accompanied by an increase in the resonant field strengths and a decrease in the intensity of the peaks of the modes that were symmetric under symmetric boundary conditions. Figure 3 depicts the change in configuration of a few first spin-wave modes when symmetric boundary conditions are replaced by asymmetric.

To explain the change in the way in which intermediate modes appear when there is a change in orientation, we allowed not only for the dissipation mechanism of spin deexcitation, but also for the dispersive and reactive properties of layers with a large α. Depending on the orientation in the range of field strengths that excite spin-wave modes, the deexcitation layer is a reactive or dispersive medium with strong dissipation for spin waves excited by standing harmonic waves localized in the excitation layer.

Figure 4 depicts the H-dependence of the wave vectors in the excitation and deexcitation layers with no allowance for dissipation. The calculations of k for each layer were done via the dispersion relations

$$\frac{\omega}{\gamma} = H + H_k^{\text{eff}} - \frac{2H_{k1}}{3} + \frac{2A}{M}k^2, \tag{2}$$

$$\left(\frac{\omega}{\gamma}\right)^2 = \left(H + \frac{2A}{M}k^2\right)\left(H - H_k^{\text{eff}} - \frac{H_{k1}}{2} + \frac{2A}{M}k^2\right), \tag{3}$$

for the perpendicular and parallel orientations, respectively. Here H_{k1} is the cubic anisotropy field, which was found by

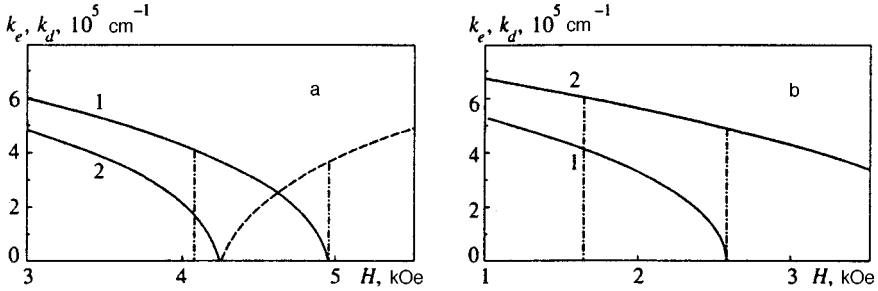

FIG. 4. The H-dependence of wave numbers in an excitation layer (k_e, *1*) and a deexcitation layer (k_d, *2*) for parallel (a) and perpendicular (b) orientations. The dashed curve corresponds to imaginary values of k_d, and the dot-dash straight lines specify an interval of field strengths corresponding to the observed spin-wave resonance spectrum.

the method described in Ref. 14, and A is the exchange constant. For perpendicular orientation, the wave vector in the deexcitation layer with fields higher than that needed for a homogeneous resonance in this layer (H_0), is imaginary:

$$\mathbf{k}_d = -i\mathbf{k}_d''$$

(the spin wave is an exponentially damped wave). For parallel orientation of the field,

$$\mathbf{k}_d = \mathbf{k}_d'$$

is a real quantity (a harmonic wave) over the entire range of field strengths.

Allowing for the dissipative properties of the deexcitation layer leads to a situation in which for perpendicular orientation the damping coefficient for the spin wave increases, while for parallel orientation $\mathbf{k}_d$ becomes complex-valued, and the spin wave becomes an exponentially damped harmonic wave.

Figure 4a implies that for perpendicular orientation, as the wave number k_e of the harmonic wave excited in the middle layer increases, the strength of the external magnetic field H approaches H_0 in the deexcitation layer, which makes the layer less reactive. Clearly, there is a sharp decrease in k_d, which corresponds to an increase in the penetration depth. For this reason the modes with the higher values of n are the most sensitive to the thickness of the deexcitation layer. Note that as k_e increases, the amplitude of the variable magnetization of the harmonic spin wave is reduced and, simultaneously, as the strength of the external magnetic field H approaches H_0 in the deexcitation layer, the variable magnetization in this layer excited directly by the microwave field grows, i.e., the difference in the magnetization precession angles of two media becomes smaller. All this causes the degree of deexcitation to decrease as the mode number increases.

For parallel orientation, the relative variation of k_d' within the range of field strengths corresponding to the observed spin-wave resonance spectrum (Fig. 4b) is moderate, and amounts to roughly 20%. The depth l of penetration of the deexcitation layer by the spin wave (determined by the dissipative properties),[12]

$$l = \frac{1}{k_d''} = \frac{4A\gamma}{\alpha\omega M} k_d', \tag{4}$$

is therefore only slightly dependent on the mode number. In contrast to the purely exponential wave (perpendicular orientation), the exponentially damped harmonic wave in the deexcitation layer for parallel orientation of the field has a node near the boundary between layers. This leads to a considerably weaker dependence of the degree of deexcitation on the mode number and a lower sensitivity of deexcitation to the thickness of the deexcitation layer. Only when the thickness h of the upper layer is diminished to values comparable to or smaller than l is there a sharp decrease in the degree of deexcitation, which is accompanied by a change in the configuration of the spin wave near the boundary between the layers.

Estimates support this conclusion. All the intermediate modes emerged when the thickness h of the upper deexcitation layer became approximately equal to 0.03 μm, which amounts to roughly half of l, or a quarter of the wavelength λ in this layer. Etching with a smaller step would probably lead not to a sudden rise in the intensities of all intermediate modes, but to a more gradual rise. Note that for parallel orientation, the spin waves are transverse–longitudinal, and are generally elliptically polarized.

The large damping coefficients for modes with a small n in the event of perpendicular orientation ensure fairly strong deexcitation for thinner deexcitation layers.

Thus, on the basis of our results, we can draw the following conclusions.

(1) The modification of the law of dispersion of the spin-wave spectrum (in particular, the appearance of a break in the dispersion curve) may be due only to fluctuations in A and M, but also to a situation in which the boundary conditions are in the ''transition'' region between symmetric and asymmetric boundary conditions.

(2) The influence of the dispersive or reactive properties of a layer with a large α, in addition to the dominant dissipation mechanism of deexcitation, leads to a qualitative change in the way in which asymmetric intermediate modes appear as the thickness of this layer decreases.

Note that an intermediate situation similar to the one discussed above can emerge for antisymmetric boundary conditions. The reason may be the small thickness of one of the layers or another factor leading to an inequality (in absolute values) between the degrees of deexcitation at the boundaries.

[1] N. M. Salanskiĭ and M. Sh. Erukhimov, *Physical Properties and Applications of Magnetic Films* [in Russian], Nauka, Novosibirsk (1975).
[2] V. M. Sokolov and B. A. Tavger, Fiz. Tverd. Tela (Leningrad) **10**, 1793 (1968) [Sov. Phys. Solid State **10**, 1412 (1968)].
[3] Yu. A. Korchagin, R. G. Khlebopros, and N. S. Chistyakov, Fiz. Tverd. Tela (Leningrad) **14**, 2121 (1972) [Sov. Phys. Solid State **14**, 1826 (1972)].
[4] V. A. Ignatchenko, R. S. Iskhakov, L. A. Chekanova, and N. S. Chistyakov, Zh. Éksp. Teor. Fiz. **75**, 653 (1978) [Sov. Phys. JETP **48**, 328 (1978)].
[5] G. Suran, H. Daver, and J. Sztern, in *AIP Conf. Proc.*, Vol. 34, AIP, New York (1976), p. 310.
[6] E. Schlömann, J. Appl. Phys. **35**, 159 (1964).
[7] C. H. Wiltz and S. Prasad, IEEE Trans. Magn. **MAG-17**, 2405 (1981).
[8] B. Hoecstra, R. P. van Stapele, and J. M. Robertson, J. Appl. Phys. **48**, 382 (1977).
[9] V. A. Ignatchenko and R. S. Iskhakov, Zh. Éksp. Teor. Fiz. **72**, 1005 (1977) [Sov. Phys. JETP **45**, 526 (1977)].
[10] R. S. Iskhakov, A. S. Chekanov, and L. A. Chekanova, Fiz. Tverd. Tela (Leningrad) **32**, 441 (1990) [Sov. Phys. Solid State **32**, 255 (1990)].
[11] A. M. Zyuzin, N. N. Kudel'kin, V. V. Randoshkin, and R. V. Telesnin, Pis'ma Zh. Tekh, Fiz. **9**, 177 (1983) [Sov. Tech. Phys. Lett. **9**, 78 (1983)].
[12] A. G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* [in Russian], Nauka, Moscow (1973).
[13] R. S. Iskhakov, M. M. Brushtunov, and A. S. Chekanov, Fiz. Tverd. Tela (Leningrad) **29**, 2699 (1987) [Sov. Phys. Solid State **29**, 1553 (1987)].
[14] A. M. Zyuzin and Al. M. Zyuzin, Fiz. Tverd. Tela (Leningrad) **29**, 3128 (1987) [Sov. Phys. Solid State **29**, 1795 (1987)].

Translated by Eugene Yankovsky

Long-range intensity correlations for the multiple scattering of waves in unordered media

D. B. Rogozkin

Moscow Engineering Physics Institute, 115409 Moscow, Russia
(Submitted 18 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1674–1716 (May 1997)

The long-range correlations in the reflected and transmitted fluxes in the case of the coherent transport of waves in an unordered medium with discrete inhomogeneities are considered. The correlator and spectrum of the intensity fluctuations are expressed in a general form in terms of the one-center scattering amplitude and the propagators of the mean radiated intensity. The random interference of the waves and the fluctuations of the number of scattering centers in a microvolume of the medium are taken into account simultaneously. Detailed calculations are performed for two limiting radiation propagation regimes, viz., spatial diffusion and small-angle multiple scattering. It is shown that the conservation of the total flux upon elastic scattering leads to the formation of a dip in the spectrum and, accordingly, a negative correlation between the intensities at large distances. In the case of spatial diffusion this feature is displayed upon reflection, and in the case of small-angle multiple scattering it is displayed upon transmission through a slab. The relative roles of the various sources of intensity fluctuations, as well as the sensitivity of the correlations to factors that influence the wave propagation regime, viz., the finite size of the scattering sample, absorption in the medium, and the presence of a frequency shift in the incident waves, are analyzed. We find that fluctuations in the distribution of the scatterers show up most strongly in a medium with strong, i.e., ''non-Born,'' centers, especially if they exhibit absorption. [S1063-7761(97)01105-0]

1. INTRODUCTION

The interference of waves upon multiple elastic scattering is known to be the cause of a whole list of unusual ''mesoscopic'' effects, which appear when electrons and classical waves are transported in unordered media.[1–3] Some examples of these effects are the universal conductivity fluctuations of small metallic samples,[4–6] the correlations in speckles, i.e., intensity distributions that fluctuate strongly in space, in the case of the multiple scattering of coherent light and microwave radiation.[3,7–9]

Interest in the analysis of the correlations of multiply scattered wave fields first arose quite long ago in connection with the study of the scintillation of coherent laser radiation and radio waves in turbulent media.[10] This question has been investigated in extreme detail within the model of a random continuous medium with large-scale (with a dimension a that is much greater than the wavelength λ) weakly refracting inhomogeneities (see the reviews in Refs. 10–13). The region of applicability of the theoretical results[10–13] is restricted by conditions which are characteristic of turbulent media (very large inhomogeneities of the refractive index, the Born approximation for single scattering, and the absence of absorption).

The analogous problem for media with discrete scatterers became a topic of investigation comparatively recently in connection with the research on the diffusive transport of electrons and classical waves (coherent light, microwave radiation) in unordered systems.[1–9,14–22] The conditions characteristic of turbulent media are not satisfied in this case, and the wave correlation regime is different. Here the Born approximation cannot work for scattering by an individual center, and the effects caused by the absorption of radiation can play a significant role.

Despite the large number of publications devoted to the study of intensity fluctuations in unordered media,[1,7–9,14–33] the corresponding general theory has been developed to only a small extent beyond the work in Refs. 10–13. The existing theoretical results (see, for example, Refs. 14 and 17) pertain to the Born approximation for one-center scattering and, with the exception of Ref. 32, were obtained directly for a particular limiting regime of wave propagation in the medium. Because of the use of the Born approximation, the question of the intensity fluctuations associated with the random microinhomogeneity of the medium, in particular, was not considered.

In the Born approximation only one source of large-scale intensity fluctuations in the observation plane is taken into account, viz., spatial spreading of a local intensity surge that appears in the bulk of the medium due to interference of the waves upon multiple scattering. Such a fluctuation source is presumed in both the Langevin[1,15,18] and diagram[14,17] descriptions of the long-range correlation between diffusion fluxes in an unordered medium. As is shown below, when a departure is made from the Born approximation for one-center scattering, another source of intensity fluctuations is included, viz., local perturbations of the spatial distribution of the intensity and the bulk speckle structure due to Poisson fluctuations of the number of scatterers in a microvolume of the medium. This source of fluctuations is of the same nature as the variation of the transmission coefficient appearing when an additional scatterer is added to the medium, which

was considered in Refs. 34 and 35. Unlike the Born approximation, in which only the pairwise correlations of the wave fields are taken into account, the inclusion of the local inhomogeneity of the medium in the treatment requires consideration of the correlations between all four fields appearing in the definition of the intensity correlator.

Below we present the solution of the problem of calculating the spectrum and the correlation function of the intensity fluctuations for the multiple scattering of coherent radiation in an unordered medium with discrete inhomogeneities. The approach previously developed to describe fluctuations in the case of small-angle scattering in a turbulent medium[10–13] is generalized to the case of an arbitrary distribution of multiply scattered waves. The transport equation for an intensity correlator free of restrictions on the one-center scattering force is derived in the ladder approximation, and a closed analytic solution of that equation is found. The solution is expressed in a general form in terms of scattering amplitudes and ladder propagators and, thus, makes it possible to reduce the problem of calculating the spectrum and the correlation function of the intensity fluctuations to the solution of the transport equation[36,37] for the mean intensity. The relations found in this work generalize the basic formulas[32,33] to the case of strong scatterers and include the previously obtained results (see, for example, Refs. 1, 3, 14, 17, and 18) as the corresponding limiting cases.

Departure from the Born approximation for one-center scattering enables us to take into account the mechanism for intensity fluctuations caused by the random microinhomogeneity of the medium along with the purely interference mechanism. On the one hand, this opens up an incoherent channel of fluctuations (in the Born approximations there was no such channel) and, on the other hand, this can have a significant effect on fluctuations of a coherent nature. Inhomogeneity effects are displayed most strongly in media with large centers (with dimensions greater than the wavelength), especially if they are absorbing centers. The contribution of the incoherent fluctuations can also be important for the scattering of frequency-shifted waves, which has the additional problem of interpreting the experimentally observed damping of the fluctuations as the frequency shift increases.[20–25]

A detailed investigation of long-range intensity correlations is performed for two limiting regimes of multiple scattering of waves, viz., spatial diffusion of radiation in a system of centers of small radius (the case of arbitrary scattering centers is considered in Appendix B) and small-angle multiple scattering in a medium with large-scale inhomogeneity. The dependence of the correlations on such factors as the restricted size of the scattering samples, absorption in the medium, and the presence of a frequency shift in the incident waves is analyzed. It is shown that the form of the fluctuation spectrum at low spatial frequencies and, accordingly, the asymptote of the intensity correlator depend significantly on the conservation of the total radiation flux. In the case of spatial diffusion this feature is displayed in the fluctuation spectrum of the intensity reflected from the medium, and in the case of small-angle multiple scattering it is displayed in the spectrum of the transmitted intensity. In both cases conservation of the flux during elastic scattering leads to the formation of a dip in the spectrum and, accordingly, to a negative correlation between the intensities at large distances. One consequence of the conservation of the flux is the fact that the fluctuations of the diffuse reflectance are determined only by long ray trajectories and therefore, unlike the conclusions in Refs. 14 and 19, are sensitive to any factors that constrain the length of the trajectories.

The phenomena considered in the present work should be observed under conditions that are typical of many experiments on the multiple scattering of coherent light and microwave radiation (see, for example, Refs. 3, 7, 9, 20–25, and 30) and can be of interest for investigating the properties of unordered media.

2. GENERAL RELATIONS

Let a plane wave of unit amplitude fall on the slab $0<z<L$.

The problem of calculating the spectrum and the correlation function of the intensity fluctuations reduces to finding the moments of the wave field $\mathscr{G}_1$, $\mathscr{G}_2$, and $\mathscr{G}_4$ averaged over the positions of the scatterers, where

$$\mathscr{G}_n=\langle \Psi(1)\ldots\Psi^*(n)\rangle. \tag{1}$$

Under the conditions of weak localization of the waves ($k_0 l\gg 1$, where l is the elastic scattering length and $k_0=2\pi/\lambda$) the main contribution to the moments of the wave field is made by the ladder diagrams. Their summation leads to a system of transport equations (Fig. 1).

The equations for the first two moments, i.e., the mean field and the mutual coherence function, are well known (see, for example, Refs. 37 and 38). The mean field in a substance satisfies the following equation (Fig. 1a):

$$\mathscr{G}_1=\mathscr{G}_1^{(0)}+G_0\sum_a \hat{\mathscr{T}}_a\mathscr{G}_1, \tag{2}$$

where $\mathscr{G}_1^{(0)}$ is the field in the incident wave, $G_0=(\nabla^2+k_0^2+i0)^{-1}$ is the Green's function of the free wave equation, and $\hat{\mathscr{T}}_a$ is the matrix for scattering on a center located at the point $\mathbf{r}_a$. In the case in which the free term in (2) corresponds to the field of a point source, i.e., $\mathscr{G}_1^{(0)}=G_0$, Eq. (2) defines the Green's function of the elastic scattering problem.[37] Knowing G, we can write the equation for the second moment in the following manner:

$$\mathscr{G}_2=\mathscr{G}_2^{(0)}+GG^+\sum_a \hat{\mathscr{T}}_a\hat{\mathscr{T}}_a^+\mathscr{G}_2, \tag{3}$$

where $\mathscr{G}_2^{(0)}=\mathscr{G}_1\mathscr{G}_1$ is the second moment of the unscattered field. Equation (3) is depicted in Fig. 1b in diagram symbolism.

It is also not difficult to derive the transport equation for the fourth moment of the wave field by successive summation of the ladder diagrams (Fig. 1c):

$$\mathscr{G}_4=\mathscr{G}_4^{(0)}+\mathscr{G}_2\mathscr{G}_2\sum_a \hat{h}_a\hat{h}_a^+\mathscr{G}_4, \tag{4}$$

where

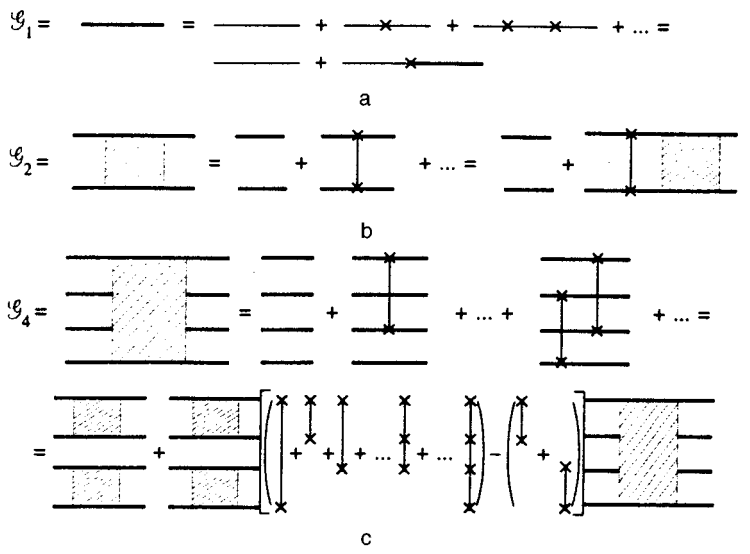

FIG. 1.

$\mathscr{G}_4^{(0)}+\mathscr{G}_2\mathscr{G}_2$, $\quad \hat{h}_a=\hat{h}_a^+=\hat{\mathscr{T}}_a(G^+)^{-1}+G^{-1}\hat{\mathscr{T}}_a^++\hat{\mathscr{T}}_a\hat{\mathscr{T}}_a^+$. (5)

Equations (3) and (4) have a similar structure. In Eq. (4) the second moment of the wave field $\mathscr{G}_2$ plays the same role as the Green's function G in Eq. (3) for the second moment. The free terms in Eqs. (3) and (4) describe the propagation of the unscattered fields and the uncorrelated intensities, respectively. In Eq. (4) the matrix $\hat{h}_a$ plays the role of the scattering matrix $\hat{\mathscr{T}}_a$. The quantity $\Sigma_a\hat{\mathscr{T}}_a\hat{\mathscr{T}}_a^+$ in (3) specifies the pair correlation of the fields, while the quantity $\hat{H}=\Sigma_a\hat{h}_a\hat{h}_a^+$ in (4) specifies the correlation of the intensities.

Equations (2)–(5) are general transport equations, which describe the correlation of the intensity values of the coherent radiation in an unordered medium. They are valid for any multiple-scattering regime of waves, including spatial diffusion and small-angle multiple scattering. In the latter case system (2)–(5) can be regarded as a generalization of the known transport equation for the fourth moment of the wave field, which was studied in detail in the context of the problem of the intensity fluctuations in the case of wave propagation in a turbulent medium,[10–13] to the case of strong discrete scatterers. In the small-angle approximation Eqs. (4) and (5) transform into the corresponding equations of the theory in Refs. 10–13, if only the terms responsible for the pair correlations of the fields are left in the quantity $\hat{H}=\Sigma_a\hat{h}_a\hat{h}_a^+$. For this purpose, the term that is quadratic in $\hat{\mathscr{T}}_a$ must be neglected in (5), and the expression for $\hat{\mathscr{T}}_a$ in the Born approximation must be inserted.

We transform Eqs. (3), (4), and (5), assuming that the scatterers are found at sufficient distances from one another, so that each successive scattering act takes place in the Fraunhofer zone of an individual scattering center, $l\gg\max(k_0^{-1}, k_0a^2)$, i.e., a condition opposite to the one adopted in Refs. 10–13 is satisfied. In this case, assuming that the scattering is multiple, we can separate the fast and slow spatial variables in the problem (see, for example, Refs. 37 and 38). The scale of variation of the fast variables is specified by the coherence radius of the scattered fields ($r_{\rm coh}\sim\lambda/\theta$, where θ is the characteristic multiple-scattering angle in the medium), and the scale of variation of the slow variables is specified by the characteristic spatial deflection of the ray trajectories in the scattering sample ($r_{\rm spat}>l\theta\gg r_{\rm coh}$). Taking the Fourier transform with respect to the fast variables, we arrive at the $(\mathbf{r},\mathbf{\Omega})$ representation ($\mathbf{r}$ denotes the coordinates, and $\mathbf{\Omega}$ denotes the directions of propagation of the waves), and we express all the observable quantities in terms of a smooth function of the directions and coordinates, i.e., the ''ray'' radiated intensity $I(\mathbf{r},\mathbf{\Omega})$.

Equation (3) can be transformed into an ordinary transport equation.[36–38] For this purpose, we must use the explicit form of the Green's function of the scattering problem:[37]

$$G(\mathbf{r},\mathbf{r}_1)=-\frac{1}{4\pi}\exp\left(ik_0|\mathbf{r}-\mathbf{r}_1|-\frac{1}{2}n\sigma_{\rm tot}|\mathbf{r}-\mathbf{r}_1|\right)\frac{1}{|\mathbf{r}-\mathbf{r}_1|}, \tag{6}$$

where n is the number of scatterers per unit volume and $\sigma_{\rm tot}$ is the total cross section of the one-center interaction, and go over to the mixed $(\mathbf{r},\mathbf{\Omega})$ representation. Taking into account the equality which is valid for $k_0l\gg1$

$$\begin{aligned}\int\int d\mathbf{r}_d\,d\mathbf{r}_d'\,G\left(\mathbf{r}+\frac{\mathbf{r}_d}{2},\mathbf{r}_1+\frac{\mathbf{r}_d'}{2}\right)G^*\left(\mathbf{r}-\frac{\mathbf{r}_d}{2},\mathbf{r}_1-\frac{\mathbf{r}_d'}{2}\right)\\ \times\exp(-i\mathbf{k}\cdot\mathbf{r}_d+i\mathbf{k}'\cdot\mathbf{r}_d')=4\pi^4\\ \times\frac{\exp(-n\sigma_{\rm tot}|\mathbf{r}-\mathbf{r}_1|)}{|\mathbf{r}-\mathbf{r}_1|^2}\,\delta\left(\mathbf{k}-k_0\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|}\right)\\ \times\delta\left(\mathbf{k}'-k_0\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|}\right)\end{aligned} \tag{7}$$

and introducing the notation $\mathbf{\Omega}=\mathbf{k}/k$ and $\mathbf{\Omega}'=\mathbf{k}'/k'$, we write Eq. (3) in the form

$$I(\mathbf{r},\mathbf{\Omega})=I^{(0)}(\mathbf{r},\mathbf{\Omega})+n\int d\mathbf{r}'\int d\mathbf{\Omega}'\,\frac{d\sigma}{d\Omega}\times(\mathbf{\Omega}\cdot\mathbf{\Omega}')\frac{\exp(-n\sigma_{\rm tot}|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|^2}$$

$$\times\delta\left(\mathbf{\Omega}-\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)I(\mathbf{r}',\mathbf{\Omega}'), \tag{8}$$

where $I(\mathbf{r},\mathbf{\Omega})\equiv\langle I(\mathbf{r},\mathbf{\Omega})\rangle$ is the radiated intensity at the point $\mathbf{r}$ in the direction $\mathbf{\Omega}$, averaged over the positions of the scatterers,

$$\frac{d\sigma}{d\Omega}(\mathbf{\Omega}\cdot\mathbf{\Omega}')=|f(\mathbf{\Omega}\cdot\mathbf{\Omega}')|^2$$

is the differential elastic scattering cross section, $f(\mathbf{\Omega}\cdot\mathbf{\Omega}')$ is the scattering amplitude, and $\hat{f}=-(4\pi)^{-1}\hat{\mathscr{T}}$. For the sake of brevity, in (8) and below we omit the averaging sign $\langle\ldots\rangle$ in the notation of the mean intensity, leaving it in explicit form only in the second moment of the intensity.

In deriving (8) we took into account the presence of δ functions in (7), which enables us to write the relation between $I(\mathbf{r},\mathbf{\Omega})$ and $\mathscr{G}_2(\mathbf{r},\mathbf{r}_1)$ in the form[37]

$$I\left(\mathbf{r},\mathbf{\Omega}=\frac{\mathbf{k}}{k}\right)\frac{\delta(k-k_0)}{k^2}=\frac{1}{(2\pi)^3}\int d\mathbf{R}\int d\mathbf{R}_1\mathscr{G}_2(\mathbf{R},\mathbf{R}_1)$$
$$\times\exp\{-i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}_1)\}\delta\left(\mathbf{r}-\frac{\mathbf{R}+\mathbf{R}_1}{2}\right). \tag{9}$$

For directed point and planar sources of unit power, the unscattered intensity in (8) equals, respectively,

$$I^{(0)}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}_1,\mathbf{\Omega}_1)=\frac{\exp(-n\sigma_{\rm tot}|\mathbf{r}-\mathbf{r}_1|)}{|\mathbf{r}-\mathbf{r}_1|^2}\delta(\mathbf{\Omega}-\mathbf{\Omega}_1)$$
$$\times\delta\left(\mathbf{\Omega}-\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|}\right),$$
$$I^{(0)}(z,\mathbf{\Omega}|z_1,\mathbf{\Omega}_1)=\frac{1}{|\Omega_z|}\exp\left(-n\sigma_{\rm tot}\frac{|z-z_1|}{|\Omega_z|}\right)\delta(\mathbf{\Omega}-\mathbf{\Omega}_1). \tag{10}$$

After applying the operator $(\mathbf{\Omega}\cdot\nabla_{\mathbf{r}}+n\sigma_{\rm tot})$ to both sides of Eq. (8), we obtain the transport equation in the ordinary integrodifferential form. For a directed point source [see (10)] it has the form[36]

$$(\mathbf{\Omega}\cdot\nabla_{\mathbf{r}}+n\sigma_{\rm tot})I(\mathbf{r},\mathbf{\Omega}|\mathbf{r}_1,\mathbf{\Omega}_1)=\delta(\mathbf{\Omega}-\mathbf{\Omega}_1)\delta(\mathbf{r}-\mathbf{r}_1)$$
$$+n\int d\mathbf{\Omega}'\frac{d\sigma}{d\mathbf{\Omega}'}I(\mathbf{r},\mathbf{\Omega}'|\mathbf{r}_1,\mathbf{\Omega}_1). \tag{11}$$

In the case in which the fields impinging on the medium differ in frequency by $\Delta\omega\ll\omega_0$ ($\omega_0=ck_0$ is the carrier frequency, and c is the speed of light), in going over to the $(\mathbf{r},\mathbf{\Omega})$ representation instead of (7) we must use the relation

$$\int\int d\mathbf{r}_d d\mathbf{r}_d' G_{\omega_0+\Delta\omega/2}\left(\mathbf{r}+\frac{\mathbf{r}_d}{2},\mathbf{r}_1+\frac{\mathbf{r}_d'}{2}\right)$$
$$\times G_{\omega_0-\Delta\omega/2}\left(\mathbf{r}-\frac{\mathbf{r}_d}{2},\mathbf{r}_1-\frac{\mathbf{r}_d'}{2}\right)\exp(-i\mathbf{k}\cdot\mathbf{r}_d+i\mathbf{k}'\cdot\mathbf{r}_d')$$
$$=4\pi^4\frac{\exp[-(n\sigma_{\rm tot}-i(\Delta\omega)/c)|\mathbf{r}-\mathbf{r}_1|]}{|\mathbf{r}-\mathbf{r}_1|^2}$$
$$\times\delta\left(\mathbf{k}-k_0\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|}\right)\delta\left(\mathbf{k}'-k_0\frac{\mathbf{r}-\mathbf{r}_1}{|\mathbf{r}-\mathbf{r}_1|}\right). \tag{12}$$

The equations for $I_{\omega_0}(\mathbf{r},\mathbf{\Omega},\Delta\omega)$ defined in accordance with (9) can then be obtained from Eq. (8) or (11), if $n\sigma_{\rm tot}-i(\Delta\omega)/c$ is substituted into them instead of $n\sigma_{\rm tot}$. Physically, $I_{\omega_0}(\mathbf{r},\mathbf{\Omega},\Delta\omega)$ is the intensity of the signal modulated with the frequency $\Delta\omega$.

To pass from Eq. (4) to the equation in the $(\mathbf{r},\mathbf{\Omega})$ representation, the fast and slow variables must be separated in the fourth moment $\mathscr{G}_4$ of the field. Under the conditions of multiple scattering $\mathscr{G}_4$ is an abruptly varying function of the difference variables and a smooth function of the summed variables, which decreases on the scale $r_{\rm coh}$. When the values of $\mathbf{r}_1-\mathbf{r}_2$ and $\mathbf{r}_3-\mathbf{r}_4$ are small, $\mathscr{G}_4$ is a smooth function of the variables $(\mathbf{r}_1+\mathbf{r}_2)/2$ and $(\mathbf{r}_3+\mathbf{r}_4)/2$. In addition, $\mathscr{G}_4$ depends on the variables which are obtained if $\mathbf{r}_2$ and $\mathbf{r}_4$ change places. This follows from the symmetry condition

$$\mathscr{G}_4(1,2,3,4)=\mathscr{G}_4(1,4,3,2), \tag{13}$$

which corresponds to transposition of the conjugate wave fields appearing in the definition (1). At small values of the difference variables the main contribution to $\mathscr{G}_4$ is made by the diagrams containing at least one scattering act in the ''outgoing'' ladder propagators. The sum of such diagrams specifies a part of the fourth moment $\mathscr{G}_4^{(\rm sc)}$ (see Appendix A). It is more convenient to work with $\mathscr{G}_4^{(\rm sc)}$ than with $\mathscr{G}_4$. Unlike $\mathscr{G}_4$, $\mathscr{G}_4^{(\rm sc)}$ does not obey the condition (13), and the fast and slow variables can be unequivocally separated in $\mathscr{G}_4^{(\rm sc)}$. In addition, under multiple-scattering conditions $\mathscr{G}_4$ can be expressed in terms of $\mathscr{G}_4^{(\rm sc)}$, and a closed equation similar to (4) can be written for $\mathscr{G}_4^{(\rm sc)}$.

In the problem of long-range spatial correlations the observable quantities can be expressed in terms of a smooth function of the coordinates and the angles, i.e., the second moment of the intensity. It is related to $\mathscr{G}_4^{(\rm sc)}$ by the equality

$$\left\langle I_{\omega_0+\Delta\omega/2}\left(\mathbf{r},\mathbf{\Omega}=\frac{\mathbf{k}}{k}\right)I_{\omega_0-\Delta\omega/2}\left(\mathbf{r}_1,\mathbf{\Omega}_1=\frac{\mathbf{k}_1}{k_1}\right)\right\rangle$$
$$\times\frac{\delta(k-k_0)}{k^2}\frac{\delta(k_1-k_0)}{k_1^2}$$
$$=\frac{1}{(2\pi)^6}\int\ldots\int d\mathbf{R}_1\ldots d\mathbf{R}_4$$
$$\times\mathscr{G}_4^{(\rm sc)}\left(\mathbf{R}_1,\omega_0+\frac{\Delta\omega}{2},\mathbf{R}_2,\omega_0\right.$$
$$\left.+\frac{\Delta\omega}{2};\mathbf{R}_3,\omega_0-\frac{\Delta\omega}{2},\mathbf{R}_4,\omega_0-\frac{\Delta\omega}{2}\right)$$
$$\times\delta\left(\mathbf{r}-\frac{\mathbf{R}_1+\mathbf{R}_2}{2}\right)\delta\left(\mathbf{r}_1-\frac{\mathbf{R}_3+\mathbf{R}_4}{2}\right)$$
$$\times\exp\{-i\mathbf{k}\cdot(\mathbf{R}_1-\mathbf{R}_2)-i\mathbf{k}_1\cdot(\mathbf{R}_3-\mathbf{R}_4)\}. \tag{14}$$

The quantity $\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega})I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1)\rangle$ satisfy the equation

$$\langle\mathrm{I}\omega0+\Delta\omega/2(\mathbf{r},\mathbf{\Omega})I\omega0-\Delta\omega/2(\mathbf{r}1,\mathbf{\Omega}1)\rangle$$

$$
\begin{aligned}
&=(I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0)I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1|\mathbf{\Omega}_0))^{(\mathrm{sc})}\\
&+n\int d\mathbf{r}'\int\int d\mathbf{\Omega}'d\mathbf{\Omega}_1'\int\int d\mathbf{\Omega}''d\mathbf{\Omega}_1''\\
&\times(I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1|\mathbf{r}',\mathbf{\Omega}_1'))^{(\mathrm{sc})}\\
&\times[h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}'')h(\mathbf{\Omega}_1',\mathbf{\Omega}_1'|\mathbf{\Omega}_1'',\mathbf{\Omega}_1'')\\
&\times(\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}'')I_{\omega_0-\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}_1'')\rangle\\
&+I^{(0)}_{\omega_0+\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}|\mathbf{\Omega}_0)I^{(0)}_{\omega_0-\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}_1|\mathbf{\Omega}_0))\\
&+h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}_1'')h(\mathbf{\Omega}_1',\mathbf{\Omega}_1'|\mathbf{\Omega}_1'',\mathbf{\Omega}'')\\
&\times\langle I_{\omega_0}(\mathbf{r}',\mathbf{\Omega}'',\Delta\omega)I_{\omega_0}(\mathbf{r}',\mathbf{\Omega}_1'',-\Delta\omega)\rangle], \qquad (15)
\end{aligned}
$$

where

$$
\begin{aligned}
(I_{\omega_0+\Delta\omega/2}I_{\omega_0-\Delta\omega/2})^{(\mathrm{sc})}&=I_{\omega_0+\Delta\omega/2}I_{\omega_0-\Delta\omega/2}\\
&\quad-I^{(0)}_{\omega_0+\Delta\omega/2}I^{(0)}_{\omega_0-\Delta\omega/2},
\end{aligned}
$$

$$
\begin{aligned}
h(\mathbf{\Omega},\mathbf{\Omega}_1|\mathbf{\Omega}',\mathbf{\Omega}_1')&=\frac{2\pi i}{k_0}[f(\mathbf{\Omega}\mathbf{\Omega}')\delta(\mathbf{\Omega}_1-\mathbf{\Omega}_1')\\
&\quad-\delta(\mathbf{\Omega}-\mathbf{\Omega}')f^*(\mathbf{\Omega}_1\mathbf{\Omega}_1')]\\
&\quad+f(\mathbf{\Omega}\mathbf{\Omega}')f^*(\mathbf{\Omega}_1\mathbf{\Omega}_1'). \qquad (16)
\end{aligned}
$$

The intensity propagators appearing in (15) have the following meaning. The quantity $I_\omega(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0)$ is the radiated intensity at the frequency ω at the point $\mathbf{r}$ in the direction $\mathbf{\Omega}$ when a plane wave impinges on the surface of the medium in the direction $\mathbf{\Omega}_0$: $I_\omega(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0)$ is normalized to the unit z component of the incident flux. The quantity $I_\omega(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')$ is the radiated intensity at the point $\mathbf{r}$ in the direction $\mathbf{\Omega}$ from a point source of unit power, which is located at the point $\mathbf{r}'$ and emits a wave in the direction $\mathbf{\Omega}'$. The quantities $I^{(0)}_\omega(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0)$ and $I^{(0)}_\omega(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')$ are the values of the unscattered intensity (10).

There will be an equation similar to (15) for the moment of the intensity $\langle I_{\omega_0}(\mathbf{r},\mathbf{\Omega},\Delta\omega)I_{\omega_0}(\mathbf{r}_1,\mathbf{\Omega}_1,-\Delta\omega)\rangle$ appearing in the second integral term in Eq. (15). Unlike (15), it will contain the intensity propagators $I_{\omega_0}(\mathbf{r},\mathbf{\Omega},\pm\Delta\omega)$ corresponding to incident waves with the frequency shift $\Delta\omega$.

The first term on the right-hand side of Eq. (15), i.e., the product of the mean values of the intensity, describes the independent propagation of two ''rays'' (each ''ray'' corresponds to an intensity and is described by a sum of ladder diagrams, see Figs. 1b and 1c), and the next two integral terms describe the correlation of the respective ''rays'' due to incoherent scattering on an individual center and interference.

As an analysis shows, under the condition $l\gg\max(k_0^{-1},k_0a^2)$, the amplitude of the correlations must be much smaller than the mean value of the intensity:

$$
\begin{aligned}
&\langle I(\mathbf{r},\mathbf{\Omega})I(\mathbf{r}_1,\mathbf{\Omega}_1)\rangle-I(\mathbf{r},\mathbf{\Omega})I(\mathbf{r}_1,\mathbf{\Omega}_1)\\
&\quad\ll I(\mathbf{r},\mathbf{\Omega})I(\mathbf{r}_1,\mathbf{\Omega}_1).
\end{aligned}
$$

Therefore, Eq. (15) can be solved by an iterative method with expansion of the second moment of the intensity into a series with respect to the multiplicity of the interaction of the ''rays.'' The small parameters in this expansion are $1/(k_0l)^2$ and σ/l^2, which characterize the coherent and incoherent interactions of the ''rays,'' respectively.

The principle contribution to the correlation function is made by the first iteration cycle. In this approximation for the second moment of the intensity we obtain the following expression:

$$
\begin{aligned}
&\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega})I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1)\rangle\\
&\quad=(I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0)I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1|\mathbf{\Omega}_0))^{(\mathrm{sc})}\\
&\qquad+n\int d\mathbf{r}'\int\int d\mathbf{\Omega}'d\mathbf{\Omega}_1'\int\int d\mathbf{\Omega}''d\mathbf{\Omega}_1''\\
&\qquad\times(I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1|\mathbf{r}',\mathbf{\Omega}_1'))^{(\mathrm{sc})}\\
&\qquad\times[h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}'')h(\mathbf{\Omega}_1',\mathbf{\Omega}_1'|\mathbf{\Omega}_1'',\mathbf{\Omega}_1'')\\
&\qquad\times I_{\omega_0+\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}''|\mathbf{\Omega}_0)I_{\omega-\Delta\omega/2}(\mathbf{r}',\mathbf{\Omega}_1''|\mathbf{\Omega}_0)\\
&\qquad+h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}_1'')h(\mathbf{\Omega}_1',\mathbf{\Omega}_1'|\mathbf{\Omega}_1'',\mathbf{\Omega}'')\\
&\qquad\times(I_{\omega_0}(\mathbf{r}',\mathbf{\Omega}''|\mathbf{\Omega}_0,\Delta\omega)I_{\omega_0}(\mathbf{r}',\mathbf{\Omega}_1''|\mathbf{\Omega}_0,-\Delta\omega))^{(\mathrm{sc})}]. \qquad (17)
\end{aligned}
$$

The expression (17) describes the long-range correlations of the intensity for an arbitrary law of single and multiple scattering of waves in an unordered medium. It enables us to express the intensity correlator in a general form in terms of the value of the mean intensity and to thereby reduce the problem of calculating the correlation function to solving the ordinary transport equation[36,37] with the respective sources.

We are interested only in values of the frequency shift that are so small that it is reasonable to refer to the interference of multiply scattered waves $[\Delta\omega\ll\omega_0/(k_0l)]$. Therefore, we can set $\omega_0\pm\Delta\omega/2\approx\omega_0$ everywhere on the right-hand side of (17), leaving the frequency difference finite only in the argument of the propagators $I_{\omega_0}(\mathbf{r},\mathbf{\Omega}|\mathbf{\Omega}_0,\pm\Delta\omega)$ in the last (interference) term.

The h function appearing in the incoherent part of (17) can be brought, using the optical theorem (see, for example, Ref. 39), into the form

$$
h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}'')=-\sigma_{\mathrm{tot}}\delta(\mathbf{\Omega}'-\mathbf{\Omega}'')+\frac{d\sigma}{d\mathbf{\Omega}}(\mathbf{\Omega}'\mathbf{\Omega}''). \qquad (18)
$$

We note that the incoherent contribution to (17) describes the intensity fluctuations caused only by Poisson fluctuations of the number of scatterers in a microvolume. This contribution could be obtained using the approach previously developed to describe fluctuations of the distribution function of the Boltzmann kinetic equation in a nonequilibrium gas.[40] For this purpose, the disturbed, locally nonuniform distribution of the scatterers $n+\delta n(\mathbf{r})$ must be substituted into the transport equation for the intensity (see (8) or (11)), and then the correlator $\langle\delta I\,\delta I_1\rangle$ can be calculated using

the known relation $\langle \delta n(\mathbf{r})\delta n(\mathbf{r}_1)\rangle = n\,\delta(\mathbf{r}-\mathbf{r}_1)$, which is valid for an uncorrelated arrangement of the scattering centers.

Knowing the second moment $\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}) \times I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1)\rangle$, we can calculate some experimentally observable quantities (see Appendix B), viz., the correlation function between the local values of the radiation flux density emerging from the medium $C(\boldsymbol{\rho})$ and its Fourier transform, i.e., the spatial fluctuation spectrum $M(\mathbf{q})$:

$$M(\mathbf{q}) = \int d^2\rho\, \exp(-i\mathbf{q}\cdot\boldsymbol{\rho}) C(\boldsymbol{\rho}), \tag{19}$$

where $\boldsymbol{\rho} = (\mathbf{r}-\mathbf{r}_1)_\parallel$ is a vector in the xy plane, which is parallel to the boundary of the slab. The asymptote of $C(\boldsymbol{\rho})$, which characterizes the long-range spatial correlations, is related to $\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r},\mathbf{\Omega}) I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\mathbf{\Omega}_1)\rangle$ by

$$\begin{aligned} C(z_f,\boldsymbol{\rho}=(\mathbf{r}-\mathbf{r}_1)_\parallel,\Delta\omega) &= \int_{(\mathbf{\Omega}\mathbf{\Omega}_n)>0} d\mathbf{\Omega} \int_{(\mathbf{\Omega}_1\mathbf{\Omega}_n)>0} d\mathbf{\Omega}_1 (\mathbf{\Omega}\mathbf{\Omega}_n)(\mathbf{\Omega}_1\mathbf{\Omega}_n) \\ &\times [\langle I_{\omega_0+\Delta\omega/2}(z_f,\mathbf{r}_\parallel,\mathbf{\Omega}) I_{\omega_0-\Delta\omega/2}(z_f,\mathbf{r}_{1\parallel},\mathbf{\Omega}_1)\rangle \\ &- (I_{\omega_0+\Delta\omega/2}(z_f,\mathbf{\Omega}|\mathbf{\Omega_0}) I_{\omega_0-\Delta\omega/2}(z_f,\mathbf{\Omega}_1|\mathbf{\Omega}_0))^{(\mathrm{sc})}], \end{aligned} \tag{20}$$

where z_f is the coordinate of the slab boundary ($z_f=0$ in the reflection geometry and $z_f=L$ in the transmission geometry), and $\mathbf{\Omega}_n$ is an external normal to the corresponding boundary. Although (20) defines the correlation function between local values of the flux density, in keeping with established tradition (see, for example, Refs. 10–13, 18, and 37) we shall use the term ''intensity correlation function.''

Henceforth it will generally be more convenient for us to work with the correlation function $C(\boldsymbol{\rho})$ and with the fluctuation spectrum $M(\mathbf{q})$. The shape of the spectrum at low spatial frequency q contains information on the long-range spatial correlations of the intensity and is very sensitive to variation of the wave propagation regime in an unordered medium. The amplitude of the spectrum at zero spatial frequency specifies the fluctuations of the total reflection or transmission coefficients:

$$\begin{aligned} \langle \delta T_{\omega_0+\Delta\omega/2}\delta T_{\omega_0-\Delta\omega/2}\rangle &= \frac{1}{A} M(z_f=L,\mathbf{q}=0,\,\Delta\omega), \\ \langle \delta R_{\omega_0+\Delta\omega/2}\delta R_{\omega_0-\Delta\omega/2}\rangle &= \frac{1}{A} M(z_f=0,\mathbf{q}=0,\,\Delta\omega), \end{aligned} \tag{21}$$

where A is the surface area of the slab. We define the reflection R and transmission T coefficients as the values of the total reflected and transmitted radiation fluxes normalized to the unit z component of the incident flux.[36,37] In particular,

$$\begin{aligned} \langle R_\omega\rangle &= \int_{(\mathbf{\Omega}\mathbf{\Omega}_n)>0} d\mathbf{\Omega}(\mathbf{\Omega}\mathbf{\Omega}_n) I_\omega(z_f=0,\mathbf{\Omega}|z_i=0,\mathbf{\Omega}_0), \\ \langle T_\omega\rangle &= \int_{(\mathbf{\Omega}\mathbf{\Omega}_n)>0} d\mathbf{\Omega}(\mathbf{\Omega}\mathbf{\Omega}_n) I_\omega(z_f=L,\mathbf{\Omega}|z_i=0,\mathbf{\Omega}_0). \end{aligned} \tag{22}$$

Using the relation between the fluctuation spectrum and the second moment of the intensity (see Appendix B), at low spatial frequencies ($q \ll k_0$) in a first approximation we obtain

$$M(z_f,\mathbf{q},\Delta\omega) = M^{\mathrm{coh}}(z_f,\mathbf{q},\Delta\omega) + M^{\mathrm{incoh}}(z_f,\mathbf{q}), \tag{23}$$

where

$$\begin{aligned} M^{\mathrm{coh}}(z_f,\mathbf{q},\Delta\omega) = n\int_0^L dz' \int\int d\mathbf{\Omega}'' d\mathbf{\Omega}_1'' \Big| \int d\mathbf{\Omega}' \\ \times I_{\omega_0}(z_f,\mathbf{q}|z',\mathbf{\Omega}') h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}_1'') \Big|^2 \\ \times I_{\omega_0}(z',\mathbf{\Omega}''|\mathbf{\Omega}_0,\Delta\omega) I_{\omega_0}(z',\mathbf{\Omega}_1''|\mathbf{\Omega}_0, \\ -\Delta\omega), \end{aligned} \tag{24}$$

$$\begin{aligned} M^{\mathrm{incoh}}(z_f,\mathbf{q}) = n\int_0^L dz' \Big| \int d\mathbf{\Omega}' \int d\mathbf{\Omega}'' I_{\omega_0}(z_f,\mathbf{q}|z',\mathbf{\Omega}') \\ \times \left[\sigma_{\mathrm{tot}}\delta(\mathbf{\Omega}'-\mathbf{\Omega}) - \frac{d\sigma}{d\mathbf{\Omega}}(\mathbf{\Omega}'\mathbf{\Omega}'')\right] \\ \times I_{\omega_0}(z',\mathbf{\Omega}''\mathbf{\Omega}_0) \Big|^2. \end{aligned} \tag{25}$$

In (24) and (25) it was taken into account that the ''incoming'' intensity propagator $I_{\omega_0}(z',\mathbf{\Omega}'|\mathbf{\Omega}_0,\pm\Delta\omega,)$ does not depend on the coordinates in the plane parallel to the boundaries of the slab and that the ''outgoing'' propagator $I_{\omega_0}(z_f,\mathbf{q}|z',\mathbf{\Omega}')$ equals

$$\begin{aligned} I_{\omega_0}(z_f,\mathbf{q}|z',\mathbf{\Omega}') &= \int d^2\rho\, \exp(-i\mathbf{q}\cdot\boldsymbol{\rho}) I_{\omega_0}(z_f,\boldsymbol{\rho}|z',\mathbf{\Omega}') \\ &= \int d^2\rho\, \exp(-i\mathbf{q}\cdot\boldsymbol{\rho}) \int_{(\mathbf{\Omega}\mathbf{\Omega}_n)>0} d\mathbf{\Omega}(\mathbf{\Omega}\mathbf{\Omega}_n) \\ &\times I_{\omega_0}(z_f,\boldsymbol{\rho},\mathbf{\Omega}|z',\mathbf{\Omega}'), \end{aligned} \tag{26}$$

where $I_{\omega_0}(z_f,\boldsymbol{\rho}|z',\mathbf{\Omega}')$ is the spatial distribution of the radiation flux density emerging from the medium at the surface $z=z_f$. The amplitude of the spectrum (23)–(25) is normalized to the unit z component of the incident flux.

The expressions (17) and (23)–(25) were derived without any assumptions regarding the form of the intensity propagators and are the most general of the results that have been proposed hitherto to describe long-range intensity correlations in unordered media. Formulas (23)–(25) generalize the known results reported in Refs. 1, 14, 17, 18, 29, 32, and 33 in two respects. On the one hand, they take into account the effects caused by strong, i.e., non-Born, one-center scattering, and, on the other hand, they are valid for any regime of multiple scattering of waves in a medium. The results obtained for the spatial diffusion of radiation[1,14,17,18] and small-angle multiple scattering[29] follow from (23)–(25) as limiting cases when the corresponding approximate expressions for the intensity propagators are plugged into them.

In the Born approximation, in which only pairwise correlations of the wave fields are taken into account in the diagrams presented in Fig. 1c, the incoherent contribution to

the fluctuation spectrum vanishes (see, for example, Ref. 41), and the coherent contribution takes the following form:[32]

$$M^{\mathrm{coh}}(z_f,\mathbf{q},\Delta\omega)=\left(\frac{2\pi}{k_0}\right)^2 n\int_0^L dz\int\int d\boldsymbol{\Omega}'\,d\boldsymbol{\Omega}_1'|\times I(z_f,\mathbf{q}|z,\boldsymbol{\Omega}')-I(z_f,\mathbf{q}|z,\boldsymbol{\Omega}_1')|^2\times\frac{d\sigma}{d\boldsymbol{\Omega}}(\boldsymbol{\Omega}'\boldsymbol{\Omega}_1')I(z,\boldsymbol{\Omega}'|\boldsymbol{\Omega}_0,\Delta\omega)\times I(z,\boldsymbol{\Omega}_1'|\boldsymbol{\Omega}_0,-\Delta\omega).\tag{27}$$

This result is obtained from (24), if we substitute the Born scattering amplitude into the h function and restrict attention to the terms that are linear with respect to the amplitude.

Expressions like (24), (25), and (27) are also valid for the correlations between values of the intensity on opposite boundaries of the slab and for the fluctuation spectrum $M_{\mathrm{tot}}(\mathbf{q})$ of the total scattered intensity, i.e., the sum of the reflected and transmitted intensities. In the latter case the sum $I(0,\mathbf{q}|z,\boldsymbol{\Omega})+I(L,\mathbf{q}|z,\boldsymbol{\Omega})$ must be substituted into (24), (25), and (27) instead of $I(z_f,\mathbf{q}|z,\boldsymbol{\Omega})$.[32,33]

Relations (23)–(27) enable us to draw several conclusions at once regarding the general properties of the fluctuation spectrum.

In the case of purely elastic backscattering of the waves by a semi-infinite medium, the incident flux is completely reflected, and the reflection coefficient $R=1$. Therefore, its dispersion $\langle(\delta R)^2\rangle=0$, and when $\mathbf{q}=0$, the expressions (23)–(25) and (27) should vanish. In fact, in this case the value of $I(0,\mathbf{q}|z,\boldsymbol{\Omega})$ for $\mathbf{q}=0$ does not depend on $\boldsymbol{\Omega}$,[36,42] and we obtain $M(z_f=0,\mathbf{q}=0)=0$ (the contribution (24) for M^{coh} vanishes by virtue of the optical theorem for the scattering amplitude[39]). When $q>0$, a dependence of $I(0,\mathbf{q}|z,\boldsymbol{\Omega})$ on $\boldsymbol{\Omega}$ appears, and $M(z_f=0,\mathbf{q})$ increases as a result. Thus, conservation of the flux upon elastic scattering leads to the appearance of a minimum on the spectrum of the reflected intensity.

Consideration of subsequent iteration cycles in Eq. (15) does not destroy the feature of the fluctuation spectrum just mentioned, since factors of the form

$$\int d\boldsymbol{\Omega}'\,I(0,\mathbf{q}|z,\boldsymbol{\Omega}')h(\boldsymbol{\Omega}',\boldsymbol{\Omega}'|\boldsymbol{\Omega}'',\boldsymbol{\Omega}_1'')$$

$$\int d\boldsymbol{\Omega}_1'\,I^*(0,\mathbf{q}|z,\boldsymbol{\Omega}_1')h(\boldsymbol{\Omega}_1',\boldsymbol{\Omega}_1'|\boldsymbol{\Omega}'',\boldsymbol{\Omega}_1'')$$

are present explicitly in all the terms of the expansion of the fluctuation spectrum into a series with respect to the multiplicity of the intensity interaction.

When waves are elastically scattered by a finite slab, the conservation of the flux has exactly the same effect on the behavior of the fluctuation spectrum of the total intensity $M_{\mathrm{tot}}(\mathbf{q})$. In this case $R+T=1$, and $\langle(\delta R+\delta T)^2\rangle=0$. Therefore, when $\mathbf{q}=0$, the spectrum $M_{\mathrm{tot}}(\mathbf{q})$ should vanish. This follows formally from the fact that when $\mathbf{q}=0$, the sum $I(0,\mathbf{q}|z,\boldsymbol{\Omega})+I(L,\mathbf{q}|z,\boldsymbol{\Omega})$ appearing in $M_{\mathrm{tot}}(\mathbf{q})$ defines the total radiation flux through both boundaries of the slab and does not depend on $\boldsymbol{\Omega}$ in the absence of absorption.

This type of behavior of the spectrum $M_{\mathrm{tot}}(\mathbf{q})$ is most pronounced in the case of small-angle multiple scattering.[10–13,29] In this case the probability of the reflection of waves is negligible, and the feature in $M_{\mathrm{tot}}(\mathbf{q})$ is observed in the fluctuation spectrum of the transmitted intensity.

We have been dealing with the long-range spatial correlations of the intensity, for which information is contained in the coordinate dependence $\langle I_{\omega_0+\Delta\omega/2}(\mathbf{r},\boldsymbol{\Omega})\times I_{\omega_0-\Delta\omega/2}(\mathbf{r}_1,\boldsymbol{\Omega}_1)\rangle$. However, knowing the second moment, we can also investigate the short-range ($\rho<r_{\mathrm{coh}}\sim\lambda/\theta$) intensity correlations in speckles. The small-scale structure of a speckle image in the observation plane is determined by the interference of waves with different angles of incidence. Therefore, information regarding the short-range correlations between the local values of the flux density can be obtained from the angular dependence of the second moment of the intensity. The corresponding correlation function is defined by the expression (see Appendix B)

$$C(z_f,\boldsymbol{\rho}=(\mathbf{r}-\mathbf{r}_1)_\parallel,\Delta\omega)=\frac{1}{4}\int d\boldsymbol{\Omega}\int d\boldsymbol{\Omega}_1[\boldsymbol{\Omega}_n\cdot(\boldsymbol{\Omega}+\boldsymbol{\Omega}_1)]^2\times\exp\{ik_0(\boldsymbol{\Omega}-\boldsymbol{\Omega}_1)\cdot\boldsymbol{\rho}\}\times\left\langle I_{\omega_0}\left(z_f,\frac{1}{2}(\mathbf{r}+\mathbf{r}_1)_\parallel,\boldsymbol{\Omega},\Delta\omega\right)I_{\omega_0}\times\left(z_f,\frac{1}{2}(\mathbf{r}+\mathbf{r}_1)_\parallel,\boldsymbol{\Omega}_1,-\Delta\omega\right)\right\rangle.\tag{28}$$

In a first approximation, from (28) there follows the formula

$$C(z_f,\boldsymbol{\rho},\Delta\omega)=\frac{1}{4}\int d\boldsymbol{\Omega}\int d\boldsymbol{\Omega}_1[\boldsymbol{\Omega}_n\cdot(\boldsymbol{\Omega}+\boldsymbol{\Omega}_1)]^2\exp\{ik_0(\boldsymbol{\Omega}-\boldsymbol{\Omega}_1)\cdot\boldsymbol{\rho}\}\times(I_{\omega_0}(z_f,\boldsymbol{\Omega}|\boldsymbol{\Omega}_0,\Delta\omega)I_{\omega_0}(z_f,\boldsymbol{\Omega}_1|\boldsymbol{\Omega}_0,-\Delta\omega))^{(\mathrm{sc})},\tag{29}$$

which, with consideration of (9), can also be written in the form

$$C(z_f,\boldsymbol{\rho},\Delta\omega)=-\frac{1}{4k_0^2}\left[\left(\frac{\partial}{\partial z}-\frac{\partial}{\partial z_1}\right)^2(|\mathscr{G}_2(\mathbf{r},\mathbf{r}_1,\Delta\omega)|^2-|\mathscr{G}_2^{(0)}(\mathbf{r},\mathbf{r}_1,\Delta\omega)|^2)\right]\Bigg|_{z=z_1=z_f}.\tag{30}$$

In the case of small-angle scattering Eqs. (29) and (30) transform into the familiar formula previously obtained in Refs. 10–13 and 37.

It can be seen by a direct calculation that the fluctuation spectrum corresponding to Eqs. (29) and (30) in the range of spatial frequencies $q\ll k_0$ coincides with the high-frequency asymptote of the spectrum (24) or (27) [at small values of q the asymptotes of (24) and (27) do not differ from one another].

The general relations (23)–(25) and (29) solve the problem of calculating the spectrum and the correlation function of the intensity fluctuations in an unordered medium with discrete scatterers. According to (23)–(25) and (29), to in-

vestigate the correlations it is sufficient to know the corresponding solutions of the transport equation for the propagators.

Below we shall consider the qualitative features of the long-range correlations for the limiting wave propagation regimes, for which approximate analytic expressions for the intensity propagators are known in explicit form. The next sections of this paper are devoted to calculating the spectrum and the correlation function of the intensity fluctuations for two types of unordered media, which consist, respectively, of centers of small radius and large-scale scatterers.

3. INTENSITY FLUCTUATIONS FOR THE SCATTERING OF WAVES ON A SYSTEM OF CENTERS OF SMALL RADIUS

Let us consider the long-range intensity correlations for the scattering of a plane wave by a slab of a medium containing unordered small-scale (with dimensions smaller than the wavelength) centers. We assume that the slab is optically thick ($L \gg l$) and that scattering on each center has an isotropic character.

The intensity correlations at the large distances $\rho \gg l$ are determined by the behavior of the spectrum at low spatial frequencies $q \ll l^{-1}$. At $q \ll l^{-1}$ (24) and (25) can be transformed in the following manner.

Let us first consider the fluctuation spectrum of the reflected intensity for purely elastic scattering in a semi-infinite medium ($L \to \infty$). In this case the intensity propagator $I(z_f=0,\mathbf{q}|z,\mathbf{\Omega})$ for $\mathbf{q}=0$ does not depend on z or on the direction $\mathbf{\Omega}$.[36,42] Here $I(0,\mathbf{q}=0|z,\mathbf{\Omega})$ is the total flux through the surface from a point source which is located at a depth z and emits in the direction $\mathbf{\Omega}$. Therefore, the statement made means that all the radiation emerges from the medium. As q increases, a weak dependence of $I(0,\mathbf{q}|z,\mathbf{\Omega})$ on $\mathbf{\Omega}$ appears at $q \ll l^{-1}$, and we can utilize the expansion of I in spherical harmonics. The first terms of such an expansion have the form[37]

$$I(0,\mathbf{q}|z,\mathbf{\Omega}) = \frac{1}{4\pi}(F(0,\mathbf{q}|z) + 3\mathbf{\Omega}\cdot\mathbf{J}(0,\mathbf{q}|z) + \ldots), \tag{31}$$

where

$$F(0,\mathbf{q}|z) = \int d\mathbf{\Omega}\; I(0,\mathbf{q}|z,\mathbf{\Omega}),$$

$$\mathbf{J}(0,\mathbf{q}|z) = \int d\mathbf{\Omega}\; \mathbf{\Omega} I(0,\mathbf{q}|z,\mathbf{\Omega}). \tag{32}$$

Substituting the expansion (31) into (24) and (25), we obtain

$$\int d\mathbf{\Omega}' I(0,\mathbf{q}|z,\mathbf{\Omega}') h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}_1'') = \frac{3}{4\pi}\frac{2\pi i}{k_0}(f\mathbf{\Omega}'' - f^*\mathbf{\Omega}_1'')\mathbf{J}(0,\mathbf{q}|z), \tag{33}$$

$$\int d\mathbf{\Omega}' I(0,\mathbf{q}|z,\mathbf{\Omega}') h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}'') = -\frac{3}{4\pi}\sigma\mathbf{\Omega}''\mathbf{J}(0,\mathbf{q}|z), \tag{34}$$

where σ is the elastic scattering cross section and $n\sigma = l^{-1}$. At small $q \ll l^{-1}$ the characteristic scale of the variation of $\mathbf{J}(0,\mathbf{q}|z)$ with respect to z is of the order of q^{-1}. Therefore, the range for integration over z in (24) and (25) will be much greater than the mean free path l. This, in turn, allows us to use an expansion like (31) for the propagator $I(z,\mathbf{\Omega}|0,\mathbf{\Omega}_0)$. As a result, for M_R^{coh} and M_R^{incoh} we find

$$M_R^{\mathrm{coh}}(\mathbf{q}) = \frac{3}{8\pi k_0^2}\int_0^\infty \frac{dz}{l}|\mathbf{J}(0,\mathbf{q}|z)|^2 |F(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2 + \frac{9}{16\pi^2}\left(\sigma - \frac{2\pi}{k_0^2}\right)\int_0^\infty \frac{dz}{l}|\mathbf{J}(0,\mathbf{q}|z) \times \mathbf{J}(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2, \tag{35}$$

$$M_R^{\mathrm{incoh}}(\mathbf{q}) = \frac{9}{16\pi^2}\sigma\int_0^\infty \frac{dz}{l}|\mathbf{J}(0,\mathbf{q}|z)\mathbf{J}(z|0,\mathbf{\Omega}_0)|^2. \tag{36}$$

The identity indicated above $M_R(\mathbf{q}=0)=0$ follows from (33)–(36), because $|\mathbf{J}(0,\mathbf{q}|z)| \sim q^2$ when $q \to 0$.

When the waves are reflected from a thick slab ($L \gg l$) or a weakly absorbing medium ($l \ll l_a$, where l_a is the absorption length), the foregoing arguments remain in force. This is because the dependence of $I(0,\mathbf{q}|z,\mathbf{\Omega})$ on $\mathbf{\Omega}$ remains weak at $q \ll l^{-1}$ in those cases and the expansion (31) is value for $I(0,\mathbf{q}|z,\mathbf{\Omega})$ at any z. In the case of absorbing centers an additional term appears in (33) and (34):

$$\int d\mathbf{\Omega}' I(0,\mathbf{q}|z,\mathbf{\Omega}') h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}_1'') = -\frac{\sigma_a}{4\pi}F(0,\mathbf{q}|z) + \frac{3}{4\pi}\frac{2\pi i}{k_0}(f\mathbf{\Omega}'' - f^*\mathbf{\Omega}_1'')\mathbf{J}(0,\mathbf{q}|z), \tag{37}$$

$$\int d\mathbf{\Omega}' I(0,\mathbf{q}|z,\mathbf{\Omega}') h(\mathbf{\Omega}',\mathbf{\Omega}'|\mathbf{\Omega}'',\mathbf{\Omega}'') = -\frac{\sigma_a}{4\pi}F(0,\mathbf{q}|z) - \frac{3\sigma_{\mathrm{tot}}}{4\pi}\mathbf{\Omega}''\mathbf{J}(0,\mathbf{q}|z), \tag{38}$$

where $\sigma_a = \sigma_{\mathrm{tot}} - \sigma$ is the cross section for one-center absorption and $n\sigma_a = l_a^{-1}$. The first term in (37) and (38) describes the influence of the inhomogeneous (due to the random arrangement of the centers) absorption in the medium on the coherent and incoherent intensity fluctuations. The presence of the additional term in (37) and (38) does not alter the trend of our reasoning. The range for integration over z in (24) and (25) remains much greater than l, and, as before, an expansion like (31) can be used for $I(z,\mathbf{\Omega}|0,\mathbf{\Omega}_0)$. Ultimately, we obtain

$$M_R^{\mathrm{coh}}(\mathbf{q}) = \frac{3}{8\pi k_0^2}\int_0^L \frac{dz}{l}|\mathbf{J}(0,\mathbf{q}|z)|^2 |F(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2 - \frac{9}{8\pi k_0^2}\int_0^L \frac{dz}{l}|\mathbf{J}(0,\mathbf{q}|z)\mathbf{J}(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2$$

$$+\frac{n}{16\pi^2}\int_0^L dz|\sigma_a F(0,\mathbf{q}|z)F(z|0,\mathbf{\Omega}_0,\Delta\omega)$$

$$+3\sigma_{\rm tot}\mathbf{J}(0,\mathbf{q}|z)\cdot\mathbf{J}(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2, \quad (39)$$

$$M_R^{\rm incoh}(\mathbf{q})=\frac{n}{16\pi^2}\int_0^L dz|\sigma_a F(0,\mathbf{q}|z)F(z|0,\mathbf{\Omega}_0)$$

$$+3\sigma_{\rm tot}\mathbf{J}(0,\mathbf{q}|z)\cdot\mathbf{J}(z|0,\mathbf{\Omega}_0)|^2. \quad (40)$$

Formulas (39) and (40) explicitly describe the physical mechanism for the appearance of large-scale fluctuations. The source is a local disturbance of the intensity distribution in the bulk of the medium. The intensity surge can be attributed to interference between the waves (i.e., to the bulk speckle[1]), as well as to the local inhomogeneity of the scattering and absorbing properties of the substance due to fluctuations of the number of centers in a small volume. The inhomogeneity of the medium disturbs both the ''mean'' intensity distribution with resultant fluctuations of incoherent origin (40) and the bulk interference pattern (see the last term in Eq. (39)). As for the first two terms in (39), they are not related to fluctuations of the density of the scattering centers and are governed directly by the bulk speckle. They correspond to the expansion (31) in Eq. (27).

We note that the possibility of representing $I(0,\mathbf{q}|z,\mathbf{\Omega})$ in form (31) is a necessary condition for going from (24) and (25) to (39) and (40).[33] In the reflection geometry this equality is valid for any z, if and only if $I(0,\mathbf{q}|z,\mathbf{\Omega})$ specifies the radiation flux through the boundary of the medium (see (26)). If, conversely, $I(0,\mathbf{q}|z,\mathbf{\Omega})$ is defined simply as the radiated energy density on the boundary, it is no longer possible to go from (24) and (25) to (39) and (40). In particular, in the case of purely elastic backscattering from a semi-infinite medium, the value of $I(0,\mathbf{q}|z,\mathbf{\Omega})$ for $q=0$ will depend on $\mathbf{\Omega}$ in the range $z\leqslant l$, and, therefore, the dip discussed above is not observed at $q=0$. The contribution to the integral over z will be determined by the near-surface region $z\leqslant l$. This does not contradict the arguments presented above regarding the dispersion $\langle(\delta R)^2\rangle$, since the radiated energy density at the surface of the medium is not related in any way to the reflection coefficient.

It is important to bear this in mind in comparing our results with Refs. 14 and 19. The analogous calculations performed in Refs. 14 and 19 do not produce the dip for $q\to 0$. The contribution of the near-surface region $z\leqslant l$ to the spectrum remains nonvanishing when $q\to 0$. On this basis Stephen and Cwilich[14] and Wang and Feng[19] drew qualitatively incorrect conclusions regarding $\langle(\delta R)^2\rangle$, the form of the fluctuation spectrum of the reflected intensity (in Ref. 14 a maximum, instead of a dip, was predicted at $q=0$), and the behavior of the intensity correlator.

There are two reasons,[33] which could have led these investigators[14,19] to conclude that the surface contribution is important for the reflection geometry. One reason is that, as was shown in Ref. 33, not all the diagrams which make contributions to the intensity correlator were taken into account in Ref. 14 and 19, and, as a result, the dispersion of the reflection coefficient does not satisfy the condition for flux conservation. The other possible reason was just discussed. Stephen and Cwilich[14] did not distinguish between the energy density at the surface and the flux density through the surface. Their calculations refer to fluctuations of the energy density, and conclusions regarding the fluctuations of the reflection coefficient cannot be drawn on their basis.

When waves are transmitted through a thick slab ($L\gg l$), the situation is simpler than in the case of reflection. The main contribution to the integrals in (24) and (25) is made by the region z, $L-z\gg l$. The contribution from the near-surface layers ($z\leqslant l, L-z\leqslant l$) to the integrals in (24) and (25) is l/L times smaller than the bulk contribution. Expansions like (31) can be used at once for the propagators appearing in (24) and (25). As a result, for $M_T(\mathbf{q})$ we obtain an expression which is distinguished from (39) and (40) only by the replacement of $F(0,\mathbf{q}|z)$ and $\mathbf{J}(0,\mathbf{q}|z)$ by $F(L,\mathbf{q}|z)$ and $\mathbf{J}(L,\mathbf{q}|z)$.

In this approximation, in which the intensity can be determined by the first two terms in the expansion (31), the flux $\mathbf{J}$ can be expressed in terms of the density gradient F,[37] and the quantities $\mathbf{J}(z|0,\mathbf{\Omega}_0)$ and $\mathbf{J}(z_f,\mathbf{q}|z)$ in (39) and (40) can be written in the following manner:

$$J_k(z|0,\mathbf{\Omega}_0,\Delta\omega)=-\frac{l}{3}\delta_{k,z}\frac{\partial}{\partial z}F(z|0,\mathbf{\Omega}_0,\Delta\omega),$$

$$J_k(z_f,\mathbf{q}|z)=\frac{l}{3}\left(\delta_{k,z}\frac{\partial}{\partial z}+iq_k\right)F(z_f,\mathbf{q}|z), \quad (41)$$

where the functions $F(z|0,\mathbf{\Omega}_0,\Delta\omega)$ and $F(z_f,\mathbf{q}|z)$ can be expressed in terms of the diffusion asymptotes of the solution of the transport equation.[33]

In the case of the transport of waves of a different nature, for example, electrons or neutrons, the relations (41) can be more complicated and contain terms that are proportional to the external fields.

With consideration of (41), from (39) and (40) we obtain

$$M^{\rm coh}(z_f,\mathbf{q})=\frac{l}{24\pi k_0^2}\int_0^L dz\left(\left|\frac{\partial}{\partial z}F(z_f,\mathbf{q}|z)\right|^2\right.$$

$$\left.+q^2|F(z_f,\mathbf{q}|z)|^2\right)|F(z|0,\mathbf{\Omega}_0,\Delta\omega)|^2$$

$$-\frac{l^3}{72\pi k_0^2}\int_0^L dz\left|\frac{\partial}{\partial z}F(z_f,\mathbf{q}|z)\right.$$

$$\left.\times\frac{\partial}{\partial z}F(z|0,\mathbf{\Omega}_0,\Delta\omega)\right|^2+\frac{n}{16\pi^2}$$

$$\times\int_0^L dz\left|\sigma_a F(z_f,\mathbf{q}|z)F(z|0,\mathbf{\Omega}_0,\Delta\omega)\right.$$

$$\left.-\frac{1}{3}\sigma l^2\frac{\partial}{\partial z}F(z_f,\mathbf{q}|z)\frac{\partial}{\partial z}F(z|0,\mathbf{\Omega}_0,\Delta\omega)\right|^2, \quad (42)$$

$$M^{\rm incoh}(z_f,\mathbf{q})=\frac{n}{16\pi^2}\int_0^L dz|\sigma_a F(z_f,\mathbf{q}|z)F(z|0,\mathbf{\Omega}_0)$$

$$-\frac{1}{3}\sigma l^2\frac{\partial}{\partial z}F(z_f,\mathbf{q}|z)\frac{\partial}{\partial z}F(z|0,\mathbf{\Omega}_0)|^2. \tag{43}$$

According to Eqs. (39) and (40) and Eqs. (42) and (43), the coherent and incoherent contributions to the fluctuation spectrum can be represented in the following form:

$$M^{\mathrm{coh}}=M^{\mathrm{coh(sp)}}+M^{\mathrm{coh}(a)}+M^{\mathrm{coh}(a,s)}+M^{\mathrm{coh}(s)}, \tag{44}$$

$$M^{\mathrm{incoh}}=M^{\mathrm{incoh}(a)}+M^{\mathrm{incoh}(a,s)}+M^{\mathrm{incoh(s)}}. \tag{45}$$

The first term in (44) describes intensity fluctuations whose source is the bulk speckle. The terms $M^{(a)}$, $M^{(s)}$, and $M^{(a,s)}$ appearing in both equalities are caused by the Poisson fluctuations of the spatial distribution of the particles in the unordered medium; these terms correspond to different fluctuations of the absorbing and scattering properties of the medium, as well as the correlation between them.

As an analysis shows, the first term in (42) is the dominant term in the expansion of the spectrum defined by (42) and (43). In fact, this term coincides with the Fourier transform of the formula derived by Pnini and Shapiro derived in Ref. 18 for the correlation function in the transmission geometry. This is because the low-frequency ($q<l^{-1}$) behavior of the fluctuation spectrum, as was shown above, is determined only by the long ray trajectories both in the case of transmission and in the case of reflection, and the contribution to the spectrum from the near-surface region $z \leqslant l$ can be neglected. The corresponding formula in Ref. 18 was derived within the Langevin approach previously proposed by Zyuzin and Spivak[1,15] to describe the correlation functions of the diffusion fluxes in an unordered substance. In this approach only long ray trajectories are considered, and it is assumed a priori that the contribution of the short trajectories is negligibly small.

The differences between the first term in (42) and the corresponding result in Ref. 18 affect only the definition of $F(z_f,\mathbf{q}|z)$. The correlations in the bulk of the medium between the values of the energy density were considered in Ref. 18, while Eq. (42) was written for the flux density through the surface. In the case of the transmission of waves through a thick slab, this difference is of no fundamental significance, but it is significant in the case of reflection. As was pointed out above, Eq. (42) is valid for the reflection geometry only if the fluctuations of the radiation flux density are considered. The statements pertaining to the influence of flux conservation on the fluctuations of the reflected and transmitted intensity remain valid in just this case, and the contribution from the region $z \leqslant l$, in which the intensity propagators are specified by the short ray trajectories, can be neglected in (24) and (25).

The second and third terms in (42) and the incoherent spectrum (43) are corrections to the dominant coherent contribution in (42). Under the condition $\min(L,l_D) < l(k_0l)^2$ ($l_D=\sqrt{ll_a/3}$ is the diffusion length[36,37]) the contribution of these terms to the fluctuation spectrum exceeds the second-order contribution with respect to the interaction of the ''rays'' (i.e., the contribution of the next iteration cycle in Eq. (15)). The latter was previously considered in the context of fluctuations of the electronic conductivity of small metal objects (see, for example, Refs. 1, 4, 5, and 41).

We note that the incoherent contribution (25) tends to zero as the value of the spatial frequency q increases ($l^{-1}<q\ll k_0$), while the coherent contribution (24) tends to a certain constant value, which coincides with the value found in Ref. 33 using Eq. (27). This asymptotic value of M^{coh} can also be obtained by taking the Fourier transform of (29) for $q\ll k_0$ and corresponds to the short-range intensity correlations previously discussed in Refs. 1, 8, and 17, and 31. Thus, we see that the spectrum $M(z_f,q)$ cannot be regarded simply as the sum[17,19] of the high- and low-frequency contributions. This is not so important for the transmission geometry, since $M_T(q\gg l^{-1})\ll M_T(q\ll l^{-1})$.[1,16–19,31] However, the situation is significantly different in the case of reflection, since the reverse inequality $M_R(q\gg l^{-1})\gg M_R(q\ll l^{-1})$ holds.[32,33]

The values of the functions $F(z|0,\mathbf{\Omega}_0,\Delta\omega)$ and $F(z_f,\mathbf{q}|z)$ in (42) and (43) can be found using relations obtained by solving the problem of the coherent backscattering of waves from a slab of an unordered medium containing centers of small radius.[43–45] In accordance with the solution in Ref. 45, the expressions for these functions have the form[33]

$$F(z|0,\mathbf{\Omega}_0,\Delta\omega)=\sqrt{3}H(\mu_0,1)\frac{\sinh[\gamma_0(L-z)]}{\sinh\ \gamma_0 L},$$

$$F(z_f,\mathbf{q}|z)=4\pi\frac{\sinh[\gamma(L-|z_f-z|)]}{\sinh\ \gamma L}, \tag{46}$$

where $H(\mu_0,1)$ is Chandrasekhar's function,[36,42] $\mu_0=|\mathbf{\Omega}_n\mathbf{\Omega}_0|$ is the cosine of the angle of incidence of the radiation to the surface of the medium,

$$\gamma_0=\sqrt{l_D^{-2}+3i\frac{\Delta\omega}{c}l^{-1}},\quad \gamma=\sqrt{l_D^{-2}+q^2},$$

and l_D is the diffusion length. We note that the function $H(\mu_0,1)$ satisfies the equalities[36,42]

$$\frac{1}{4\pi}\int_{\mathbf{\Omega}_n\mathbf{\Omega}>0}H(\mathbf{\Omega}_n\mathbf{\Omega},1)d\mathbf{\Omega}=1,$$

$$\frac{1}{4\pi}\int_{\mathbf{\Omega}_n\mathbf{\Omega}>0}(\mathbf{\Omega}_n\mathbf{\Omega})H(\mathbf{\Omega}_n\mathbf{\Omega},1)d\mathbf{\Omega}=\frac{1}{\sqrt{3}}.$$

The latter equality ensures fulfillment of the reciprocity theorem[36] for the functions $F(z|0,\mathbf{\Omega}_0)$ and $F(z_f=0,\mathbf{q}=0|z)$.

Substituting the functions (46) into (42) and (43), we arrive at the following results.

3.1. Nonabsorbing medium

When monochromatic coherent radiation ($\Delta\omega=0$) is reflected from a thick slab of an unordered medium, the coherent contribution to the fluctuation spectrum is always the dominant contribution. It is described by the expression

$$M_R^{\mathrm{coh}}(q)=H^2(\mu_0,1)\left\{\frac{2\pi}{k_0^2}\left[\frac{l}{L}\mathscr{F}_R(qL)-\frac{1}{3}\left(\frac{l}{L}\right)^3\mathscr{F}(qL)\right.\right.$$

$$+\ldots\Big]+\frac{1}{3}\sigma\left(\frac{l}{L}\right)^3\mathscr{F}(qL)+\ldots\Big\}, \tag{47}$$

where

$$\mathscr{F}_R(x)=\frac{1}{4x\ \sinh^2 x}(\sinh 2x-2x\ \cosh 2x + 2x^2\sinh 2x), \tag{48}$$

$$\mathscr{F}(x)=\frac{x}{4\ \sinh^2 x}(\sinh 2x+2x). \tag{49}$$

The terms in square brackets in (47) describe the contribution of purely interference origin, and the last term in (47) describes the contribution due to the disturbance of the bulk interference pattern by the local inhomogeneity of the medium. The expression (47) is an expansion in powers of l/L. In the case of scatterers of small radius the inequality $\sigma \leqslant 4\pi/k_0^2$ holds, and, therefore, the main contribution to the spectrum (47) is always made by the first term in the square brackets.

The expression (47) is valid in the range $q<l^{-1}$. At $q>l^{-1}$ the $M_R^{\rm coh}(q)$ curve has a plateau, where $M_R^{\rm coh}(q)\sim 1/k_0^2$.[32,33]

The spectrum $M_R(q)$ has a minimum at $q=0$. According to Eq. (21), $M_R(q=0)$ specifies the dispersion of the reflection coefficient:

$$\langle(\delta R)^2\rangle^{\rm coh}=\frac{1}{A}M_R(q=0)=\frac{(1-\langle R\rangle)^2}{A}\times\left[\frac{2\pi}{k_0^2}\left(\frac{L}{l}-\frac{l}{L}+\ldots\right)+\sigma\frac{l}{L}+\ldots\right], \tag{50}$$

where

$$\langle R\rangle=1-\frac{l}{\sqrt{3}L}H(\mu_0,1)$$

is the mean reflection coefficient of the slab for elastic scattering.[42] The value of the dispersion (50) is significantly smaller than $\langle(\delta R)^2\rangle\sim 1/Ak_0^2$, which follows from Refs. 14 and 19.

In the limit of a semi-infinite medium ($L\to\infty, R\to 1$) the fluctuations of the reflection coefficient vanish ($\langle(\delta R)^2\rangle\to 0$). In this case the dip in the spectrum acquires a triangular shape:[32,33]

$$M_R(q)=H^2(\mu_0,l)\frac{2\pi}{k_0^2}|\mathbf{q}|l. \tag{51}$$

For a slab of finite thickness the linear law (51) holds at $q>L^{-1}$ (see Eq. (47)).

The intensity correlator corresponding to the spectrum (47) can be represented in the form

$$C_R^{\rm coh}(\rho)=H^2(\mu_0,1)\left\{\frac{2\pi}{k_0^2}\left[\frac{l}{L^3}\nu_R\left(\frac{\rho}{L}\right)-\frac{1}{3}\frac{l^3}{L^5}\nu\left(\frac{\rho}{L}\right)+\ldots\right]+\frac{1}{3}\sigma\frac{l^3}{L^5}\nu\left(\frac{\rho}{L}\right)+\ldots\right\}. \tag{52}$$

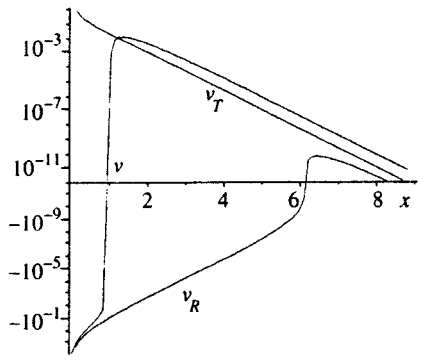

FIG. 2. Plots of the functions $\nu_R(x)$, $\nu_T(x)$, and $\nu(x)$.

The main contribution to (52) is made by the term proportional to ν_R. When $l/L<x<1$, the function $\nu_R(x)$ is a power law:[32,33]

$$\nu_R(x)=-\frac{1}{2\pi x^3}. \tag{53}$$

When $x\approx 6$, it changes sign and then decreases exponentially:

$$\nu_R(x)=\sqrt{\frac{x}{8}}\exp(-\pi x). \tag{54}$$

The function $\nu(x)$ appearing in (52) equals $\nu(x)=-1/(4\pi x^3)$ at small x, changes sign at $x\approx 1$, and decreases at large x as

$$\nu(x)=\pi^2\sqrt{\frac{x}{8}}\exp(-\pi x).$$

Plots of $\nu_R(x)$ and $\nu(x)$ are shown in Fig. 2.

We note that the shape of the dip in the spectrum of the spatial functions (47) and (51) is reminiscent of an inverted coherent backscattering peak.[3,7,43–47] This similarity is a consequence of the long-range character of the correlations. At the same time, the physical factors determining the features in the intensity fluctuation spectrum and in the angular distribution of backscattered waves differ significantly. The fact that the triangular feature (51) corresponds to a minimum in $M_R(q)$ is a consequence of the conservation of the total energy flux upon purely elastic scattering in the medium. For the same reason, an anticorrelation effect should be observed between the local values of the intensity in the range of distances $l<\rho<L$ (see (52)–(54)).

The incoherent contribution to the fluctuation spectrum, $M_R^{\rm incoh}(q)$, coincides with the term in $M_R^{\rm coh}(q)$ that is caused by the local inhomogeneity of the medium

$$M_R^{\rm incoh}(q)=\frac{1}{3}\sigma H^2(\mu_0,1)\left(\frac{l}{L}\right)^3\mathscr{F}(qL), \tag{55}$$

where the function $\mathscr{F}(x)$ is defined by (49). The correlation function corresponding to (55) equals

$$C_R^{\rm incoh}(\rho)=\frac{1}{3}\sigma H^2(\mu_0,1)\frac{l^3}{L^5}\nu\left(\frac{\rho}{L}\right) \tag{56}$$

and is $(\sigma k_0^2)(l/L)^2$ times smaller than $C_R^{\rm coh}(\rho)$ (52). Therefore, the role of the incoherent fluctuations for the case of reflection is insignificant.

When the waves incident upon the medium differ in frequency, additional weakening of the long-range correlations should be observed, at variance with the conclusions in Ref. 14.

In this case the range of depths z where the incident waves form the bulk speckle structure is restricted by $l_{\Delta\omega}=(2lc/\Delta\omega)^{1/2}$. An appreciable dependence of the spectrum of $M_R^{\rm coh}$ on $\Delta\omega$ appears when $l_{\Delta\omega}<L$. The value of $M_R^{\rm coh}$ at the minimum $M_R^{\rm coh}(q=0)$ decreases with increasing $\Delta\omega$ proportionally to $l_{\Delta\omega}/L$. Equation (51) becomes valid only for $q>l_{\Delta\omega}^{-1}$. As for the spatial correlations, $\nu_R(x,\Delta\omega)$ is specified, as before, by (53) at $l/L<x<l_{\Delta\omega}/L$, and it decreases more rapidly at $l_{\Delta\omega}/L<x<1$:

$$\nu_R(x,\Delta\omega)=-\frac{3}{2\pi}\frac{l_{\Delta\omega}^2}{L^2}\frac{1}{x^5}. \tag{57}$$

The tail of the correlation function at large x decreases according to the same law as for $\Delta\omega=0$ (see (54)); however, the amplitude of the correlations decreases (when $x=\rho/L\gg 1$, $\nu_R(x,\Delta\omega)=(\pi^2 l_{\Delta\omega}/3^{1/2}L)\,\nu_R(x,\Delta\omega=0)$). The incoherent intensity correlations do not depend on $\Delta\omega$, but as long as $l_{\Delta\omega}>l$, their amplitude remains smaller than the amplitude of the correlations of interference origin.

When the waves are transmitted through the slab, the expression for the coherent contribution to the spectrum is specified, as before, by Eq. (47), in which $\mathscr{T}_R(x)$ must now be replaced by the function

$$\mathscr{T}_T(x)=\frac{1}{4x\,\sinh^2 x}(\sinh 2x-2x). \tag{58}$$

A contribution to the fluctuation spectrum proportional to $\mathscr{T}_T(x)$ has been obtained by different methods in many studies (see, for example, Refs. 14, 17, 18, and 31). The spectrum $M_T^{\rm coh}(q)$ decreases with increasing q according to the law $M_T^{\rm coh}(q)\sim k_0^{-2}(l/L)^2(ql)^{-1}$ down to $M_T^{\rm coh}(q\sim l^{-1})\sim k_0^{-2}(l/L)^2$, which corresponds to the short-range intensity correlations.[1,8,17,31]

The spatial correlations of the radiation transmitted through the slab always remain positive. The correlation function $C_T^{\rm coh}(\rho)$ is obtained from (52) by replacing $\nu_R(x)$ by $\nu_T(x)$, where $\nu_T(x)=1/(4\pi x)$ when $x<1$, while $\nu_T(x)$ is specified by the same asymptotic expression as $\nu_R(x)$ (see (54)) when $x\gg 1$; a plot of $\nu_T(x)$ is shown in Fig. 2. The fit between the asymptotes of the intensity correlators of the reflected and transmitted waves can be attributed to the fact that in the case of elastic scattering in the slab the wave propagation regime at large distances, $\rho\gg L$, does not depend on the position of the source and has a universal character.

The coherent and incoherent intensity fluctuations caused by the local inhomogeneity of the medium are described by the same relations in the case of transmission as in the case of reflection: in particular, $M_T^{\rm incoh}=M_R^{\rm incoh}$ and $C_T^{\rm incoh}=C_R^{\rm incoh}$. Since the corresponding contribution to the spectrum increases with increasing q, the role of the local inhomogeneity of the medium is more appreciable in the transmission geometry. Comparing (47) and (58) with (55) and (49), we obtain $M_T^{\rm incoh}/M_T^{\rm coh}\sim k_0^2\sigma(ql)^2>M_R^{\rm incoh}M_R^{\rm coh}$.

As a consequence of the conservation of the total radiation flux for purely elastic scattering $(R+T=1)$, the dispersion of the fluctuations of the transmission coefficient satisfies the identity

$$\langle(\delta T)^2\rangle=-\langle(\delta T)(\delta R)\rangle=\langle(\delta R)^2\rangle. \tag{59}$$

Therefore, $\langle(\delta T)^2\rangle^{\rm coh}$ coincides with (50).

Some attention must be focused here on one fundamental difference between the cases of wave transmission and reflection. In the transmission geometry $C_T^{\rm coh}(\rho)$ decreases as $1/\rho$, and the main contribution to $\langle(\delta T)^2\rangle^{\rm coh}$ is made by the long-range correlations. The contribution of the short-range correlations is of the next smaller order of magnitude with respect to l/L and can be neglected.

The situation in the reflection geometry is different. The long-range and short-range correlations make contributions of the same order, $1/Ak_0^2$, but of different sign, to $\langle(\delta R)^2\rangle^{\rm coh}$. In the case of reflection from a semi-infinite medium, these contributions compensate one another exactly, and in the case of a finite slab the total contribution is of the order of $l/(Ak_0^2L)$ and is specified by the variation of the correlations at large ρ.

When waves differing in frequency pass through the medium, the amplitude of the intensity correlations caused by interference decreases. In this situation the incoherent correlation mechanism becomes especially important. When $l_{\Delta\omega}<L$, it can become the main mechanism.

The frequency shift $\Delta\omega$ has different effects on different portions of the fluctuation spectrum $M_T^{\rm coh}(q)$. The equality $M_R(q=0)=M_T(q=0)$ remains valid for nonzero values of $\Delta\omega$, since the total radiated energy flux is conserved. The height of the maximum of $M_T^{\rm coh}$ is lowered by a factor of $\sqrt{3}l_{\Delta\omega}/2L$. As q increases, the spectral dependence of $M_T^{\rm coh}$ at first decreases exponentially,

$$M_T^{\rm coh}\sim k_0^{-2}q^2 l_{\Delta\omega} l\,\exp(-2qL),$$

and then at $q>l_{\Delta\omega}^{-1}$ it decreases as $1/q$:

$$M_T^{\rm coh}\sim k_0^{-2}(ql)^{-1}\left[\frac{l}{l_{\Delta\omega}}\exp\left(-\frac{\sqrt{3}L}{l_{\Delta\omega}}\right)\right]^2.$$

Comparing these results with the contribution $M_T^{\rm incoh}=M_R^{\rm incoh}$, which does not depend on $\Delta\omega$ (see (55)), we find that the incoherent contribution to the spectrum will be the main contribution at $q>l_{\Delta\omega}^{-1}$, if $l_{\Delta\omega}<L/\ln(L/l\sqrt{k_0^2\sigma})$.

Similar conclusions can be drawn by analyzing the correlation function $C_T(\rho,\Delta\omega)=C_T^{\rm coh}(\rho,\Delta\omega)+C_T^{\rm incoh}(\rho)$. When $l_{\Delta\omega}<L$, a dependence of $C_T^{\rm coh}$ on the frequency shift $\Delta\omega$ is displayed for any ρ. In the range $l<\rho<L$ the function $C_T^{\rm coh}(\rho,\Delta\omega)$ decreases very slowly with increasing ρ:[33]

$$\nu_T(x,\Delta\omega)\approx\frac{\sqrt{3}}{8\pi}\frac{l_{\Delta\omega}}{L}\zeta(3)+\frac{3}{\pi x}\left[\frac{L}{l_{\Delta\omega}}\exp\left(-\frac{\sqrt{3}L}{l_{\Delta\omega}}\right)\right]^2, \tag{60}$$

where $\zeta(3)=1.202$ and $\zeta(z)$ is the Riemann zeta function.[48] When $\rho>L$, the behavior of $C_T^{\rm coh}(\rho,\Delta\omega)$ is the same as in the case of reflection (i.e., when $x\gg 1$, $\nu_T(x,\Delta\omega)$

$=\nu_R(x,\Delta\omega)=(\pi^2 l_{\Delta\omega}/\sqrt{3}L)\,\nu_R(x,\Delta\omega=0)$, where $\nu_R \times(x,\Delta\omega=0)$ is determined from Eq. (54)). A comparison of $C_T^{\rm coh}(\rho,\Delta\omega)$ with $C_T^{\rm incoh}(\rho)=C_R^{\rm incoh}(\rho)$ (56) shows that for $\rho<L(l^2k_0^2\sigma/Ll_{\Delta\omega})^{1/3}$ the correlation function $C_T(\rho)$ is determined by the incoherent contribution.

3.2. Absorbing medium

There are two mechanisms by which absorption influences the intensity fluctuation spectrum. On the one hand, absorption influences the wave propagation regime in the medium. It restricts the spatial scale of the long-range correlations and leads to violation of the conservation of the radiation flux. On the other hand, if the scattering centers themselves exhibit absorption, the local spatial inhomogeneity of the absorption properties of the medium becomes significant.[49]

The influence of absorption on the propagation regime is reflected in the value of the functions (46) appearing in (42) and (43). The inhomogeneity of the absorption is responsible for the appearance of new terms, which are proportional to the absorption cross section σ_a, in the fluctuation spectrum described by (42) and (43).

When waves are reflected from a relatively thin slab ($L\ll l_D$), the role of absorption is essentially imperceptible. The laws governing the intensity correlations that were considered above for the case of purely elastic scattering remain unchanged. Absorption has a significant influence on the intensity fluctuations only when the thickness of the scattering slab exceeds the characteristic length for weakening of the radiation flux in the medium, $L>l_D$. In this case the various contributions to the spectrum of spatial fluctuations of the reflected intensity (see (44) and (45)) are specified by the expressions

$$M_R^{\rm coh(sp)}=3\frac{\pi}{k_0^2}\frac{l_D}{l}(1-\langle R\rangle)^2\frac{1}{a(x)+1}\left[2a^2(x)-1-\frac{1}{3}\frac{l^2}{l_D^2}a^2(x)\right], \tag{61}$$

$$M_R^{\rm coh(a)}=M_R^{\rm incoh(a)}=\frac{3}{2}\sigma_a\frac{l_D}{l}(1-\langle R\rangle)^2\frac{1}{a(x)+1}, \tag{62}$$

$$M_R^{\rm coh(a,s)}=M_R^{\rm incoh(a,s)}=-3\sigma_a\frac{l_D}{l}(1-\langle R\rangle)^2\frac{1}{a(x)+1}, \tag{63}$$

$$M_R^{\rm coh(s)}=M_R^{\rm incoh(s)}=\frac{1}{2}\sigma\frac{l}{l_D}(1-\langle R\rangle)^2\frac{a^2(x)}{a(x)+1}, \tag{64}$$

where $x=ql_D$, $a(x)=\sqrt{1+x^2}$, and

$$\langle R\rangle=1-\frac{l}{\sqrt{3}l_D}H(\mu_0,1)$$

is the reflection coefficient from an absorbing medium.[42]

In Eqs. (61)–(64) passage from the case of discrete absorbing particles to bulk absorption in the medium is attained in the limit $\sigma_a\to 0$: $n\sigma_a=l_a^{-1}=\text{const}$.

In the case of centers of small radius under consideration, the contribution of purely interference origin (61) is always dominant. The contributions caused by nonuniformity of the medium, (62)–(64), have additional smallness of the order of $\sigma k_0^2(l/l_D)^2$. In (61)–(64) the diffusion length l_D qualitatively plays the same role as the thickness L in the case of purely elastic scattering. When $q>l_D^{-1}$, the dominant contribution in the spectrum (61) transforms into (51).

As for the behavior of the intensity correlation function, the analogy to a slab of finite thickness is completely suited to it, since the anticorrelation of the intensities is also maintained for large ρ ($\rho\gg l_D$):

$$C_R(\rho)\approx C_R^{\rm coh(sp)}(\rho)=-H^2(\mu_0,1)\frac{l}{k_0^2\rho^3}\times\begin{cases}1, & l<\rho<l_D,\\ \dfrac{\rho}{2l_D}\exp\left(-\dfrac{\rho}{l_D}\right), & \rho>l_D.\end{cases} \tag{65}$$

When waves are transmitted through a thick absorbing slab ($L\gg l_D$), the terms in the fluctuation spectrum equal

$$M_T^{\rm coh(sp)}=\frac{3\pi}{4k_D^2}\frac{l_D}{l}\langle T\rangle^2\frac{1}{a(x)-1}\left\{[2a^2(x)-1]\times\left[\frac{2}{a(x)[a(x)+1]}-\exp\left(-\frac{2L}{l_D}[a(x)-1]\right)\right]-\frac{1}{3}\frac{l^2}{l_D^2}a^2(x)\left[1-\exp\left(-\frac{2L}{l_D}[a(x)-1]\right)\right]\right\}, \tag{66}$$

$$M_T^{\rm coh(a)}=M_T^{\rm incoh(a)}=\frac{3}{8}\sigma_a\frac{l_D}{l}\langle T\rangle^2\frac{1}{a(x)-1}\times\left[1-\exp\left(-\frac{2L}{l_D}[a(x)-1]\right)\right], \tag{67}$$

$$M_T^{\rm coh(a,s)}=M_T^{\rm incoh(a,s)}=\frac{3}{4}\sigma_a\frac{l_D}{l}\langle T\rangle^2\frac{a(x)}{a(x)-1}\times\left[1-\exp\left(-\frac{2L}{l_D}[a(x)-1]\right)\right], \tag{68}$$

$$M_T^{\rm coh(s)}=M_T^{\rm incoh(s)}=\frac{1}{8}\sigma\frac{l}{l_D}\langle T\rangle^2\frac{a^2(x)}{a(x)-1}\times\left[1-\exp\left(-\frac{2L}{l_D}[a(x)-1]\right)\right], \tag{69}$$

where

$$\langle T\rangle=\frac{2l}{\sqrt{3}l_D}H(\mu_0,1)\exp\left(-\frac{L}{l_D}\right)$$

is the transmission coefficient through a thick absorbing slab ($L\gg l_D$, $\langle T\rangle\ll 1-\langle R\rangle$).[42]

The main contribution to the fluctuation spectrum of the transmitted intensity is made by (66). The spectrum $M_T(q)$ can be divided into three sections. At $q<(l_DL)^{-1/2}$ the amplitude of $M_T(q)$ remains essentially unchanged, in the range $(l_DL)^{-1/2}<q<l_D^{-1}$ the fluctuation spectrum decreases as $1/q^2$, and, finally, at $q>l_D^{-1}$ it decreases as $1/q$.

The correlation function corresponding to the spectrum (66) behaves in the following manner:

$$C_T^{\mathrm{coh(sp)}}(\rho)=\frac{3\langle T\rangle^2}{2(k_0 l)^2}\frac{l}{l_D}\times\begin{cases}\dfrac{l_D}{\rho}+\dfrac{1}{4}\ln\dfrac{L}{l_D}, & l<\rho<l_D,\\ \dfrac{Ll_D}{\rho^2}\exp\left(-\dfrac{\rho^2}{4Ll_D}\right), & \rho>\sqrt{Ll_D}.\end{cases}\tag{70}$$

In the range $l_D<\rho<\sqrt{Ll_D}$ the function $C_T^{\mathrm{coh(sp)}}(\rho)$ remains essentially constant. Such behavior of the correlation function differs significantly from the predictions in Ref. 18.

The contribution to the correlation function due to the inhomogeneity of the medium can be important only for small ρ, $l<\rho<l_D$, where

$$C_T^{\mathrm{coh}(s)}(\rho)=C_T^{\mathrm{incoh}(s)}(\rho)=-\frac{1}{16\pi}\langle T\rangle^2\frac{\sigma}{l_D l}\left(\frac{l}{\rho}\right)^3.\tag{71}$$

It is noteworthy that, according to Eqs. (63) and (68), simultaneous fluctuations of the scattering and absorption properties of the medium act differently in the reflection geometry and in the transmission geometry. Local increases in the scattering and absorption coefficients cause changes of opposite sign in the reflected intensity: an increase in the scattering power of the medium decreases the value of the intensity, while an increase in absorption coefficient reduces the intensity. Therefore, in the reflection geometry the fluctuations caused by the local inhomogeneity of the medium are suppressed (see the sum of contributions (62)–(64)). The dispersion of the reflection coefficient in the approximation under consideration is determined by a purely interference mechanism:

$$\langle(\delta R)^2\rangle=\frac{3\pi}{2Ak_0^2}(1-\langle R\rangle)^2\frac{l_D}{l}\left(1-\frac{l^2}{3l_D^2}+\ldots\right).\tag{72}$$

The situation is different in the case of transmission. Local increases in the scattering and absorption coefficients lead to effects of the same sign, i.e., they decrease the value of the transmitted intensity. Therefore, the fluctuations only increase in the presence of absorbing particles [see (67)–(69)]. The dispersion of the transmission coefficient is

$$\langle(\delta T)^2\rangle=\frac{3}{2A}\langle T\rangle^2\frac{L}{l}\left[\frac{\pi}{k_0^2}\left(1-\frac{l^2}{3l_D^2}+\ldots\right)+3\left(\sigma_a+\sigma\frac{l^2}{9l_D^2}+\ldots\right)\right].\tag{73}$$

In contrast to (59), $\langle(\delta R)^2\rangle\gg\langle(\delta T)^2\rangle$. However, the relative fluctuations of the transmission coefficient are much greater, $\langle(\delta R)^2\rangle/\langle R\rangle^2\ll\langle(\delta T)^2\rangle/\langle T\rangle^2$, and they increase linearly with thickness L.

When waves differing in frequency propagate in an absorbing medium, the coherent contribution to the fluctuation spectrum varies precisely as in the case of elastic scattering. In particular, $M^{\mathrm{coh}}(q=0,\Delta\omega)$ acquires the additional factor $2l_{\Delta\omega}/(\sqrt{3}l_D)$ in the case of reflection or $2l_{\Delta\omega}/(\sqrt{3}L)$ in the case of transmission.

Although the results presented above were obtained for an unordered medium consisting of centers of small radius, many qualitative conclusions regarding the character of the long-range intensity correlations remain valid, as a whole, for scatterers of arbitrary size. It is only necessary that the wave propagation regime in the medium be a diffusion regime. The condition $\sigma\leqslant 4\pi/k_0^2$ cannot be satisfied for large scatterers; therefore, the role of the terms resulting from the local inhomogeneity of the medium is more appreciable for them. This point is discussed in greater detail in the next section.

4. INTENSITY FLUCTUATIONS IN A MEDIUM WITH LARGE SCATTERERS

In the case of the multiple scattering of waves in a medium with large particles (having dimensions greater than the wavelength), two limiting propagation regimes are possible.

In the case in which the scattering slab is thick ($L\gg l_{\mathrm{tr}}$, where $l_{\mathrm{tr}}=l/(1-\langle\cos\vartheta\rangle)$ is the transport length in elastic scattering and $\langle\cos\vartheta\rangle$ is the mean cosine of the single-scattering angle) and the absorption in the medium is weak ($l_{\mathrm{tr}}\ll l_a$), the angular distribution of the scattered waves differs only slightly from an isotropic distribution,[36,37] and a regime of spatial diffusion of the radiation similar to the one considered in Sec. 3 for scatterers of small radius is realized. The structure of the expressions for the fluctuation spectrum remains the same as in Eqs. (42) and (43); however, the coefficients of the various contributions are somewhat altered (see Appendix C).

At the same time, the coefficients of the dominant coherent and incoherent contributions to the fluctuation spectrum differ from the corresponding coefficients in Eqs. (42) and (43) only by the replacement of l and σ by l_{tr} and $\sigma_{\mathrm{tr}}=\sigma(1-\langle\cos\vartheta\rangle)$.

It is significant that there can be an arbitrary relation between k_0^{-2} and σ_{tr} for large particles. In particular, for non-Born scatterers, $k_0^2\sigma_{\mathrm{tr}}\sim k_0^2a^2|n_0-1|^2\gg 1$ (n_0 is the refractive index of the particles), and the incoherent contribution becomes just as important as the dominant coherent contribution to the fluctuation spectrum.

When $k_0^2\sigma_{\mathrm{tr}}\gg 1$, a comparison of the expressions for the coherent and incoherent contributions to the correlation function (see Eqs. (52), (56), (57), (60), (65), (70), and (71)) with consideration of the replacement of l and σ by l_{tr} and σ_{tr} leads to the following conclusions.

In the case of the transmission of waves through a slab either with purely elastic scattering or with absorption the contribution to the correlation function $C_T(\rho)$ dominates in the range $l_{\mathrm{tr}}<\rho<l_{\mathrm{tr}}\sqrt{k_0^2\sigma_{\mathrm{tr}}}$. If the incident waves differ in frequency by $\Delta\omega$, the upper bound of this range shifts to

$$l_{\mathrm{tr}}\sqrt{k_0^2\sigma_{\mathrm{tr}}}\left(\frac{\Delta\omega L^4}{cl_{\mathrm{tr}}^3k_0^2\sigma_{\mathrm{tr}}}\right)^{1/6}>l_{\mathrm{tr}}\sqrt{k_0^2\sigma_{\mathrm{tr}}}.$$

In the case of the reflection of monochromatic radiation from the medium, the coherent and incoherent contributions to the correlation function $C_R(\rho)$ have the same dependence on ρ, and therefore the incoherent contribution dominates only for relatively thin slabs and for sufficiently strong ab-

sorption: $\min(L,l_D)<l_{\rm tr}\sqrt{k_0^2\sigma_{\rm tr}}$, where $l_D=\sqrt{l_a l_{\rm tr}/3}$ is the diffusion length for particles of arbitrary size.[37,42] If the frequencies of the incident waves differ, the coherent contribution begins to decrease more rapidly with increasing ρ (see (57)), and the incoherent contribution becomes dominant at the sufficiently large distances $\rho>L(\Delta\omega l_{\rm tr}k_0^2\sigma_{\rm tr}/c)^{-1/2}$ when the frequency shift $\Delta\omega>c/(l_{\rm tr}k_0^2\sigma_{\rm tr})$.

In the case in which the thickness of the scattering slab is small ($L<l_{\rm tr}$), as well as under the conditions of fairly strong absorption ($l_a<l_{\rm tr}$), another limiting wave propagation regime, viz., small-angle multiple scattering, is realized for any value of L.[37,50–53] In contrast to the spatial-diffusion regime for small-angle multiple scattering, the angular distribution is highly anisotropic, i.e, extended in the direction in which the waves originally propagated. In this case the intensity of the forward-scattered radiation can be neglected.[37]

The intensity fluctuations for small-angle multiple scattering were investigated in extremely great detail in reference to the problem of wave propagation in a turbulent medium.[10–13] The applicability of the results in Refs. 10–13 is restricted by the same conditions which are characteristic of turbulent media: very large spatial inhomogeneities of the refractive index, the absence of absorption, and Born single scattering.

As the analysis in Ref. 29 shows, the results obtained in Refs. 10–13 are applicable as long as the transverse displacement of the ray trajectories does not exceed the maximum dimension of the inhomogeneities (an ''external'' turbulence scale). In an inhomogeneous medium with discrete scatterers the assumptions used in Refs. 10–13 generally do not hold, and a different scheme must be employed to solve the equation for $\mathscr{G}_4$. The question of going from the approximations proposed in Refs. 10–13 to the case considered here, in which successive collisions occur in the Fraunhofer zone of an individual scatterer, was analyzed in Refs. 29 and 49.

Let us consider the general expressions (24) and (25) under the conditions of multiple scattering at small angles. To avoid cumbersome calculations, we assume that the waves impinge on the medium along a normal to the surface. In this case it is convenient to go over from the unit vectors $\boldsymbol{\Omega}$ to the two-dimensional vectors $\boldsymbol{\theta}=\boldsymbol{\Omega}_\parallel$ and to assume that the components of $\boldsymbol{\theta}$ vary over an infinite range. In the new variables the general expressions (24) and (25) for the intensity fluctuation spectrum of the radiation transmitted through the slab take the form

$$M_T^{\rm coh}(\mathbf{q})=n\int_0^L dz\int d^2\theta''\int d^2\theta_1''\left|\int d^2\theta' I(L,\mathbf{q}|z,\boldsymbol{\theta}')\times h(\boldsymbol{\theta}',\boldsymbol{\theta}'|\boldsymbol{\theta}'',\boldsymbol{\theta}_1'')|^2 I(z,\boldsymbol{\theta}''|\boldsymbol{\theta}_0=0,\Delta\omega)\times I(z,\boldsymbol{\theta}_1''|\boldsymbol{\theta}_0=0,-\Delta\omega),\right. \tag{74}$$

$$M_T^{\rm incoh}(\mathbf{q})=n\int_0^L dz|\int d^2\theta'\int d^2\theta'' I(L,\mathbf{q}|z,\boldsymbol{\theta}')\times\left[\sigma_{\rm tot}\delta(\boldsymbol{\theta}'-\boldsymbol{\theta}'')-\frac{d\sigma}{d\Omega}(|\boldsymbol{\theta}'-\boldsymbol{\theta}''|)\right]\times I(z,\boldsymbol{\theta}''|\boldsymbol{\theta}_0=0)|^2, \tag{75}$$

where

$$h(\boldsymbol{\theta}',\boldsymbol{\theta}'|\boldsymbol{\theta}'',\boldsymbol{\theta}_1'')=\frac{2\pi i}{k_0}[f(\boldsymbol{\theta}'-\boldsymbol{\theta}'')\delta(\boldsymbol{\theta}'-\boldsymbol{\theta}_1'')-f^*(\boldsymbol{\theta}'-\boldsymbol{\theta}_1'')\delta(\boldsymbol{\theta}'-\boldsymbol{\theta}'')]+f(\boldsymbol{\theta}'-\boldsymbol{\theta}'')f^*(\boldsymbol{\theta}'-\boldsymbol{\theta}_1''). \tag{76}$$

In contrast to the case of the spatial diffusion of radiation, the fluctuation spectrum in a medium with large-scale centers is sensitive to the form of the single-scattering amplitude. Therefore, calculations based on Eqs. (74) and (75) must be performed with consideration of the specific angular dependence $f(\boldsymbol{\theta})$.

The qualitative features of the intensity correlations for small-angle scattering can be understood within the simplest models corresponding to the Fokker–Planck approximation or the diffusion approximation with respect to the angular variable.[37,42,50–53] In this approximation, on the one hand, the different contributions to the fluctuation spectrum can be separated fairly simply, and, on the other hand, it is possible to avoid the computational difficulties associated with a specific form of the scattering amplitude.

Let us assume that the deflection angle in one collision is small compared with the characteristic multiple-scattering angle and that the intensity propagators appearing in (74) and (75) are smoother functions of the angles than is the amplitude $f(\boldsymbol{\theta})$. Then in the transport equation (8) (or (11)), as well as in Eqs. (74) and (75), the intensity propagators can be expanded into a series in the small single-scattering angle.

The transport equation (11) is transformed in this approximation into the Fokker–Planck equation or into an angular diffusion equation.[42] Its solution was found with consideration of absorption in the medium in Refs. 50 and 51. Using the results in Ref. 51, we can write the propagators appearing in (74) and (75) in the form

$$I(z,\boldsymbol{\theta}|0,\Delta\omega=0)=\left(2\pi\sinh\frac{z}{\sqrt{l_a l_{\rm tr}}}\right)^{-1}\sqrt{\frac{l_{\rm tr}}{l_a}}\exp\left[-\frac{z}{l_a}-\frac{1}{2}\theta^2\sqrt{\frac{l_{\rm tr}}{l_a}}\coth\frac{z}{\sqrt{l_a l_{\rm tr}}}\right], \tag{77}$$

$$I(L,\mathbf{q}|z,\boldsymbol{\theta})=\left(\cosh\frac{L-z}{\sqrt{l_a l_{\rm tr}}}\right)^{-1}\exp\left[-\frac{L-z}{l_a}-\frac{\theta^2}{2}\sqrt{\frac{l_{\rm tr}}{l_a}}\tanh\frac{L-z}{\sqrt{l_a l_{\rm tr}}}-i\mathbf{q}\boldsymbol{\theta}\sqrt{l_a l_{\rm tr}}\tanh\frac{L-z}{\sqrt{l_a l_{\rm tr}}}-\frac{1}{2}q^2 l_a\left(L-z-\sqrt{l_a l_{\rm tr}}\tanh\frac{L-z}{\sqrt{l_a l_{\rm tr}}}\right)\right]. \tag{78}$$

The value of $I(z,\boldsymbol{\theta}|0,\Delta\omega)$ for a finite value of the frequency shift $\Delta\omega$ can be obtained from (77) by replacing l_a by $l_a/[1-i(\Delta\omega/c)l_a]$.

We transform the coherent contribution to the fluctuation spectrum (74) by expanding $I(L,\mathbf{q}|z,\boldsymbol{\theta}')$ in series in the vicinity of the direction $\boldsymbol{\theta}''_+=(\boldsymbol{\theta}''+\boldsymbol{\theta}''_1)/2$. As a result, we obtain

$$\int d^2\theta' I(L,\mathbf{q}|z,\boldsymbol{\theta}')h(\boldsymbol{\theta}'|\boldsymbol{\theta}'',\boldsymbol{\theta}''_1)$$
$$=I(L,\mathbf{q}|z,\boldsymbol{\theta}_+)\frac{d\sigma_a}{d\boldsymbol{\Omega}}(\boldsymbol{\theta}_-)+\left\{-\frac{\pi i}{k_0}[f(-\boldsymbol{\theta}_-)\right.$$
$$+f^*(\boldsymbol{\theta}_-)]\boldsymbol{\theta}_-+\int d^2\theta'(\boldsymbol{\theta}'-\boldsymbol{\theta}_+)f\left(\boldsymbol{\theta}'-\boldsymbol{\theta}_+\right.$$
$$\left.-\frac{1}{2}\boldsymbol{\theta}_-\right)$$
$$\left.\times f^*\left(\boldsymbol{\theta}'-\boldsymbol{\theta}_++\frac{1}{2}\boldsymbol{\theta}_-\right)+\ldots\right\}$$
$$\times\frac{\partial}{\partial\boldsymbol{\theta}_+}I(L,q|z,\boldsymbol{\theta}_+)+\ldots, \tag{79}$$

where $\boldsymbol{\theta}_-=\boldsymbol{\theta}''-\boldsymbol{\theta}''_1$,

$$\frac{d\sigma_a}{d\boldsymbol{\Omega}}(\boldsymbol{\theta}_-)=\frac{2\pi i}{k_0}[f(-\boldsymbol{\theta}_-)-f^*(\boldsymbol{\theta}_-)]$$
$$+\int d^2\theta' f\left(\boldsymbol{\theta}'-\frac{1}{2}\boldsymbol{\theta}_-\right)f^*\left(\boldsymbol{\theta}'+\frac{1}{2}\boldsymbol{\theta}_-\right)$$
$$\approx\frac{1}{k_0^2}\int_0^\infty d^2p\exp(i\mathbf{p}\cdot\boldsymbol{\theta})(1-|S_p|^2) \tag{80}$$

is the ''differential'' absorption cross section in the small-angle approximation,

$$S_p=|S_p|\exp(2i\delta_p)=1+2ik_0f_p,$$

and S_p and f_p are the expansion coefficients of the S-matrix and the partial-wave scattering amplitude.[39] Substituting the expansion (79) into (74), we arrive at the following series for the coherent contribution to the fluctuation spectrum:

$$M_T^{\mathrm{coh}}(\mathbf{q})=n\int_0^z dz\int d^2\theta\left[\alpha_a|I(L,\mathbf{q}|z,\boldsymbol{\theta})|^2\right.$$
$$+\alpha_{sp}\left|\frac{\partial}{\partial\boldsymbol{\theta}}I(L,\mathbf{q}|z,\boldsymbol{\theta})\right|^2$$
$$-\alpha_{a,s}\frac{\partial^2}{\partial\boldsymbol{\theta}^2}|I(L,\mathbf{q}|z,\boldsymbol{\theta})|^2+\alpha_s\left(\frac{\partial^2}{\partial\theta_i\partial\theta_k}\right.$$
$$\left.+\frac{\delta_{ik}}{2}\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\right)\frac{\partial}{\partial\theta_i}I(L,\mathbf{q}|z,\boldsymbol{\theta})\frac{\partial}{\partial\theta_k}I^*(L,\mathbf{q}|z,\boldsymbol{\theta})$$
$$\left.+\ldots\right]I^2(z,\boldsymbol{\theta}|0,\Delta\omega). \tag{81}$$

Equation (81) is the sum of the contributions (44) corresponding to different mechanisms for the appearance of intensity fluctuations.

The first term in (81) is the contribution $M_T^{\mathrm{coh}(a)}$ due to the disturbance of the bulk speckle structure by fluctuations of the number of absorbing particles in a small volume of the medium. When $\Delta\omega=0$, the expansion of the spectrum of $M_T^{\mathrm{coh}(a)}$ at small q to first order in l/L has the form

$$M_T^{\mathrm{coh}(a)}(\mathbf{q})$$
$$=n\alpha_a\int_0^L dz\int d^2\theta|I(L,\mathbf{q}|z,\boldsymbol{\theta})|^2I^2(z,\boldsymbol{\theta}|0,\Delta\omega=0)$$
$$=\frac{n\alpha_a}{4\pi}\langle T\rangle^2l_{\mathrm{tr}}\begin{cases}ln\dfrac{L}{l}\left(1-\dfrac{1}{3}q^2\dfrac{L^3}{l_{\mathrm{tr}}}+\ldots\right), & L<\sqrt{l_al_{\mathrm{tr}}},\\ 2\dfrac{L}{\sqrt{l_al_{\mathrm{tr}}}}\left(1-\dfrac{1}{2}q^2Ll_a+\ldots\right), & L>\sqrt{L_al_{\mathrm{tr}}}.\end{cases} \tag{82}$$

where

$$\alpha_a=\int d^2\theta\left[\frac{d\sigma_a}{d\boldsymbol{\Omega}}(\boldsymbol{\theta})\right]^2=\frac{(2\pi)^2}{k_0^4}\int d^2p(1-|S_p|^2)^2, \tag{83}$$

$$\langle T\rangle=\left(\cosh\frac{L}{\sqrt{l_al_{\mathrm{tr}}}}\right)^{-1}\exp\left(-\frac{L}{l_a}\right) \tag{84}$$

is the transmission coefficient through a slab of thickness L.[50,51] The expression for the spectrum at zero, $M_T^{\mathrm{coh}(a)}(\mathbf{q}=0)$, was obtained in a somewhat different form in Ref. 49.

The coefficient α_a (93) is proportional to the difference between the total cross section for the absorption of two ''rays'' propagating independently and the cross section for the absorption of two ''rays'' passing through one center:

$$\alpha_a=\frac{4\pi^2}{k_0^2}(2\sigma_a-\sigma_a^{(2)}),\quad \sigma_a=\frac{1}{k_0^2}\int d^2p(1-|S_p|^2),$$
$$\sigma_a^{(2)}=\frac{1}{k_0^2}\int d^2p(1-|S_p|^4). \tag{85}$$

We note that for ''black'' spherical particles of radius a

$$\sigma_a=\sigma_a^{(2)}=\pi a^2,\quad \alpha_a=\left(\frac{2\pi}{k_0}\right)^2\sigma_a.$$

The second term in (81) is the contribution $M_T^{\mathrm{coh(sp)}}$, which originates from the intensity peaks in the bulk speckle structure. In the same approximation in which (82) was obtained, the contribution $M_T^{\mathrm{coh(sp)}}$ has the form

$$M_T^{\mathrm{coh(sp)}}(\mathbf{q})$$
$$=n\alpha_{\mathrm{sp}}\int_0^L dz\int d^2\theta\left|\frac{\partial}{\partial\boldsymbol{\theta}}I(L,\mathbf{q}|z,\boldsymbol{\theta})\right|^2I^2(z,\boldsymbol{\theta}|0,\Delta\omega=0)$$
$$=\frac{n\alpha_{\mathrm{sp}}}{4\pi}\langle T\rangle^2l_{\mathrm{tr}}\begin{cases}\dfrac{1}{3}\dfrac{L^3}{l_a^2l_{\mathrm{tr}}}+q^2L^2\ln\dfrac{L}{l}+\ldots, & L<\sqrt{l_al_{\mathrm{tr}}},\\ \dfrac{L}{l_a}\left(1-\dfrac{1}{2}q^2Ll_a+\ldots\right), & L>\sqrt{L_al_{\mathrm{tr}}}.\end{cases} \tag{86}$$

where

$$\alpha_{\rm sp}=\left(\frac{2\pi}{k_0}\right)^2\sigma_{\rm tr}=2\frac{(2\pi)^2}{k_0^4}\int d^2p\left(\frac{d\delta}{dp}\right)^2|S_p|^2. \tag{87}$$

Equation (86) was previously obtained in Ref. 29 in the Born approximation for the single-scattering cross section using the solution of the small-angle transport equation for the fourth moment.

The ''cross'' contribution to the spectrum, $M_T^{\rm coh(a,s)}$, is specified by the third term in (81). It is governed by the correlation between disturbances to the bulk speckle structure by the fluctuations of the absorbing and scattering properties of the medium. The expression for $M_T^{\rm coh(a,s)}$ has the form

$$\begin{aligned}M_T^{\rm coh(a,s)}(\mathbf{q})&=-n\alpha_{a,s}\int_0^L dz\int d^2\theta\\&\times\left(\frac{\partial^2}{\partial\boldsymbol{\theta}^2}|I(L,\mathbf{q}|z,\boldsymbol{\theta})|^2\right)I^2(z,\boldsymbol{\theta}|0,\Delta\omega=0)\\&=\frac{n\alpha_{a,s}}{4\pi}\langle T\rangle^2 l_{\rm tr}\begin{cases}4\dfrac{L}{l_a}\ln\dfrac{L}{l}-\dfrac{8}{3}\dfrac{L^3}{l_a^2 l_{\rm tr}}-\dfrac{4}{3}q^2\dfrac{L^4}{l_a l_{\rm tr}}\ln\dfrac{L}{l}+\ldots,\\ L<\sqrt{l_a l_{\rm tr}},\\ 4\dfrac{L}{l_a}\left(1-\dfrac{1}{2}q^2Ll_a+\ldots\right),\\ L>\sqrt{l_a l_{\rm tr}},\end{cases}\end{aligned} \tag{88}$$

where

$$\alpha_{a,s}=\frac{(2\pi)^2}{k_0^4}\int d^2p\left(\frac{d\delta}{dp}\right)^2(1-|S_p|^2). \tag{89}$$

The fourth term in (81) corresponds to $M_T^{\rm coh(s)}$, which is associated with the disturbance of the bulk speckle by fluctuations of the scattering properties of the medium. The following expansion is valid for $M_T^{\rm coh(s)}$:

$$\begin{aligned}M_T^{\rm coh(s)}(\mathbf{q})&=n\alpha_s\int_0^L dz\int d^2\theta\left(\frac{\partial^2}{\partial\theta_i\partial\theta_k}+\frac{\delta_{ik}}{2}\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\right)\\&\times\frac{\partial}{\partial\theta_i}I(L,\mathbf{q}|z,\boldsymbol{\theta})\frac{\partial}{\partial\theta_k}I^*(L,\mathbf{q}|z,\boldsymbol{\theta})I^2(z,\boldsymbol{\theta}|0,\Delta\omega=0)\\&=\frac{n\alpha_s}{4\pi}\langle T\rangle^2 l_{\rm tr}\begin{cases}4\ln\dfrac{L}{l}\left[\left(\dfrac{L}{l_a}\right)^2+q^2\dfrac{L^3}{l_a}+\dfrac{3}{8}q^4L^4+\ldots\right],\\ L<\sqrt{l_a l_{\rm tr}},\\ 2\dfrac{L}{l_a}\sqrt{\dfrac{l_a}{l_{\rm tr}}}\left(1-\dfrac{1}{2}q^2Ll_a+\ldots\right),\\ L>\sqrt{l_a l_{\rm tr}},\end{cases}\end{aligned} \tag{90}$$

where

$$\alpha_s=\frac{(2\pi)^2}{k_0^4}\int d^2p\left(\frac{d\delta}{dp}\right)^4. \tag{91}$$

The contribution $M_T^{\rm coh(s)}$ is proportional to the fourth power of the scattering amplitude and is not included in the usual Born approximation for single scattering.[29]

Equation (81) represents only the dominant terms corresponding to contributions of different nature to the fluctuation spectrum. For this reason, the term $(\partial I/\partial\boldsymbol{\theta})\times(\partial^3 I^*/\partial\boldsymbol{\theta}^3)+$c.c., which has the same origin as the contribution (86) and is only a correction to it, is not written out separately in (81). To see this, it is sufficient to perform the small-angle expansion in Eq. (27).

Using the relations obtained above for the various contributions to the fluctuation spectrum and Eq. (21), we find

$$\begin{aligned}\frac{\langle(\delta T)^2\rangle^{\rm coh}}{\langle T\rangle^2}&=\frac{\pi n}{Ak_0^4}l_{\rm tr}\\&\times\int d^2p\begin{cases}\dfrac{2}{3}\left(\dfrac{d\delta}{dp}\right)^2\dfrac{L^3}{l_a^2 l_{\rm tr}}+\left[(1-|S_p|^2)+2\left(\dfrac{d\delta}{dp}\right)^2\dfrac{L}{l_a}\right]^2\ln\dfrac{L}{l},\\ L\ll\sqrt{l_a l_{\rm tr}},\\ 2\left(\dfrac{d\delta}{dp}\right)^2\dfrac{L}{l_a}+2\left[(1-|S_p|^2)+\left(\dfrac{d\delta}{dp}\right)^2\sqrt{\dfrac{l_{\rm tr}}{l_a}}\right]^2\dfrac{L}{\sqrt{l_a l_{\rm tr}}},\\ L\gg\sqrt{l_a l_{\rm tr}}.\end{cases}\end{aligned} \tag{92}$$

The first term in (92) is associated with the formation of a random interference pattern in the bulk of the medium, and the term in square brackets is associated with the disturbance of this pattern by Poisson fluctuations of the distribution of the scatterers. The contribution to the dispersion $\langle(\delta T)^2\rangle^{\rm coh}$ caused by the local inhomogeneity of the medium is actually an expansion in powers of a small parameter, viz., the mean multiple-scattering angle in the slab $\langle\theta^2\rangle_L$ ($\langle\theta^2\rangle_L=2L/l_{\rm tr}$ for $L<\sqrt{l_a l_{\rm tr}}$ and $\langle\theta^2\rangle_L=2\sqrt{l_a/l_{\rm tr}}$ for $L>\sqrt{l_a l_{\rm tr}}$ (Refs. 50–53))].

If the absorption is caused by the particles of the medium, the first of the terms associated with Poisson fluctuations, which originates from fluctuations of the absorbing properties of the medium, will be dominant. As a result, only two contributions, viz., $M_T^{\rm coh(sp)}$ and $M_T^{\rm coh(a)}$, can compete in Eqs. (81) and (92). A comparison of these contributions gives the following universal estimate for the ratio between them:

$$\frac{M_T^{\rm coh(sp)}}{M_T^{\rm coh(a)}}\sim\frac{\langle\theta^2\rangle_L^3}{\langle\vartheta^2\rangle},$$

where $\langle\vartheta^2\rangle\approx 2(1-\langle\cos\vartheta\rangle)$ is the mean square of the single-scattering angle. It follows from this estimate that the ratio between $M_T^{\rm coh(sp)}$ and $M_T^{\rm coh(a)}$ can take any value. However, as the thickness increases, the role of $M_T^{\rm coh(sp)}$ increases in the general case. In particular, when $L>\sqrt{l_a l_{\rm tr}}$, this contribution is always dominant, if the scattering cross section is not very great, $\sigma_a<\sigma(\sigma_{\rm tr}/\sigma)^{1/3}<\sigma$.

If the particles only scatter radiation, and the continuum between them absorbs radiation, only two contributions, viz., $M_T^{\rm coh(sp)}$ and $M_T^{\rm coh(s)}$, remain in the expansions (81) and (92). The ratio between them in order of magnitude equals

$$\frac{M_T^{\mathrm{coh(sp)}}}{M_T^{\mathrm{coh}(s)}}\sim\frac{\langle\theta^2\rangle_L}{\langle\vartheta^2\rangle}\gg 1.$$

In this case all the qualitative conclusions obtained in the Born approximation for the cross sections[29] remain valid.

The following asymptotes of the correlation functions correspond to the main contribution to the spectrum:

$$C_T^{\mathrm{coh(sp)}}(\rho)=\frac{\langle T\rangle^2}{k_0^2}\times\begin{cases}4\dfrac{L^2}{\rho_L^4}\ \ln\dfrac{L}{l}\left(1-\dfrac{\rho^2}{\rho_L^2}\right)\exp\left(-\dfrac{\rho^2}{\rho_L^2}\right), & L<\sqrt{l_a l_{\mathrm{tr}}},\\ \dfrac{2}{\rho_L^2}\sqrt{\dfrac{l_{\mathrm{tr}}}{l_a}}\left(\dfrac{\rho_L^4}{8l_a^{3/2}l_{\mathrm{tr}}^{1/2}\rho^2}-1\right)\exp\left(-\dfrac{\rho^2}{\rho_L^2}\right), & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{93}$$

$$C_T^{\mathrm{coh}(a)}(\rho)=\frac{n\alpha_a}{(2\pi)^2}\langle T\rangle^2 l_{\mathrm{tr}}\begin{cases}\dfrac{1}{\rho_L^2}\ \ln\dfrac{L}{l}\exp\left(-\dfrac{\rho^2}{\rho_L^2}\right), & L<\sqrt{l_a l_{\mathrm{tr}}},\\ \dfrac{2L}{l_a^{1/2}l_{\mathrm{tr}}^{1/2}}\dfrac{l}{\rho_L^2}\exp\left(-\dfrac{\rho^2}{\rho_L^2}\right), & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{94}$$

where $\rho_L^2=4L^3/(3l_{\mathrm{tr}})$ for $L<\sqrt{l_a l_{\mathrm{tr}}}$ and $\rho_L^2=4Ll_a$ for $L>\sqrt{l_a l_{\mathrm{tr}}}$.

Let us now consider the incoherent contribution to the fluctuation spectrum. Assuming that the intensity propagators appearing in (75) are smoother functions of the angles than the differential single-scattering cross section, we write the expression for $M_T^{\mathrm{incoh}}(\mathbf{q})$ in the following form:

$$M_T^{\mathrm{incoh}}(\mathbf{q})=n\int_0^L dz\left|\int d^2\theta\left(\sigma_a I(L,\mathbf{q}|z,\boldsymbol{\theta})I(z,\boldsymbol{\theta}|0)+\frac{\sigma_{\mathrm{tr}}}{2}\frac{\partial}{\partial\boldsymbol{\theta}}I(L,\mathbf{q}|z,\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}}I(z,\boldsymbol{\theta}|0)\right)\right|^2.\tag{95}$$

The first term under the absolute-value sign in (95) corresponds to fluctuations of the absorption properties of the medium, and the second term corresponds to the scattering properties of the medium. For this reason, Eq. (95) can be written as the sum of contributions (45). Substituting (77) and (78) into (95), we obtain

$$M_T^{\mathrm{incoh}}(\mathbf{q})=n\langle T\rangle^2[\sigma_a^2 m_a(\mathbf{q})+\sigma_a\sigma_{\mathrm{tr}}m_{a,s}(\mathbf{q})+\sigma_{\mathrm{tr}}^2 m_s(\mathbf{q})],\tag{96}$$

where

$$m_a(\mathbf{q})=\begin{cases}L-\dfrac{1}{6}q^2\dfrac{L^4}{l_{\mathrm{tr}}}+\ \ldots, & L<\sqrt{l_a l_{\mathrm{tr}}},\\ L\left(1-\dfrac{1}{2}q^2Ll_a+\ \ldots\right), & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{97}$$

$$m_{a,s}(\mathbf{q})=\begin{cases}\dfrac{L^2}{l_a}+\dfrac{1}{3}q^2L^3+\ \ldots, & L<\sqrt{l_a l_{\mathrm{tr}}},\\ L\sqrt{\dfrac{l_{\mathrm{tr}}}{l_a}}\left(1-\dfrac{1}{2}q^2Ll_a+\ \ldots\right), & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{98}$$

$$m_s(\mathbf{q})=\begin{cases}\dfrac{1}{3}\dfrac{L^3}{l_a^2}+\dfrac{1}{4}q^2\dfrac{L^4}{l_a}+\dfrac{1}{20}q^4L^5+\ \ldots, & L<\sqrt{l_a l_{\mathrm{tr}}},\\ \dfrac{1}{4}L\dfrac{l_{\mathrm{tr}}}{l_a}\left(1-\dfrac{1}{2}q^2Ll_a+\ \ldots\right), & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{99}$$

When $q=0$, Eqs. (96)–(99) specify the dispersion of the transmission coefficient across the slab [see Eq. (21)]:

$$\frac{\langle(\delta T)^2\rangle^{\mathrm{incoh}}}{\langle T\rangle^2}=\frac{\sigma_a}{A}\frac{L}{l_a}\begin{cases}1+\dfrac{L}{l_{\mathrm{tr}}}+\dfrac{1}{3}\left(\dfrac{L}{l_{\mathrm{tr}}}\right)^2, & L<\sqrt{l_a l_{\mathrm{tr}}},\\ 1+\sqrt{\dfrac{l_a}{l_{\mathrm{tr}}}}+\dfrac{1}{4}\dfrac{l_a}{l_{\mathrm{tr}}}, & L>\sqrt{l_a l_{\mathrm{tr}}}.\end{cases}\tag{100}$$

It follows from (96)–(100) that if absorption is caused by the particles of the unordered medium themselves, the fluctuations of the absorption coefficient of the medium are the source of the incoherent intensity fluctuations. Conversely, if the particles of the medium only scatter radiation, and the continuum between them absorbs radiation, there are no fluctuations of the absorption, and only the term proportional to $m_s(\mathbf{q})$ remains in (96). In this case the dispersion of the transmission is determined by the third terms in Eq. (100).

The asymptote of the correlation function for the spectrum (96)–(99) has the form:

$$C_T^{\mathrm{incoh}}(\rho)=n\langle T\rangle^2\frac{\exp(-\rho^2/\rho_L^2)}{\pi\rho^2}L\times\begin{cases}\left(\sigma_a^2-3\dfrac{\rho^2}{\rho_L^2}\dfrac{l_{\mathrm{tr}}}{L}\sigma_a\sigma_{\mathrm{tr}}+9\dfrac{\rho^4}{\rho_L^4}\dfrac{l_{\mathrm{tr}}}{L}\sigma_{\mathrm{tr}}^2\right)\dfrac{1}{2}\sqrt{\dfrac{\pi}{3}}\dfrac{\rho}{\rho_L}, & L<\sqrt{l_a l_{\mathrm{tr}}},\\ \sigma_a^2+\sqrt{\dfrac{l_{\mathrm{tr}}}{l_a}}\sigma_a\sigma_{\mathrm{tr}}+\dfrac{1}{4}\dfrac{l_{\mathrm{tr}}}{l_a}\sigma_{\mathrm{tr}}^2, & L>\sqrt{l_a l_{\mathrm{tr}}},\end{cases}\tag{101}$$

where the meaning of ρ_L^2 is the same as in (93) and (94), and $\rho>\rho_L$ is assumed. As follows from (101), at small thicknesses ($L<\sqrt{l_a l_{\mathrm{tr}}}$) the long-range asymptote $C_T^{\mathrm{incoh}}(\rho)$ is determined by the fluctuations of the scattering properties of the medium (when $\rho>\rho_L(l_{\mathrm{tr}}L/l_a^2)^{1/4}$, the last term dominates in (101)). When $L>\sqrt{l_a l_{\mathrm{tr}}}$, the correlation function, as well as the fluctuation spectrum, are determined primarily by the fluctuations of the absorption.

In the case in which the incident waves differ in frequency, the fluctuations of interference origin weaken, and the incoherent fluctuations remain unchanged. The spectrum

$M_T^{\mathrm{coh}}(\mathbf{q},\Delta\omega)$ can be calculated using Eq. (81) after substituting into it the value of $I(z,\boldsymbol{\theta}|0,\Delta\omega)$, which is obtained from (77) by replacing l_a by $l_a(1-i(\Delta\omega/c)l_a)^{-1}$. We shall not dwell on the analysis of the spectrum, and we shall consider only the correlations between the values of the total flux at different frequencies. In connection with the experiments in Refs. 20–25 it is interesting to understand how sensitive the frequency correlations of the intensity are to the features of the scattering of the waves in the unordered medium.

The asymptote of the correlation function

$$C_T(\Delta\omega)=\langle(T_{\omega_0+\Delta\omega/2}-\langle T_{\omega_0+\Delta\omega/2}\rangle)(T_{\omega_0-\Delta\omega/2}-\langle T_{\omega_0-\Delta\omega/2}\rangle)\rangle^{\mathrm{coh}}=\frac{1}{A}M_T^{\mathrm{coh}}(\mathbf{q}=0,\Delta\omega)$$

has the following form:

$$C_T(\Delta\omega)=\frac{\pi}{A}\frac{n}{k_0^4}\langle T\rangle^2 l_{\mathrm{tr}}\int d^2p\left[3\left(\frac{d\delta}{dp}\right)^2\frac{l_{\Delta\omega}}{l_a}\tanh^2\frac{L}{\sqrt{l_a l_{\mathrm{tr}}}}+(1-|S_p|^2)^2\left(\ln\frac{l_{\Delta\omega}}{l}+2\frac{l_{\Delta\omega}}{\sqrt{l_a l_{\mathrm{tr}}}}\tanh\frac{L}{\sqrt{l_a l_{\mathrm{tr}}}}+\ldots\right)\right]. \tag{102}$$

In (102) $l_{\Delta\omega}=\sqrt{2cl_{\mathrm{tr}}/\Delta\omega}$, and it is assumed that $l<l_{\Delta\omega}<\min(L,\sqrt{l_a l_{\mathrm{tr}}})$. If the particles of the medium do not absorb radiation, from (102) we obtain

$$C_T(\Delta\omega)=\frac{3\pi}{2Ak_0^2}\langle T\rangle^2\frac{l_{\Delta\omega}}{l_a}\tanh^2\frac{L}{\sqrt{l_a l_{\mathrm{tr}}}}. \tag{103}$$

Comparing (102) and (103), we can conclude that the fluctuations of the absorption properties of the medium lead to the appearance of a contribution to $C(\Delta\omega)$ that depends weakly on the frequency. It is fairly simple to account for this effect. At relatively large values of $\Delta\omega$ the region in space where the waves interfere and form the bulk speckle is bounded by the depths $z<l_{\Delta\omega}$. Therefore, in the case under consideration the length $l_{\Delta\omega}$ plays the same role as the slab thickness L in the case $\Delta\omega=0$ for $L<\sqrt{l_a l_{\mathrm{tr}}}$ (see (82) and (92)).

We note that when waves which differ fairly strongly in frequency ($l_{\Delta\omega}<l_{\mathrm{tr}}$) propagate, the general formula (81) can be used to calculate the fluctuation spectrum not only for small-angle scattering, but also in the case of the diffusive transport of the waves in a thick ($L\gg l_{\mathrm{tr}}$) slab. In this situation the ''incoming'' propagators in (81) describe the small-angle scattering of waves in the case of propagation in relatively thin (with a thickness of the order of $l_{\Delta\omega}$) slabs, while the ''outgoing'' propagators correspond to the propagation of waves in a spatial-diffusion regime. The main contribution will be made by the second term in (81), since the diffusion regime presumes fulfillment of the conditions of weak absorption [$l_{\mathrm{tr}}\ll l_a$ or $1-|S_p|^2\ll(d\delta/dp)^2$]. In the transmission geometry

$$C_T(\Delta\omega)=\frac{\pi}{2Ak_0^2}\langle T\rangle^2\frac{l_{\Delta\omega}}{l_{\mathrm{tr}}}, \tag{104}$$

where $\langle T\rangle$ is the transmission coefficient of radiation through a thick ($L\gg l_{\mathrm{tr}}$) slab of the medium under normal incidence. A similar formula can be obtained for the case of reflection, if the values of the ''outgoing'' propagators at $z=0$, rather than $z=L$, are substituted into (81). When $L\ll\sqrt{l_a l_{\mathrm{tr}}}$, the results of the calculations do not differ from (104), as a consequence of the conservation of the flux.

It would be interesting to compare Eq. (104) with the value of $C_T(\Delta\omega)$ for small frequency shifts ($l_{\Delta\omega}\gg l_{\mathrm{tr}}$), which can easily be obtained using the formulas in Sec. 3 after replacing l by l_{tr} in them. The corresponding results differ only with respect to a common numerical factor (it is three times smaller in (104)) because of the differences in the propagation regime of the interfering waves in the region $z<l_{\Delta\omega}$.

5. DISCUSSION

General relations were derived above, which make it possible to reduce the problem of calculating the fluctuation spectrum to the solution of the transport equation for the mean intensity, and the results of calculations performed for two characteristic regimes for the multiple scattering of waves in an unordered medium, viz., spatial diffusion and small-angle scattering, were presented. The results obtained provide a fairly complete picture of the long-range intensity correlations when coherent radiation propagates in a medium with discrete scatterers.

In the general case there are two sources of intensity fluctuations, viz., random interference of the multiply scattered waves and random spatial inhomogeneity of the scattering and absorption coefficients of the medium, which is caused by fluctuations of the number of scatterers per unit volume. Both sources make contributions to the coherent intensity fluctuations, and only the latter source makes a contribution to the incoherent fluctuations.

The relative role of each of the mechanisms for the appearance of fluctuations indicated above, as well as the features of the behavior of the correlation function and the intensity fluctuation spectrum depend on the regime for the multiple scattering of waves in the medium, the presence of absorption, the dimensions of the scattering particles, and the degree of monochromaticity of the incident radiation.

When waves propagate in the spatial-diffusion regime ($l_{\mathrm{tr}}\ll\min(L,l_a)$), the purely interference mechanism for the appearance of intensity fluctuations predominates. The fluctuations caused by the random inhomogeneity of the medium generally comprise a correction to the purely interference contribution. Their role is most important for the diffusion of waves in a medium with strong (''non-Born,'' $k_0a|n_0-1|>1$) scatterers. In this situation the incoherent contribution to the spectrum becomes appreciable, and when $q>l_{\mathrm{tr}}^{-1}\sqrt{k_0^2\sigma_{\mathrm{tr}}}$, it exceeds the contribution associated with interference in the transmission geometry (see Fig. 3).

We note that when the length of the ray trajectories is bounded ($\min(L,\sqrt{l_a l_{\mathrm{tr}}})<l_{\mathrm{tr}}(k_0 l_{\mathrm{tr}})^2(k_0^2\sigma_{\mathrm{tr}})$), the fluctuations caused by the random inhomogeneity of the medium, which

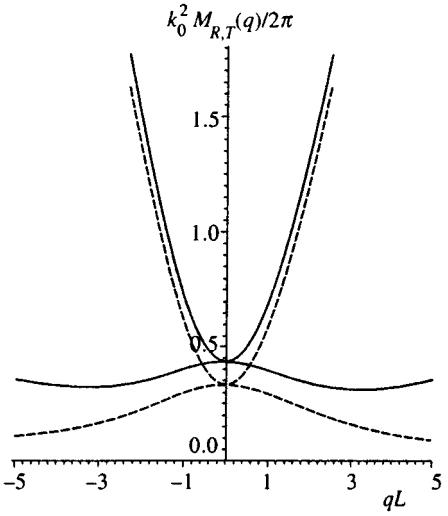

FIG. 3. Fluctuation spectra of the reflected (upper curves) and transmitted (lower curves) intensity for elastic scattering in a layer of an unordered medium. The thickness of the slab $L=10l_{\rm tr}$, and absorption is absent. Solid curves — strong scatterers ($k_0a|n_0-1|=2$), dashed lines — limiting case of ''Born'' centers ($k_0a|n_0-1|\ll 1$).

are primarily of incoherent origin, are greater than the previously discussed (see, for example, Refs. 16, 17, and 19) second-order interference contribution to the interaction of the ''rays.''

The incoherent contribution to the fluctuation spectrum plays the principal role when the incident waves differ in frequency, and the dominant coherent contribution decreases as the frequency shift $\Delta\omega$ increases. Consideration of the incoherent contribution can account for the experimentally observed horizontal segment[23,24] on the decay of the correlation curve with increasing $\Delta\omega$ and can alter the quantitative results for the intensity correlator. For this reason, the procedure used in Refs. 23 and 24 to extract the values of the transport coefficients of the medium from the decay of the correlation curve, as well as the interpretation of the experimental data in Refs. 23 and 24, should be re-examined.

The occurrence of absorption when waves propagate in the spatial-diffusion regime is manifested mainly in the alteration of the distribution of the multiply scattered radiation in the medium. The additional fluctuations appearing because of absorption in the scattering particles themselves can be appreciable only for fairly large particles, if the condition $k_0^2\sigma_a\sim(k_0a)^3{\rm Im}\ n_0>1$ is satisfied (the assumption for a diffusion regime, $\sigma_a\ll\sigma_{\rm tr}$, must be satisfied simultaneously). As an analysis shows, similar conditions were achieved in the experiments in Refs. 23 and 24.

In the case of the strongly anisotropic multiple scattering of waves ($l_{\rm tr}\gg\min(L,l_a)$), the purely interference source of fluctuations remains important, as before, but the role of the effects caused by fluctuations of the number of scatterers become more appreciable than in the spatial-diffusion regime.

If radiation is absorbed in the continuum, and the scattering particles themselves do not absorb radiation, the ratio between the dominant interference and incoherent contributions to the spectrum depends on the value of $k_0^2\sigma_{\rm tr}$; for weakly refractive particles of large radius

$$k_0^2\sigma_{\rm tr}\simeq 8\pi(k_0a|n_0-1|)^2\ \ln\frac{1}{|n_0-1|}$$

(Ref. 53). In the Born case ($k_0a|n_0-1|\ll 1$) the interference contribution predominates, and in the opposite case of strong scatterers ($k_0a|n_0-1|>1$) the incoherent contribution to the fluctuation spectrum predominates. For strong scatterers the interference contribution is important only in thin slabs ($L<\sqrt{l_al_{\rm tr}}$), where it determines the dependence of the spectrum on the spatial frequency q at $q<(k_0^2\sigma_{\rm tr}L^3/l_{\rm tr})^{-1/2}$.

If the particles of the unordered medium themselves absorb radiation, the role of the fluctuations of the number of particles per unit volume becomes especially important. The spatial fluctuations of the absorption coefficient appearing in this case is manifested in both the coherent and incoherent contributions to the fluctuation spectrum, but the coherent contribution is determined to the same extent by the random interference of the waves and by the disturbance of the interference pattern by fluctuations of the absorption in the medium. In the situation under consideration the ratio between the coherent and incoherent contributions to the spectrum changes. The incoherent contribution can turn out to be dominant even in the Born case ($k_0^2\sigma_{\rm tr}\sim(k_0a|n_0-1|)^2\ll 1$) provided the absorption cross section is large enough ($k_0^2\sigma_a\sim(k_0a)^3{\rm Im}\ n_0>1$).

We note that the various terms in the coherent and incoherent contributions to the fluctuation spectrum should be comparable to one another under the intermediate conditions $\sigma\sim\sigma_{\rm tr}\sim\sigma_a\sim k_0^{-2}$ ($k_0a\sim|n_0-1|\sim{\rm Im}\ n_0\sim 1$), under which the wave propagation regime does not reduce to diffusion or small-angle scattering.

As follows from the results obtained above, an increase in absorption (when the other optical parameters of the medium are left unchanged) generally leads to enhancement of the intensity fluctuations. The simplest reason for this is the fluctuations of the absorption coefficient in the case in which absorption takes place in the particles of the unordered medium. If the continuum between the particles absorbs radiation, only the distribution of the radiation in the medium varies as the absorption increases. The effective lengths of the ray trajectories decrease, and the fluctuation spectrum narrows. However, the relative magnitude of the dispersion of the transmission coefficient fluctuations $\langle(\delta T)^2\rangle/\langle T\rangle^2$ exhibits nonmonotonic behavior as the absorption in the medium increases. For large-scale scatterers $\langle(\delta T)^2\rangle/\langle T\rangle^2$ can be estimated by taking into account only the purely interference and incoherent contributions (the additional terms of the coherent contribution are smaller than the incoherent contribution in this case). As a result, we obtain

$$\frac{\langle(\delta T)^2\rangle}{\langle T\rangle^2}=\frac{L}{Al_{\rm tr}}\begin{cases}\dfrac{2\pi}{k_0^2}+\sigma_{\rm tr}\left(\dfrac{l_{\rm tr}}{L}\right)^2, & l_{\rm tr}<L<\sqrt{l_a l_{\rm tr}},\\ \dfrac{3\pi}{2k_0^2}+\sigma_{\rm tr}\dfrac{l_{\rm tr}}{l_a}, & l_{\rm tr}<\sqrt{l_a l_{\rm tr}}<L,\\ \dfrac{\pi}{k_0^2}+\sigma_{\rm tr}\dfrac{l_{\rm tr}}{l_a}, & \sqrt{l_a l_{\rm tr}}<l_{\rm tr}<L.\end{cases}\tag{105}$$

It follows from (105) that in the diffusion regime the interference contribution decreases with increasing absorption, and after going over to the propagation regime with a strongly anisotropic angular distribution of the radiation it begins to increase. As for the incoherent contribution in (105), it increases monotonically with increasing absorption in the medium, and when $L>\sqrt{l_a l_{\rm tr}}$, the formulas obtained for the cases of spatial diffusion and small-angle multiple scattering coincide.

One more important qualitative law is associated with the influence of the conservation of the flux on the form of the intensity fluctuation spectrum. If almost the entire incident flux is reflected ($1-\langle R\rangle\ll 1$), the fluctuation spectrum of the reflected intensity contains a dip (Fig. 3). There is a similar feature in the spectrum of the transmitted radiation scattered at small angles when almost the entire flux passes through the medium ($1-\langle T\rangle\ll 1$, see Eqs. (82), (86), and (96)).

If absorption occurs in the continuum, the dip in the fluctuation spectrum at $q=0$ is maintained to large depths. The form of the spectrum in latter case will now depend on the value of the ''reduced'' transmission coefficient $\langle\tilde{T}\rangle$ ($\langle T\rangle=\exp(-L/l_a)\langle\tilde{T}\rangle$), which describes the decrease in the transmission of radiation by the layer of the medium due to the bending of the ray trajectories that appears upon scattering. As long as $1-\langle\tilde{T}\rangle\ll 1$, random redistribution of the intensity in the observation plane will occur at an assigned transmitted flux equal to the incident flux reduced by the factor $\exp(-L/l_a)$. When $\langle\tilde{T}\rangle<1$, the condition associated with the conservation of the flux becomes insignificant, and a maximum appears in the spectrum instead of the dip. A similar maximum is always found in the transmitted intensity spectrum when the waves propagate in the spatial-diffusion regime, since in that case $\langle T\rangle\ll 1$.

The behavior of the spectrum in the vicinity of $q=0$, like the form of the coherent backscattering peak,[3,7,43–47] is determined by the length of the ray trajectories in the medium. In particular, in the case of reflection from a semi-infinite nonabsorbing medium, in which the trajectory length is unbounded, the dip in the spectrum has a triangular shape.

In conclusion, we note that the results obtained above can also be used to describe the correlations between time-shifted intensity values $\langle I(\mathbf{r},\mathbf{\Omega},t)I(\mathbf{r}_1,\mathbf{\Omega}_1,t_1)\rangle$ in an unordered medium with slowly moving particles. The corresponding changes affect only the coherent contribution to the fluctuation spectrum. To calculate $M^{\rm coh}(z_f,\mathbf{q},t-t_1)$ the additional ''effective absorption'' $n\sigma_a^{\rm add}(t-t_1)$, which describes the breakdown of the coherence of the waves due to the motion of the scatterers, must be substituted into the ''incoming'' propagators instead of $-i\Delta\omega/c$ in Eqs. (24), (42), and (74) (see, for example, Refs. 3, 37, 54, and 55). In the most interesting case of small values of $t-t_1$, in which the interference is still not suppressed,

$$n\sigma_a^{\rm add}(t-t_1)=n\sigma_{\rm tr}k_0^2\langle(\Delta x)^2\rangle_{t-t_1}\ll n\sigma_{\rm tr},$$

where $\langle(\Delta x)^2\rangle_t$ is the mean square displacement of a scattering particle during the time t. For Brownian motion $\langle(\Delta x)^2\rangle_t=2Dt$, where D is the diffusion coefficient of the particles.

We thank M. Yu. Cherkasov for his cooperation in the initial stage of this work. We also thank E. E. Gorodnichev, S. L. Dudarev, and A. I. Kuzovlev for their interest in this work and valuable advice, as well as S. Feng for a useful discussion of some problems pertaining to this work. This research was carried out with the support of the International Science Foundation (Grants Nos. N3U000 and N3U300) and the Russian Foundation for Fundamental Research (Grant No. 95.02.05530).

APPENDIX A

The series of diagrams specifying the fourth moment of the field (Fig. 1c) can be regrouped, and $\mathscr{G}_4$ can be represented in the form of a sum of three contributions (Fig. 4):

$$\mathscr{G}_4(1,2,3,4)=\mathscr{G}_4^{(\rm sc)}(1,2;\,3,4)+\mathscr{G}_4^{(\rm sc)}(1,4;\,3,2)+\mathscr{G}_4^{(\rm non)}(1,2,3,4),\tag{A1}$$

where $\mathscr{G}_4^{(\rm sc)}$ is specified by diagrams containing at least one scattering event in the ''outgoing'' ladder propagators, and $\mathscr{G}_4^{(\rm non)}$ is specified by diagrams in which the ''outgoing'' propagators correspond to the unscattered field. The representation (A1) does not violate the symmetry condition (13), although the terms $\mathscr{G}_4^{(\rm sc)}$, taken individually, do not satisfy that condition.

Under multiple-scattering conditions ($l\ll L$) the main contribution to the fourth moment of the field is made by the $\mathscr{G}_4^{(\rm sc)}$. This allows us to write a closed equation for $\mathscr{G}_4^{(\rm sc)}$ (Fig. 5):

$$\begin{aligned}\mathscr{G}_4^{(\rm sc)}(1,2;3,4)=&\,\mathscr{G}_2(1,2)\mathscr{G}_2(3,4)-\mathscr{G}_2^{(0)}(1,2)\mathscr{G}_2^{(0)}(3,4)\\&+[\mathscr{G}_2(1,2)\mathscr{G}_2(3,4)-\mathscr{G}_2^{(0)}(1,2)\mathscr{G}_2^{(0)}(3,4)]\\&\times\sum_a\hat{h}_a(1,2)\hat{h}_a(3,4)[\mathscr{G}_4^{(\rm sc)}(1,2;3,4)+\mathscr{G}_4^{(\rm sc)}\\&\times(1,4;3,2)+\mathscr{G}_1(1)\mathscr{G}_1^*(2)\mathscr{G}_1(3)\mathscr{G}_1^*(4)],\end{aligned}\tag{A2}$$

where the numbers $1,\ldots,4$ denote the variables related to the fields $\Psi(1),\ldots,\Psi^*(4)$ appearing in the definition (1). Equation (A2) follows from (4), if $\mathscr{G}_4$ in the form (A1) is substituted into the integral term in Eq. (4) and the approximation $\mathscr{G}_4^{(\rm non)}=\mathscr{G}_1(1)\mathscr{G}_1^*(2)\mathscr{G}_1(3)\mathscr{G}_1^*(4)$ is adopted.

By definition, $\mathscr{G}_4^{(\rm sc)}(1,2;3,4)$ is an abruptly varying function of the difference variables $\mathbf{r}_1-\mathbf{r}_2$ and $\mathbf{r}_3-\mathbf{r}_4$ (the scale of variation of $\mathscr{G}_4^{(\rm sc)}$ with respect to these variables is of the

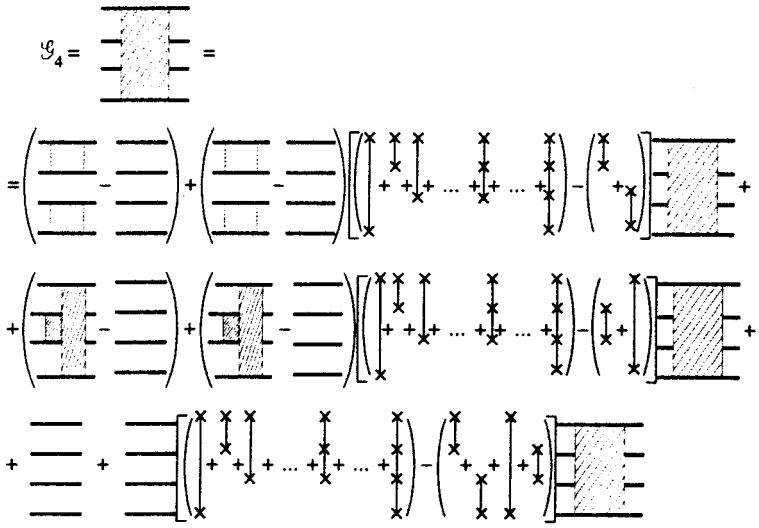

FIG. 4.

order of r_{coh}) and a smooth function of the $(\mathbf{r}_1+\mathbf{r}_2)/2$ and $(\mathbf{r}_3+\mathbf{r}_4)/2$. Therefore, to treat $\mathscr{G}_4^{(\mathrm{sc})}$, we can go over to the $(\mathbf{r},\boldsymbol{\Omega})$ representation. For this purpose, the Fourier transform with respect to the difference variables must be taken in (A2). As a result, for the second moment of the intensity, which is related to $\mathscr{G}_4^{(\mathrm{sc})}$ by Eq. (14), we obtain Eq. (15) from (A2).

APPENDIX B

We define the flux density at the point $\mathbf{r}$ as[37]

$$J_k(\mathbf{r})=\frac{i}{2k_0}\left[\Psi(\mathbf{r})\frac{\partial}{\partial r_k}\Psi^*(\mathbf{r})-\Psi^*(\mathbf{r})\frac{\partial}{\partial r_k}\Psi(\mathbf{r})\right]$$
$$=\frac{i}{2k_0}\left[\left(\frac{\partial}{\partial r_k}-\frac{\partial}{\partial r_k'}\right)\Psi(\mathbf{r}')\Psi^*(\mathbf{r})\right]\Bigg|_{\mathbf{r}=\mathbf{r}'}. \qquad \text{(B1)}$$

Then for the correlation function between the values of the radiation flux density in the observation plane $z=z_f$ we obtain the formula

$$C(z_f,\boldsymbol{\rho}=(\mathbf{r}-\mathbf{r}_1)_\parallel)=\langle J_z(z_f,\mathbf{r}_\parallel)J_z(z_f,\mathbf{r}_{1\parallel})\rangle-\langle J_z(z_f)\rangle^2$$
$$=-\frac{1}{4k_0^2}\left\{\left(\frac{\partial}{\partial z}-\frac{\partial}{\partial z'}\right)\left(\frac{\partial}{\partial z_1}-\frac{\partial}{\partial z_1'}\right)\right.$$
$$\times[\mathscr{G}_4(z,\mathbf{r}_\parallel;z',\mathbf{r}_\parallel;z_1,\mathbf{r}_{1\parallel};z_1',\mathbf{r}_{1\parallel})$$
$$\left.-\mathscr{G}_2(z,\mathbf{r}_\parallel;z',\mathbf{r}_\parallel)\mathscr{G}_2(z_1,\mathbf{r}_{1\parallel};z_1',\mathbf{r}_{1\parallel})]\right\}\Bigg|_{z=z'=z_1=z_1'=z_f}. \qquad \text{(B2)}$$

If we substitute the expressions for the components of $\mathscr{G}_4$ (see (A1)) into (B2), we obtain the correlation function $C(z_f,\boldsymbol{\rho})$ in the form of a sum of three contributions. The substitution of

$$\mathscr{G}_4^{(\mathrm{sc})}(\mathbf{r},\mathbf{r}';\mathbf{r}_1,\mathbf{r}_1')-[\mathscr{G}_2(\mathbf{r},\mathbf{r}')\mathscr{G}_2(\mathbf{r}_1,\mathbf{r}_1')$$
$$-\mathscr{G}_2^{(0)}(\mathbf{r},\mathbf{r}')\mathscr{G}_2^{(0)}(\mathbf{r}_1,\mathbf{r}_1')]$$

into (B2) with consideration of (9) and (14) leads to (20), which defines the long-range ($\rho\gg l\theta$, where θ is the characteristic multiple-scattering angle) intensity correlations. The substitution of $\mathscr{G}_4^{(\mathrm{sc})}(\mathbf{r},\mathbf{r}_1';\mathbf{r}_1,\mathbf{r}')$ into (B2) gives the contribution (28), which describes short-range correlations that de-

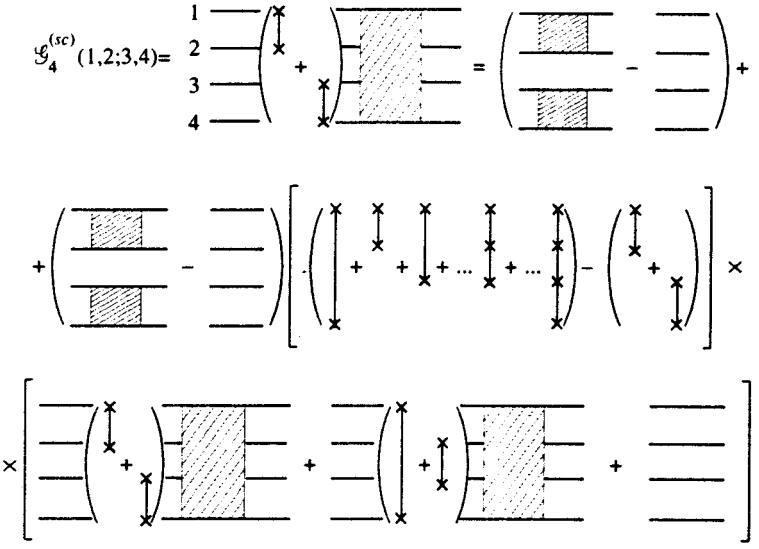

FIG. 5.

crease on the scales $\rho \sim \lambda/\theta$. Finally, $\mathscr{G}_4^{(\mathrm{non})}(\mathbf{r},\mathbf{r}',\mathbf{r}_1,\mathbf{r}_1') - \mathscr{G}_2^{(0)}(\mathbf{r},\mathbf{r}')\mathscr{G}_2^{(0)}(\mathbf{r}_1,\mathbf{r}_1')$ defines the part of the correlation function $C(z_f,\boldsymbol{\rho})$ which must be taken into account when $\rho < l\theta$, although it does not play an independent role like the first two contributions.

The Fourier transform of the correlation function, i.e., the fluctuation spectrum (19), is expressed in terms of the second moment of the intensity in the following manner.

At low spatial frequencies ($q \ll k_0\theta$) contributions to the fluctuation spectrum are made by all the components of the correlation function. Substituting the relation

$$\mathscr{G}_4(1,2,3,4) - \mathscr{G}_2(1,2)\mathscr{G}_2(3,4) = \mathscr{G}_2(1,2)\mathscr{G}_2(3,4) \times \sum_a \hat{h}_a(1,2)\hat{h}_a(3,4)[\mathscr{G}_4^{(\mathrm{sc})}(1,2;3,4) + \mathscr{G}_4^{(\mathrm{sc})}(1,4;3,2) + \mathscr{G}_1(1)\mathscr{G}_1^*(2)\mathscr{G}_1(3)\mathscr{G}_1^*(4)]$$

into (B2) and using Eqs. (9) and (14), we obtain

$$\begin{aligned} M(z_f,\mathbf{q}) &= \int_{(\boldsymbol{\Omega}\boldsymbol{\Omega}_n)>0} d\boldsymbol{\Omega} \int_{(\boldsymbol{\Omega}_1\boldsymbol{\Omega}_n)>0} d\boldsymbol{\Omega}_1 \int d^2\rho \\ &\times \exp(-i\mathbf{q}\cdot\boldsymbol{\rho})(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_n)(\boldsymbol{\Omega}_1\cdot\boldsymbol{\Omega}_n) \\ &\times \Bigg\{ n \int d\mathbf{r}' \int\int d\boldsymbol{\Omega}' d\boldsymbol{\Omega}_1' \\ &\times \int\int d\boldsymbol{\Omega}'' d\boldsymbol{\Omega}_1'' I_{\omega_0+\Delta\omega/2}\left(z_f, \frac{1}{2}(\mathbf{r}+\mathbf{r}_1)_\parallel + \frac{1}{2}\boldsymbol{\rho}, \boldsymbol{\Omega} | \mathbf{r}', \boldsymbol{\Omega}'\right) I_{\omega_0-\Delta\omega/2}\left(z_f, \frac{1}{2}(\mathbf{r}+\mathbf{r}_1)_\parallel - \frac{1}{2}\boldsymbol{\rho}, \boldsymbol{\Omega}_1 | \mathbf{r}', \boldsymbol{\Omega}_1'\right) [h(\boldsymbol{\Omega}',\boldsymbol{\Omega}'|\boldsymbol{\Omega}'',\boldsymbol{\Omega}'') \\ &\times h(\boldsymbol{\Omega}_1',\boldsymbol{\Omega}_1'|\boldsymbol{\Omega}_1'',\boldsymbol{\Omega}_1'') \\ &\times \langle I_{\omega_0+\Delta\omega/2}(\mathbf{r}',\boldsymbol{\Omega}'') I_{\omega_0-\Delta\omega/2}(\mathbf{r}',\boldsymbol{\Omega}_1'')\rangle \\ &+ h(\boldsymbol{\Omega}',\boldsymbol{\Omega}'|\boldsymbol{\Omega}'',\boldsymbol{\Omega}_1'') h(\boldsymbol{\Omega}_1',\boldsymbol{\Omega}_1'|\boldsymbol{\Omega}_1'',\boldsymbol{\Omega}'') \\ &\times \langle I_{\omega_0}(\mathbf{r}',\boldsymbol{\Omega}'',\Delta\omega) I_{\omega_0}(\mathbf{r}',\boldsymbol{\Omega}_1'',-\Delta\omega)\rangle \\ &+ h(\boldsymbol{\Omega}',\boldsymbol{\Omega}'|\boldsymbol{\Omega}'',\boldsymbol{\Omega}'') h(\boldsymbol{\Omega}_1',\boldsymbol{\Omega}_1'|\boldsymbol{\Omega}_1'',\boldsymbol{\Omega}_1'') \\ &\times I^{(0)}_{\omega_0+\Delta\omega/2}(\mathbf{r}',\boldsymbol{\Omega}''|\boldsymbol{\Omega}_0) I^{(0)}_{\omega_0-\Delta\omega/2}(\mathbf{r}',\boldsymbol{\Omega}_1''|\boldsymbol{\Omega}_0)]\Bigg\}. \end{aligned} \tag{B3}$$

At high spatial frequencies ($q > k_0\theta$) contributions to the fluctuation spectrum are made only by the short-range correlations, and $M(z_f,\mathbf{q})$ is specified as the Fourier transform of (28).

APPENDIX C

In the case of weak absorption ($l_{\mathrm{tr}} \ll l_a$) the regime of spatial diffusion of waves in a medium with scatterers of arbitrary size operates if the thickness of the slab $L \gg l_{\mathrm{tr}}$. Substituting the intensity propagators in the diffusion approximation[37] (see Eqs. (31) and (41), where l must be replaced by l_{tr}) into the general expressions (24) and (25), we obtain the following expansion for the fluctuation spectrum:

$$\begin{aligned} M_R^{\mathrm{coh}}(z_f,\mathbf{q}) &= \frac{l_{\mathrm{tr}}}{24\pi k_0^2}\int_0^L dz \left[\left| \frac{\partial}{\partial z} F(z_f,q|z) \right|^2 + q^2 |F(z_f,q|z)|^2 \right] \\ &\times |F(z|0,\boldsymbol{\Omega}_0,\Delta\omega)|^2 + \frac{1}{16\pi^2} n\varepsilon_a \\ &\times \int_0^L dz |F(z_f,q|z)|^2 |F(z|0,\boldsymbol{\Omega}_0,\Delta\omega)|^2 - \frac{l_{\mathrm{tr}}^2}{96\pi^2} n\varepsilon_{a,s} \\ &\times \int_0^L dz \frac{\partial}{\partial z}|F(z_f,q|z)|^2 \frac{\partial}{\partial z}|F(z|0,\boldsymbol{\Omega}_0,\Delta\omega)|^2 \\ &+ \frac{l_{\mathrm{tr}}^4}{144\pi^2} n \int_0^L dz \left[\varepsilon_a^{(0)} \left| \frac{\partial}{\partial z} F(z_f,q|z) \right|^2 \right. \\ &\left. + q^2 \varepsilon_s^{(1)} |F(z_f,q|z)|^2 \right] \left| \frac{\partial}{\partial z} F(z|0,\boldsymbol{\Omega}_0,\Delta\omega), \right|^2 \end{aligned} \tag{C1}$$

$$M^{\mathrm{incoh}}(z_f,\mathbf{q}) = \frac{n}{16\pi^2}\int_0^L dz |\sigma_a F(z_f,q|z) F(z|0,\boldsymbol{\Omega}_0) - \frac{1}{3}\sigma_{\mathrm{tr}} l_{\mathrm{tr}}^2 \frac{\partial}{\partial z} F(z_f,q|z) \frac{\partial}{\partial z} F(z|0,\boldsymbol{\Omega}_0)|^2. \tag{C2}$$

The functions in (C1) and (C2) can be calculated using (46), in which l must be replaced by l_{tr}. The coefficients in the expansion (C1) are defined by the expressions

$$\varepsilon_a = \sum_{p=0}^{\infty} (2p+1)(\sigma_a^p)^2, \tag{C3}$$

$$\varepsilon_{a,s} = \sum_{p=0}^{\infty} [(p+1)\sigma_a^{p+1} + p\sigma_{\mathrm{tr}}^p]\sigma_a^p, \tag{C4}$$

$$\varepsilon_s^{(0,1)} = \frac{3}{5}\sum_{p=0}^{\infty} [A_p^{(0,1)}(\Delta_p\Delta_{p-1}^* + \mathrm{c.c.}) + B_p^{(0,1)}(\Delta_p^2 + \mathrm{c.c.})], \tag{C5}$$

where σ_a^p and σ_{tr}^p are the partial absorption and transport cross sections,

$$\sigma_a^p = \frac{\pi}{k_0^2}(1-|S_p|^2), \quad \sigma_{\mathrm{tr}}^p = \frac{\pi}{k_0^2}|S_p - S_{p-1}|^2, \tag{C6}$$

$$A_p^{(0)} = \frac{4p(p-1)}{2p-1}, \quad B_p^{(0)} = \frac{p(4p^2+1)}{4p^2-1}, \tag{C7}$$

$$A_p^{(1)} = \frac{3p(p-1)}{2p-1}, \quad B_p^{(1)} = -\frac{2p(p^2-1)}{4p^2-1}, \tag{C8}$$

$$\Delta_p = \frac{2\pi i}{k_0}(f_p - f_{p-1}^*) + 4\pi f_p f_{p-1}^* = \frac{\pi}{k_0^2}(S_p S_{p-1}^* - 1), \tag{C9}$$

S_p and f_p are the coefficients in the expansions of the S-matrix and the partial wave scattering amplitude, and $S_p = 1 + 2ik_0 f_p$.[39]

In the case of scatterers of small radius, one term corresponding to the s wave remains in each of the sums (C3)–(C5). As a result, we have

$$\sigma_{\rm tr} = \sigma, \quad \varepsilon_a = \sigma_a^2, \quad \varepsilon_{a,s} = \sigma\sigma_a, \quad \varepsilon_s^{(0)} = \sigma^2 - \frac{2\pi\sigma}{k_0^2},$$
$$\varepsilon_s^{(1)} = 0, \tag{C10}$$

and formulas (C1) and (C2) transform into (42) and (43).

In the case of large scatterers, in which deflection by a small angle occurs in each individual collision, a large number of terms must be retained in the sums (C3)–(C5). If the terms appearing in (C3)–(C5) are assumed to be smooth functions of p and the summation in (C3)–(C5) is replaced by integration, the following relations can be obtained for the coefficients ε_a, $\varepsilon_{a,s}$, and ε_s:

$$\varepsilon_a = \frac{\pi}{k_0^4}\int d^2p(1-|S_p|^2)^2, \tag{C11}$$

$$\varepsilon_{a,s} = \frac{4\pi}{k_0^4}\int d^2p(1-|S_p|^2)\left(\frac{d\delta}{dp}\right)^2, \tag{C12}$$

$$\varepsilon_s^{(0)} = \frac{4\pi}{k_0^4}\int d^2p\left[\frac{3}{5}\left(\frac{d\delta}{dp}\right)^2 + \left(\frac{d\delta}{dp}\right)^4\right], \tag{C13}$$

$$\varepsilon_s^{(1)} = \frac{4\pi}{k_0^4}\int d^2p\left[\frac{6}{5}\left(\frac{d\delta}{dp}\right)^2 - \left(\frac{d\delta}{dp}\right)^4\right]. \tag{C14}$$

It was assumed during the derivation of these relations that $S_p = |S_p|\exp(2i\delta)$ and that the condition for weak absorption $1-|S_p|^2 \ll (d\delta/dp)^2$ is satisfied.

[1] A. Yu. Zyuzin and B. Z. Spivak, Zh. Éksp. Teor. Fiz. **93**, 994 (1987) [Sov. Phys. JETP **66**, 560 (1987)].
[2] J. Rammer, Rev. Mod. Phys. **63**, 781 (1994).
[3] V. L. Kuz'min and V. P. Romanov, Usp. Fiz. Nauk **166**, 247 (1996) [Phys. Usp. **39**, 231 (1996)].
[4] B. L. Al'tshuler, JETP Lett. **41**, 648 (1985).
[5] P. Lee and A. D. Stone, Phys. Rev. Lett. **55**, 1622 (1985).
[6] C. P. Umbach, S. Washburn, R. B. Laibowitz, and R. A. Webb, Phys. Rev. B **30**, 4048 (1984); R. A. Webb, S. Washburn, C. P. Umbach, and R. B. Laibowitz, in *Localization, Interaction and Transport Phenomena in Impure Metals*, G. Bergmann, Y. Bruynseraede, and B. Kramer (eds.), Springer-Verlag, New York (1985).
[7] S. Etemad, R. Thompson, and M. J. Andrejko, Phys. Rev. Lett. **57**, 575 (1986); M. Kaveh, M. Rosenbluh, I. Edrei, and I. Freund, Phys. Rev. Lett. **57**, 2049 (1986).
[8] R. Shapiro, Phys. Rev. Lett. **57**, 2168 (1986).
[9] A. Z. Genack, Phys. Rev. Lett. **58**, 2042 (1987).
[10] A. M. Prokhorov, F. V. Bunkin, K. S. Gochelashvili, and V. I. Shishov, Usp. Fiz. Nauk **114**, 415 (1974) [Sov. Phys. Usp. **17**, 826 (1975)].
[11] V. I. Tatarskii and V. U. Zavorotnyi, Prog. Opt. **18**, 204 (1980).
[12] R. L. Fante, Proc. IEEE **68**, 1424 (1980).
[13] J. L. Codona, D. B. Creamer, S. M. Flatte, R. G. Frehlich, and F. S. Henyey, Radio Sci. **21**, 929 (1986).
[14] M. J. Stephen and G. Cwilich, Phys. Rev. Lett. **59**, 285 (1987).
[15] R. Z. Spivak and A. Yu. Zyuzin, Solid State Commun. **65**, 311 (1988).
[16] P. A. Mello, E. Akkermans, and B. Shapiro, Phys. Rev. Lett. **61**, 459 (1988).
[17] S. Feng, C. Kane, P. A. Lee, and A. D. Stone, Phys. Rev. Lett. **61**, 834 (1988).
[18] R. Pnini and B. Shapiro, Phys. Rev. B **39**, 6986 (1989); Phys. Lett. A **157**, 265 (1991).
[19] L. Wang and S. Feng, Phys. Rev. B **40**, 8284 (1989).
[20] N. Garcia and A. Z. Genack, Phys. Rev. Lett. **63**, 1678 (1989).
[21] M. P. van Albada, J. F. de Boer, and A. Lagendijk, Phys. Rev. Lett. **64**, 2787 (1990).
[22] A. Z. Genack, N. Garcia, and W. Polkosnik, Phys. Rev. Lett. **65**, 2129 (1990).
[23] N. Garcia and A. Z. Genack, Phys. Rev. Lett. **66**, 1850 (1991).
[24] A. Z. Genack and N. Garcia, Phys. Rev. Lett. **66**, 2064 (1991).
[25] J. F. de Boer, M. P. van Albada, and A. Lagendijk, Phys. Rev. B **45**, 658 (1992).
[26] E. Kogan and M. Kaveh, Phys. Rev. B **45**, 1049 (1992).
[27] A. A. Lisyansky and D. Livdan, Phys. Lett. A **170**, 53 (1992); Phys. Rev. B **47**, 14 157 (1993).
[28] N. Garcia, A. Z. Genack, and A. A. Lisyansky, Phys. Rev. B **46**, 14 475 (1992).
[29] D. B. Rogozkin and M. Yu. Cherkasov, Laser Phys. **2**, 727 (1992); Phys. Lett. A **178**, 431 (1993).
[30] N. Garcia, A. Z. Genack, R. Pnini, and B. Shapiro, Phys. Lett. A **176**, 458 (1993).
[31] M. C. W. van Rossum and T. M. Nieuwenhuizen, Phys. Lett. A **177**, 452 (1993).
[32] D. B. Rogozkin and M. Yu. Cherkasov, JETP Lett. **58**, 585 (1993).
[33] D. B. Rogozkin and M. Yu. Cherkasov, Phys. Rev. B **51**, 12 256 (1995).
[34] R. Berkovits and S. Feng, Fiz. Rev. Lett. **65**, 3120 (1990).
[35] T. M. Nieuwenhuizen and M. C. W. van Rossum, Phys. Lett. A **177**, 102 (1993).
[36] K. M. Case and P. F. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, Mass. (1967).
[37] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Academic Press, New York (1978).
[38] L. A. Apresyan and Yu. A. Kravtsov, *Theory of Radiation Transfer: Wave and Statistical Aspects* [in Russian], Nauka, Moscow (1983).
[39] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[40] E. M. Lifshitz, L. P. Pitaevskii, and L. D. Landau, *Physical Kinetics*, Pergamon Press, Oxford (1981).
[41] C. L. Kane, R. A. Serota, and P. A. Lee, Phys. Rev. B **37**, 6701 (1988).
[42] V. V. Sobolev, *Light Scattering in Planetary Atmospheres*, Pergamon Press, Oxford–New York (1975).
[43] Yu. N. Barabanenkov and V. D. Ozrin, Zh. Éksp. Teor. Fiz. **94** (6), 56 (1988) [Sov. Phys. JETP **67**, 1117 (1988)].
[44] M. B. van der Mark, M. P. van Albada, and A. Lagendijk, Phys. Rev. B **37**, 3575 (1988).
[45] R. E. Gorodnichev, S. L. Dudarev, D. B. Rogozkin, Zh. Éksp. Teor. Fiz. **96**, 847 (1989) [Sov. Phys. JETP **69**, 481 (1989)]; Phys. Lett. A **144**, 48 (1990).
[46] P. E. Wolf, G. Maret, E. Akkermans, and R. Maynard, J. Phys. (Paris) **49**, 63 (1988).
[47] Yu. N. Barabanenkov, Yu. A. Kravtsov, V. D. Ozrin, and A. I. Saichev, Prog. Opt. **29**, 67 (1991).
[48] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun (eds.), Dover, New York (1976).
[49] D. B. Rogozkin and M. Yu. Cherkasov, Laser Phys. **2**, 913 (1992).
[50] L. S. Dolin, Dokl. Akad. Nauk SSSR **260**, 1344 (1981) [Sov. Phys. Dokl. **26**, 976 (1981)].
[51] V. S. Remizyunich, D. B. Rogozkin, M. M. Ryazanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **24**, 891 (1982).
[52] J. Tessendorf, Phys. Rev. A **35**, 872 (1987).
[53] E. E. Gorodnichev and D. B. Rogozkin, Zh. Éksp. Teor. Fiz. **107**, 209 (1995) [JETP **80**, 112 (1995)].
[54] G. Maret and P. E. Wolf, Z. Phys. B **65**, 409 (1987); Phys. Scr. T **29**, 223 (1989).
[55] F. C. MacKintosh and S. John, Phys. Rev. B **40**, 2383 (1989).

Translated by P. Shelnitz

Generalized Jordan–Wigner transformations and the Ising–Onsager problem

M. S. Kochmański

Institute of Physics, Pedagogical University, 35-310, Rzeszów, Poland
(Submitted 6 March 1996; resubmitted 26 July 1996)
Zh. Éksp. Teor. Fiz. **111**, 1717–1731 (May 1997)

Another possible method for obtaining Onsager's solution for the $2D$ Ising model is presented. In contrast to previous methods, the method proposed here makes it possible to study the problem in a weak external magnetic field. Generalized Jordan–Wigner transformations in the form introduced by Kochmański [J. Tech. Phys. **36**, 485 (1995)] are employed. © *1997 American Institute of Physics.* [S1063-7761(97)01205-5]

1. INTRODUCTION

Methods for solving the two-dimensional Ising model in the absence of an external magnetic field are well known (see, for example, Refs. 2–6 and the references cited there). Unfortunately, essentially none of these approaches has led to success in solving the Ising–Onsager problem in an external magnetic field or to a solution of the three-dimensional Ising model. More precisely, these methods did not make possible further advances in the solution of the Ising–Onsager problem, despite the great efforts of physicists and mathematicians.

The work by Schultz, Mattis, and Lieb[7] should be especially noted. In my opinion, this work is one of the most beautiful and elegant in the theoretical physics of the Ising–Onsager problem (another example of the deep relationship between theoretical physics and art). The method employed in Ref. 7 is based on the application of a transfer matrix followed by a transition to the Fermi representation by means of one-dimensional Jordan–Wigner transformations.[8] On the other hand, combinatorial methods have been used[9–11] in one form or another to calculate the partition function for the $2D$ Ising model using a diagrammatic representation for the initial sum. I realize and ''remember'' perfectly well that a large number of approaches and methods for solving the three-dimensional problem have now been accumulated but nonetheless have led nowhere. However, as is well known, most of these approaches and methods (see, for example, Ref. 12) are limited in scope, which makes it impossible even to study the problem in a weak external magnetic field in two dimensions, to say nothing of the three-dimensional problem.

The present paper develops another possible approach to the Ising–Onsager problem and gives as an example the solution for the $2D$ Ising model, using ideas from Ref. 7 as well as Refs. 9–11. The idea is to formulate the problem in three dimensions in the second-quantization representation and then let one of the interaction constants $J_{1,2,3}$ go to zero. Generalized Jordan–Wigner transformations are employed in the form introduced in Ref. 1. It is hoped that this approach will make it possible to advance substantially in the solution of the Ising–Onsager problem in an external magnetic field. More will be said about this in the concluding section.

2. PARTITION FUNCTION

2.1. Analytic representation

Consider a simple cubic lattice with N rows, M columns, and K planes (layers) at whose sites the ''spins'' σ_{nmk} are given and assume two values $\sigma_{nmk}=\pm 1$. The Hamiltonian for the $3D$ Ising model with nearest-neighbor interaction is

$$\mathscr{H}=-\sum_{n,m,k=1}^{NMK}(J_1\sigma_{nmk}\sigma_{n+1,mk}+J_2\sigma_{nmk}\sigma_{n,m+1,k}+J_3\sigma_{nmk}\sigma_{nm,k+1}), \tag{2.1}$$

where the collective index nmk enumerates the sites of a simple $N\times M\times K$ cubic lattice and the constants $J_j>0$ account for the anisotropy of the interaction of the Ising spins. As usual, periodic boundary conditions are imposed on the variables σ_{nmk}. We write the partition function Z_3 of the system in the form

$$Z_3=\sum_{\sigma_{111}=\pm 1}\cdots\sum_{\sigma_{NMK}=\pm 1}e^{-\beta\mathscr{H}}=\sum_{\{\sigma_{nmk}=\pm 1\}}\exp\left[\sum_{nmk}(K_1\sigma_{nmk}\sigma_{n+1,mk}+K_2\sigma_{nmk}\sigma_{n,m+1,k}+K_3\sigma_{nmk}\sigma_{nm,k+1})\right], \tag{2.2}$$

where the quantities K_i are defined as (T is the temperature)

$$K_{1,2,3}=\beta J_{1,2,3},\quad \beta=1/k_BT. \tag{2.3}$$

(Here and below, summation over nmk (or nm) and a product over nm will involve a summation or a product over a complete set of integers from 1 to N, M and K for each index.)

Using the well-known transfer-matrix method,[4,5] the partition function Z_3 in Eq. (2.2) can easily be written in the form

$$Z_3=\mathrm{Tr}\,(T)^K, \tag{2.4}$$

where T is a transfer matrix whose elements are

$$T^{\{\sigma_{nmk}\}}_{\{\sigma_{nm,k+1}\}}=\exp\left[\sum_{nm}(K_1\sigma_{n+1,mk}+K_2\sigma_{n,m+1,k})\sigma_{nmk}\right]\times\exp\left[K_3\sum_{nm}\sigma_{nmk}\sigma_{nm,k+1}\right]. \tag{2.5}$$

 1063-7761/97/050940-08$10.00

The transfer matrix elements of the layer–layer Ising model can also be written in a somewhat different form,[4] but these representations are actually all equivalent. According to (2.5), the matrix T can be represented as a product of matrices $T_{1,2,3}$ of the same dimension $(2^{NM}\times 2^{NM})$,

$$T=T_3T_2T_1, \tag{2.6}$$

where[4]

$$T_1=\exp\left(K_1\sum_{nm}\ \tau^z_{nm}\tau^z_{n+1,m}\right),$$

$$T_2=\exp\left(K_2\sum_{nm}\ \tau^z_{nm}\tau^z_{n,m+1}\right), \tag{2.7}$$

$$T_3=(2\sinh 2K_3)^{NM/2}\exp\left(K_3^*\sum_{nm}\ \tau^x_{nm}\right), \tag{2.8}$$

where $\tau^{x,y,z}_{nm}$ are three sets of 2^{NM}-dimensional matrices of the form

$$\tau^{x,y,z}_{nm}=1\otimes 1\otimes\ldots\otimes\tau^{x,y,z}\otimes\ldots\otimes 1\otimes 1, \tag{2.9}$$

in which the Pauli matrices $\tau^{x,y,z}$ appear in the nmth place. In Eq. (2.8) the quantities K_3 and K_3^* are related to one another by

$$\tanh(K_3)=\exp(-2K_3^*)\ \ \text{or}\ \ \sinh(2K_3)\sinh(2K_3^*)=1. \tag{2.10}$$

The Pauli spin matrices $\tau^{x,y,z}_{nm}$ (2.9) commute with one another if $nm\neq n'm'$, while for given nm they satisfy the standard properties.[4] It is easy to see that the matrices $T_{1,2}$ (2.7) commute with one another, but they do not commute with T_3 (2.8). The transition to the $2D$ Ising model with respect to the interaction constants K_1 or K_2 corresponds to setting $K_1=0$ or $K_2=0$ and removing the summation over n $(N=1)$ or m $(M=1)$, respectively. Then the standard expressions[5,7] for the $2D$ Ising model are obtained, the operator T_1 (2.7) being identically 1 $(T_1\equiv 1)$ in the first case, and $T_2\equiv 1$ in the second.

A somewhat different situation obtains a transition to the $2D$ Ising model with respect to the interaction constant K_3. In this case we must set $K_3=0$, $K=1$, i.e., the summation over k must be removed, as a result of which it is easy to obtain for the operator T_3 (2.8)

$$T_3^*\equiv T_3(K_3=0)=\prod_{nm}\ (1+\tau^x_{nm}), \tag{2.11}$$

where the relations (2.10) were used. This structure of the operator T_3^* makes it possible ultimately to see a somewhat different way to solve the Ising problem, and to advance substantially in the solution of the Ising–Onsager problem in an external magnetic field H. Thus, according to Eq. (2.4), the partition function of the $2D$ Ising model can be expressed as

$$Z_2=\mathrm{Tr}(T_3^*T_2T_1), \tag{2.12}$$

where the matrices $T_{1,2}$ are given by (2.7) and T_3^* is given by (2.11). The advantage of representing the partition function in the form (2.12), it seems to me, is in some sense obvious; I shall have more to say about this below.

Schultz, Mattis, and Lieb[7] showed that the T matrix in its standard representation can be expressed in terms of Fermi second-quantization operators. They employed the well-known Jordan–Wigner transformations[8] which make it possible to express the Fermi operators c_m^+ and c_m of a one-dimensional system in terms of the Pauli operators $\tau_m^\pm$:[5]

$$c_m=\exp\left(i\pi\sum_{j=1}^{m-1}\ \tau_j^+\tau_j^-\right)\tau_m^-,$$

$$c_m^+=\exp\left(i\pi\sum_{j=1}^{m-1}\ \tau_j^+\tau_j^-\right)\tau_m^+. \tag{2.13}$$

In Ref. 1 Jordan–Wigner type transformations (2.13), extended to two-, three-, and d-dimensional systems, were introduced in a form convenient for application to lattice systems. Introducing, just as in the one-dimensional case,[5] the two-index Pauli operators

$$\tau^\pm_{nm}=\frac{1}{2}\,(\tau^z_{nm}\pm i\tau^y_{nm}), \tag{2.14}$$

which satisfy anticommutation relations at any one site and commutation relations for different sites, we write the matrices $T_{1,2}$ and T_3^* in the form

$$T_1=\exp\left[K_1\sum_{nm}\ (\tau^+_{nm}+\tau^-_{nm})(\tau^+_{n+1,m}+\tau^-_{n+1,m})\right], \tag{2.15}$$

$$T_2=\exp\left[K_2\sum_{nm}\ (\tau^+_{nm}+\tau^-_{nm})(\tau^+_{n,m+1}+\tau^-_{n,m+1})\right], \tag{2.16}$$

$$T_3^*=\prod_{nm}\ [1+(1-2\tau^+_{nm}\tau^-_{nm})]. \tag{2.17}$$

As noted above, the Pauli operators $\tau^\pm_{nm}$ behave as Fermi operators at one site and Bose operators at different sites. Next, using the two-dimensional Jordan–Wigner transformations[1]

$$\alpha^+_{nm}=\exp\left(i\pi\sum_{k=1}^{n-1}\sum_{l=1}^{M}\ \tau^+_{kl}\tau^-_{kl}+i\pi\sum_{l=1}^{m-1}\ \tau^+_{nl}\tau^-_{nl}\right)\tau^+_{nm},$$

$$\alpha_{nm}=\exp\left(i\pi\sum_{k=1}^{n-1}\sum_{l=1}^{M}\ \tau^+_{kl}\tau^-_{kl}+i\pi\sum_{l=1}^{m-1}\ \tau^+_{nl}\tau^-_{nl}\right)\tau^-_{nm}, \tag{2.18}$$

$$\beta^+_{nm}=\exp\left(i\pi\sum_{k=1}^{N}\sum_{l=1}^{m-1}\ \tau^+_{kl}\tau^-_{kl}+i\pi\sum_{k=1}^{n-1}\ \tau^+_{km}\tau^-_{km}\right)\tau^+_{nm},$$

$$\beta_{nm}=\exp\left(i\pi\sum_{k=1}^{N}\sum_{l=1}^{m-1}\ \tau^+_{kl}\tau^-_{kl}+i\pi\sum_{k=1}^{n-1}\ \tau^+_{km}\tau^-_{km}\right)\tau^-_{nm}, \tag{2.19}$$

after a series of transformations[13] the operators (2.14)–(2.16) assume the form

$$T_1=\exp\left\{K_1\sum_{m=1}^{M}\left[\sum_{n=1}^{N-1}\ (\beta^+_{nm}-\beta_{nm})(\beta^+_{n+1,m}+\beta_{n+1,m})\right.\right.$$

$$\left.\left.-\hat g_m(\beta^+_{Nm}-\beta_{Nm})(\beta^+_{1,m}+\beta_{1,m})\right]\right\}, \tag{2.20}$$

$$T_2=\exp\left\{K_2\sum_{n=1}^{N}\left[\sum_{m=1}^{M-1}(\alpha_{nm}^{+}-\alpha_{nm})(\alpha_{n,m+1}^{+}+\alpha_{n,m+1})\right.\right.$$
$$\left.\left.-\hat{g}_n(\alpha_{nM}^{+}-\alpha_{nM})(\alpha_{n,1}^{+}+\alpha_{n,1})\right]\right\}, \tag{2.21}$$

$$T_3^{*}=\prod_{nm}\left[1+(-1)^{\alpha_{nm}^{+}\alpha_{nm}}\right]=\prod_{nm}\left[1+(-1)^{\beta_{nm}^{+}\beta_{nm}}\right], \tag{2.22}$$

where

$$\hat{g}_n\equiv(-1)^{\hat{M}_n},\quad \hat{M}_n=\sum_{m=1}^{M}\alpha_{nm}^{+}\alpha_{nm},$$

$$\hat{g}_m\equiv(-1)^{\hat{N}_m},\quad \hat{N}_m=\sum_{n=1}^{N}\beta_{nm}^{+}\beta_{nm}.$$

The commutation relations between the α and β operators have the form[1]

$$\{\alpha_{nm}^{+},\beta_{nm}\}_{+}=\{\beta_{nm}^{+},\alpha_{nm}\}_{+}=(-1)^{\varphi_{nm}},$$

$$[\alpha_{nm},\beta_{n'm'}]_{-}=\ldots=[\alpha_{nm}^{+},\beta_{n'm'}^{+}]_{-}=0,$$
$$\begin{pmatrix} n'\leqslant n-1, & m'\geqslant m+1\\ n'\geqslant n+1, & m'\leqslant m-1\end{pmatrix},$$
$$\{\alpha_{nm},\beta_{n'm'}\}_{+}=\ldots=\{\alpha_{nm}^{+},\beta_{n'm'}^{+}\}_{+}=0, \tag{2.23}$$

and in all other cases, when the operators are defined as

$$\alpha_{nm}^{+}=\exp(i\pi\varphi_{nm})\beta_{nm}^{+},\quad \alpha_{nm}=\exp(i\pi\varphi_{nm})\beta_{nm},$$
$$\varphi_{nm}=\left[\sum_{k=n+1}^{N}\sum_{p=1}^{m-1}+\sum_{k=1}^{n-1}\sum_{p=m+1}^{M}\right]\alpha_{kp}^{+}\alpha_{kp}$$
$$=[\cdots]\beta_{kp}^{+}\beta_{kp}, \tag{2.24}$$

and $\alpha_{nm}^{+}\alpha_{nm}=\beta_{nm}^{+}\beta_{nm}$, the operators φ_{nm} obviously commute with the operators α_{nm}^{+}, α_{nm} and β_{nm}^{+}, β_{nm}, i.e.,

$$[\varphi_{nm},\alpha_{nm}^{+}]_{-}=\ldots=\ldots=[\varphi_{nm},\beta_{nm}]_{-}=0.$$

Introducing, just as in the one-dimensional case,[7] a basis in the occupation number representation for α and β fermions (2^{NM}-dimensional space in the Fock representation[14]) and then calculating the corresponding matrix elements $\langle l|T|l\rangle$ it is easy to see that on account of the multiplicative character of the operator T_3^{*} (2.22) all matrix elements with the exception of the vacuum matrix element $\langle 0|T|0\rangle$ equal zero. For the vacuum matrix element the contribution of the operator T_3^{*} simply equals 2^{NM}, and Z_2 (2.12) can be written in the second-quantization representation as

$$Z_2=2^{NM}\langle 0|(T_2T_1)|0\rangle, \tag{2.25}$$

where the vacuum state $|0\rangle$ is defined in the standard manner:

$$\alpha_{nm}|0\rangle=\beta_{nm}|0\rangle=0,\quad n=1,2,\ldots,N,\quad m=1,2,\ldots M, \tag{2.26}$$

and the operators $T_{1,2}$ are determined by the formulas (2.20)–(2.21). Here I call attention especially to the fact that the vacuum states (2.26) for the α and β fermions can differ from one another by at most a constant phase factor, which in the case at hand can always be chosen to be 1. Finally, it is easy to show that in the case of ''free end'' boundary conditions for the α and β operators ($\alpha_{n,M+1}^{+}=0$ and so on), the expression (2.25) for the partition function Z_2 can finally be represented as

$$Z_2=2^{NM}[(1-z_1^2)(1-z_2^2)]^{-NM/2}\langle 0|T_2^{*}T_1^{*}|0\rangle,$$
$$z_1=\tanh(K_1),\quad z_2=\tanh(K_2), \tag{2.27}$$

where the operators $T_{1,2}^{*}$ are given by

$$T_1^{*}=\exp\left\{\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{l=1}^{N-n}z_1^{l}\beta_{nm}^{+}\beta_{n+l,m}^{+}\right\}, \tag{2.28}$$

$$T_2^{*}=\exp\left\{\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{M-m}z_2^{k}\alpha_{n,m+k}\alpha_{nm}\right\}. \tag{2.29}$$

In deriving the expressions (2.27), all creation operators α_{nm}^{+} were ''passed through'' the bra vector $\langle 0|$ and all annihilation operators β_{nm} were passed through the ket vector $|0\rangle$, using the equalities $\langle 0|\alpha_{nm}^{+}=0$ and $\beta_{nm}|0\rangle=0$ for all n,m.

2.2. Diagrammatic representation

Let S denote the vacuum matrix element in the formula (2.27) for Z_2 and let G denote the product of the operators $T_2^{*}T_1^{*}$, i.e.

$$S\equiv\langle 0|T_2^{*}T_1^{*}|0\rangle\equiv\langle 0|G|0\rangle. \tag{2.30}$$

Therefore we must calculate the vacuum matrix element S (2.30) of a sum of products of Fermi creation and annihilation operators. The operator G in Eq. (2.30) is a polynomial in z_1, z_2, α_{nm}, and β_{nm}^{+}. Since G appears in Eq. (30) in the brackets $\langle 0|G|0\rangle$, not all terms of the polynomial give a nonzero contribution to the matrix element S. Writing out the product G and substituting into Eq. (2.30), S can be represented as a sum of vacuum matrix elements $\Sigma_\nu S_\nu$, where S_ν is the vacuum matrix element of the νth term of the polynomial G. As follows from Eqs. (2.28) and (2.29), each term of the polynomial G is a product of different pairs $z_2^{k}\alpha_{n,m+k}\alpha_{nm}$ and $z_1^{l}\beta_{nm}^{+}\beta_{n+l,m}^{+}$, which will be conventionally termed below as α pairs and β pairs. It is obvious that all terms of the polynomial G with unequal numbers of α and β pairs make a zero contribution to S and not all terms of the polynomial G with equal numbers of α and β pairs will give a nonzero contribution to S. Indeed, only terms of the polynomial G with equal numbers of α and β pairs in which each annihilation operator α_{nm} is paired with a corresponding creation operator $\beta_{n'm'}^{+}$ with identical indices ($n=n'$, $m=m'$) will give a nonzero contribution to S. In the opposite case, such a term will give a zero contribution to S.

In this manner, we arrive at a diagrammatic representation, noting that to each vacuum matrix element S_ν there can be associated a unique collection of lines (bonds) connecting pairs of lattice sites. For example, the diagrams in Figs. 1a–1d correspond to the matrix elements

a) $z_1^2z_2^3\langle 0|\alpha_{n,m+3}\alpha_{nm}\beta_{n,m+3}^{+}\beta_{n+2,m+3}^{+}|0\rangle$,

b) $z_1^4z_2^4\langle 0|\alpha_{n,m+1}\alpha_{nm}\alpha_{n+1,m+1}\alpha_{n+1,m-1}$

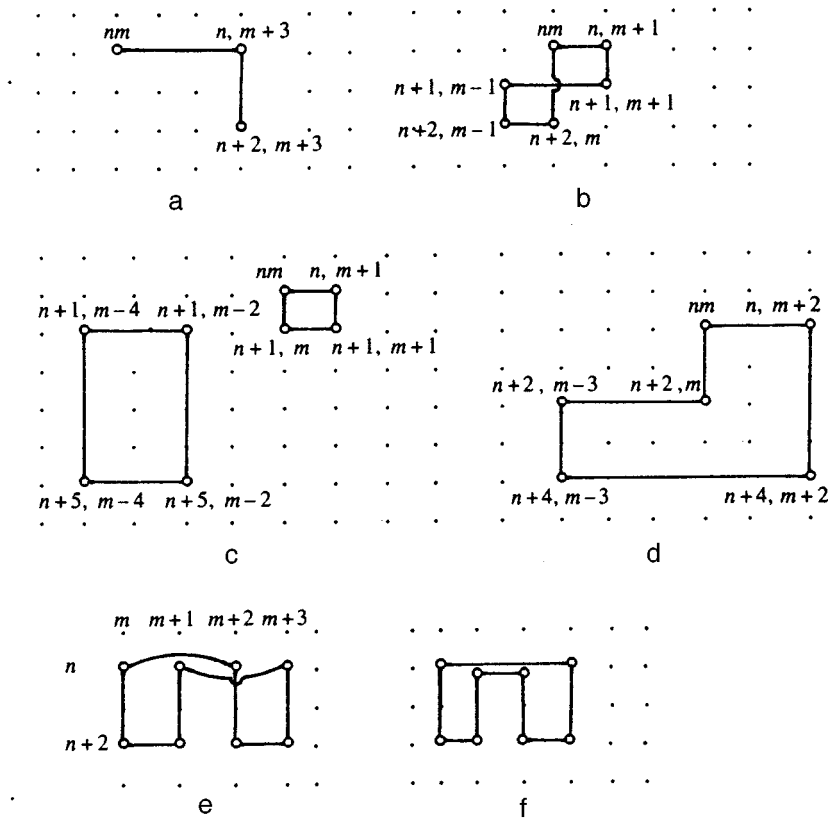

FIG. 1. Hamiltonian circuits on a simple $N\times M$ square lattice with variable step length: a—Example of a Hamiltonian chain; b—example of a simple loop with intersection of horizontal and vertical bonds (edges); c, d—examples of simple loops with no self-intersections of the bonds; e, f—simple loops with intersection of two superposed horizontal or two superposed vertical bonds.

$$\times\alpha_{n+2,m}\alpha_{n+2,m-1}\beta^{+}_{n+1,m-1}\beta^{+}_{n+2,m-1}$$
$$\times\beta^{+}_{nm}\beta^{+}_{n+2,m}\beta^{+}_{n,m+1}\beta^{+}_{n+1,m+1}|0\rangle,$$

c) $z_1^{10}z_2^{6}\langle 0|\alpha_{n,m+1}\alpha_{nm}\alpha_{n+1,m+1}\alpha_{n+1,m}$
$$\times\alpha_{n+1,m-2}\alpha_{n+1,m-4}\alpha_{n+5,m-2}\alpha_{n+5,m-4}$$
$$\times\beta^{+}_{n+1,m-4}\beta^{+}_{n+5,m-4}\beta^{+}_{n+1,m-2}$$
$$\times\beta^{+}_{n+5,m-2}\beta^{+}_{nm}\beta^{+}_{n+1,m}\beta^{+}_{n,m+1}\beta^{+}_{n+1,m+1}|0\rangle,$$

d) $z_1^{8}z_2^{10}\langle 0|\alpha_{n,m+2}\alpha_{nm}\alpha_{n+2,m}\alpha_{n+2,m-3}$
$$\times\alpha_{n+4,m+2}\alpha_{n+4,m-3}\beta^{+}_{n+2,m-3}\beta^{+}_{n+4,m-3}$$
$$\times\beta^{+}_{nm}\beta^{+}_{n+2,m}\beta^{+}_{n,m+2}\beta^{+}_{n+4,m+2}|0\rangle. \qquad (2.31)$$

As the expressions (2.28), (2.29), and (2.31) show, a factor z_2^k is associated with each horizontal line of ''length'' k and a factor z_1^l is associated with each vertical line of ''length'' l. It was shown above that only matrix elements S_ν with equal numbers of α and β pairs give a nonzero contribution to S, each annihilation operator α_{nm} being paired with a corresponding creation operator β^{+}_{nm}. Geometrically, this means that from the entire collection of possible diagrams the only diagrams giving a nonzero contribution to S are the ones in which zero or two lines (bonds) meet at a ''right angle'' at each vertex of a diagram. In other words, diagrams in which two horizontal lines (bonds) or two vertical bonds meet at any vertex are not permitted. The simplest examples of such diagrams are shown in Figs. 1a–e. Therefore all diagrams giving a nonzero contribution to S must be closed, self-intersection of lines (bonds) being prohibited at any site since $\alpha^2_{nm}=(\beta^{+}_{nm})^2=0$. From the standpoint of graph theory, the closed diagrams described above correspond to unoriented Hamiltonian circuits (the vertices having valences $\delta=0,2$) on a simple square lattice.[15–17]

Therefore the vacuum matrix element S (2.30) can be represented in the form

$$S=\sum_{\nu} S_\nu, \qquad (2.32)$$

and in the calculation any multiply-connected diagram is considered to be one diagram (for example, the diagram in Fig. 1c). Every closed diagram gives the contribution

$$(\pm 1)\prod_{j=1}^{s} z_1^{l_j}z_2^{k_j}, \qquad (2.33)$$

where s is the number of horizontal bonds, which is equal to the number of vertical bonds. Further, using the relation between the α and β operators (2.23) and Wick's theorem,[18] it can be shown that an arbitrary vacuum matrix element giving a nonzero contribution to the sum S (2.31) decomposes into a product of matrix elements corresponding to the connected parts of a diagram (which, for brevity, will be termed below simple loops without self-intersections at the lattice sites). It can be verified directly, taking account of the commutation relations (2.23) between the α and β operators, that, for example, the diagrams in Figs. 1b–f enter with a plus sign. Other diagrams, for example, the diagram in Fig. 1e, can enter with a minus sign as well. The commutation relations

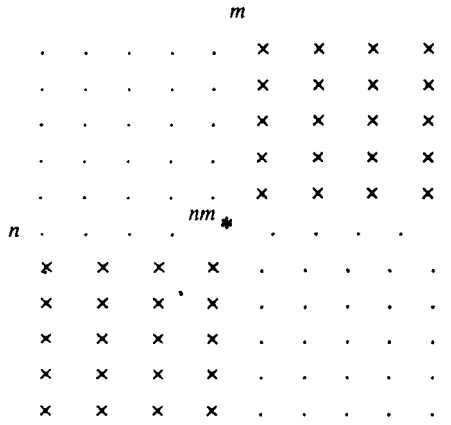

FIG. 2. ''Geometry'' of the commutation relations for the α and β operators: *—α operator, $\times$ —β operator.

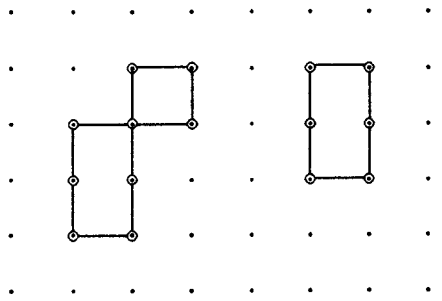

FIG. 3. Simple example of a diagram (Euler circuits) contributing to the partition function $Z(z_1,z_2)$.

between the α and β operators (2.23) are illustrated very clearly in Fig. 2, where the distinguished operator $\alpha_{nm}(*)$ for a fixed site (nm) commutes with the β operators at the $(n'm')$ sites, marked with a cross $(\times)$. For all other sites the α and β operators anticommute with one another. Therefore the contribution of each diagram decomposes into a product of contributions from simple loops, the contribution of a simple loop with s horizontal and s vertical bonds being

$$\Pi_s=(\pm 1)\prod_{j=1}^{s} z_1^{l_j}z_2^{k_j}. \tag{2.34}$$

Now S (2.32) can be expressed as

$$S=1+\sum_{\{s\}}\Pi_s+\sum_{\{s\},\{q\}}\Pi_s\Pi_q+\ldots\equiv\Gamma^{(h)}(z_1,z_2,), \tag{2.35}$$

where the loop Π_s is determined by the expression (2.34). The expression (2.35) contains, besides a summation over the number s of bonds, a summation over all admissable lengths of these bonds $\{k\}$ and $\{l\}$ with fixed s. It is easy to see that the summation in Eq. (2.35) over the lengths of the horizontal $\{k\}$ and vertical $\{l\}$ bonds is performed independently. In graph theory[15,16] the function (2.35) is called a generating function, which we noted above, introducing for it the notation $\Gamma^{(h)}(z_1,z_2)$, where the superscript (h) indicates that the generating function refers to Hamiltonian circuits. Therefore the problem reduces to summing over all Hamiltonian circuits with a step (edge) of variable length on a square lattice of the type described above.

We note here that the above-described diagrammatic representation of Z_2 is very reminiscent of the diagrammatic representation for the partition function of the $2D$ Ising model in zero magnetic field $(h=0)$ (see, for example, Refs. 11, 15, and 19). In this case, as is well known,[11,19] the partition function can be represented as

$$Z(K_1,K_2)=(2\cosh K_1\cosh K_2)^{NM}\times\sum_{\alpha,\beta}g_{\alpha,\beta}\tanh^{\alpha}K_1\tanh^{\beta}K_2, \tag{2.36}$$

where $g_{\alpha,\beta}$ is the number of closed diagrams consisting of β horizontal and α vertical bonds, the bonds connecting the nearest sites of a square lattice, so that each α bond is associated with a factor (weight) $\tanh K_1$ and each β bond is associated with a factor $\tanh K_2$. At some vertices of a diagram a single self-intersection is possible, i.e., zero, two, or four lines meet at one vertex of a diagram; this corresponds to unoriented Eulerian circuits of degree $\delta\leqslant 4$.[15,17] Figure 3 shows one of the simplest diagrams contributing to the sum (2.36) for $Z_2(K_1,K_2)$. This last circumstance makes the present case substantially different from the one described above, since in our case only zero or two lines (horizontal and vertical) can meet at a single vertex. As mentioned above, this corresponds to unoriented Hamiltonian circuits in a square lattice.[15,17] Another difference is that the α and β bonds in Eq. (2.35) can connect not only the nearest-neighbor lattice sites, which is manifested as a dependence of the weighting factors $z_1^{l_j}$ and $z_2^{k_j}$ on the distances l and k between the lattice sites in the vertical and horizontal directions, respectively. As mentioned above, in the terminology of graph theory[15] the problem of calculating the sum (2.35) is a problem of summing over Hamiltonian circuits (simple circuits) on a rectangular lattice with $N\times M$ sites and variable edge ''length'' in the horizontal and vertical directions. At the same time, the problem of calculating the sum in Eq. (2.36) is a problem of summing over all possible Eulerian circuits of the type described above $(\delta\leqslant 4)$ on the same lattice. As is well known,[15] there is a close relation between Eulerian and Hamiltonian graphs, and for some types of Eulerian graphs it is possible to switch to Hamiltonian graphs, but not vice versa. A number of examples of a nontrivial relation between the generating functions for Eulerian and Hamiltonian cirsuits on a simple rectangular lattice are presented in Refs. 20 and 21. In Ref. 21 it was shown by comparing that the generating function $\Gamma^{(h)}(z_1,z_2)$ for Hamiltonian circuits, described above, equals exactly the generating function $\Gamma^{(e)}(z_1,z_2)$ for Eulerian circuits $(\delta\leqslant 4)$ of the $2D$ Ising model,[15] so that

$$\Gamma^{(h)}(z_1,z_2)=\prod_{n,m=0}^{N,M}\left[(1+z_1^2)(1+z_2^2)-2z_1(1-z_2^2)\times\cos\left(\frac{2\pi n}{N}\right)-2z_2(1-z_1^2)\cos\left(\frac{2\pi m}{M}\right)\right]^{1/2}, \tag{2.37}$$

where $z_{1,2} \equiv \tanh K_{1,2}$. In what follows, the sum (2.35) will be calculated directly and it will be shown that the equality (2.37) is satisfied. Here I note only that a number of examples of the calculation of generating functions for Hamiltonian circuits with weights different from those examined above are presented in Ref. 20.

3. SOLUTION

The Kac–Ward solution,[9] an excellent exposition of which can be found in Ref. 22, contains a well-known topological argument. Specifically, if in a circuit around an arbitrary closed diagram (here we are talking about Eulerian graphs on a lattice) a weight $\alpha = \exp(i\pi/4)$ is assigned to every turn to the left and a weight $\alpha^{-1} = \exp(-i\pi/4)$ is assigned to every turn to the right, then the closed diagrams (i.e., the diagrams which we wish to take into account) will be taken into account and the forbidden diagrams will be canceled, if these diagrams are transversed along different paths. The complete proof of this theorem was given by Sherman.[23] A similar assertion also holds in the case of Hamiltonian graphs on a lattice with a step of variable length, which were described above; this will be proved below for simple cases. However, our arguments will follow Refs. 11 and 24.

First, recall that some Hamiltonian loops (for example, Fig. 1e) enter with a minus sign in the expression (2.35) for S. Specifically, using the commutation relations (2.23), it can be verified directly that each ''double crossing bond'' of the type shown in Fig. 1e contributes a minus sign to the total sign of the simple loop (2.34) for all admissible diagrams. This is also true of the vertical ''double crossing bonds.'' At the same time, each ''simple double bond'' of the type shown in Fig. 1f contributes a plus sign to the total sign of the simple loop (2.34). All other simple loops without ''double bonds'' of the type shown in Fig. 1b–d appear in the sum (2.35) for S with a plus sign. (I note here in passing that every Eulerian graph on a lattice can be put into a one-to-one correspondence with a Hamiltonian graph with a step of variable length and no ''double bonds,'' the Hamiltonian graph consisting of one, two, and more simple loops. For this, it is necessary to ''cross out'' in the Eulerian graph all intermediate vertices, together with vertices where the horizontal and vertical edges of an Eulerian graph self-intersect.)

Now it is easy to see that if in the expression (2.35) for S all simple loops (2.34) are taken with a plus sign and a weight $\alpha = \exp(i\pi/4)$ $(\alpha^{-1} = \exp(-i\pi/4))$ is assigned to each turn to the left (right) in a circuit around a simple loop, then the problem of calculating the sum S (2.35) actually reduces to that of a ''random walk'' of a point along a lattice with a step of variable length.[11,22,24] Indeed, for this method of traversing simple loops, all loops with ''double bonds'' are cancelled (for example, loops in Fig. 1e and 1f), as should happen. By this method it is possible to traverse all loops with ''double bonds'' and show that their total contribution will cancel.

Further, using arguments similar to those presented in Refs. 11, 22, and 24, it is easy to show for a number of specific examples that if all Hamiltonian-type loops with a step of variable length and no ''double bonds'' are traversed along different paths (assigning to each path corresponding weights α and α^{-1} in turning left and right, respectively), then all allowed diagrams will be taken into account and all forbidden diagrams will be cancelled. Here it should be especially emphasized that such complete cancellation of the forbidden diagrams in any order will occur only if factorized weights (z_1^l, z_2^k) are assigned to steps of length l and k, respectively. In Ref. 20 it is shown that in an external magnetic field H the weights z_1^l and z_2^k are renormalized by $a(l)$ and $b(k)$, where $a(l)$ and $b(k)$ are known functions of the interaction parameters $K_{1,2}$, the external field $h = \beta\mu H$, and the positive integers l and k. For this case, the above-indicated complete cancellation of the forbidden diagrams does not occur.

Returning to our problem and following now the arguments of Refs. 11 and 24, we can write S (2.35) as

$$S = \exp\left[-\sum_{r=1}^{\infty} f_r \right], \tag{3.1}$$

where f_r is a sum over all single simple loops of length $r = 2s$, i.e., consisting of s horizontal and s vertical bonds. Every horizontal bond enters with weight $z_2^k e^{i\varphi/2}$ and every vertical bond enters with weight $z_1^l e^{i\varphi/2}$, where $\varphi = \pm\pi/2$ is the change in direction on passing to the next bond. Introducing the quantity $W_r(n,m,\nu)$—the sum over all possible transitions with $r = s_1 + s_2$ bonds from an initial site (n_0, m_0, ν_0) into the site (n,m,ν), where ν is an additional index associated with the four possible directions (1,2,3,4) on a simple square lattice, we write

$$f_r = \frac{1}{2r} \sum_{n_0, m_0, \nu_0} W_r(n_0, m_0, \nu_0). \tag{3.2}$$

It is easy to obtain the following recurrence relations ($\alpha \equiv \exp(i\pi/4)$) for $W_r(n,m,\nu)$:

$$\begin{aligned}
W_{r+1}(n,m,1) &= 0 + \alpha^{-1} \sum_{l=1}^{N} z_1^l W_r(n-l,m,2) + 0 \\
&\quad + \alpha \sum_{l=1}^{N} z_1^l W_r(n+l,m,4), \\
W_{r+1}(n,m,2) &= \alpha \sum_{k=1}^{M} z_2^k W_r(n,m-k,1) + 0 \\
&\quad + \alpha^{-1} \sum_{k=1}^{M} z_2^k W_r(n,m+k,3) + 0, \\
W_{r+1}(n,m,3) &= 0 + \alpha \sum_{l=1}^{N} z_1^l W_r(n-l,m,2) + 0 \\
&\quad + \alpha^{-1} \sum_{l=1}^{N} z_1^l W_r(n+l,m,4), \\
W_{r+1}(n,m,4) &= \alpha^{-1} \sum_{k=1}^{M} z_2^k W_r(n,m-k,1) + 0 \\
&\quad + \alpha \sum_{k=1}^{M} z_2^k W_r(n,m+k,3) + 0.
\end{aligned} \tag{3.3}$$

The meaning of the recurrence relations (3.3) is completely obvious. For example, site $(n,m,1)$ can be reached from sites $(n',m,2)$ and $(n'',m,4)$, i.e., from above and below (we chose the direction ''to the right'' as the direction 1), where $n'=n-l$, $n''=n+l$, and l runs through, strictly speaking, the values from 1 to $N-1$. However, for large N the summation over l can be extended to N, which was done in the expressions (3.3), since these boundary effects vanish in the thermodynamic limit. The Hamiltonian character of the simple loops shows up clearly in the structure of the recurrence relations (3.3); this should be compared with the case of the Eulerian graphs.[22,24] Writing now the recurrence relations (3.3) in the matrix representation

$$W_{r+1}(n,m,\nu)=\sum_{n',m',\nu'}\Lambda(n,m,\nu|n',m',\nu')W_r(n',m',\nu'), \tag{3.4}$$

it is easy to see[24] that

$$\mathrm{Tr}\ \Lambda^r=\sum_{n_0,m_0,\nu_0}W_r(n_0,m_0,\nu_0), \tag{3.5}$$

and

$$f_r=\frac{1}{2r}\ \mathrm{Tr}\ \Lambda^r=\frac{1}{2r}\sum_i\ \lambda_i^r. \tag{3.6}$$

Using Eqs. (3.2) and (3.6), we can now write for S (3.1) the expression

$$S=\prod_i\ \sqrt{1-\lambda_i}, \tag{3.7}$$

where λ_i are the eigenvalues of the matrix $\Lambda(n,m,\nu)$, $i=1,2,\ldots,4NM$. The matrix $\Lambda(n,m,\nu)$ can be easily diagonalized with respect to the indices n, m by switching to a different representation with the aid of the Fourier transformation:

$$W_r(n,m,\nu)=\sum_{q,p=0}^{N,M}\ \exp\left(\frac{2\pi i}{N}\,nq+\frac{2\pi i}{M}\,mp\right)\times W_r(q,p,\nu). \tag{3.8}$$

Substituting (3.8) into Eq. (3.3), we obtain for fixed q, p

$$\Lambda(q,p,\nu|q,p,\nu')=\begin{bmatrix} 0 & \alpha^{-1}\sum_{l=1}^{N} z_1^l\varepsilon^{-lq} & 0 & \alpha\sum_{l=1}^{N} z_1^l\varepsilon^{lq} \\ \alpha\sum_{k=1}^{M} z_2^k\omega^{-kp} & 0 & \alpha^{-1}\sum_{k=1}^{M} z_2^k\omega^{kp} & 0 \\ 0 & \alpha\sum_{l=1}^{N} z_1^l\varepsilon^{-lq} & 0 & \alpha^{-1}\sum_{l=1}^{N} z_1^l\varepsilon^{lq} \\ \alpha^{-1}\sum_{k=1}^{M} z_2^k\omega^{-kp} & 0 & \alpha\sum_{k=1}^{M} z_2^k\omega^{kp} & 0 \end{bmatrix}, \tag{3.9}$$

where $\alpha\equiv\exp(i\pi/4)$, $\varepsilon\equiv\exp(2\pi i/N)$, and $\omega\equiv\exp(2\pi i/M)$.

For fixed q, p it is obviously sufficient in our case to calculate the determinant of a 4×4 matrix

$$\prod_{j=1}^{4}(1-\lambda_j)=\mathrm{Det}(\delta_{\nu\nu'}-\Lambda_{\nu\nu'})\equiv A(q,p), \tag{3.10}$$

and after simple calculations the following expression is obtained for $A(q,p)$ (3.10):

$$A(q,p)=\frac{(1+z_1^2)(1+z_2^2)-2z_1(1-z_2^2)\cos(2\pi q/N)-2z_2(1-z_1^2)\cos(2\pi p/M)}{[1-2z_1\ \cos(2\pi q/N)+z_1^2][1-2z_2\ \cos(2\pi p/M)+z_2^2]}. \tag{3.11}$$

In deriving (3.11) terms proportional to z_1^N and z_2^M were dropped, since for asymptotically large values of N and M at $z_{1,2}<1$, $z_1^N\approx0$ and $z_2^M\approx0$. Finally, for asymptotically large N and M, substituting the expression (3.11), we obtain for S (3.7)

$$S=\prod_i\ \sqrt{1-\lambda_i}=\prod_{q,p=0}^{N,M}A^{1/2}(q,p)=\prod_{q,p=0}^{N,M}\left[\frac{(1+z_1^2)(1+z_2^2)-2z_1(1-z_2^2)\cos(2\pi q/N)-2z_2(1-z_1^2)\cos(2\pi p/M)}{[1-2z_1\ \cos(2\pi q/N)+z_1^2][1-2z_2\ \cos(2\pi p/M)+z_2^2]}\right]^{1/2}. \tag{3.12}$$

For asymptotically large N and M, (3.12) obviously passes into the expression (2.37), since

$$\prod_{q=0}^{N}\left(1-2z_1\ \cos\frac{2\pi q}{N}+z_1^2\right)=1,$$

$$\prod_{p=0}^{M}\left(1-2z_2\ \cos\frac{2\pi p}{N}+z_2^2\right)=1$$

where N, $M\to\infty$, $z_{1,2}<1$. Finally, taking account of Eq.

(2.30) and substituting (3.12) into Eq. (2.27), we obtain for the free energy f per Ising spin in the thermodynamic limit the well-known Onsager expression

$$-\beta f=\lim_{N,M\to\infty}\frac{1}{NM}\ln Z_2=\ln 2+\frac{1}{2\pi^2}\int_0^{\pi}\int_0^{\pi}d\varphi_1\,d\varphi_2 \times\ln[\cosh(2K_1)\cosh(2K_2) -\sinh(2K_1)\cos\varphi_1-\sinh(2K_2)\cos\varphi_2], \quad (3.13)$$

where $\beta=1/k_BT$ is the reciprocal temperature.

4. CONCLUSIONS

Despite its relative inconvenience, the method proposed in this paper for obtaining the Onsager solution makes it possible to study analytically the Ising–Onsager problem in an external magnetic field for a number of limiting cases in both two and three dimensions. The last part of the proposed solutions still differs from the classical Sherman–Vdovichenko solution, which reduces to summing over closed diagrams. A substantial difference is that in our case the summation is performed over closed Hamiltonian graphs, while in the case of the Sherman–Vdovichenko solution the summation extends overclosed Eulerian graphs. This circumstance (together with the representation of the partition function in the form (2.25)) makes it possible to take into account an external magnetic field effectively. In Refs. 13 and 20 it is shown, on the basis of this approach, how the operator corresponding to the interaction of Ising spins with an external magnetic field can be maximally simplified in the second-quantization representation. This in turn makes it possible to renormalize effectively the interaction constants $K_{1,2}$ taking account of the external magnetic field H. Such a renormalization is made possible by the representation of the partition function in the form (2.25). Therefore, in my opinion, the formalism presented in this paper and the solution of the $2D$ Ising problem are not only of interest in themselves but they also make possible a substantial advance in the solution of the Ising–Onsager problem. These results can be applied to analysis of the thermodynamics of an Ising magnet in an external magnetic field. I have yet to be able to give a complete proof of the analog of the Kac–Ward topological theorem,[22] as was done by Sherman[23] for the case of Eulerian graphs on a simple square lattice.

I am deeply grateful to Professor Yan Mostovskiĭ for his questions and doubts as well as his extraordinary patience in listening to my ''short'' stories about the Ising–Onsager problem. In large measure, this helped me to understand more deeply the essence and difficulties of this problem.

[1] M. S. Kochmański, J. Tech. Phys. **36**, 485 (1995).
[2] B. McCoy and T. T. Wu, *Two-Dimensional Ising Models*, Harvard University Press, Cambridge (1973).
[3] J. M. Ziman, *Models of Disorder*, Cambridge University Press, New York (1979).
[4] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, New York (1982).
[5] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems*, Consultants Bureau, New York (1988).
[6] C. J. Thompson, *Classical Equilibrium Statistical Mechanics*, Clarendon Press, Oxford (1988).
[7] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).
[8] P. Jordan and E. Wigner, Z. Phys. **47**, 631 (1928).
[9] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952).
[10] P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[11] N. V. Vdovichenko, Zh. Éksp. Teor. Fiz. **47**, 715 (1964) [Sov. Phys. JETP **20**, 477 (1965)].
[12] R. J. Baxter and I. G. Enting, J. Phys. A **11**, 2463 (1978).
[13] M. S. Kochmański, submitted to Phys. Rev. E.
[14] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, New York (1977).
[15] F. Harary (ed.). *Graph Theory and Theoretical Physics*, Academic Press, New York (1967).
[16] F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, New York (1973).
[17] W. T. Tutte, *Graph Theory*, Cambridge University Press, New York (1984).
[18] G. C. Wick, Phys. Rev. **80**, 268 (1950).
[19] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1961).
[20] M. S. Kochmański, J. Tech. Phys. **37**, 67 (1996).
[21] M. S. Kochmański, submitted to Ann. Polon. Math. (1996).
[22] R. P. Feynman, *Statistical Mechanics*, W. A. Benjamin, Inc., Reading, Massachusetts (1972).
[23] S. Sherman, J. Math. Phys. **1**, 202 (1960); *ibid*. **4**, 1213 (1963).
[24] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, New York (1980).

Translated by M. E. Alferieff

Anomalous magnetic properties of ferrites of the $CuFe_{2-x}Cr_xO_4$ system with high Cr^{3+} content

L. G. Antoshina, A. N. Goryaga, E. N. Kukudzhanova, and I. A. Filgus

M. V. Lomonosov Moscow State University, 119890 Moscow, Russia
(Submitted 5 May 1996)
Zh. Éksp. Teor. Fiz. **111**, 1732–1737 (May 1997)

The paper reports on a comprehensive study of magnetic properties of $CuFe_{2-x}Cr_xO_4$ ferrites ($x=0.0$, 0.2, 0.3, 1.0, 1.4, 1.6, and 2.0). The curves of magnetization versus temperature, $\sigma_s(T)$, of the ferrites with $x=1.0$, 1.4, and 1.6 have anomalous shapes: the magnetization begins to fall off at lower temperatures than the Curie point T_C. The experimental results and analysis of exchange interactions suggest that in ferrites with high Cr^{3+} content, magnetic transitions to either a frustrated magnetic structure or a clustered spin glass can take place.
 [S1063-7761(97)01305-X]

The system of copper ferrites–chromites $CuFe_{2-x}Cr_xO_4$ have been studied for a long time, but researchers' attention has been focused mostly on their structure,[1–3] whereas little information on magnetic and electric properties of these system is available. For example, the Curie temperature T_C has been measured only in the two limiting cases of $CuFe_2O_4$ and $CuCr_2O_4$.[4] It seemed interesting to us, therefore, to undertake a comprehensive study of magnetic parameters (such as magnetization, coercive force, and magnetoresistance) of $CuFe_{2-x}Cr_xO_4$ system.

We selected for our study samples of the $CuFe_{2-x}Cr_xO_4$ system with $x=0.0$, 0.2, 0.3 (distortion parameter $c/a>1$), $x=1.0$ ($c/a=1$), and $x=1.4$, 1.6, 2.0 ($c/a<1$).[3] The samples were prepared using ceramic technology. The first anneal was performed at 750 °C for 20 h, the second at 900 °C also for 20 h. Both anneals were performed in air, and the samples were then cooled gradually. X-ray diffraction patterns recorded at room temperature indicated that the samples were single-phase spinels.

Magnetization was measured by the ballistic technique, and magnetoresistance using a resistance bridge, since the sensitivity of this method is higher than that of the potentiometer technique. The remanent magnetization σ_r and coercive force H_c were derived from the shapes of hysteresis loops. Contacts were fabricated on the samples from indium–gallium conducting paste. Magnetic fields of up to 10 kOe were generated by an electromagnet.

Our measurements indicate that throughout the investigated temperature range, the magnetization of copper ferrites–chromites in strong magnetic fields, $\sigma(H)$, is not saturated, i.e., the paraprocess (true magnetization) is observed. In this case, the spontaneous magnetization σ_s at all temperatures was determined by extrapolating the linear part of the $\sigma(H)$ curve to $H=0$. The values of σ_s determined by this method were plotted against temperature. It turned out that the $\sigma_s(T)$ curves for samples with high chromium content ($x=1.0$, 1.4, 1.6, and 2.0) have anomalous shapes. As the temperature rises, the falloff in magnetization σ_s takes place at temperatures lower than those at which the coercive force H_c decreases.

Figure 1 shows measurements of the spontaneous magnetization σ_s, coercive force H_c, and the magnitudes of longitudinal $(\Delta R/R)_\parallel$ and transverse $(\Delta R/R)_\perp$ galvanomagnetic effects for the $CuFeCrO_4$ sample as functions of the temperature. One can see that the coercive force remains fairly high at temperatures where the spontaneous magnetization is low. As concerns galvanomagnetic effects, their magnitudes drop at the same temperature as the decrease in magnetization $\sigma_s(T)$. The negative signs of the two galvanomagnetic effects at $H=10.5$ kOe and their approximate equality indicate that the sample behavior in strong magnetic fields can be described in terms of a paraprocess. Similar effects were detected in ferrites–chromites with $x=1.4$ and 1.6. The spontaneous magnetization $\sigma_{s(1.6)}$ of the $CuFe_{0.4}Cr_{1.6}O_4$ sample versus temperature is plotted as a dashed line in Fig. 1.

These results suggest that there are two magnetic phase transitions in ferrites–chromites with high Cr^{3+} content, i.e., in addition to the transition at the Curie temperature T_C, another magnetic transition occurs at the lower temperature $T\simeq T_t$. The Curie temperature of such samples was defined as the temperature at which both the magnetization σ_s and coercive force H_c vanish, and the temperature T_t was determined by extrapolating the linear part of the magnetization curve $\sigma_s(T)$ to its intercept with the temperature axis.

Figure 2 shows $(\sigma_s/\sigma_{s0})(T/T_C)$ for three samples and demonstrates that for $CuFe_2O_4$ and $CuFe_{1.8}Cr_{0.2}O_4$ ferrites these curves are Q-type (according to Néel's classification), whereas the behavior for the $CuFeCrO_4$ sample is anomalous: the ratio σ_s/σ_{s0} begins to fall off at a temperature considerably lower than the Curie temperature. Note that curves of $(\sigma_s/\sigma_{s0})(T/T_C)$ for the samples with $x=1.4$ and 1.6 are not given in Fig. 2 since they are essentially identical to the curve for the sample with $x=1.0$.

Measurements of the remanent magnetization σ_r in samples with $x=1.0$, 1.4, and 1.6 have demonstrated that the sign of σ_r does not change anywhere within the temperature range studied. We thus conclude that the observed drop in magnetization is not due to the compensation temperature, but has a different cause.

In the reported work, we attempted to discover the cause

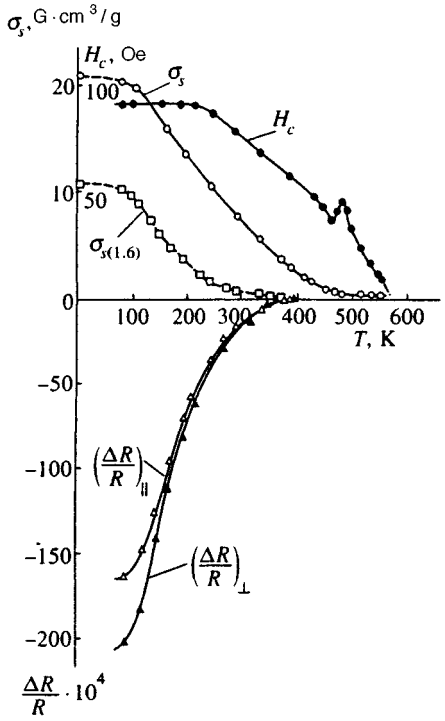

FIG. 1. Spontaneous magnetization σ_s, coercive force H_c, longitudinal $(\Delta R/R)_\parallel$ and transverse $(\Delta R/R)_\perp$ magnetoresistances versus temperature measured in $CuFeCrO_4$ in a magnetic field $H=10.5$ kOe. The spontaneous magnetization $\sigma_{s(1.6)}$ of $CuFe_{0.4}Cr_{1.6}O_4$ is shown by the dashed curve.

of the anomalous behavior of magnetic properties of copper ferrites–chromites with high Cr^{3+} content.

Recently a lot of attention has been focused on materials with frustrated magnetic structures (spero-, speri-, and asperomagnetics) and on spin-glass-like structures. According to the theoretical work of Van Hemmen,[5,6] if a transition from a paramagnetic to spin-glass state occurs in a magnetic material, a second transition from the spin glass to a mixed magnetic phase should take place as the temperature continues to fall. The mixed phase, for example, might be either ferromagnet plus spin glass or ferrimagnet plus spin glass.

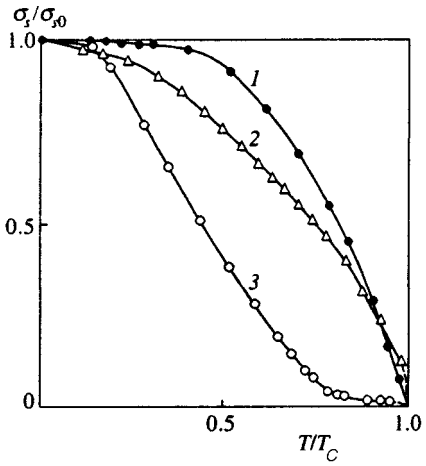

FIG. 2. Relative spontaneous magnetization versus relative temperature, $(\sigma_s/\sigma_{s0})(T/T_C)$, for the samples (*1*) $CuFe_2O_4$; (*2*) $CuFe_{1.8}Cr_{0.2}O_4$; (*3*) $CuFeCrO_4$.

An ideal spin glass is a magnetic structure which has zero spontaneous magnetization at $H=0$. No ideal spin glass, however, has been detected in real materials, but clustered spin glasses are observed. In this case, there are small magnetized regions due to short-range exchange, i.e., short-range magnetic ordering occurs. Therefore, materials with the magnetic structure of clustered spin glass have at $H=0$ a spontaneous magnetization σ_s considerably lower than that of materials with frustrated magnetic structures. In frustrated magnetic structures, regions with spontaneous magnetization are fairly large and have long-range magnetic order. According to Coey,[7] the necessary condition for formation of such anomalous magnetic structures is the presence of two or more types of magnetic ions, with the exchange interactions among them of different magnitudes and signs.

It is known that ferrites with the spinel structure usually contain two or more types of magnetic cations, and direct BB-exchange is possible, beside the indirect AB-, BB-, and AA-interactions. Therefore, we assume that the probability of forming both a frustrated magnetic structure and a clustered spin glass in these magnetic materials is fairly high.

It is of interest to check whether the prerequisites for forming such anomalous magnetic structures are satisfied in the studied ferrites–chromites. Using the Goodenough–Kanamori rules[8–10] and the distribution of cations in copper ferrites–chromites described by Ohnishi and Teranishi,[3] we have qualitatively estimated the exchange interactions that might occur in $CuFeCrO_4$, $CuFe_{0.6}Cr_{1.4}O_4$, and $CuFe_{0.4}Cr_{1.6}O_4$.

According to the Goodenough–Kanamori rules,[8–10] the negative exchange coupling between $Fe_A^{3+}(e_g^2t_{2g}^3)$ and $Fe_B^{3+}(t_{2g}^3e_g^2)$ ions should be strongest, since it involves the half-filled t_{2g} and e_g orbitals. The second strongest should be the positive interaction between the $Fe_A^{3+}(e_g^2t_{2g}^3)$ and $Cr_B^{3+}(t_{2g}^3e_g^0)$ ions. Since the $Cu_A^{2+}(e_g^4t_{2g}^5)$ and $Cu_B^{2+}(t_{2g}^6e_g^3)$ ions have orbitals that are more than half full, all the other indirect AB-exchange interactions (all of them negative) will be weaker than the first two. Along with strong negative indirect BB-exchanges like Fe_B^{3+}–O^{2-}–Fe_B^{3+} and Fe_B^{3+}–O^{2-}–Cr_B^{3+}, negative direct cation–cation couplings Cr_B^{3+}–Cr_B^{3+} and Fe_B^{3+}–Fe_B^{3+} can occur in spinel structures. Since the angle between cations at A-sites in spinels is small, indirect AA-exchanges are weak and are usually omitted from calculations.

It follows from the data given above that both frustrated magnetic structures and clustered spin glasses can be formed in the ferrites–chromites discussed here, since they contain magnetic cations of three types, namely Fe^{3+}, Cr^{3+}, and Cu^{2+}, with exchange interactions of different signs and magnitudes among them.

Since ferrimagnetic order in ferrites with the spinel structure is due to the negative AB-interaction between the sublattices, a positive AB-interaction and negative BB-interaction tend to disrupt this magnetic ordering. Therefore, we assume that the transition to the clustered spin glass at the temperature $T\simeq T_t$ during sample heating occurs because the strong negative direct AB-interactions Fe_A^{3+}–O^{2-}–Fe_B^{3+} and Fe_A^{3+}–O^{2-}–Cu_B^{2+} are counteracted by the negative direct Cr_B^{3+}–Cr_B^{3+} and Fe_B^{3+}–Fe_B^{3+}, negative indirect

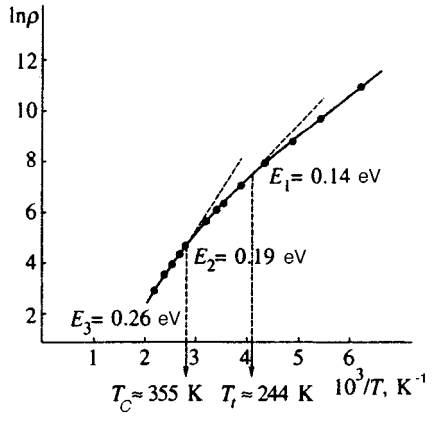

FIG. 3. The plot of $\ln \rho(1/T)$ for the $CuFe_{0.4}Cr_{1.6}O_4$ sample.

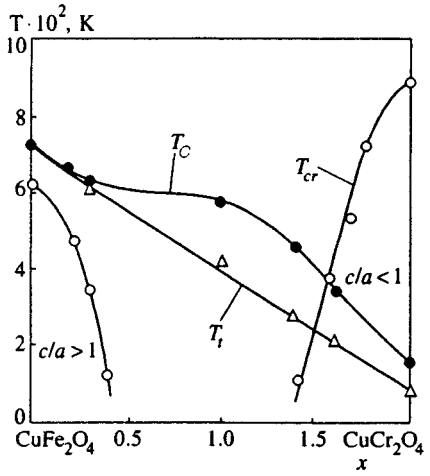

FIG. 4. Temperatures of structural transition T_{cr}, Curie temperature T_C, and temperature of magnetic transition T_t as functions of the $CuFe_{2-x}Cr_xO_4$ system composition x.

Fe_B^{3+}–O^{2-}–Fe_B^{3+} and Fe_B^{3+}–O^{2-}–Cr_B^{3+} BB-exchange interactions, and by the positive Fe_A^{3+}–O^{2-}–Cr_B^{3+} AB-exchange.

Thus we assume that in copper ferrites–chromites with high Cr^{3+} content, a transition from a paramagnetic to clustered spin-glass state occurs at T_C, and at T_t the material transforms to a mixed (ferrimagnetic plus clustered spin glass) or purely ferrimagnetic phase.

It is known that neither galvanomagnetic effects nor magnetostriction (even-order effects) are observed in the state of clustered spin glass. Therefore, we assume that the lack of galvanomagnetic effects $(\Delta R/R)_{\parallel}$ and $(\Delta R/R)_{\perp}$ in $CuFeCrO_4$ ferrites (Fig. 1) at $T > T_t$ also provides evidence in favor of the formation of a clustered spin glass magnetic phase.

Good confirmation of the existence of two phase transitions in these ferrites–chromites is the change in the activation energy detected in $CuFe_{0.4}Cr_{1.6}O_4$. We measured the resistivity of this sample and concluded from the shape of the $\ln \rho(1/T)$ curve that at a temperature of about 244 K the activation energy jumps from 0.14 to 0.19 eV, and one more jump occurs at $T \approx 355$ K (from 0.19 to 0.26 eV) (Fig. 3). (Note the behavior of $\sigma_{s(1.6)}(T)$ for the composition with $x = 1.6$ in Fig. 1.) Unfortunately, the transition temperatures T_t and T_C for the $CuFeCrO_4$ sample are too high to measure changes in activation energy using indium–gallium conducting paste contacts.

Figure 4 shows $T_t(x)$, $T_C(x)$, and $T_{cr}(x)$ for $CuFe_{2-x}Cr_xO_4$ ferrite systems, where T_{cr} is the temperature of the tetragonal-to-cubic transition taken from Ref. 3. One can see that the largest difference between the temperatures T_C and T_t was measured in $CuFeCrO_4$ $(c/a = 1)$ and $CuFe_{0.6}Cr_{1.4}O_4$ $(c/a < 1)$, and this difference drops with increasing concentration of Cr^{3+} ions up to the composition $CuCr_2O_4$ $(c/a < 1)$. It is interesting that the ratio between T_t and the Curie temperature, $\gamma = T_t/T_C = 0.71 \pm 0.03$, is approximately the same for the samples with $x = 1.0$, 1.4, and 1.6.

We would like to express our gratitude to Prof. P. N. Stetsenko for valuable remarks made in discussing the results of this work. This work was supported by the Russian Fund for Fundamental Research, Project No. 96-02-19684a.

[1] V. I. Nikolaev, N. N. Oleinikov, V. S. Rusakov, and A. M. Shipilin, Fiz. Tverd. Tela (Leningrad) **29**, 1523 (1987) [Sov. Phys. Solid State **29**, 872 (1987)].
[2] N. T. Malafaev, A. A. Murakhovskii, Yu. A. Popkov, and V. V. Vorob'ev, Ukrainskiĭ Fiz. Zh. **29**, 286 (1984).
[3] H. Ohnishi and T. Teranishi, J. Phys. Soc. Jap. **16**, 35 (1961).
[4] S. Krupička, *Physik der Ferrite und der Verwandten Magnetische Oxide* [in German], Academia, Praha (1973).
[5] J. L. Van Hemmen, Phys. Rev. Lett. **49**, 409 (1982).
[6] A. C. D. Van Enter and J. L. Van Hemmen, Phys. Rev. A **29**, 355 (1984).
[7] J. M. D. Coey, J. Appl. Phys. **49**, 1646 (1978).
[8] J. B. Goodenough, *Magnetism and the Chemical Bond*, Interscience, New York (1963).
[9] J. Kanamori, Phys. Chem. Sol. **10**, 87 (1959).
[10] M. Kataoka and J. Kanamori, J. Phys. Sol. Jap. **31**, 113 (1972).

Translation provided by the Russian Editorial office.

Effect of electron–electron interaction on thermoelectric power in impure metals

K. D. Belashchenko

Kurchatov Institute, 123182 Moscow, Russia

D. V. Livanov

Moscow Institute of Steel and Alloys, 117936 Moscow, Russia

A. V. Sergeev

Moscow State Pedagogical University, 119435 Moscow, Russia

(Submitted 13 June 1996)
Zh. Éksp. Teor. Fiz. **111**, 1738–1747 (May 1997)

A new kinetic phenomenon related to the effect of electron–electron scattering on the thermoelectric coefficient η in a conductor with a small electron mean free path l is considered. The effect is proportional to the electron–hole asymmetry factor $(\epsilon_F\tau)^{-1}$ and the real part of the diffusion-enhanced Coulomb propagator with characteristic wave vectors of up to l^{-1}. Unlike weak localization effects, in the two-dimensional case this effect results in a logarithmic temperature dependence of η and yields the major contribution to the differential thermoelectric power. © *1997 American Institute of Physics.* [S1063-7761(97)01405-4]

1. INTRODUCTION

Numerous theoretical and experimental studies have shown that the interference between electron–electron and electron–impurity interactions in an impure conductor results in a radical modification of both thermodynamic (electron density of states and specific heat) and kinetic (conductivity, thermoelectric power, thermal conductivity, etc.) parameters of the system.[1,2] In low-dimensional conductors, this interference yields nontrivial corrections that differ from those due to Fermi-liquid effects.[2] In particular, it is presently known that the interference corrections to the conductivity have the form

$$\frac{\Delta\sigma^{3d}_{e-e}}{\sigma_0}\sim\frac{(T\tau)^{1/2}}{(\epsilon_F\tau)^2},\qquad \frac{\Delta\sigma^{2d}_{e-e}}{\sigma_0}\sim\frac{1}{\epsilon_F\tau}\ \ln T\tau, \tag{1}$$

where σ_0 is the residual conductivity, T is the temperature, τ is the electron momentum relaxation time due to impurity scattering, and ϵ_F is the Fermi energy. Several studies[3–5] were dedicated to the effect of the electron–electron interaction on the thermoelectric coefficient η in an impure conductor, and attention was focused on the most interesting two-dimensional case. In order to estimate the thermoelectric coefficient starting from the correction to the conductivity, one must calculate two parameters, namely, the typical electron energy ϵ^* measured with respect to the Fermi energy, and the electron–hole asymmetry factor $b_{\rm as}$, which equals the ratio of the difference between the number of electrons and holes and the number of electrons. For noninteracting electrons, $\epsilon^*\sim T$ and $b_{\rm as}\sim T/\epsilon_F$, so $\eta_0\sim(\epsilon^*/eT)b_{\rm as}\sigma_0\sim(T/e\epsilon_F)\sigma_0$. The expression obtained by Ting *et al.*[3] and Fabrizio *et al.*[5] is similar to the formula for noninteracting electrons, namely, $\Delta\eta_{\rm int}\sim T/(e\epsilon_F)\Delta\sigma_{\rm int}$. Hsu *et al.*[4] derived a formula containing a steeper function. The difficulties in the calculations of the thermoelectric coefficient using the Kubo–Greenwood technique (linear response)[4] derive from the necessity to take into account corrections to the operator of heat flux due to interaction of all orders. The operator itself can be expressed in terms of the time derivatives of electron field operators (i.e., in terms of the electron frequency). In another representation of the heat-flux operator, these derivatives can be recalculated through the equations of motion, and the electron frequency can be expressed as the sum of the kinetic and potential electron energies.[6] Hsu *et al.*[4] added terms to the first representation of the heat-flux operator that included the potential energy of the electron–electron interaction in the second representation, and their results was erroneously large.

In this paper we also discuss the effect of electron–electron interaction on the thermoelectric coefficient. We have studied the effect which is fundamentally different from that analyzed previously[3–5] and described by a steeper function of $(T\tau)^{-1}$ than in Refs. 3 and 5. Unlike the authors of these papers, who determined the electron–hole asymmetry factor by expanding all parameters near the Fermi surface, we separate out the asymmetry by retaining in the first-order correction the factor

$$c_{\rm as}=\frac{1}{\pi\nu\tau}\int d\mathbf{p}\ G^A(\mathbf{p},\epsilon)G^A(\mathbf{p},\epsilon)=\frac{i}{\epsilon_F\tau}$$

(in what follows, we assume that the conductor has a single parabolic energy band). For a conductor with equal numbers of electrons and holes, the latter expression, obviously, is equal to zero by virtue of the analytic properties of the Green's functions. Given the additional imaginary part of $c_{\rm as}$, we have to take into account in our calculations the real part of the diffusion-enhanced Coulomb propagator, whereas in the earlier studies[3–5] the imaginary part of this propagator was taken into account.

The new effects can be defined as ''renormalization'' of the thermoelectric coefficient due to electron–electron scattering in an impure conductor. In our opinion, the similarity to the renormalization of the electron density of states due to the electron–phonon scattering is quite to the point. Recall

that the latter is proportional to the real part of the phonon propagator (the kinetic coefficients are proportional to its imaginary part) and is determined by the entire volume occupied by phonons in the quasimomentum space, i.e., by virtual phonons with all wave vectors up to the Debye wave vector. The effect in question is also proportional to the real part of the boson propagator. The diffusion enhancement of the Coulomb propagator is limited by the wave vectors $q \sim 1/l$, where l is the electron mean free path, hence the effect magnitude is determined by the volume of this region in conductors of various dimensionalities. Similarly to the case of renormalization of the density of states due to electron–phonon interaction, the characteristic electron energy is much smaller than the boson energy; in calculations of the thermoelectric power $\epsilon^* \sim T$.

The correction to the thermoelectric coefficient can be estimated using the correction to the density of states, $\Delta\nu_{\rm int}(\epsilon)$, resulting from electron–electron interaction[2]:

$$\Delta\eta_{\rm int} \sim \frac{\epsilon^*}{T}\frac{c_{\rm as}}{b_{\rm as}}\frac{\Delta\nu_{\rm int}}{\nu_0}\eta_0 .$$

Given the expressions for the correction to the density of states due to the diffusion-enhanced Coulomb interaction, we obtain in the case of a two-dimensional conductor $\Delta\eta_{\rm int}^{2d} \sim \eta_0 (T\tau)^{-1}(\epsilon_F\tau)^{-1}\ln(T\tau)$. As noted above, the effect of renormalization exceeds the kinetic effect studied previously[3,5] by the parameter $(T\tau)^{-1}$ resulting from the difference in the electron–hole asymmetry factors.

In order to calculate the corrections to the thermoelectric coefficient, we use the quantum kinetic equation technique, which has, in the context of the discussed problem, several important advantages over the linear response method. In calculating the electric current as a response to the temperature gradient, there is no problem involving corrections to the heat-flux operator.

2. CALCULATION OF THERMOELECTRIC POWER

In calculating corrections to the thermoelectric power in an impure metal due to electron–electron interactions, we use the quantum kinetic equation method based in Keldysh's diagrammatic technique.[7] This method has been applied to calculations of conductivity in impure conductors induced by electron–electron[8] and electron–phonon interaction,[9] and was also used in calculations of the phonon renormalization of thermoelectric power[10] and corrections to the thermal conductivity due to electron–electron interactions.[6]

In Keldysh's technique, the Green's function and electron self-energy, as well as the electron–electron interaction potential, are expressed in the form of matrices:

$$\hat{G} = \begin{pmatrix} 0 & G^A \\ G^R & G^C \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \Sigma^C & \Sigma^R \\ \Sigma^A & 0 \end{pmatrix}, \quad \hat{V} = \begin{pmatrix} 0 & V^A \\ V^R & V^C \end{pmatrix}, \tag{2}$$

while the interaction vertices are tensors.

If the electron–electron interaction is neglected, the electron Green's function averaged over the impurity coordinates is

$$G_0^R(\mathbf{p},\epsilon) = (\epsilon - \xi_p + i/2\tau_\epsilon)^{-1} = [G_0^A(\mathbf{p},\epsilon)]^*,$$

$$\xi_p = (p^2 - p_F^2)/2m, \tag{3}$$

where τ_ϵ is the momentum relaxation time of an electron with a frequency ϵ due to impurity scattering, and p_F is the Fermi momentum.

The screened electron–electron interaction potential in the case of small momenta and energies, $ql \ll 1$, $\omega\tau \ll 1$ (but $p_F l \gg 1$, where $l = v_F\tau$), can be expressed as[2]

$$\begin{aligned} V^R(\mathbf{q},\omega) &= [V^A(\mathbf{q},\omega)]^* \\ &= 4\pi e^2 \left(q^2 + \frac{k_3^2 D q^2}{-i\omega + Dq^2} \right)^{-1}, \quad d = 3, \\ V^R(\mathbf{q},\omega) &= [V^A(\mathbf{q},\omega)]^* \\ &= 2\pi e^2 \left(|q| + \frac{k_2 D q^2}{-i\omega + Dq^2} \right)^{-1}, \quad d = 2. \end{aligned} \tag{4}$$

Here d is the dimensionality of the electron system, $D = v_F^2\tau/d$ is the electron diffusion factor, $k_3^2 = 4\pi e^2\nu_3$, $k_2 = 2\pi e^2\nu_2$ ($\nu_3 = mp_F/\pi^2$, $\nu_2 = m/2\pi$, $\nu_1 = \nu_3 a^2$), a is the characteristic conductor dimension and $k_2 = k_3^2 a/2$ in the quasi-two-dimensional case.

If the electron system is in equilibrium,

$$\begin{aligned} &G^C(\mathbf{q},\omega) = S_0(\epsilon)[G^A(\mathbf{p},\epsilon) - G^R(\mathbf{p},\epsilon)], \\ &V^C(\mathbf{q},\omega) = -(2N_\omega + 1)[V^A(\mathbf{q},\omega) - V^R(\mathbf{q},\omega)], \\ &S_0(\epsilon) = -\tanh(\epsilon/2T), \quad N_\omega = (\exp(\omega/T) - 1)^{-1}. \end{aligned} \tag{5}$$

Since the quantum kinetic equation method was previously described in detail,[7] as well as its application to the problem of the thermoelectric power,[10] we give only the equations necessary for this specific calculation. As in the earlier publication,[10] our goal is to calculate the electron current resulting from the temperature gradient ∇T.

The effects of electron–electron interaction and the nonequilibrium of the system are taken into account in first-order perturbation theory. Without electron–electron interaction, the nonequilibrium correction to the distribution function is well known:

$$\phi_0(\mathbf{p},\epsilon) = \tau_\epsilon \mathbf{v}\nabla T \frac{\partial S_0(\epsilon)}{\partial\epsilon}\frac{\epsilon}{T}. \tag{6}$$

The system inhomogeneity in the momentum space naturally yields corrections in the form of Poisson brackets:

$$\{A,B\} = \left(\frac{\partial A}{\partial \mathbf{R}}\frac{\partial B}{\partial \mathbf{p}} - \frac{\partial A}{\partial \mathbf{p}}\frac{\partial B}{\partial \mathbf{R}} \right), \tag{7}$$

and $\nabla_R = \nabla T \partial/\partial T$ in the presence of a temperature gradient.

The correction to the thermoelectric power is related to both the additional terms in the distribution function and various corrections to the electron density of states:

$$\begin{aligned} \Delta\eta = \frac{2e}{|\nabla T|}\int \frac{d^3p\,d\epsilon}{(2\pi)^4}(\mathbf{v}\cdot\mathbf{n})\{&\phi_1 \,\mathrm{Im}\, G_0^A \\ &+ \phi_0 \,\mathrm{Im}[\delta_{\rm int} G^A(S_0)] + S_0 \,\mathrm{Im}[\delta_{\rm int} G^A(\phi_0)] \\ &+ S_0 \,\mathrm{Im}[\delta G^A(S_0)]\}, \end{aligned} \tag{8}$$

FIG. 1. Diagram of electron self-energy taking into account electron–electron interaction in the first order of perturbation theory.

FIG. 2. Electron–electron interaction vertex in which electron–impurity scattering is taken into account in the ladder approximation.

where **n** is the unit vector aligned with ∇T, $\phi_1(\mathbf{p},\epsilon)$ is the correction to the electron distribution function in first-order perturbation theory with respect to the interaction:

$$\phi_1(\mathbf{p},\epsilon)=\tau_\epsilon[I_{e-e}(S_0+\phi_0)+\delta_{\rm int}I_{e-\rm imp}(S_0+\phi_0)], \quad (9)$$

where I_{e-e} is the electron collision integral with an effective matrix element modified by the electron–electron interaction, $\delta_{\rm int}I_{e-\rm imp}$ is the additional term in the impurity collision integral due to renormalization of the electron density of states. The latter is determined by $\delta_{\rm int}G^A$, which is directly related to the self-energy component:

$$\delta_{\rm int}G^A=(G_0^A)^2\Sigma^A_{e-e}(S_0+\phi_0). \quad (10)$$

In the second term on the right-hand side of Eq. (8), Σ^A_{e-e} is calculated using the equilibrium electron distribution function $S_0(\epsilon)$, and the third term includes the correction to the distribution function ϕ_0 (6). The fourth term in Eq. (8) is related to the corrections to the self-energy in the form of Poisson brackets $\delta\Sigma^A_{e-e}$:

$$\delta G^A=(G_0^A)^2\delta\Sigma^A_{e-e}. \quad (11)$$

As in the calculation of corrections to the electric and thermal conductivity due to electron–electron interaction, the main correction to the thermoelectric power comes both from terms with a singularity in the form of a diffusion pole $(-i\omega+Dq^2)^{-1}$ in the third power with the additional momentum transfer squared in the numerator and terms with the diffusion pole squared.

In contrast to the conductivity, the thermoelectric power contains the electron frequency ϵ in the power one unit higher, so integrating the product of electron distribution functions over ϵ yields an even function of ω. Thus, in calculating the thermoelectric power instead of $\mathrm{Im}\, V^{A,R}(\mathbf{q},\omega)\times(-i\omega+Dq^2)^{-n}$, which is an odd function of ω, one must separate out terms in this function with a real part that is even in ω. When this is done, expanding the integrals of the products of the electron Green's functions in the parameter $(\epsilon_F\tau)^{-1}$ yields an additional imaginary term.

The self-energy diagram, allowing for electron–electron interaction in the random phase approximation, is shown in Fig. 1. The corresponding equation has the form

$$\Sigma^{ij}_{e-e}=i\int\frac{d^dq\,d\omega}{(2\pi)^{d+1}}V^{kl}(\mathbf{q},\omega)\Big\{\Gamma^k_{ii'}G_{i'j'}(\mathbf{p}+\mathbf{q},\epsilon+\omega)\Gamma^l_{j'j}+\frac{i}{2}\Gamma^k_{ii'}\{G_{i'j'}(\mathbf{p}+\mathbf{q},\epsilon+\omega),\Gamma^l_{j'j}\}+\frac{i}{2}\{\Gamma^k_{ii'},G_{i'j'}(\mathbf{p}+\mathbf{q},\epsilon+\omega)\}\Gamma^l_{j'j}\Big\}, \quad (12)$$

where the vertex $\hat\Gamma^l=\hat\Gamma^l_0+\delta\hat\Gamma^l$ is exact in terms of the electron–impurity interaction. The wavy line in Fig. 1 denotes the screened potential of electron–electron interaction.

The electron–electron interaction vertex $\hat\Gamma^k_0$ renormalized for impurities (without corrections in the form of Poisson brackets) can be calculated in the ladder approximation by solving the following matrix equation corresponding to the diagrams in Fig. 2:

$$\hat\Gamma^k_0=\hat\gamma^k+\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dp}{(2\pi)^3}\,\hat\sigma_x\hat G(\mathbf{p},\epsilon)\hat\Gamma^k_0\times\hat G(\mathbf{p}+\mathbf{q},\epsilon+\omega)\hat\sigma_x, \quad (13)$$

where $\hat\gamma^k$ is the bare Coulomb vertex and $\hat\sigma_x$ is the Pauli matrix. The components of the renormalized vertex are expressed to first order in $(\epsilon_F\tau)^{-1}$ as

$$\begin{aligned}
\Gamma^k_{22}&=\frac{\gamma^k_{22}}{1-\zeta},\quad \Gamma^k_{12}=\frac{\gamma^k_{12}}{1-\zeta^R}+S(\epsilon)\gamma^k_{22}\left[\frac{1}{1-\zeta}-\frac{1}{1-\zeta^R}\right],\\
\Gamma^k_{21}&=\frac{\gamma^k_{21}}{1-\zeta^A}-S(\epsilon+\omega)\gamma^k_{22}\left[\frac{1}{1-\zeta}-\frac{1}{1-\zeta^A}\right],\\
\Gamma^k_{11}&=\frac{\gamma^k_{11}}{1-\zeta^*}+S(\epsilon+\omega)\gamma^k_{12}\left[\frac{1}{1-\zeta^R}-\frac{1}{1-\zeta^*}\right]-S(\epsilon)\gamma^k_{21}\\
&\quad\times\left[\frac{1}{1-\zeta}-\frac{1}{1-\zeta^R}\right]+S(\epsilon)S(\epsilon+\omega)\gamma^k_{22}\\
&\quad\times\left[\frac{1}{1-\zeta}+\frac{1}{1-\zeta^*}-\frac{1}{1-\zeta^*}\left(\frac{1}{1-\zeta^A}+\frac{1}{1-\zeta^R}\right)\right],
\end{aligned} \quad (14)$$

where

$$\begin{aligned}
\zeta&=\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dp}{(2\pi)^d}G^A(\mathbf{p},\epsilon)G^R(\mathbf{p}+\mathbf{q},\epsilon+\omega)\\
&=1+i\omega\tau-Dq^2\tau,\\
\zeta^A&=(\zeta^R)^*=\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^2p}{(2\pi)^d}G^A(\mathbf{p},\epsilon)G^A(\mathbf{p}+\mathbf{q},\epsilon+\omega)\\
&=(2-d)\frac{i}{4\epsilon_F\tau}
\end{aligned} \quad (15)$$

provided that $ql\ll 1$ and $\omega\tau\ll 1$. As follows from Eq. (14), the terms in the vertex $\hat\Gamma_0$ of the first order in $(\epsilon_F\tau)^{-1}$ do not contain diffusion-pole singularities, and therefore can be neglected in our calculations.

It follows from Eq. (13) that the equation for the corrections to the vertex $\hat\Gamma^k_0$ due to Poisson brackets is

$$\delta\hat\Gamma^k=\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dp}{(2\pi)^d}\hat\sigma_x\Big[\hat G(\mathbf{p},\epsilon)\delta\hat\Gamma^k\hat G(\mathbf{p}+\mathbf{q},\epsilon+\omega)$$

$$+\frac{i}{2}\{\hat{G}(\mathbf{p},\epsilon),\hat{\Gamma}_0^k\}\hat{G}(\mathbf{p}+\mathbf{q},\epsilon+\omega)$$

$$+\frac{i}{2}\hat{G}(\mathbf{p},\epsilon)\{\hat{\Gamma}_0^k,\hat{G}(\mathbf{p}+\mathbf{q},\epsilon+\omega)\}\Big]\hat{\sigma}_x. \tag{16}$$

Proceeding to the final stages of our calculation, note that the first order in $(\epsilon_F\tau)^{-1}$ in the integral of the product of Green's functions over the electron momentum appears due to expansion of the electron velocity and density of states near the Fermi level, the finite lower limit in the integral over ξ_p, and expansion in powers of q/p_F in terms where angular integration would yield terms of order $ql\ll 1$ without this expansion. Expressions required for subsequent integrals are:

$$\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dp}{(2\pi)^d}(\mathbf{vn})^2[G^A(\mathbf{p},\epsilon)]^2G^R(\mathbf{p}+\mathbf{q},\epsilon+\omega)$$
$$=\frac{i}{d}v_F^2\tau\left(1+\frac{3i}{4\epsilon_F\tau}\right),$$
$$\frac{1}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dp}{(2\pi)^d}(\mathbf{vn})^2[G^A(\mathbf{p},\epsilon)]^2[G^R(\mathbf{p}+\mathbf{q},\epsilon+\omega)]^2 \tag{17}$$
$$=\frac{2}{d}v_F^2\tau^2.$$

As noted above, one must take into account both terms containing the cube of the diffusion pole with additional powers of q in the numerator and terms with the diffusion pole squared. It was shown in Ref. 7 that the correction with the triple diffusion pole is due to $\delta\Gamma_{11}^2$, i.e., a correction to the equilibrium component of the vertex Γ_{11}^2 proportional to the diffusion pole squared, which can be derived from Eq. (16):

$$\delta\Gamma_{11}^2=\frac{i}{2^{2/3}}\frac{\epsilon}{T}\frac{\partial}{\partial\epsilon}(S(\epsilon+\omega)-S(\epsilon))\frac{1}{(1-\zeta)^2}\frac{1}{\pi\nu_\epsilon\tau_\epsilon}$$
$$\times\int\frac{d^dp}{(2\pi)^d}\mathbf{n}\Big(G^A(\mathbf{p}+\mathbf{q},\epsilon+\omega)\frac{\partial}{\partial\mathbf{p}}G^R(\mathbf{p},\epsilon)$$
$$-G^A(\mathbf{p},\epsilon)\frac{\partial}{\partial\mathbf{p}}G^R(\mathbf{p}+\mathbf{q},\epsilon+\omega)\Big). \tag{18}$$

Using the integrals in Eq. (17), we find that $\delta\Gamma_{11}^2$ does not contain any corrections of the first order in $(\epsilon_F\tau)^{-1}$. As a result,

$$\delta\Gamma_{22}^1=-\frac{i\zeta}{2^{1/2}(1-\zeta)^2}\frac{\partial\zeta}{\partial q}|\nabla T|\Big(\frac{\partial S_0(\epsilon)}{\partial\epsilon}\frac{\epsilon}{T}$$
$$+\frac{\partial S_0(\epsilon+\omega)}{\partial\epsilon}\frac{\epsilon+\omega}{T}\Big). \tag{19}$$

Nonlocal corrections to the collision integral δI containing $\delta\Gamma_{11}^2$ cancel, as in the calculation of electric conductivity. Finally, the third term in Eq. (8) — a correction to the density of states in the form of Poisson brackets, which made the main contribution to the conductivity when $\delta\Gamma_{11}^2$ was taken into account — contains the following integral of the product of the electron Green's functions:

FIG. 3. Diagram of electron self-energy, illustrating renormalization of the impurity collision integral due to electron–electron interaction. The double solid line represents the electron Green's function with due account of electron–electron interaction.

$$\int\frac{d^dp}{(2\pi)^d}(\mathbf{v}\cdot\mathbf{n})(\mathbf{q}\cdot\mathbf{n})[G^A(\mathbf{p},\epsilon)]^2G^R(\mathbf{p}+\mathbf{q},\epsilon+\omega). \tag{20}$$

The latter expression does not contain imaginary terms proportional to $(\epsilon_F\tau)^{-1}$, hence the corrections containing the third power of the diffusion pole do not contribute to the thermoelectric power.

Now let us consider the terms containing the diffusion pole squared. The collision integral corresponding to the self-energy diagram in Fig. 1 can be expressed as

$$I_{e-e}(S)=\phi_0(\epsilon,\mathbf{p})\int\frac{d^dq\ d\omega}{(2\pi)^{d+1}}$$
$$\times(S(\epsilon+\omega)-S(\epsilon))\ \mathrm{Re}\frac{V^A(\mathbf{q},\omega)G^A(\mathbf{p}+\mathbf{q},\epsilon+\omega)}{(1-\zeta^*)^2}. \tag{21}$$

Substituting Eq. (21) into Eq. (9) and integrating over electron momentum, we obtain, taking into account Eq. (20),

$$\Delta\eta_1'=-\frac{\pi e\nu\tau v^2}{4\epsilon_F a^{3-d}}\mathrm{Re}\int\frac{d\epsilon d\omega}{(2\pi)^2}(S_0(\epsilon+\omega)$$
$$-S_0(\epsilon))\frac{\partial S_0(\epsilon)}{\partial\epsilon}\frac{\epsilon}{T}\int\frac{d^dq}{(2\pi)^d}\frac{V^R(\mathbf{q},\omega)}{(1-\zeta)^2}. \tag{22}$$

The first term in Eq. (8) also contains a nonequilibrium distribution function due to the correction $\Delta\eta_1''$ to the impurity collision integral resulting from electron–electron interaction (the corresponding self-energy diagram is shown in Fig. 3):

$$\delta_{e-e}I_{e-\mathrm{imp}}=\frac{2}{\pi\nu_\epsilon\tau_\epsilon}\int\frac{d^dk}{(2\pi)^d}\frac{d\epsilon}{2\pi}[S(\mathbf{k},\epsilon)-S(\mathbf{p},\epsilon)]$$
$$\times\delta_{\mathrm{int}}G^A(\mathbf{k},\epsilon). \tag{23}$$

Substituting Eq. (23) into Eq. (9) and then into Eq. (8), and integrating over $\mathbf{p}$, taking into account Eq. (17), we have

$$\Delta\eta_1''=\frac{1}{3}\Delta\eta_1'. \tag{24}$$

The second term in Eq. (8)—an equilibrium correction to the density of states due to the interaction—also contributes to the thermoelectric power:

$$\Delta\eta_2=-\Delta\eta_1'. \tag{25}$$

Note also that the third term in Eq. (8)—a nonequilibrium, nonlocal contribution to the density of states, which is important in calculating the conductivity and is proportional to the integral (20)—also contains no imaginary part proportional to $(\epsilon_F\tau)^{-1}$.

After integrating over ϵ, we obtain the total contribution to the thermoelectric coefficient in the form

$$\Delta\eta^d=\frac{e\nu_d D_d}{16\pi\epsilon_F\tau a^{3-d}T}\int d\omega\left(\omega^2\frac{d}{d\omega}(2N_\omega+1)+2T\right)$$
$$\times\mathrm{Re}\int\frac{d^dq}{(2\pi)^d}\frac{V^R(\mathbf{q},\omega)}{(-i\omega+Dq^2)^2}. \tag{26}$$

In accordance with (4), in the case of three-dimensional interaction at $q\ll k_3$ and $q\sim(\omega/D)^{1/2}$,

$$\frac{Dq^2}{-i\omega+Dq^2}V^R(\mathbf{q},\omega)\approx\frac{1}{\nu_3}. \tag{27}$$

We now integrate Eq. (22) over q and ω. Note that the integral over ω is formally divergent at the upper limit. Nonetheless, as in the calculation of corrections to the density of states and conductivity due to electron–electron interaction, the divergence should be cut off at $1/\tau$, which is the limit of applicability of the diffusion approximation. Finally, we have

$$\Delta\eta^{3d}=\frac{6^{1/2}}{16\pi^2}\frac{e}{l\epsilon_F\tau}. \tag{28}$$

The electron gas with a two-dimensional spectrum has no electron–hole asymmetry, and in accordance with (15), the correction to the thermoelectric coefficient is zero. We connect the dimensionality reduction in this problem with that of the Coulomb interaction Eq.(4), and assume that the electron spectrum remains three-dimensional.

Let us dwell on the two-dimensional case. Note that since the characteristic energy $\omega\sim1/\tau$, the distinction between the three-dimensional and two-dimensional cases is determined by the critical sample thickness $d\sim(D/\omega)^{1/2}\sim l$. As in the calculations of corrections to the density of states and thermal conductivity due to electron–electron interaction, the integral over q is divergent at the lower limit if the approximation defined by Eq. (27) is used. In this case, $|q|$ in the denominator in Eq. (4) cannot be omitted. But since the parity of the integrand with respect to ω is different from the parity in the calculation of thermal conductivity,[2,6] we need, as was noted above, the real part of the integral of the diffusion pole squared times the screened Coulomb potential. For this reason, we cannot use the logarithmic approximation in the integral over q, as was done in Ref. 6. An exact calculation of the required integral yields in the lowest order in $(\epsilon_F\tau)^{-1}\ll1$

$$\mathrm{Re}\int_0^\infty\frac{dq}{(-i\omega+Dq^2)(-i\omega+Dk_2|q|)}=-\frac{\pi}{4k_2D|\omega|}. \tag{29}$$

Finally,

$$\Delta\eta^{2d}=\frac{e}{64\pi}\frac{1}{a\epsilon_F\tau}\ln T\tau. \tag{30}$$

Note that in contrast to the calculation of conductivity[2] in both the three- and two-dimensional configurations, the result strongly depends on the upper limit of integration over ω. The reason is that in the three-dimensional case, this integration contributed only a temperature-independent constant to the conductivity, i.e., only the residual conductivity was renormalized. In calculating the thermoelectric power, one must bear in mind that the diffusion component is linear in temperature, and therefore the correction described by Eq. (28) is nontrivial. Certainly, the range of applicability of first-order perturbation theory with respect to the electron–electron interaction constrains the correction magnitude: $\Delta\eta\ll\eta_0$, where $\eta_0=-(2/9)eT\tau p_F$ is the thermoelectric coefficient of a metal without electron–electron interaction. Thereby one can determine the applicability range of Eq. (28): $1/(\epsilon_F\tau)^2\ll T\tau\ll1$ (the second inequality is due to the diffusion approximation).

3. CONCLUSIONS

The main result of this work is the calculation of the multiparticle correction to the thermoelectric coefficient due to diffusion-enhanced electron–electron interaction. The ''renormalization'' effect proportional to the real part of the Coulomb propagator with the electron–hole asymmetry factor $i/(\epsilon_F\tau)$ is a steeper function of $(T\tau)^{-1}$ than the ''kinetic'' effect studied previously.[3,5]

The parameter usually measured in experiments is the differential thermoelectric power (Seebeck factor) $S=-\eta/\sigma$, which is expressed, with due account of the calculated corrections, as

$$S=S_0\left(1-\frac{\Delta\sigma}{\sigma_0}+\frac{\Delta\eta}{\eta_0}\right), \tag{31}$$

where $S_0=-\eta_0/\sigma_0=\pi^2T/3e\epsilon_F$. In a disordered metal, the multiparticle corrections to the thermoelectric coefficient and conductivity in Eq. (31) are due to both electron–electron interaction and weak localization. In order to compare their contributions, we write the corrections to the thermoelectric coefficient calculated in this paper:

$$\frac{\Delta\eta^{3d}}{\eta_0}\approx-\frac{1}{(T\tau)(\epsilon_F\tau)^2},\quad\frac{\Delta\eta^{2d}}{\eta_0}\approx-\frac{1}{(T\tau)(\epsilon_F\tau)}\ln T\tau. \tag{32}$$

In the two-dimensional case, the relative correction is enhanced by the large logarithmic factor. For comparison, we give the expressions for corrections to the conductivity and thermoelectric coefficient due to weak localization[11,12]:

$$\frac{\Delta\sigma^{3d}_{\mathrm{loc}}}{\sigma_0}\sim-\frac{1}{(\epsilon_F\tau)(p_FL_\phi)},\quad\frac{\Delta\sigma^{2d}_{\mathrm{loc}}}{\sigma_0}\sim-\frac{1}{\epsilon_F\tau}\ln\frac{L_\phi}{l},$$
$$\frac{\Delta\eta^{3d}_{\mathrm{loc}}}{\eta_0}\sim\frac{1}{(\epsilon_F\tau)(p_FL_\phi)},\quad\frac{\Delta\eta^{2d}_{\mathrm{loc}}}{\eta_0}\sim\frac{1}{\epsilon_F\tau}, \tag{33}$$

where L_ϕ is the dephasing range. Comparing the corrections to the thermoelectric coefficient due to electron–electron interaction Eq. (32), weak localization Eq. (33), and the corrections to the conductivity Eqs. (1) and (33), we conclude that in the two-dimensional case the correction to the thermoelectric coefficient calculated in this paper yields the largest contribution to the thermoelectric power, whereas in the three-dimensional case the localization corrections can play

an important role in the dependence on L_ϕ. Note that although the calculated correction is independent of temperature in the three-dimensional case, it can easily be extracted from measurements of the diffusion thermoelectric power, which is linear in temperature. In the two-dimensional case, unlike the correction due to weak localization, the calculated correction contains a logarithmic factor and can be also derived from experimental data.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-16877). One of the authors (A.V.S.) is grateful to the *Physics of Solid-State Nanostructures* Program.

[1] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[2] B. L. Altshuler and A. G. Aronov, *Electron–Electron Interaction in Disordered Systems*, A. L. Efros and M. Pollak (eds.), North Holland, Amsterdam–Oxford–N. Y.–Tokyo (1985), p. 1.
[3] C. S. Ting, A. Houghton, and J. R. Senna, Phys. Rev. B **25**, 1439 (1982).
[4] J. P. M. Hsu, A. Kapitulnik, and M. Yu. Reizer, Phys. Rev. B **38**, 5260 (1989).
[5] M. Fabrizio, C. Castellani, and G. Strinati, Phys. Rev. B **43**, 11088 (1991).
[6] D. V. Livanov, M. Yu. Reĭzer, and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **99**, 1360 (1991) [*sic*].
[7] B. L. Al'tshuler, Zh. Éksp. Teor. Fiz. **75**, 1330 (1978) [Sov. Phys. JETP **48**. 670 (1978)].
[8] B. L. Al'tshuler, A. G. Aronov, JETP Lett. **30**, 482 (1979).
[9] M. Yu. Reĭzer and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **92**, 2291 (1987) [Sov. Phys. JETP **65**, 1291 (1987)].
[10] M. Yu. Reĭzer and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **93**, 2191 (1987) [Sov. Phys. JETP **66**, 1250 (1987)].
[11] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979); L. P. Gor'kov, A. I. Larkin, and D. E. Khmel'nitskiĭ, JETP Lett. **30**, 228 (1979).
[12] V. V. Afonin, Yu. M. Gal'perin, and V. L. Gurevich, Zh. Éksp. Teor. Fiz. **87**, 335 (1984) [Sov. Phys. JETP **60**, 194 (1984)].

Translation provided by the Russian Editorial office.

Luminescence detection of multiphoton ionization–fragmentation of the molecular CrO_4^{2-} anions adsorbed on the surface of dispersed SiO_2

Yu. D. Glinka

Institute of Surface Chemistry, Ukrainian National Academy of Sciences, 252650 Kiev, Ukraine
(Submitted 27 June 1996; resubmitted 30 October 1996)
Zh. Éksp. Teor. Fiz. **111**, 1748–1774 (May 1997)

It is shown that luminescence detection of multiphoton ionization–fragmentation of the molecular CrO_4^{2-} anions adsorbed on the surface of dispersed SiO_2 is possible under excitation with the fundamental frequency of a Nd:YAG pulsed laser ($\lambda = 1.064$ μm). The structure and the process of formation of the adsorbed complexes under thermal activation of the surface and the nature of luminescence transitions in CrO_4^{2-} anions are studied in detail. It is shown that luminescence is excited as a result of the recombination of photoelectrons and ionized chromate ions. Multiphoton ionization of the ions occurs under three-photon resonance conditions. The resonance level is an antibonding state of the adsorption complex formed with the participation of an oxygen vacancy on the SiO_2 surface. The dynamics of the multiphoton luminescence excitation process includes autoionization (stimulated by intercomplex electronic excitation) in superexcited states, fragmentation of chromate anions, and annealing of surface oxygen vacancies. The rate equations for three-photon-resonance multiphoton ionization are studied. The cross sections for two- and one-photon transitions on the nonresonance steps of multiphoton absorption are obtained. It is concluded that the nonlinear polarizability of the donor–acceptor adsorption bond in ''chromate anion–oxygen vacancy'' complexes is very important. © *1997 American Institute of Physics.* [S1063-7761(97)01505-9]

1. INTRODUCTION

In recent years, substantial progress has been made in photoionization laser spectroscopy of molecules.[1–3] This is explained, first, by the progress made in laser technology, especially UV- and VUV-range excimer lasers. Second, the combination of the method of resonance laser photoionization of molecules with mass-spectroscopic identification of the photoions produced has been found to be very effective.

Simple inorganic molecules are of greatest interest for the investigation of resonance multiphoton ionization processes which are followed by fragmentation, since for such molecules the energy structure of the electronic states has been studied quite well and some states absorb in the visible region of the spectrum, thereby eliminating a number of experimental difficulties. For example, NO_2 (Refs. 4 and 5) and NO (Refs. 6 and 7) molecules have been found to be very convenient objects for such investigations.

However, there are few such examples. Moreover, since this method is usually employed in combination with mass-spectrometric detection, the molecules investigated must be in a gaseous state. This makes it impossible to study by laser photoionization spectroscopy a host of inorganic molecular ions that are difficult to produce in the gaseous state. On the other hand, molecular ions such as, for example, CrO_4^{2-} and UO_2^{2+}, can be easily recorded on the surface of wide-gap oxides.[8,9] Such molecular ions are luminescence probes and have been used to assess adsorption activity of the surfaces of dispersed substances.[10–12] Since the adsorption energy for molecular ions ($\leqslant 1.0$ eV) is much lower than their ionization potential and the band gap for the oxides (~ 10 eV), the multiphoton ionization parameters of such ions will correspond roughly to the gas-state approximation. We must also allow for photoionization processes in solids with the participation of excitons and local levels of impurities and lattice defects[13] as well as the resonance photoionization of absorbing centers on a surface.[14]

We recently proposed a method of multiphoton laser photoionization spectroscopy of molecular CrO_4^{2-} ions adsorbed on the surface of dispersed SiO_2 in a monochromatic laser field ($\lambda = 1.064$ μm). In this method, intracenter luminescence of the ions, which is excited in a recombination process, was used to identify the photoionization processes.[15–17] Just as in the case of mass-spectrometric detection of organic molecules,[18] the intensity of the luminescence bands is characterized by power-law functions of excitation radiation intensity. The exponent lies in the range 5–7.[15] This paper reports the results of a further investigation of the luminescence properties of the SiO_2–CrO_4^{2-} adsorption system under excitation at the fundamental frequency of a pulsed Nd:YAG laser ($\lambda = 1.064$ μm). Special attention will be devoted below to analysis of the power dependence (dependence of the luminescence yield on the power density of the laser radiation) and dose dependence (dependence of the luminescence yield on the number of laser pulses). The results of an investigation of the luminescence spectra of the SiO_2–CrO_4^{2-} adsorption system under UV excitation (third and fourth harmonics of the Nd:YAG laser, $\lambda = 0.3547$ and 0.266 μm, respectively) as well as IR- and Raman scattering spectroscopy data will be used to construct a model of the adsorption complexes.

2. EXPERIMENT

The experimental samples were prepared by the following procedure. Dispersed SiO_2 (Aerosil) with a specific sur-

 1063-7761/97/050957-14$10.00

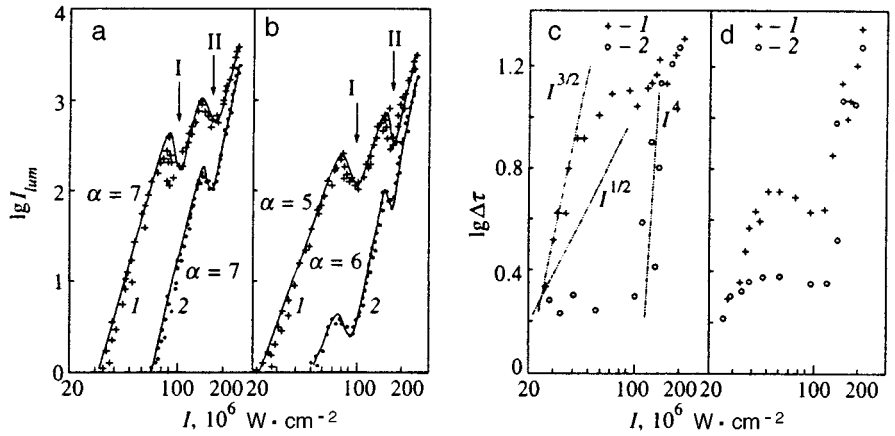

FIG. 1. Power dependences (a, b) and dependences of the time shift $\Delta\tau$ (c,d) on the intensity of excitation with ($\lambda_{exc}=1.064$ μm) for luminescence bands at 540 nm (a, c) and 640 nm (b, d) for SiO_2–CrO_4^{2-} samples ($T_{h.t.}=900$ K): *1*—Initial sample; *2*—after irradiation ($\lambda_{exc}=1.064$ μm, $I=5\times10^7$ W·cm^{-2}) with a dose of 100 pulses.

face area of 220 $m^2\cdot g^{-1}$ was dehydrated at 700 K for 2 h in air. CrO_4^{2-} molecular ions were adsorbed onto the dehydrated surface from a water solution of the salt Na_2CrO_4. Next, the modified sorbent was dehydrated in air at a heat-treatment temperature of $T_{h.t.}=350-1300$ K for 2 h and compacted into tablets under a pressure of 4.0×10^6 $N\cdot m^{-2}$. According to our estimates, the surface density of the CrO_4^{2-} ions was $\leqslant 1.0$ nm^{-2}.

A computer-controlled laser spectrometer was used to measure the luminescence spectra. A pulsed Nd:YAG laser with an amplifier ($\lambda=1.064$ μm, $I=10^5-2\times10^8$ W·cm^{-2}, $\tau=20$ ns) and harmonic generators ($\lambda=0.3547$ μm, $I=10^5-10^7$ W·cm^{-2} and $\lambda=0.266$ μm, $I=10^5-10^6$ W·cm^{-2}) were used as a source of excitation. A S1-70 oscillograph was used to investigate the decay kinetics of the luminescence. The IR absorption spectra were recorded with Specord M-80 and LAFS-1000 spectrometers, and the Raman scattering spectra were recorded with a DFS-24 spectrometer and an LGN-503 laser ($\lambda=0.5145$ μm). All measurements were performed in air at room temperature ($T=300$ K).

3. EXPERIMENTAL RESULTS

First we study the luminescence properties of the SiO_2–CrO_4^{2-} adsorption system under intense IR laser excitation ($\lambda=1.064$ μm). Models of adsorption complexes will be constructed to interpret these results. The models are based on data from luminescence spectroscopy with UV excitation, IR absorption spectroscopy, and Raman scattering spectroscopy. Then the energy structure of adsorption complexes will be studied and the radiative electronic transitions will be systematized. The experimental results on the multiphoton ionization of adsorption complexes will be interpreted on the basis of the proposed models, after which in the last section the rate equations for such a system will be studied and the cross sections for two- and one-photon transitions for nonresonance steps of multiphoton ionization will be determined.

3.1. Luminescence properties of the SiO_2–CrO_4^{2-} adsorption system under intense IR laser excitation ($\lambda=1.064$ μm)

As previously determined,[15] two emission bands with maxima at ~540 and ~640 nm are observed in the luminescence spectrum of the SiO_2–CrO_4^{2-} samples under intense IR laser excitation. The emission band at 540 nm can also be excited for pure dispersed SiO_2.[19] The emission band at 640 nm, however, is due to an electronic transition in the molecular ion and can also be excited by UV radiation.[8,15] These bands are characterized by power-law dependences of the luminescence response of the system on the power density of the exciting radiation (power dependences) (Figs. 1a and 1b, curves *1*). The exponents are $\alpha_1=7\pm0.2$ and $\alpha_2=5\pm0.2$ for the 540 and 640 nm bands, respectively.

It is well known[3,13] that the probability of α-photon ionization can be represented as

$$W_i^{(\alpha)}=\sigma_i^{(\alpha)}I^{\alpha}, \tag{1}$$

where $\sigma_i^{(\alpha)}$ is the cross section for α-photon ionization and I is the laser radiation intensity. For this reason, we previously proposed[15] that the power dependences are a result of multiphoton absorption. The exponent here is the photonicity of the process. Note that if the indicated process were not associated with multiphoton absorption and were determined, for example, by laser-induced heating of the surface with a corresponding desorption process, then the power dependences would be described by an exponential function.[13]

In our case, however, first, the wavelength of the laser radiation falls in the transmission range of the substrate and in the range where there are no absorption bands of the molecular ion. Since the reagents employed are highly pure, absorption of laser radiation by uncontrollable impurities can be neglected. Therefore laser-induced heating of the surface, if it contributes to the interaction of radiation with the adsorption system, does not dominate.

Second, the power-law dependence presented is observed over a wide range of laser radiation intensities, which rules out the possibility that it has been improperly identified. The same high photonicities were observed in the case

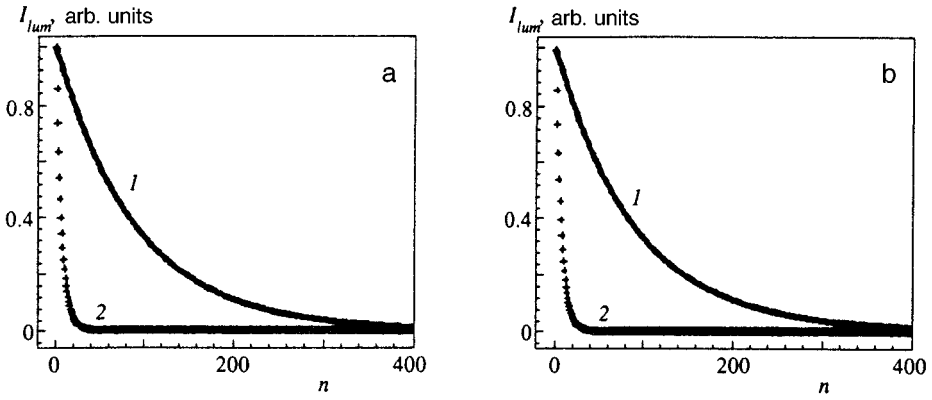

FIG. 2. Dose dependence of the luminescence bands at 540 nm (a) and 640 nm (b) for SiO_2–CrO_4^{2-} samples ($T_{h.t.}=900$ K), $\lambda_{exc}=1.064$ μm, $I=5\times10^7$ (*1*) and 9×10^7 (*2*) W·cm^{-2}.

of mass-spectrometric detection of organic molecules desorbed by powerful laser radiation.[18]

However, it must be noted that the exponent in these experiments is a formal quantity, not at all a consequence of real multiphoton processes but rather reflecting primarily the heating of surfaces by the laser radiation.[20] When the laser radiation reaches an intensity $\sim 8\times10^7$ W·cm^{-2}, the intensity of the emission bands reaches a plateau (saturation of a multiphoton transition), and as the laser intensity increases further, the intensity of the bands drops sharply and then once again increases. However, this time the photonicities are $\alpha_1=7\pm0.2$ and $\alpha_2=6\pm0.2$.

Thus, a dip forms in the power dependence (Fig. 1a and 1b). This dip is due to a sharp resonance decrease, as a function of the intensity of the laser radiation, in the density of luminescing centers. Then the situation repeats, and a second dip forms in the power dependence (the dips are marked in Figs. 1a and 1b by arrows). However, after the second dip the photonicity of the processes is $\alpha_1=\alpha_2=7\pm0.2$.

Note that a time delay $\Delta\tau$ of the luminescence pulse with respect to the laser excitation pulse is also observed (Figs. 1c and 1d); the magnitude of this delay also depends on the power density of the laser radiation. Since at room temperature the afterglow time of the CrO_4^{2-} molecular ion is less than 20 ns, the luminescence response has the same form as the laser pulse but it is delayed with respect to the laser pulse by $\Delta\tau$.

Moreover, dips are also observed in this dependence at positions corresponding to the dips in the power dependences. As established previously,[15] the intensity of the luminescence bands decreases with each subsequent excitation pulse (dose dependence). However, when the laser intensity corresponds to the position of the dips, the rate of ''fading'' of the luminescence bands increases by more than a factor of 10 (Fig. 2). The dose dependence can be described well by an exponential function of the form

$$I_{\text{lum}}=I_{\text{lum}}^{(0)}e^{-n/N}, \tag{2}$$

where $I_{\text{lum}}^{(0)}$ is the initial luminescence intensity, n is the number of laser pulses, and N characterizes the rate of fading of the luminescence bands. For pure SiO_2 samples, $N=3$.[15,19] For the SiO_2–CrO_4^{2-} adsorption system, $N=90$, and when the laser intensity corresponds to the position of the dips, $N=6.5$. Therefore the experimental results attest to multiphoton absorption of intense IR radiation in the SiO_2–CrO_4^{2-} adsorption system, as well as to the complicated relaxation scheme of the multiphoton excitation energy.

3.2. Models of adsorption complexes for the SiO_2–CrO_4^{2-} system

Dispersed SiO_2 is a complicated heterogeneous system, whose surface contains both intrinsic and impurity defects.[21] The surface of the initial dispersed silicon dioxide is hydroxylated ($\equiv$Si–OH groups) with a high content of adsorbed water molecules. The $\equiv$Si–OH groups are the primary water adsorption centers, and the diversity of the variants of adsorption complexes gives rise to various forms of adsorbed water and, in consequence, to the complicated form of the IR absorption spectra in the region of the stretching vibrations of the $\equiv$Si–OH groups and H_2O molecules.

The structure of the surface layer is a function of the heat-treatment temperature of the dispersed silicon dioxide. We therefore investigated the spectroscopic properties of the modified dispersed materials as a function of their preliminary heat treatment temperature ($T_{h.t.}$).

The dynamics of the variation of the spectral and luminescence properties of SiO_2–CrO_4^{2-} samples under UV excitation as a function of the heat-treatment temperature is displayed in Fig. 3a. For the original sample, a luminescence band with a maximum at 510 nm dominates the luminescence spectrum. As the preliminary heat-treatment temperature of the samples increases, bands with maxima at $\sim$620 and $\sim$690 nm appear in the luminescence spectrum. The appearance of red luminescence is a characteristic indicator of the fact that the chromate ions form complexes with active centers on the surface, i.e., thermal activation of the surface of dispersed SiO_2 occurs, as a result of which adsorption complexes are formed. Since the CrO_4^{2-} molecular ions are selective luminescence probes (they luminescence only when adsorption complexes form on the surface),[8] the possibility of radiative transitions in the chromate ion must be attributed to a lowering of its local symmetry.

The intensity of the radiation bands with maxima at 510, 620, and 690 nm is plotted in Fig. 3b. Note the inverse pro-

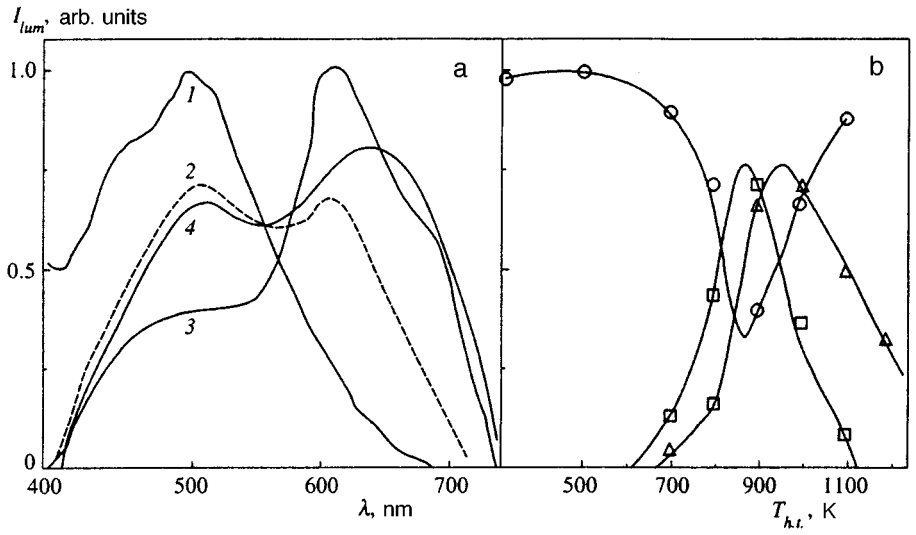

FIG. 3. a) Luminescence spectra of SiO_2–CrO_4^{2-} samples with $\lambda_{exc}=266$ nm; $T_{h.t.}=300$ (*1*), 800 (*2*), 900 (*3*), and 1000 K (*4*). b) Intensities of the 510 (○) 620 (□), and 690 nm (△) luminescence bands versus $T_{h.t.}$.

portionality of the intensities of the 510 and 620 nm radiation bands. This indicates that these bands belong to luminescence centers that interact with the same active center on the surface. The decrease in the intensity of the 620 and 690 nm emission bands at high $T_{h.t.}$ is due to the desorption of the CrO_4^{2-} molecular ion. The intensity of the luminescence bands is proportional to the density of adsorption complexes, which in turn is proportional to the probability $1-W_{\rm des}$, where $W_{\rm des}$ is the probability of desorption per unit time and is given by the Arrhenius formula

$$\log W_{\rm des}=\log \Omega - E/2.303RT_{h.t.}, \tag{3}$$

where Ω is a frequency factor, E is the adsorption energy of CrO_4^{2-} on the surface at absolute temperature $T_{h.t.}$, and R is the gas constant.

The following values of the adsorption energy and frequency factor were obtained for the adsorption complexes I (radiation band at 620 nm) and II (radiation band at 690 nm): $E_{\rm I}=-88.18$ kJ·mole^{-1}, $\Omega_{\rm I}=6.39\times10^{-6}$ s^{-1} and $E_{\rm II}=-39.03$ kJ·mole^{-1}, $\Omega_{\rm II}=6.64\times10^{-3}$ s^{-1}. We note that the adsorption energies are of the same order of magnitude as the adsorption energies for different forms of adsorbed water and are approximately half the binding energies for ≡Si–OH groups.[21]

Therefore thermal activation of the surface is required for the formation of chemically bound adsorption complexes involving CrO_4^{2-} ions. Prior to thermal activation of the surface, the CrO_4^{2-} ions are secured to the surface by weak hydrogen bonds (physisorbed form). Their structure and local symmetry are essentially undistorted, and therefore the molecular ions do not radiate. Thermal activation results in the formation of active centers, with which the CrO_4^{2-} ions form a chemical bond (chemisorption of CrO_4^{2-} ions), on the surface. The local symmetry of the molecular ions is lowered with respect to the unperturbed ion, and this lifts the forbiddenness for radiative electronic transitions.

The presence of physically sorbed CrO_4^{2-} ions on the surface also affects the efficiency of dehydroxylation of the surface during heat treatment, i.e., the molecular ions catalyze the dehyroxylation process. Figure 4 displays the integrated intensity of the IR absorption band that peaks at 3748 cm^{-1} (free ≡Si–OH group) versus the preliminary heat-treatment temperature for pure SiO_2 and the adsorption system SiO_2–CrO_4^{2-}. For pure SiO_2, the destruction of the hydroxyl cover starts at $T_{h.t.}=900$ K,[22] while for the modified sorbent the onset of this process is shifted in the direction of lower temperatures by ~100 K (Fig. 4, curve *2*).

Thus, the CrO_4^{2-} ions stimulate dehyroxylation, and it is therefore logical to assume that they will form an adsorption complex with the products of the dehydroxylation reaction. The fact that the CrO_4^{2-} ions form a strong chemical bond with surface active centers is confirmed by the fact that the vibrational frequencies of CrO_4^{2-} ions as a whole relative to the surface of the dispersed SiO_2 appear in the long-wavelength IR absorption spectrum (Fig. 5). The frequencies 162 and 206 cm^{-1} (marked by arrows) correspond to such vibrations for adsorption complex I. This is in good agreement with low-temperature (4.2 K) investigations of luminescence spectra, where the frequency 170 cm^{-1} appeared.[8]

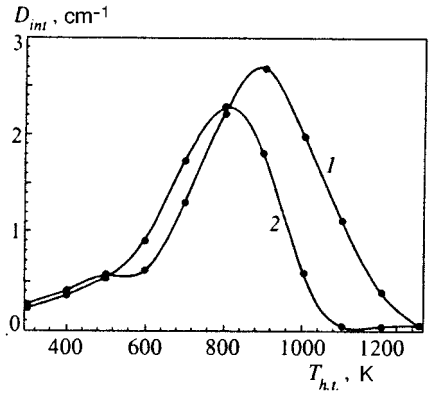

FIG. 4. Integrated optical density D_{int} of the 3748 cm^{-1} IR absorption band for pure dispersed SiO_2 (*1*) and the SiO_2–CrO_4^{2-} adsorption system (*2*) as a function of the heat-treatment temperature $T_{h.t.}$.

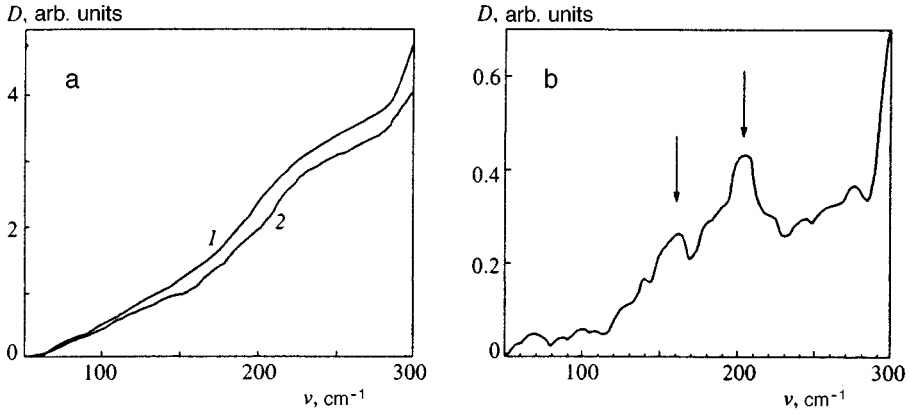

FIG. 5. a) IR absorption spectra for the SiO_2–CrO_4^{2-} adsorption system (*1*) and pure dispersed SiO_2 (*2*) ($T_{h.t.}=900$ K). b) Difference of the spectra *1* and *2*.

To determine the local symmetry of the CrO_4^{2-} ions in the I and II complexes, we also investigated the Raman scattering spectra in the region of the stretching vibrations of the molecular ions (Fig. 6). We selected samples with the same values of $T_{h.t.}$, where according to Fig. 3 type I ($T_{h.t.}=900$ K) and type II ($T_{h.t.}=1200$ K) adsorption complexes dominate. The Raman structure due to the formation of complexes with the local symmetry of the CrO_4^{2-} ions lowered from T_d to C_{3v} and C_{2v} was identified by the method described in Ref. 23. The results of the identification are presented in Table I.

According to these results, two types of complexes with C_{3v} local symmetry can be distinguished for samples prepared with $T_{h.t.}=900$ K and complexes of one type with C_{2v} local symmetry can be distinguished for samples prepared with $T_{h.t.}=1200$ K. We designate these adsorption complexes C'_{3v}-I, C''_{3v}-I, and C_{2v}-II, respectively. Note that the impurity Raman scattering efficiency was low for the $T_{h.t.}=900$ K samples. This resulted in the appearance of vibrational frequencies of the SiO_2 matrix in the spectrum. For this reason, the difference of the Raman spectra for the SiO_2–CrO_4^{2-} system and pure SiO_2 is presented in Fig. 5a. The impurity scattering efficiency in the $T_{h.t.}=1200$ K samples was so high that it was best to neglect Raman scattering by the SiO_2 matrix.

The effect of intense IR radiation ($\lambda=1.064$ μm) on the luminescence of adsorption complexes excited by UV radiation was demonstrated in a two-frequency experiment (Fig. 7a). The luminescence intensity of type-I complexes (C'_{3v}-I, C''_{3v}-I) decreases with increasing IR radiation dose, just as in the case of one-frequency IR excitation (Figs. 2b and 7b); the intensity of the radiation from IR excitation could be neglected, since the radiation from UV excitation was an order of magnitude higher. The emission band with a maximum at 690 nm (type-II complex (C_{2v}-II)) does not change appreciably under IR irradiation (Fig. 7b). Therefore the CrO_4^{2-} ions form chemically bound adsorption complexes at the surface with dehydroxylation reaction products. Two types of complexes with C_{3v} local symmetry (C'_{3v}-I and C''_{3v}-I) and complexes of one type with C_{2v} local symmetry (C_{2v}-II) are formed.

Low-temperature investigations (4.2 K) of the luminescence properties of adsorption complexes in the SiO_2–CrO_4^{2-} system have established[8] that a distinct electronic–vibrational structure of the luminescence spectra is only typical of emission bands with a maximum of 620 nm; this was attributed to the weak electron–phonon interaction in a complex. On the other hand, experiments on the luminescence properties of tetrahedral impurity molecular

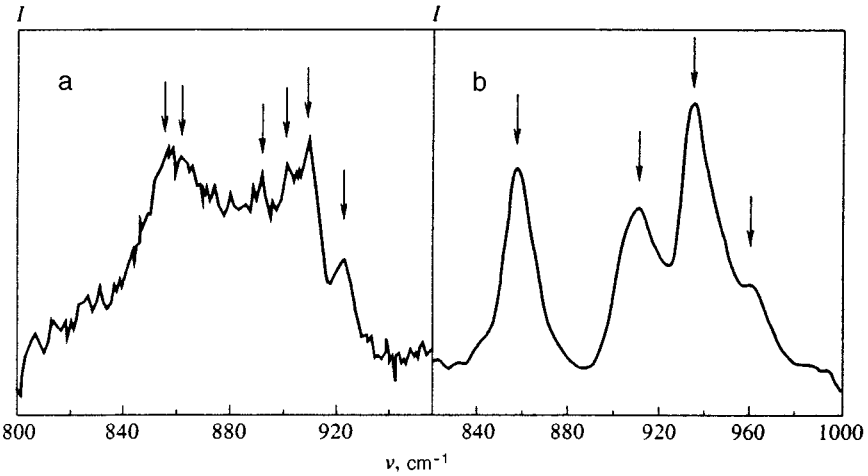

FIG. 6. Raman scattering spectra for the SiO_2–CrO_4^{2-} absorption system: a—$T_{h.t.}=900$ K; b—$T_{h.t.}=1200$ K. The arrows mark the positions of the vibrational peaks (see Table I).

TABLE I. Positions ν_i [cm^{-1}] of Raman bands and their interpretation for the SiO_2–CrO_4^{2-} adsorption system.

Type of center	Interpretation	$T_{h.t.}=900$ K	$T_{h.t.}=1200$ K
	$\nu_1(A_1)$	855	-
C'_{3v}-I	$\nu_3(A_1)$	900	-
	$\nu_3(E)$	922	-
	$\nu_1(A_1)$	861	-
C''_{3v}-I	$\nu_3(A_1)$	908	-
	$\nu_3(E)$	891	-
	$\nu_1(A_1)$	-	858
	$\nu_3(B_1)$	-	911
C_{2v}-II	$\nu_3(A_1)$	-	934
	$\nu_3(B_2)$	-	961

ions in crystals established[23] that complexes with the maximum possible local symmetry of the ions (C_{3v}) are characterized by the weakest electron–phonon interaction (the electronic–vibrational structure is clearly manifested in the luminescence spectra). Therefore, multiphoton absorption in the SiO_2–CrO_4^{2-} adsorption system is due to the formation of adsorption complexes with C_{3v} local symmetry.

As mentioned above, chromate ions form complexes with dehydration and dehydroxylation reaction products. We now examine in greater detail the thermal activation of the surface during heat treatment, basing the analysis on IR and luminescence spectroscopy data for pure and modified dispersed SiO_2.

As shown above (Fig. 4), the density of surface hydroxyl groups decreases at $T_{h.t.}=900$ K for pure SiO_2 and $T_{h.t.}=800$ K for the SiO_2–CrO_4^{2-} system. For pure SiO_2 this process is accompanied by an increase in the intensity of the 685 nm luminescence band (Fig. 8), which is typical of nonbridge oxygen atoms.[24,25] We note that the 620 and 690 nm emission bands for the SiO_2–CrO_4^{2-} adsorption system are much more intense than the 685 nm luminescence band for pure SiO_2. Therefore the contribution of this emission band in the case of an impurity system can be neglected. The antibatic character of the increase in the intensity of the 685 nm luminescence band for pure SiO_2 with respect to the total intensity of the IR absorption band, corresponding to free surface hydroxyl groups, attests to the occurrence of the process

$$\equiv\mathrm{Si{-}OH} \xrightarrow{T_{h.t.}} \equiv\mathrm{Si{-}O\bullet} + \mathrm{H}^0 \tag{4}$$

where three bars denote a triple bond, the black dot denotes an unpaired spin, and the superscript denotes the charge state of the atom.

The 510 nm emission band for the SiO_2–CrO_4^{2-} system (Fig. 3) also appears in the luminescence spectrum of pure SiO_2.[19] The nonmonotonic dependence of the intensity of this band on $T_{h.t.}$ is displayed in Fig. 8. This dependence can be explained by assuming that the 510 nm emission band is due to $\equiv$Si–H centers. Then its decrease as $T_{h.t.}$ increases to 900 K is a result of the reaction

$$\equiv\mathrm{Si{-}H} \xrightarrow{T_{h.t.}} \equiv\mathrm{Si\bullet} + \mathrm{H}^0. \tag{5}$$

However, after the density of atomic hydrogen increases as a result of the reaction (4), the reverse process

$$\equiv\mathrm{Si\bullet} + \mathrm{H}^0 \xrightarrow{T_{h.t.}} \equiv\mathrm{Si{-}H} \tag{6}$$

is also possible. This explains the simultaneous increase in the density of nonbridge oxygen atoms and $\equiv$Si–H centers at $T_{h.t.} \geqslant 1000$ K (Fig. 8). Dehyroxylation of the surface can also occur by the traditional channel[26]

$$2(\equiv\mathrm{Si{-}OH}) \xrightarrow{T_{h.t.}} (\equiv\mathrm{Si{-}O{-}Si}\equiv) + \mathrm{H_2O} \tag{7}$$

or

$$2(\equiv\mathrm{Si{-}OH}) \xrightarrow{T_{h.t.}} \equiv\mathrm{Si{-}O\cdot} + \equiv\mathrm{Si\bullet} + \mathrm{H_2O}. \tag{8}$$

The destruction of $\equiv$Si–H centers can also occur in the process[27]

$$2(\equiv\mathrm{Si{-}H}) \xrightarrow{T_{h.t.}} (\equiv\mathrm{Si{-}Si}\equiv) + \mathrm{H_2}, \tag{9}$$

where ($\equiv$Si–Si$\equiv$) is an oxygen vacancy. According to Ref. 28, for ''wet'' glassy SiO_2, whose spectroscopic characteris-

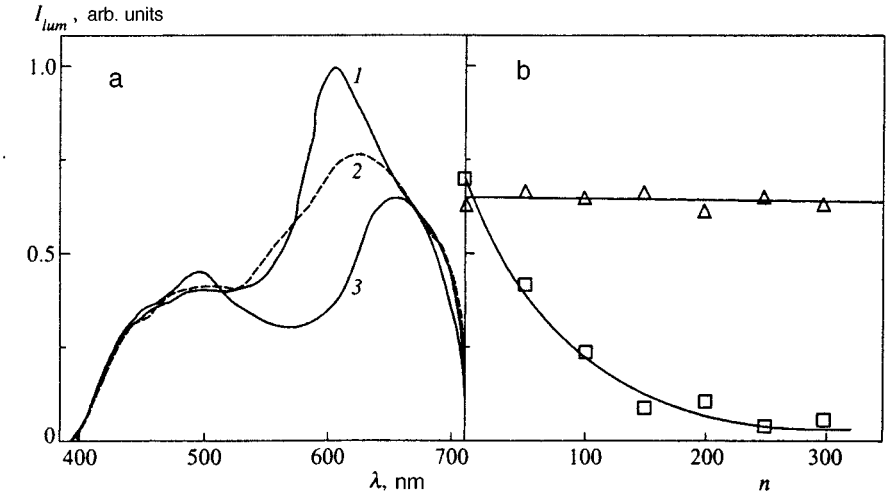

FIG. 7. a) Luminescence spectra for the SiO_2–CrO_4^{2-} absorption system ($T_{h.t.}=900$ K) with $\lambda_{exc}=266$ nm as a function of the IR irradiation dose ($\lambda_{exc}=1.064$ μm, $I=5\times10^7$ W·cm^{-2}): *1*—0, *2*—50, *3*—300 pulses. b) Intensities of the 620 (□) and 690 nm (△) luminescence bands versus the IR irradiation dose.

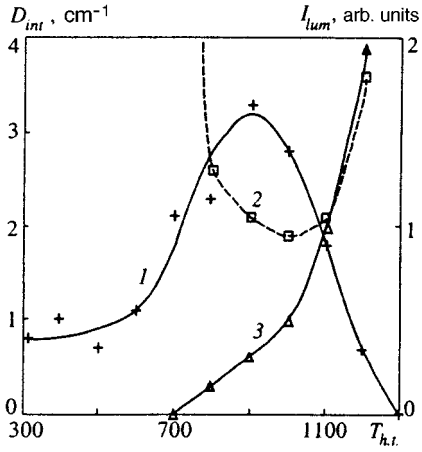

FIG. 8. Integrated optical density of the 3748 cm^{-1} IR absorption band (*1*), intensities of the 510 (*2*) and 680 nm (*3*) luminescence bands for pure dispersed SiO_2 as a function of $T_{h.t.}$.

tics are closest to those of dispersed silicon dioxide, heat treatment releases H_2 and H_2O molecules. Therefore the processes (7)–(9) as well as a combined reaction of the type

$$(\equiv\text{Si–OH})+(\equiv\text{Si–H}) \xrightarrow{T_{h.t.}} (\equiv\text{Si–Si}\equiv)+\text{H}_2\text{O} \tag{10}$$

must be considered first and foremost.

In summary, dehydration and dehyroxylation of the surface during heat treatment of dispersed SiO_2 result in the appearance of centers with unpaired spins as well as oxygen vacancies, which are characterized by a high reactivity,[21] on it. We regard such centers as active in the adsorption of CrO_4^{2-} ions. The two electrons of the chromate ion that define its charge state occupy a nonbonding t_1 orbital.[29] This orbital has the same energy as an atomic orbital in an oxygen atom (atomic type, p_x or p_y). In my opinion, the interaction of these orbitals with hybrid sp^3 orbitals, which form an oxygen vacancy, results in the formation of an adsorption complex with C_{3v} local symmetry of the molecular ion. Obviously there can be two types of such complexes (Fig. 9). The interaction of the t_1 orbitals of the CrO_4^{2-} ion with the p_x or p_y orbitals of two nonbridge oxygen atoms results in the formation of a complex with C_{2v} local symmetry of the ion. We note that in the case of both the type-I complexes (C'_{3v}-I) and type-II complexes, the interacting orbitals are characterized by identical symmetry properties. This maximizes their overlap and gives rise to the formation of a π-type adsorption bond.

The characteristic features of the interaction of intense IR laser radiation with the SiO_2–CrO_4^{2-} adsorption system can be understood on the basis of the proposed models of adsorption complexes. For example, the excitation of luminescence is due to multiphoton ionization of molecular ions and oxygen vacancies. The ''fading'' of the luminescence is due to laser annealing of the vacancies (restoration of the $\equiv$Si–O–Si$\equiv$ bond), which are stabilized by CrO_4^{2-} ions. The parameter N in Eq. (2) characterizes the stabilization level of the vacancies. The characteristic dips in the power dependences are, in my opinion, due to fragmentation of the CrO_4^{2-} ion accompanying saturation of a multiphoton transition. Obviously, for laser radiation intensity corresponding to the position of the dips, the rate of ''healing'' of the oxygen vacancies should increase rapidly, as is in fact observed experimentally.

The existence of two dips attests to the existence of several channels for fragmentation of CrO_4^{2-} ions. We also note that complete annealing of the surface vacancies can be achieved under high laser doses ($\sim$400 pulses), after which weak wide-band luminescence correlated with multiphoton generation of elementary electronic excitations in particles of dispersed silicon dioxide is observed in the luminescence spectrum under intense IR laser excitation.[30]

3.3. Energy structure of adsorption complexes and nature of radiative electronic transitions

The luminescence of CrO_4^{2-} ions was discovered comparatively recently, so before examining the characteristic features of multiphoton absorption in the SiO_2–CrO_4^{2-} adsorption system and the excitation-energy relaxation chan-

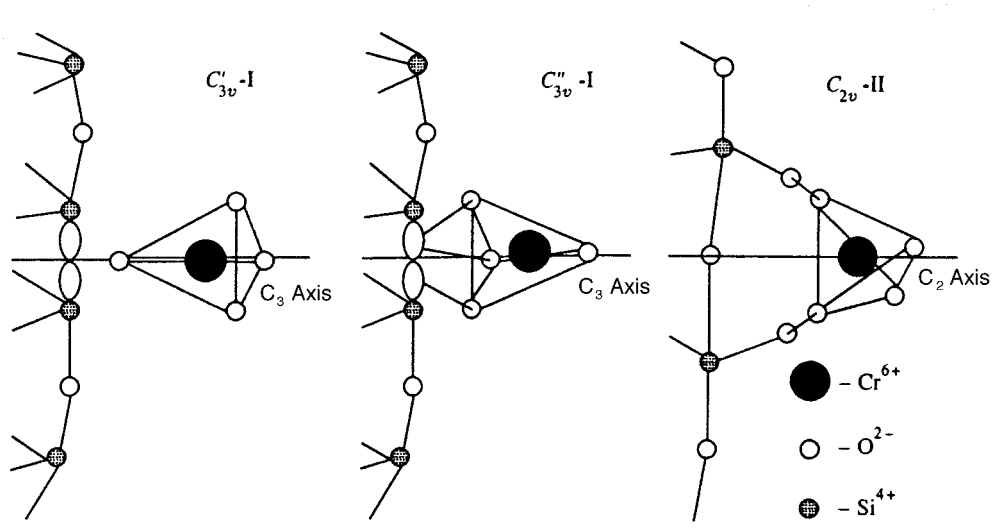

FIG. 9. Proposed models of adsorption complexes on the surface of dispersed SiO_2.

nels, we examine the energy structure of the adsorption complexes and the nature of radiative transitions in them.

The low-temperature (4.2 K) luminescence of the chromate salt $CaCrO_4$ and of the $KBr–K_2CrO_4$ impurity system was first observed in Ref. 31. The weak and structureless luminescence bands at ~620 and ~650 nm with decay constants ~100 μs and ~10 ms for $CaCrO_4$ and $KBr–K_2CrO_4$, respectively, were referred to the spin- and symmetry-forbidden electronic transition $^3T_1 \rightarrow {}^1A_1$ in the CrO_4^{2-} molecular ion, which has T_d symmetry. The salt K_2CrO_4 and the $MgO–K_2CrO_4$ impurity system did not luminesce.

To determine the characteristic features of the radiative dynamics in the CrO_4^{2-} ion, a number of inorganic crystals doped with chromate ions were investigated. As is well known, molecular ions in inorganic crystals are convenient model systems for investigating the spectra of impurity centers and their interaction with their environment.[32] For example, luminescence spectra with a distinct vibronic structure were first obtained for the $CsCaCl_3–Na_2CrO_4$ impurity system[33,34] and then also for alkali–halide crystals doped with chromate salts.[23,35]

Luminescence spectra with vibronic structure have also been observed for O_3CrO^- and CrO_3Cl^- complexes in different salts, where $Cr_2O_7^{2-}$ ions were present,[36,37] but the nature of the luminescence centers was not definitively established. A comparative analysis of the vibrational spectra of impure crystals and their radiative characteristics[23,38] established that the luminescence is due to electronic transitions in CrO_4^{2-} ions. Furthermore, the differences in the spectroscopic and kinetic characteristics for the luminescence spectra of different chromate salts and impure crystals have a logical explanation on the basis of the concept that several types of complexes with impurity and intrinsic defects in crystals form a molecular ion. The formation of complexes results in a lowering of the local symmetry of an ion and, as already mentioned, lifting of the forbiddenness for radiative electronic transitions. The triplet nature of the red luminescence for $CaCrO_4$ and $SrCrO_4$ salts and a water solution of K_2CrO_4 was checked in optically detected magnetic resonance experiments[39] and in kinetic experiments.[40] Here we present another experimental fact confirming the triplet nature of luminescence transitions in the CrO_4^{2-} ion.

It is well known that nonradiative transitions between energy states of different multiplicity are due to the spin–orbit mechanism.[41] On account of the spin selectivity of the relaxation processes, the distribution of population among the spin sublevels is ordinarily different from a Boltzmann distribution.[42] This feature of triplet states should obviously be manifested in nonradiative excitation energy relaxation processes.

The temperature dependence of the relative quantum yield of luminescence for adsorption complexes of the types C_{3v}'-I, C_{3v}''-I, and C_{2v}-II is presented in Fig. 10. One typical feature of such temperature dependences is a sharp drop in intensity at low temperatures (4.2–15 K), followed by stabilization of the yield and another decrease at higher temperatures. The low-temperature drop in the luminescence intensity is due to the low activation barrier for nonradiative

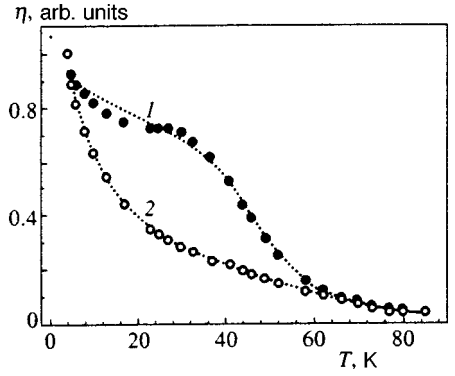

FIG. 10. Temperature dependences of the quantum yield at 620 (*1*) and 690 nm (*2*) for the $SiO_2–CrO_4^{2-}$ adsorption system: $\lambda_{exc}=377$ nm.

transitions, to which it is logical to attribute the spin–lattice relaxation between the spin sublevels of the triplet state. The stabilization of the yield in the temperature range 15–30 K is due to the spin selectivity of relaxation processes. We note that on account of the low density of excited centers, the weakness of the spin–orbit interaction, and the large splitting between the triplet levels and the singlet levels, the conventional Waller and van Vleck spin–phonon coupling mechanisms are negligible.

We now examine radiative and nonradiative transitions in adsorption complexes in greater detail. According to a group-theoretic analysis,[8] the C_{3v}'-I and C_{3v}''-I type complexes are characterized by an e^2 ground state and $e^1 2e^1$ and $e^1 e^{*1}$ excited-state configurations. The following types of wave-function symmetries were obtained for the e^2 ground-state configuration: $^1A_1$, ^1E, and $^3A_2$. For the $e^1 2e^1$ excited state configuration the wave-function symmetries are $^1A_1$, $^1A_2$, ^1E, $^3A_1$, $^3A_2$, and ^3E. The $e^1 e^{*1}$ excited-state configuration is not characterized by equilibrium states, and therefore there are no symmetry types for it (this corresponds to a transition from a bonding e orbital to an antibonding e^* orbital of the adsorption complex). Therefore the orbital symmetry of the excited state with the lowest energy is E.

The types of symmetry of the total wave functions of the triplet spin sublevels can be easily obtained by taking into account the symmetry of the spin wave functions (A_2, E (Ref. 43)): $A_1(z)$, $E(x,y)$, where the symbols in parentheses indicate the polarization of the radiative transitions. Since the spin–orbit operator mixes states possessing identical total symmetry,[44] the nearest singlet state of the excited configuration ^1E will interact only with the lowest spin sublevels $E(x,y)$. This fact gives rise to spin selectivity of both nonradiative and radiative intercombination transitions. Using the selection rules for radiative transitions,[43] it can be shown that all three spin sublevels of the excited triplet state with the lowest energy must be active for radiative and nonradiative transitions. However, the probability of the radiative transitions $E(x,y) \rightarrow {}^1A_1$ will be much higher than that of the transitions $A_1(z) \rightarrow {}^1A_1$; as already mentioned above, this is due to the spin selectivity of the intercombination transitions.

Therefore two processes will participate in the relaxation of the excitation energy of the lowest triplet level. One is

associated with spin–lattice relaxation and will dominate at low temperatures. The other is the standard relaxation process associated with the excitation of the phonon subsystem of the matrix on account of the electron–phonon interaction and appears at higher temperatures.

As is well known,[45] in a direct one-phonon process the dependence of the spin–lattice relaxation rate on the energy splitting between the spin sublevels $|i\rangle$ and $|j\rangle$ has the form

$$W_{ij} \sim \frac{(\hbar\omega_{ij})^3}{1-\exp(-\hbar\omega_{ij}/k_BT)}, \tag{11}$$

where k_B is Boltzmann's constant and T is the absolute temperature. The relaxation process associated with the excitation of phonons is conventionally taken into account by the Mott formula

$$\eta = \frac{\eta_0}{1+a\ \exp(-E_A/k_BT)}, \tag{12}$$

where η_0 is the quantum yield at $T=0$, a is a frequency factor, and E_A is the activation energy. Therefore the temperature dependence of the relative quantum yield for such a system can be written in the form

$$\eta = \eta_0\left[1+a\ \exp\left(-\frac{E_A}{k_BT}\right)+\frac{b(\hbar\omega_{ij})^3}{1-\exp(-\hbar\omega_{ij}/k_BT)}\right]^{-1}, \tag{13}$$

where the constants a and b determine the contributions of the two indicated relaxation processes to the overall relaxation process.

As one can see from Fig. 10, the temperature dependence of the relative quantum yield can be satisfactorily approximated by Eq. (13). Then $E_A = 247\ \text{cm}^{-1}$ and $\hbar\omega_{ij} = 2.6\ \text{cm}^{-1}$ for complex I and $E_A = 279\ \text{cm}^{-1}$ and $\hbar\omega_{ij} = 5.0\ \text{cm}^{-1}$ for complex II. Comparing the values obtained for E_A with the theoretically computed spectrum of the photon density of states in quartz glass,[46] it can be stated that for complex I the activation barrier for nonradiative transitions is associated with the excitation of silicon vibrations, and for complex II it is associated with deflections of a nonbridge oxygen atoms.

The theory of spin–lattice relaxation in local triplet-excited centers of organic crystals can be used to interpret the low values of $\hbar\omega_{ij}$.[47–49] According to this theory, acoustic-phonon-induced mixed translational–rotational motions of molecules are responsible for spin–lattice relaxation. In our case, however, in which the lowest components of a triplet are degenerate with energy $E(x,y)$, thermal activation of the component $|y\rangle$ will cause the CrO_4^{2-} ion to rotate by 90° (the x and y orbitals are orthogonal), while the complex as a whole possesses C_{3v} symmetry (the complex transforms into itself under a rotation by 60° around the $C_3(z)$ axis), i.e., the thermal population of the component $|y\rangle$ will result in a modulation of the fine structure of the levels and transitions between the magnetic sublevels with the emission of resonance phonons.

In summary, the experimental results presented in this section can be explained on the basis of the triplet nature of the lowest excited electronic state of the CrO_4^{2-} ion. We note that previous investigations of triplet excitations were performed exclusively for pure and impure molecular crystals.[50] For this reason, the present results extend the range of objects suitable for such investigations.

3.4. Characteristic features of multiphoton absorption in the SiO_2–CrO_4^{2-} adsorption system and excitation-energy relaxation channels

Analysis of the experimental results shows that the luminescence response of the system is a result of the recombination between the photoelectron and ionized chromate ion (the emission band at 640 nm) and an ionized oxygen vacancy (luminescence band at 540 nm). In the general case the luminescence intensity can be represented as

$$I_{\text{lum}} = N_s W_i^{(\gamma)} \tau_t^{-1} \int_\omega I_{\text{lum}}(\omega)\, d\omega, \tag{14}$$

where N_s is the density of adsorption complexes, τ_t^{-1} is the probability that a photoelectron is captured by an ionized center, and $I_{\text{lum}}(\omega)$ is the intensity of the intrinsic luminescence of a CrO_4^{2-} ion excited in a recombination process. Then the excitation of luminescence can be represented in the form

$$(\equiv\text{Si}-\text{Si}\equiv)-\text{CrO}_4 \xrightarrow{5\hbar\omega} \equiv\text{Si}-\text{Si}\equiv + (\text{CrO}_4^{2-})^+ + e^-, \tag{15}$$

$$\equiv\text{Si}-\text{Si}\equiv \xrightarrow{7\hbar\omega} \equiv\text{Si}\bullet + \equiv\text{Si}^+ + e^-. \tag{16}$$

Here we took account of the fact that the molecular ions form a strongly polar π-type adsorption bond, and the two electrons occupying a t_1 nonbonding molecular orbital that form this bond also participate in the absorption of radiation.[8]

The antibonding component of the π-bond lies ~ 3.5 eV above the bonding component.[8] The total energy of the three laser photons likewise equals ~ 3.5 eV. Therefore we assume that multiphoton absorption followed by ionization proceeds under conditions of excitation of a three-photon resonance. Here the nonbonding state of the adsorption complex appears as a resonance level.

The presence of this antibonding state is, in my opinion, a decisive factor in the multiphoton ionization process. For this reason, desorbed molecular ions, as well as ions that are part of a type-II complex, do not participate in multiphoton absorption. We also note that according to Ref. 51 the absorption band at 7.6 eV for highly pure quartz glasses, whose properties are closest to those of dehydrated dispersed silicon dioxide, is due to $\equiv$Si–Si$\equiv$ bonds.

In our case the total seven-photon absorption energy is ~ 8.15 eV, thereby confirming the luminescence excitation scheme proposed above. Now it becomes understandable that annealing of the surface oxygen vacancies (restoration of the silicon–oxygen–silicon bridge bonds) is possible only under the processes (15) and (16). We also note that two-photon ionization of the oxygen vacancies under excitation by an excimer laser was observed in Refs. 51–55.

The luminescence response of the adsorption system can be represented as being the result of the processes

$$\equiv\mathrm{Si}\bullet+(\equiv\mathrm{Si}^{+}+e^{-})\xrightarrow{540\ \mathrm{nm}}\equiv\mathrm{Si}-\mathrm{Si}\equiv, \tag{17}$$

$$(\mathrm{CrO}_4^{2-})^{+}+e^{-}\rightarrow(\mathrm{CrO}_4^{2-})^{*}\xrightarrow{640\ \mathrm{nm}}\mathrm{CrO}_4^{2-}, \tag{18}$$

$$\equiv\mathrm{Si}-\mathrm{Si}\equiv+\mathrm{CrO}_4^{2-}\rightarrow(\equiv\mathrm{Si}-\mathrm{Si}\equiv)-\mathrm{CrO}_4, \tag{19}$$

where the asterisk denotes an excited state. As already mentioned above, the formation of an adsorption complex requires thermal activation of the surface of the dispersed silicon dioxide. Multiphoton ionization destroys type-I complexes, resulting in the formation of photoions and photoelectrons. Photoions are retained on the surface by the weaker hydrogen bonds, i.e., multiphoton absorption changes the type of interaction between a molecular ion and the surface of the dispersed oxide. The fixation of a photoion on the surface increases the probability of recombination of the ion with a photoelectron, followed by excitation and deexcitation, after which the adsorption bond is restored in the type-I complex. The processes described above comprise the feature that distinguishes multiphoton ionization of molecular ions adsorbed on the surface of wide-gap dispersed materials from multiphoton ionization in the gaseous state.

We now turn to an analysis of the dips in the power dependences. As already mentioned, these dips are due to a rapid increase in the probability of annealing of surface oxygen vacancies. In my opinion, this process is due to fragmentation of CrO_4^{2-} ions accompanying saturation of a multiphoton transition. In the process, the density of luminescing centers decreases because only CrO_4^{2-} ions bound to the surface of an adsorption complex can luminesce and participate in multiphoton absorption. However, as the density of type-I complexes decreases, a photoelectron can absorb an additional photon from the laser field, as explained by the following model.

A photoelectron passes into a superexcited state in a multiphoton absorption process. There exist two channels for relaxation of the excitation energy in this superexcited state—autoionization and fragmentation. As is well known,[54] superexcited states are characterized by large electronic orbitals ($\sim$10 nm). I therefore suggest that the interaction between these orbitals in a system of adsorption complexes results in mutual perturbation of the orbitals, which stimulates autoionization. Ionization therefore results from multiphoton absorption and electronic perturbation between complexes in their superexcited states. For this reason, when the density of luminescing centers decreases, the intercomplex electronic perturbation parameter becomes too small to stimulate autoionization, and the intensity of the corresponding luminescence response of the system decreases. However, a photoelectron can still be transferred to the next higher (in energy) superexcited state by the absorption of an additional photon (increase in the size of the electronic orbital), where the intercomplex electronic perturbation parameter will be large enough to stimulate autoionization.

This model explains the unit increase in the photonicity of the processes after each dip in the power dependence. The scheme of multiphoton luminescence excitation accompanying fragmentation of a molecular ion can be represented as follows:

$$(\equiv\mathrm{Si}-\mathrm{Si}\equiv)-\mathrm{CrO}_4\xrightarrow{5\hbar\omega}\equiv\mathrm{Si}-\mathrm{Si}\equiv+\mathrm{O}^{2-}+\mathrm{CrO}_3\rightarrow\equiv\mathrm{Si}-\mathrm{O}-\mathrm{Si}\equiv+\mathrm{CrO}_3, \tag{20}$$

$$(\equiv\mathrm{Si}-\mathrm{Si}\equiv)-\mathrm{CrO}_4\xrightarrow{6\hbar\omega}\equiv\mathrm{Si}-\mathrm{Si}\equiv+2\mathrm{O}^{2-}+\mathrm{CrO}_2^{2+}\rightarrow\equiv\mathrm{Si}-\mathrm{O}-\mathrm{Si}\equiv+\mathrm{O}^{2-}+\mathrm{CrO}_2^{2+}. \tag{21}$$

The luminescence response of the system obviously depends on the lifetime of a photoelectron in the conduction band,[55]

$$\Delta\tau=(v\sigma_t n_s)^{-1}, \tag{22}$$

where v is the velocity of the photoelectron, σ_t is the cross section for photoelectron capture by an ionized center, and n_s is the density of ionized centers. For the gas-kinetic case, which is a good approximation to such a system, the capture cross section is given by

$$\sigma_t=\pi R_0^2, \tag{23}$$

where R_0 is the interaction length between an ionized center on the surface and a photoelectron. The interaction length is ordinarily found by equating the kinetic energy of a free electron to the interaction potential energy between the electron and an ionized center,

$$\frac{1}{2}mv^2=\frac{e^2}{R_0}, \tag{24}$$

where m and e are the electron mass and charge. Then Eqs. (23) and (24) yield

$$\sigma_t=\pi\left(\frac{2e^2}{mv^2}\right)^2, \tag{25}$$

yielding

$$\Delta\tau=\frac{m^2v^3}{4\pi n_s e^4}\propto v^3. \tag{26}$$

The velocity of a photoelectron can be represented as the sum

$$v^2=v_0^2+\frac{2}{m}\mathscr{E}_{\max}, \tag{27}$$

where v_0 is the initial velocity of the photoelectron after photoionization. This velocity does not depend on the intensity of the laser radiation, and is solely determined by the parameters of the photoionized system:[3]

$$v_0^2=\frac{2}{m}[k\hbar\omega-E_{\mathrm{ion}}-(I_i-E_{\mathrm{res}})], \tag{28}$$

where I_i is the ionization potential, E_{res} is the energy of the resonance excited state, E_{ion} is the vibrational energy of the ionic state, $\hbar\omega$ is the laser photon energy, and $k=2$ is the

number of laser photons required for ionization from the resonance state. In (27), $\mathscr{E}_{\max}$ is the maximum energy of an electron heated by the laser radiation,[56]

$$\mathscr{E}_{\max} \approx \frac{M}{m} \frac{I\lambda^2 r_e}{2\pi c} \propto I, \tag{29}$$

where r_e is the classical electron radius, λ is the wavelength of the laser radiation of intensity I, M is the mass of the ion or regular surface atoms, and c is the speed of light. Then (26) and (27) yield

$$\Delta\tau \propto \left(v_0^2 + \frac{2}{m} \mathscr{E}_{\max} \right)^{3/2}. \tag{30}$$

It is well known[54] that $v_0 \sim 0$ for vibrationally autoionized photoelectrons. Therefore $v_0^2 \ll 2\mathscr{E}_{\max}/m$ and (30) can be rewritten in the form

$$\Delta\tau = \Delta\tau_1 + \Delta\tau_2, \tag{31}$$

where

$$\Delta\tau_1 \propto \left(\frac{2}{m} \mathscr{E}_{\max} \right)^{3/2} \propto I^{3/2}, \tag{32}$$

$$\Delta\tau_2 \propto \frac{3}{2} v_0^2 \left(\frac{2}{m} \mathscr{E}_{\max} \right)^{1/2} \propto I^{1/2}. \tag{33}$$

Note that (32) describes most accurately the dependence of the time delay of the luminescence response of the adsorption system on the intensity of the laser radiation (Fig. 1c); this confirms the validity of the condition $v_0 = 0$. However, after the first dip, the time delay also increases as a power-law function, but with the exponent ~ 4. This is probably due to the nonlinear dependence of $\mathscr{E}_{\max}$ on laser intensity, which is possible in the case of strong heating of free electrons in a laser field.[56]

3.5. Rate equations for multiphoton resonance ionization of adsorption complexes

In the present section we examine the rate equations for 5–7-photon ionization of type-I complexes excited under three-photon resonance conditions. As noted above, multiphoton absorption of laser radiation can transfer the system of adsorption complexes into a multitude of superexcited states in which the complexes can autoionize, dissociate, or absorb an additional laser photon. Therefore ionization of CrO_4^{2-} ions followed by excitation of the ions in a recombination process and radiation emission will depend on the nonradiative relaxation rate of the resonance state, as well as on the fragmentation rate of the molecular ions in the superexcited states.

Photoionization of molecular ions requires at least five laser photons. I therefore assume that resonance level 1 (Fig. 11) is populated from the ground state 0 at a rate $\sigma_1^{(3)} I^3$, and ionizes at a rate $\sigma_2^{(2)} I^2$. Nonradiative relaxation of the resonance state at rate γ_1 efficiently depletes this level, since the level is due to the antibonding orbital of the adsorption complex, which is typified by nonactivational nonradiative processes. In the case of six-photon ionization (transition to the next superexcited state at the rate $\sigma_3^{(3)} I$), the state excited in

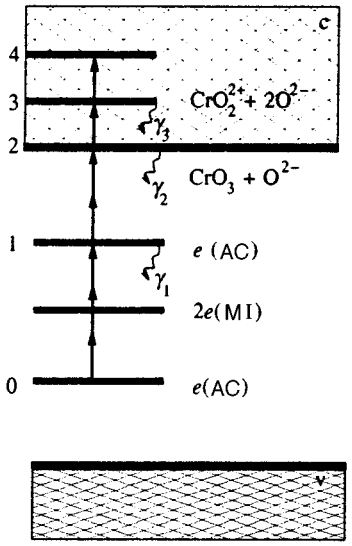

FIG. 11. Energy level scheme for the SiO_2–CrO_4^{2-} adsorption system and observed multiphoton ionization–fragmentation processes (AC—adsorption complex, MI—molecular ion).

a five-photon process plays the role of a resonant state. The relaxation rate γ_2 of this process is a total constant characterizing all loss mechanisms, including dissociation. For the seven-photon process (transition to the next superexcited state at a rate $\sigma_4^{(1)} I$), there are three resonance levels: an antibonding state of the adsorption complex and two superexcited states with relaxation rates γ_1 and γ_2, γ_3.

Thus, to describe the variation of the intensity of the luminescence of a CrO_4^{2-} ion as a function of laser power density, we model the process as seven-photon ionization with three intermediate states, of which one is an antibonding state of the adsorption complex and the other two are superexcited states of the molecular ion. The time-dependent behavior of the system is described by the system of equations

$$\begin{aligned}
\frac{dN_0(t)}{dt} &= \sigma_1^{(3)} I^3 [N_1(t) - N_0(t)], \\
\frac{dN_1(t)}{dt} &= \sigma_1^{(3)} I^3 [N_0(t) - N_1(t)] - \sigma_2^{(2)} I^2 N_1(t) \\
&\quad - \gamma_1 N_1(t), \\
\frac{dN_2(t)}{dt} &= \sigma_2^{(2)} I^2 [N_1(t) - N_2(t)] - \sigma_3^{(1)} I N_2(t) - \gamma_2 N_2(t), \\
\frac{dN_3(t)}{dt} &= \sigma_3^{(1)} I [N_2(t) - N_3(t)] - \sigma_4^{(1)} I N_3(t) - \gamma_3 N_3(t), \\
\frac{dN_4(t)}{dt} &= \sigma_4^{(1)} I N_3(t),
\end{aligned} \tag{34}$$

where $N_n(t)$ is the population of the corresponding levels (Fig. 11). This system of equations simplifies in the case of five-photon ionization:

$$\frac{dN_0(t)}{dt}=\sigma_1^{(3)}I^3[N_1(t)-N_0(t)],$$

$$\frac{dN_1(t)}{dt}=\sigma_1^{(3)}I^3[N_0(t)-N_1(t)]-\sigma_2^{(2)}I^2N_1(t)-\gamma_1N_1(t),$$

$$\frac{dN_2(t)}{dt}=\sigma_2^{(2)}I^2N_1(t), \tag{35}$$

where $N_2(t)$ is the number of ionized chromate ions in the five-photon ionization process.

The lifetime of the adsorption complex in the resonance level, which is its antibonding state, will be of the order of the vibrational relaxation time. Therefore $\gamma_1\sim10^{13}$ s^{-1}, which according to Ref. 53 suggests quasistationary variation of the population of the resonance level $(dN_1(t)/dt=0)$. The solution of the system of equations (35) will be

$$\frac{dN_2(t)}{dt}=\frac{\sigma_1^{(3)}\sigma_2^{(2)}I^5N_0(t)}{\sigma_1^{(3)}I^3+\sigma_2^{(2)}I^2+\gamma_1}. \tag{36}$$

In the absence of saturation of two- and three-photon transitions, taking account of the fact that the typical cross sections of such transitions are $\sim10^{-50}$ cm$^4\cdot$s and $\sim10^{-82}$ cm$^6\cdot$s^2, respectively,[3] and the intensity of the laser radiation $\sim10^{27}$ cm$^{-2}\cdot$s^{-1}, it can be assumed that $\sigma_1^{(3)}I^3\ll\gamma_1$ and $\sigma_2^{(2)}I^2\ll\gamma_1$. Then the number of ionized chromate ions over the time of the laser pulse in the process of five-photon ionization is

$$N_2=\int_0^{\tau_p}\frac{\sigma_1^{(3)}\sigma_2^{(2)}I^5N_0dt}{\gamma_1}=\sigma_1^{(3)}\sigma_2^{(2)}I^5N_0\,\frac{\tau_p}{\gamma_1}. \tag{37}$$

As already mentioned, however, when the laser intensity reaches $\sim8\times10^7$ W$\cdot$cm^{-2}, the five-photon ionization process saturates. In my opinion, this is due to saturation of the second step ($\sigma_1^{(3)}I^3\ll\sigma_2^{(2)}I^2$ and $\gamma_1\ll\sigma_2^{(2)}I^2$), since it is impossible to saturate the first step because of the antibonding nature of the resonance level. With this in mind, (36) can be rewritten in the form

$$\frac{dN_2(t)}{dt}=\sigma_1^{(3)}I^3N_0(t), \tag{38}$$

and therefore

$$N_2=\sigma_1^{(3)}I^3N_0\tau_p. \tag{39}$$

The second step saturates with the transition from (37) to (39) at the point with intensity $I_{\mathrm{sat}(1)}^2=\gamma_1/\sigma_2^{(2)}$, where $I_{\mathrm{sat}(1)}$ is the laser intensity before the first dip in the power dependence. Hence the cross section for two-photon ionization from a resonance level of the adsorption complex becomes

$$\sigma_2^{(2)}=\frac{\gamma_1}{I_{\mathrm{sat}(1)}^2}. \tag{40}$$

Arguing similarly and using the general system of rate equations (34), we can obtain formulas for the number of ionized molecular ions in the case of six- and seven-photon ionization. If there is no saturation of the transitions, we have, respectively,

$$N_3=\sigma_1^{(3)}\sigma_2^{(2)}\sigma_3^{(1)}I^6N_0\,\frac{\tau_p}{\gamma_1\gamma_2}, \tag{41}$$

$$N_4=\sigma_1^{(3)}\sigma_2^{(2)}\sigma_3^{(1)}\sigma_4^{(1)}I^7N_0\,\frac{\tau_p}{\gamma_1\gamma_2\gamma_3}. \tag{42}$$

In deriving Eqs. (41) and (42), high rates of nonradiative deactivation were assumed, but this time for the resonance levels 2 (for six-photon ionization) and for 2, 3 (for seven-photon ionization) on account of the possibility of fast dissociation. The saturation of multiphoton transitions for these cases is governed by the saturation of the last excitation steps, since they are due to one-photon transitions. Therefore, in the case of saturation, (41) and (42) become

$$N_3=\sigma_1^{(3)}\sigma_2^{(2)}I^5N_0\,\frac{\tau_p}{\gamma_1}, \tag{43}$$

$$N_4=\sigma_1^{(3)}\sigma_2^{(2)}\sigma_3^{(1)}I^6N_0\,\frac{\tau_p}{\gamma_1\gamma_2}, \tag{44}$$

which is equivalent to the case of five- or six-photon ionization, respectively, in the absence of saturation. As in the case of five-photon ionization, saturation of the last steps for six- and seven-photon ionization occurs with the transitions from (41) to (43) and (42) to (44) at points with intensities $I_{\mathrm{sat}(2)}=\gamma_2/\sigma_3^{(1)}$ and $I_{\mathrm{sat}(3)}=\gamma_3/\sigma_4^{(1)}$, respectively, where $I_{\mathrm{sat}(2)}$ and $I_{\mathrm{sat}(3)}$ are the laser intensities before the first dip and at maximum power density, where, in my opinion, onset of saturation for the fourth excitation step occurs. We obtain for the cross sections of the one-photon transitions

$$\sigma_3^{(1)}=\gamma_2/I_{\mathrm{sat}(2)}, \tag{45}$$

$$\sigma_4^{(1)}=\gamma_3/I_{\mathrm{sat}(3)}. \tag{46}$$

Note that since the luminescence intensity is proportional to the number of ionized molecular ions, Eqs. (37) and (41), (42) describe the dependence of the intensity of the luminescence response on the laser power density under five-, six-, and seven-photon ionization, respectively.

Using (40), (45), and (46), we now estimate the cross sections for multiphoton ionization at nonresonance steps. Since the photoionization efficiency in this case is determined by the relaxation rates of the intermediate resonance levels, the validity of the estimates will depend on the chosen values of γ_1, γ_2, and γ_3. As already mentioned, γ_1 can be set equal to the vibrational relaxation rate, $\gamma_1\sim10^{13}$ s^{-1}. The values of γ_2 and γ_3 can be estimated as the average dissociation rate of OsO_4 and SF_6 molecules in an infrared laser field: $\sim10^8-10^9$ s^{-1}.[57] The estimates are presented in Table II.

4. DISCUSSION

In the present multiphonon ionization experiment the radiation intensity is the main variable in tuning the laser to resonance. In five-photon ionization the process occurs in a one-frequency field, with one resonance step and two-photon

TABLE II. Saturation number N_s in the power dependence corresponding to laser intensity $I_{\mathrm{sat}(N_s)}$ and cross section σ for two- and one-photon transitions.

N_s	$I_{\mathrm{sat}(N_s)}$	$\sigma_2^{(2)}$, cm^4·s	$\sigma_3^{(1)}$, cm^2	$\sigma_4^{(1)}$, cm^2
1	7.5×10^7	7.1×10^{-41}	-	-
2	1.4×10^8	-	10^{-18}–10^{-19}	-
3	2.4×10^8	-	-	10^{-18}–10^{-19}

excitation at the second nonresonance step. In six-photon ionization the process proceeds with two resonance steps (three and two photons, respectively), and one-photon excitation at the third nonresonance step. In seven-photon ionization there are three resonance steps (three and two photons and one photon, respectively) and a one-photon excitation at the fourth nonresonance step.

Note that the resonance nature of the second and third excitation steps in the case of six- and seven-photon ionization is conditional, since the energy levels of the superexcited states are determined by the energy of the laser photon. However, as noted above, the superexcited states are characterized by the total relaxation rate constant, including dissociation of the molecular ions in these states, i.e., the molecular ions in the superexcited states have a certain lifetime, and in the general case, they can therefore be regarded as real resonance levels in a multiphoton absorption process followed by ionization.

Note that adsorption systems of this type are unique for photoionization spectroscopy, since multiphoton ionization of molecular ions occurs as a result of a strongly polar donor–acceptor adsorption π-bond, whose antibonding state fulfills the function of a resonance level. Since the resonance state is essentially an electron–vibrational state (large width), the stringent requirements on the selectivity of the multistep photoionization are lifted. Such multiphoton processes can therefore be detected in a one-frequency laser field. In this sense, a three-photon luminescence excitation process should also be observed for type-I adsorption complexes; the luminescence would not ''fade'' and its parameters would be identical to those of the luminescence with excitation by the third harmonic of the Nd:YAG laser. However, the probability of radiative transitions in this case will be determined by the specific nature of the channels of energy relaxation from the resonance level. For example, when nonradiative relaxation occurs without an interaction with excited valence levels of the molecular ions (the corresponding adiabatic potentials do not cross), the probability of radiative transitions will be low and vice versa. These processes require detailed analysis, which falls outside the scope of the present paper.

As noted in the Introduction, the production of CrO_4^{2-} ions in the gaseous state is problematic, so there is no information on their ionization potential. It follows from our experiments that the ionization potential for such ions should be $\leqslant 5\hbar\omega = 5.8$ eV. Therefore this method can be used to estimate the ionization potential of the CrO_4^{2-} ion.

The estimated absorption cross sections for one-photon transitions at the nonresonant multiphoton absorption steps show, despite their approximate character, that the cross sections correspond to typical values for one-photon processes. In two-photon transitions, however, the derived absorption cross section is much greater than the typical values. As a result, the typical probability of multiphoton transitions in an adsorption system is high. This anomalous phenomenon, in my opinion, can be explained by two factors.

1) The large nonlinear polarizability of the adsorption bond, which, as shown above, arises during thermal activation of the surface and is of a donor–acceptor nature. Therefore such a bond should be characterized by high asymmetry of the electronic cloud (there is no center of inversion) and, in consequence, the displacement of the electrons along the bond in different directions will be characterized by different force constants. As a result, vibrations will be highly anharmonic, and therefore the probability of nonlinear processes will be high.

Thus, the proposed method makes it possible to obtain information about strongly polar bonds on a surface, i.e., about bonds in chemisorbed complexes. Note that this method is selective with respect to the intensity of the incident radiation. Since the probability of observing a response of the system for polar and nonpolar bonds can differ by ten orders of magnitude,[54] the response of the system will only contain information about the chemisorbed complexes. Structural defects are also characterized by strongly polar bonds on a surface.[25] Such defects will also contribute to the nonlinear polarizability of the dispersed medium.

2) The specific nature of the interaction of intense laser radiation with dispersed wide-gap oxides, manifested as an intensification of the local fields on account of the induced dipole moments of the particles of the dispersed medium. These effects must be very important in our dispersed system, since the wavelength of the laser radiation in our case is much greater than the size of the dispersed particles (10–50 nm (Ref. 21)). Moreover, such processes, which lead to the generation of surface polaritons, can be excited according to a stimulated scheme (stimulated Raman scattering by surface polaritons).[56] The latter circumstance will result in a substantial increase in the local electric fields in the dispersed matrix with respect to the applied laser field; this is a decisive factor in multiphoton absorption of intense laser radiation. As is well known,[58] such processes sharply increase the intensity of the system response, as has been demonstrated for various organic molecules adsorbed on metal surfaces. However, for the adsorption system studied here these questions also require further experimental and theoretical study.

I also call attention to the fact that luminescence observation of multiphoton ionization–fragmentation of CrO_4^{2-} molecular ions and, on the whole, room-temperature luminescence spectra for such an adsorption system are unusual from the standpoint of the triplet nature of the luminescence. For example, previously, the luminescence spectra of such ions were recorded only at low temperatures.[8,23,33–37,39] Obviously, in the case at hand, intense laser beams were used, which made it possible to increase substantially the intensity of the luminescence response. Note also, however, that a dispersed matrix makes it possible to investigate luminescence spectra with laser intensities for which optical break-

down followed by damage to the material is observed in crystals with CrO_4^{2-} impurity ions and ions of the corresponding chromate salts. Therefore the fixation of inorganic molecular ions on the surfaces of wide-gap dispersed oxides is a promising method for investigating spectroscopic characteristics of such ions.

In conclusion, I thank my colleague S. N. Naumenko of the Optics Department at Kiev State University for assisting in the luminescence measurements.

[1] S. J. Smith, J. H. Eberly, and J. W. Gallagher (eds.), *Multiphoton Bibliography 1983–1986*, NBS LP-92, Suppl. 5, 1989.
[2] M. N. R. Ashford, S. G. Clement, J. D. Howe, and C. M. Western, J. Chem. Soc. Faraday Trans. **89**, 1153 (1993).
[3] V. S. Letokhov, *Laser Photoionization Spectroscopy*, Academic Press, N.Y. (1987).
[4] V. S. Letokhov, Commun. Atomic Mol. Phys. **7**, 107 (1977).
[5] V. S. Antonov, I. N. Knyazev, V. S. Letokhov *et al.*, Opt. Lett. **3**, 37 (1978).
[6] P. Esherick and R. J. M. Anderson, Chem. Phys. Lett. **70**, 621 (1980).
[7] A. S. Sudbo and M. M. T. Loy, Chem. Phys. Lett. **82**, 135 (1981).
[8] Yu. D. Glinka, Zh. Prikl. Spektrosk. **57**, 289 (1992).
[9] I. Ya. Kushnirenko, Yu. D. Glinka, T. B. Krak *et al.*, Zh. Prikl. Spektrosk. **59**, 286 (1993).
[10] Yu. D. Glinka, T. B. Krak, and Yu. N. Belyak, J. Mol. Struct. **349**, 215 (1995).
[11] Yu. D. Glinka and T. B. Krak, Phys. Rev. B **52**, 14985 (1995).
[12] Yu. D. Glinka and T. B. Krak, Fresenius J. Anal. Chem. **355**, 647 (1996).
[13] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[14] V. S. Letokhov and S. K. Sekatskiĭ, Opt. Spektrosk. **76**, 303 (1994) [Opt. Spectrosc. **76**, 271 (1994)].
[15] Yu. D. Glinka and S. N. Naumenko, Izv. Ross. Akad. Nauk, Ser. Fiz., No. 12, 55 (1992).
[16] Yu. D. Glinka, S. N. Naumenko, V. M. Ogenko, and A. A. Chuiko, in *Proceedings of the 4th International Conference on the Fundamentals of Absorption*, Kyoto, 1992, M. Suzuki (ed.), Kodansha, Tokyo (1993), p. 217.
[17] Yu. D. Glinka and S. N. Naumenko, in *Proceedings of the 5th International Conference on the Fundamentals of Absorption*, M. D. LeVan (ed.), Kluwer Academic Publishing Co., Boston (1996), p. 313.
[18] S. E. Egorov, V. S. Letokhov, and A. N. Shibanov, in *Surface Studies with Lasers*, F. Aussenegg, A. Leither, and M. E. Lippisch (eds.), Springer-Verlag, N.Y. (1983), p. 156.
[19] Yu. D. Glinka and S. N. Naumenko, Materials and Manufacturing Processes **10**, 571 (1995).
[20] S. E. Egorov, V. S. Letokhov, and A. N. Shibanov, Kvant. Élektron. **11**, 1393 (1984) [Sov. J. Quantum Electron. **14**, 940 (1984)].
[21] A. A. Chuĭko and Yu. I. Gorlov, *The Surface Chemistry of Silicon Dioxide: Surface Structure, Active Centers, and Adsorption Mechanisms* [in Russian], Naukova Dumka, Kiev (1992).
[22] Yu. N. Belyak, Yu. D. Glinka, S. N. Naumenko *et al.*, Zh. Pirkl. Spektrosk. **59**, 77 (1993).
[23] M. U. Belyĭ, V. V. Boĭko, Yu. D. Glinka *et al.*, Izv. Akad. Nauk SSSR, ser. fizich. **53**, 1778 (1989).
[24] D. L. Griscom, J. Ceram. Soc. Jpn. **99**, 923 (1991).
[25] A. R. Silin' and A. N. Trukhin, *Point Defects and Elementary Excitations in Crystalline and Glassy* SiO_2 [in Russian], Zinatne, Riga (1985).
[26] A. J. Moulson and J. P. Roberts, J. Chem. Soc. Faraday Trans. **57**, 1208 (1961).
[27] H. Imai, K. Arai, H. Imagava, and Y. Abe, Phys. Rev. B **38**, 12772 (1988).
[28] Yu. Morimoto, I. Igarashi, H. Sagahara, and S. Nasu, J. Non-Crys. Sol. **139**, 35 (1992).
[29] A. B. P. Lever, *Inorganic Electronic Spectroscopy*, Elsevier, N.Y. (1968).
[30] Yu. D. Glinka, V. Ya. Degoda, and S. N. Naumenko, J. Non-Crys. Sol. **152**, 219 (1993).
[31] G. A. M. Dalhoeven and G. Blass, Chem. Phys. Lett. **76**, 27 (1980).
[32] K. K. Rebane, *Impurity Spectra of Solids: Elementary Theory of Vibrational Structure*, Plenum Press, N.Y. (1970).
[33] M. U. Belyĭ, Yu. D. Glinka, I. Ya. Kushnirenko *et al.*, in *Abstracts of Reports at the 30th All-Union Conference on Luminescence* [in Russian], Rovno (1984), p. 87.
[34] M. U. Belyĭ, Yu. D. Glinka, I. Ya. Kushnirenko, and V. R. Kumeskiĭ, Dokl. Akad. Nauk Ukr. SSR, Ser. A, No. 1, 39 (1987).
[35] M. U. Belyĭ, Yu. D. Glinka, and I. Ya. Kushnirenko, Dokl. Akad. Nauk Ukr. SSR, Ser. A, No. 4, 41 (1988).
[36] R. M. Miller and D. S. Tinti, Mol. Phys. **56**, 923 (1985).
[37] R. M. Miller and D. S. Tinti, Chem. Phys. Lett. **130**, 352 (1985).
[38] Yu. D. Glinka, I. Ya. Kushnirenko, and G. I. Salevon, Opt. Spektrosk. **74**, 153 (1993) [Opt. Spectrosc. **74**, 94 (1993)].
[39] R. M. Miller and D. S. Tinti, J. Luminescence **36**, 143 (1986).
[40] G. B. Porter, A. D. Kirk, and D. K. Sharma, J. Chem. Phys. **90**, 1781 (1996).
[41] É. S. Medvedev and V. I. Osherov, *The Theory of Nonradiative Transitions in Polyatomic Molecules* [in Russian], Nauka, Moscow (1983).
[42] K. H. Hausser and H. C. Wolf, in *Advances in Magnetic Resonance*, Academic Press, N.Y. (1976), Vol. 8, p. 85.
[43] G. Herzberg, *Molecular Spectra amd Molecular Structure: Electronic Spectra and Electronic Structure of Polyatomic Molecules*, Krieger, N.Y. (1991).
[44] M. A. El-Sayed, in *Excited States*, E. C. Lim (ed.), Academic Press, N.Y. (1974), Vol. 1, p. 35.
[45] A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions*, Clarendon Press, Oxford (1970).
[46] R. J. Bell, Rep. Prog. Phys. **35**, 1315 (1972).
[47] V. I. Sugakov, Fiz. Tverd. Tela (Leningrad) **19**, 1877 (1977) [Sov. Phys. Solid State **19**, 1099 (1977)].
[48] V. I. Sugakov, Izv. Akad. Nauk SSSR **47**, 1389 (1983).
[49] V. I. Sugakov and Yu. D. Shtepa, Phys. Stat. Sol. B **116**, 633 (1983).
[50] *Collected Scientific Papers on the Physics of Molecular Crystals* [in Russian], Naukova Dumka, Kiev (1986).
[51] K. Arai, H. Imai, H. Hosono *et al.*, Appl. Phys. Lett. **53**, 1891 (1988).
[52] J. H. Stathis and M. A. Kastner, Mat. Res. Symp. Proc. **61**, 161 (1986).
[53] V. N. Bagratashvili, A. O. Rybaltovskii, and S. I. Tsypina, Spectrochim. Acta A **46**, 665 (1990).
[54] D. S. Kliger (ed.), *Ultrasensitive Laser Spectroscopy*, Academic Press, N.Y. (1983).
[55] É. D. Aluker, D. Yu. Lusis, and S. A. Chernov, *Electronic Excitations and Radioluminescence of Alkali–Halide Crystals* [in Russian], Zinatne, Riga (1979).
[56] N. I. Koroteev and I. L. Shumaĭ, *The Physics of High-Power Laser Radiation* [in Russian], Nauka, Moscow (1991).
[57] Y. R. Shen, *The Principles of Nonlinear Optics*, Wiley, N.Y. (1984).
[58] R. K. Chang and T. E. Furtak (eds.), *Surface-Enhanced Raman Scattering*, Plenum Press, N.Y. (1982).

Translated by M. E. Alferieff

Kinetics of vortex structure formation in magnetic materials

A. É. Filippov

Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine
(Submitted 24 June 1996)
Zh. Éksp. Teor. Fiz. **111**, 1775–1786 (May 1997)

A kinetic scenario for the formation of a vortex phase in magnetic materials is discussed. It is found that such a phase can be generated from fluctuations at the kinetic stage of evolution and can subsequently be fixed as a thermodynamically stable phase. © *1997 American Institute of Physics.* [S1063-7761(97)01605-3]

1. INTRODUCTION

Until recently, the feasibility of describing phase transitions was severely limited to obtaining analytic solutions for equilibrium configurations of the (average) order parameter φ under well-known thermodynamic assumptions.[1,2] This made it impossible to study processes that take place at intermediate times and spatial scales, as well as processes involving several moderate interactions. Later it was found that the interaction of fluctuations can lead to stabilization of phases that differ from those of the mean-field theory.[3–9] The theory of critical phenomena that predicts the existence of such phases is based on coarsening the description by applying the renormalization-group method. Such a theory is unable to ''follow'' the structuring of the fluctuating field at the mesoscopic scale,[10] which in the kinetic picture actually results in the emergence of anomalously ordered phases,[11] in the generation of nonlocal structures in the Ginzburg–Landau functional, and in anomalous effective dimensionalities.[12]

On the other hand, statistical theory does not allow for ''virtual phases''[13,14] (see also Ref. 15) that appear as the system proceeds to equilibrium. The formation of such phases is of a purely kinetic nature.

In the simplest case the corresponding kinetic effects can be obtained by describing the order-parameter relaxation with the following equation (for a review of this topic see Refs. 16–21):

$$\frac{\partial\varphi}{\partial t}=-\frac{\delta\mathscr{F}[\varphi]}{\delta\varphi}+\xi(t,\mathbf{r}). \tag{1.1}$$

Usually the structure of the spectrum of the noise source, $\xi(t,\mathbf{r})$, is unimportant[18,19], so that for the sake of simplicity we assume that we are dealing with white noise:

$$\langle\xi(t,\mathbf{r})\rangle=0,\quad \langle\xi(t,\mathbf{r})\xi(t,\mathbf{r}')\rangle=D\,\delta(\mathbf{r}-\mathbf{r}')\,\delta(t-t'). \tag{1.2}$$

Generally, the functional $\mathscr{F}[\varphi]$ is a nonlocal form in powers of the fluctuating field $\varphi(\mathbf{r})$, a form that for the magnetic systems considered here can be derived from the microscopic theory.[22–25] To get a qualitative idea of the problem we keep only the lowest-order terms, with nonlocality allowed only in terms quadratic in φ. The corresponding form then reduces to

$$\mathscr{F}[\varphi]=\int d^dr\left[\frac{1}{2}\varphi(\mathbf{r})\int d^dr'\;J(\mathbf{r}-\mathbf{r}')\varphi(\mathbf{r}')+F(\varphi(\mathbf{r}))\right], \tag{1.3}$$

which we use below.

Under relaxation of the system, the ordered structures of the order parameter φ are selected from the fluctuation noise.[16,17] These structures are attractors for the field φ, attracting it in all directions of the function space with the exception of a denumerable (finite) set of directions, in which they are unstable under a transition to an absolutely stable state.[18,19] The system rapidly relaxes to such attractors and then slowly evolves on them up to the point where a new instability develops and a transition to a lower-lying attractor occurs.

The latter feature makes it possible, at least in principle, to develop various kinetic scenarios by varying the external parameters in the transition. In this paper we examine, among other things, the possibility of a vortex structure being formed in magnetic materials. For an analogy that has been found useful, we first recall certain facts of the theory of superconducting systems.

Vortex configurations of the order parameter play an important role in the kinetics of the phase transition to superconductivity. This is especially true of a planar (or layered) superconductor, where because of suppression by large long-range order fluctuations,[26] the presence of a phase transition is usually related to the production of vortex–antivortex pairs and their possible unpairing under certain conditions.[27–29] Analytic vortex-like solutions for a superconductor in an external field (Abrikosov vortices)[30] are well known, and the concept of vortex pairs has been corroborated by numerous experiments.

Mesoscopic structures (and a pair is such a structure) cannot form instantly. The formation of such a structure constitutes a regular stage in the relaxation of a relatively large order-parameter region, in which the consistent evolution of the order parameter and the vector potential develops in the process of fluctuation pair production. However, vortices of opposite signs attract each other, so that the pair being produced is unstable against collapse. Nevertheless, as shown by recent research,[11,31] a vortex structure emerges in a natural and probably inevitable manner. The researchers found that the corresponding configurations of the fluctuating field

are attractors. In the context of the present paper, the vortex phase of a superconductor in many respects can be considered a virtual phase. Vortices are stably formed from arbitrary (relatively small-scale) fluctuations, produce a minimum of dissipation, and constitute the longest stage in the ''journey'' of the order parameter to equilibrium. And because of this, actually, a vortex state manifests itself as a stable state in the thermodynamic properties of real superconductors.

2. KINETICS OF FORMATION OF A VORTEX STRUCTURE. BOGDANOV–HUBERT PHASE

From a formal viewpoint, the vortex configuration of $\varphi(r)$ in superconductors emerges because of the interaction of the order-parameter current and fluctuations of the gauge electromagnetic field $(\mathbf{j}\cdot\mathbf{A})$, fluctuations whose contribution to the Ginzburg–Landau functional of the system has a rotational structure: $(\mathrm{curl}\,\mathbf{A})^2$. In magnetic systems, the Dzyaloshinskiĭ–Morya interaction

$$W_D = J_k \frac{\partial J_j}{\partial r} - J_j \frac{\partial J_k}{\partial r} \tag{2.1}$$

can serve as a sort of analogy to the energy terms (see Refs. 32–36). Over the years many substances have been shown to contain numerous one-dimensional modulated structures that are products of such interaction. Recently Bogdanov and Hubert[37,38] established that in addition to exhibiting one-dimensional order-parameter configurations, systems with the interaction (2.1) may have two-dimensional structures of the vortex type (see Figs. 2 and 3 in Ref. 38, which depict the corresponding configurations of the magnetization vector).

According to Bogdanov and Hubert,[38] in contrast to periodic magnetic structures, which can exist only with a fairly strong Dzyaloshinskiĭ interaction, isolated magnetic vortices can exist for arbitrarily small strengths of this interaction. This fact was mentioned in the papers cited above as being preferable from the standpoint of applications when used in interpreting the experimental data, since a formal solution for a solitary vortex line can exist even when there is no external field. Bogdanov and Hubert[37] compared the energies of the modulated phase and the phase formed by a hexagonal lattice of magnetic vortices. They also established how the regions of thermodynamic stability of these phases depend on the external field strength and the relative contribution of the Dzyaloshinskiĭ energy (see Fig. 10 in Ref. 37). Estimates of the free energy have shown that for certain values of the external field and for certain ratios of the parameters of the interaction and anisotropy, vortex structures are preferable to one-dimensional.

Here there are at least two big problems. First, thus far only one-dimensional structures have been observed in experiments. In addition, Bogdanov and Hubert[37,38] assumed that magnetic vortices form a regular hexagonal lattice, which (if the lattice had already been formed in some way) are thermodynamically stable.

Here, however, we are confronted with the same problem as in the case of superconducting vortices. More precisely, a vortex (and even more so, a lattice consisting of vortices) is a mesoscopic structure, i.e. it cannot emerge fully formed. One can easily imagine and reproduce kinetically the growth of such a simple topological structure as a wave of spin rotations, but it is much more difficult to do the same in connection with the spontaneous growth of a system of vortices formed by rotations of moments in all three directions. Nevertheless, as demonstrated below, such a process is possible.

For our subsequent analysis it is convenient to briefly list some results describing the static radial structure of a solitary vortex. The functional for the system's energy in the notation adopted in Ref. 37 has the form

$$W = \int d^d r \left\{ A(\nabla\cdot\mathbf{m})^2 - K m_z^2 - J_z H_z^0 - \frac{1}{2} J H_d + D W_D \right\}, \tag{2.2}$$

where the reduced magnetization $m^2(\mathbf{r})=1$, JH_d is the demagnetization energy, and it is convenient to write the Dzyaloshinskiĭ energy density in terms of spherical coordinates:

$$\mathbf{m} = (\sin\theta\cos\psi,\ \sin\theta\sin\psi, \cos\theta),$$

$$\mathbf{r} = (r\cos\varphi,\ r\sin\varphi, z).$$

Since we intend to analyze a lattice consisting of localized magnetic structures, we limit ourselves to the case of a strongly localized vortex with C_{nv} symmetry (see Refs. 35–38). Here the Dzyaloshinskiĭ energy in terms of the new variables has the form

$$W_D = \cos(\varphi-\psi)\theta_r + \sin\theta\cos\theta\sin(\varphi-\psi)\psi_r + \frac{\sin(\varphi-\psi)\theta_\varphi}{r} + \frac{\sin\theta\cos\theta\cos(\varphi-\psi)\psi_\varphi}{r}. \tag{2.3}$$

We are in fact dealing with a lattice consisting of hexagonal cells in which the magnetization vector is parallel to the external field at the boundaries of the cells and is antiparallel at the center of each cell. Such a magnetization distribution suggests a certain analogy to a lattice consisting of Abrikosov vortices, and has been examined in Ref. 37.

If we assume that the magnetic moment distribution is characterized by strictly radial symmetry, the problem reduces to solving only one equation,

$$\frac{d^2\theta}{dr^2} + \frac{1}{r}\frac{d\theta}{dr} - \frac{\sin\theta\cos\theta}{r^2} + \frac{4\tilde{\kappa}\sin^2\theta}{\pi r} - \sin\theta\cos\theta = 0, \tag{2.4}$$

which describes the rotation of a moment vector of fixed length in relation to the angle θ. Here, as in Ref. 38, Eq. (2.4) is the Euler equation for the functional (2.1) written in terms of the variables r, φ, and θ, and we have introduced the notation

$$\tilde{\kappa} = \frac{\pi D}{4\sqrt{A(K+J_z^2)/2\mu_0}}$$

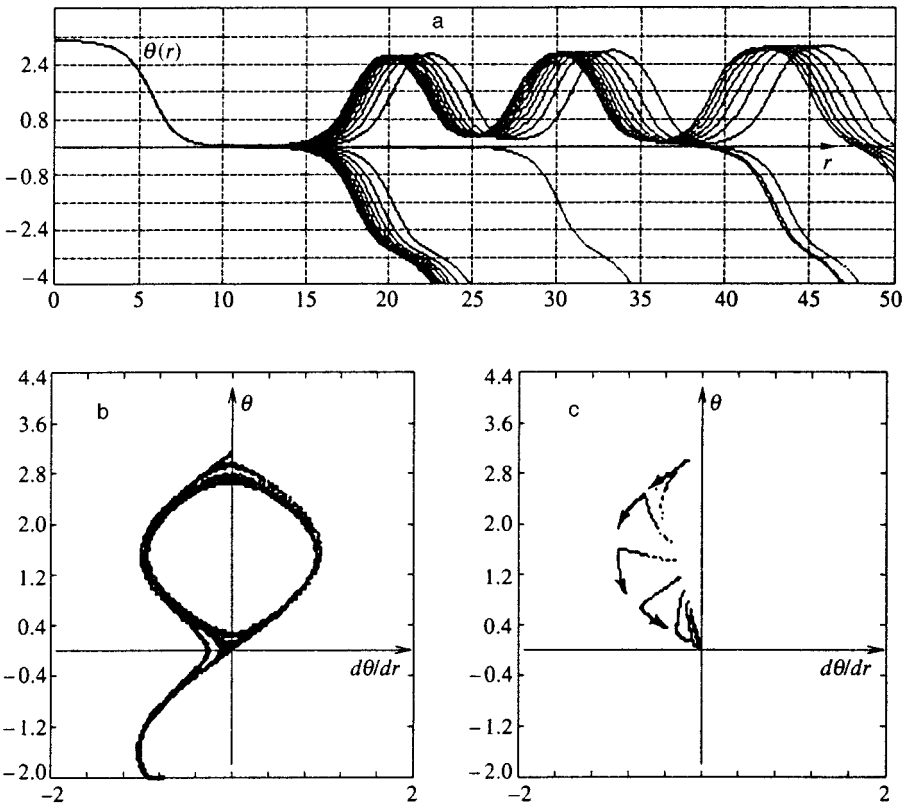

FIG. 1. Solution of Eqs. (2.4) and (2.5). a—The dependence of phase θ on radius r for physically interesting solutions close to the saddle-point localized distribution $\theta(r)$. b—The projections of the same solutions on the phase portrait in the $(\partial\theta/\partial r, \theta)$ plane. c—Projections on the phase portrait of the trajectories describing evolution of an (arbitrary) distribution $\theta(r)$ in accordance with Eq. (2.5).

for the parameter $\tilde{\kappa}$ describing the relative contribution of the Dzyaloshinskiĭ energy.

Localized solutions of Eq. (2.4) are saddle-point trajectories in the $(\partial\theta/\partial r, \theta)$ phase plane. They can be found by the method of successive approximations, in which one of the constant terms for the second-order equation is fixed as the initial condition, while the other is used as a fitting parameter (shooting method).

The solution of this equation with the boundary condition $\theta(0)=\pi$ is summarized in Fig. 1. Figure 1a depicts several branches of the solutions for the phase $\theta(r)$, solutions that are close to the desired (saddle-point) localized distribution and that demonstrate the successive approximations of the shooting method. The corresponding saddle is clearly visible in the projection of these solutions on the phase portrait in the $(\partial\theta/\partial r, \theta)$ plane depicted in Fig. 1b.

Note that the localized solutions whose phase trajectories in the $(\partial\theta/\partial r, \theta)$ plane are saddle-points do not necessarily correspond to a saddle point (in the function space) for the free energy being solved for. The desired energy calculated in such a realization of the distribution $\theta(r)$ is finite. This distribution $\theta(r)$ is only a point that can be suspected of being a minimum, and its stability must be studied separately by employing a time-dependent relaxation version of the equation for the order parameter. Such an equation (sometimes called the time-dependent Ginzburg–Landau model) reflects the real evolution of the order-parameter field, an evolution that among all formally admissible solutions of the static equation selects only those that correspond to physically realizable structures.

In terms of the variables θ and r, such a solution has the form

$$\frac{d\theta}{dt}=-\left(\frac{d^2\theta}{dr^2}+\frac{1}{r}\frac{d\theta}{dr}-\frac{1}{r^2}\sin\theta\cos\theta+\frac{4\tilde{\kappa}}{\pi r}\sin^2\theta-\sin\theta\cos\theta\right). \tag{2.5}$$

In Fig. 1c the trajectories describing the evolution of an (arbitrary) distribution $\theta(r)$ in accordance with the relaxation equation are projected onto the same phase plane $(\partial\theta/\partial r, \theta)$ as the solutions of the static equation. Formally this approach generalizes the study of the stability of the static solutions, which in a special case amounts to analyzing the equation linearized near some special solution of the static equation. In our case such an equation coincides with the Schrödinger equation with some effective potential. The general form of the relaxation equation makes it possible not only to study stability against small deviations, but also to analyze the global stability of the desired localized solutions.

Figure 1c shows that a strictly localized solution of the static equation satisfying the condition $\theta(\infty)=0$ is also a saddle-point solution, with the result that it is unstable. The typical solutions are those that in their behavior are close to this saddle-point solution for small θ and degenerate at in-

finity into an essentially periodic wave $m_z(r)$, which, incidentally, is quite natural from a physical standpoint. It can be shown (see, e.g., the phase portrait in Fig. 1c) that a fairly arbitrary configuration $\theta(r,t=0)$ rapidly relaxes to the localized solution described above, which, however, is unstable and in time collapses to $r=0$.

The fact that a solitary vortex is unstable and generates a periodic wave at infinity does not mean, however, that a lattice consisting of such vortices is also unstable. In this context, further study of Eq. (2.5) is of little interest, since it does not allow for a study of the stability against deviation from radial distributions at fixed $m^2=1$. With the aim of analyzing the spontaneous onset of a vortex structure, we go back to the initial equations for the components of the order parameter $\mathbf{m}=\{m_x,m_y,m_z\}$. Note that when the system proceeds from the paraphase to the ordered phase, as the temperature decreases, the absolute value of the moment automatically becomes fixed and cannot be achieved by formally stating that $m^2=\text{const}$. This means that in the equations for the components $\{m_x,m_y,m_z\}$ we must keep the nonlinear (local) contributions $\mathscr{F}[\mathbf{m}]$ to the free energy, which in and of itself ensures that m^2 is fixed in the final stages of ordering of the system.

In other words, we again study an equation of the extremely general form

$$\frac{\partial \mathbf{m}}{\partial t}=\alpha\Delta\mathbf{m}-\beta\ \nabla\times\mathbf{m}+h-\frac{\delta\mathscr{F}[\mathbf{m}]}{\delta\mathbf{m}}+\xi(\mathbf{r};t). \qquad (2.6)$$

This equation is similar to the original (scalar) equation (1.1), but is written for the three interacting components of the order parameter, m_x, m_y, and m_z. The gradient terms in Eq. (2.6) are specified explicitly, while the local form of $\mathscr{F}[\mathbf{m}]$ is restricted only by the requirement that at equilibrium m^2 must be fixed.

The choice of the form of $\mathscr{F}[\mathbf{m}]$ is fairly arbitrary within the model. For real systems, $\mathscr{F}[\mathbf{m}]$ can be derived from the microscopic Hamiltonian of Refs. 22–25. Usually only the lower-order terms in the expansion in powers of m are kept in the local form of $\mathscr{F}[\mathbf{m}]$, e.g.,

$$\frac{\delta\mathscr{F}[\mathbf{m}]}{\delta\mathbf{m}}=\mathbf{m}(\tau-am+bm^2).$$

In this sense the microscopic theory actually fixes the values of the constants $\tau=(T-T_c)/T_c$, a, and $b>0$ in this expansion. The theory also determines the values of the parameters α and β fixed by the spatial dispersion of the exchange interaction and the Dzyaloshinskiĭ energy, respectively.

Above all, it is advisable to reproduce the process of nucleation of a solitary vortex from a local nucleation center in the absence of noise. Figure 2 depicts the corresponding results, which were obtained at the intermediate stage of growth of such a vortex in space. The results for each component m_j are shown by the corresponding changes in the shades of gray. The distribution of m_z is radially symmetric, while the other two components change sign at the center of the vortex. This indeed corresponds to a ''fountain'' of the vector $\mathbf{m}(\mathbf{r})$ at the center, described in Refs. 37 and 38, with concentrically diverging waves far from $\mathbf{r}=0$. For the sake of comparison, in Fig. 2d we depict the surface $m_z(\mathbf{r})$ at an earlier stage in the growth of the vortex.

Naturally, the idea of a solitary vortex growing indefinitely is an extreme idealization. In reality, vortices begin to grow out of many local fluctuations of $\mathbf{m}(\mathbf{r})$ essentially simultaneously. As they expand, they come into contact with one another, the circular distributions of $\mathbf{m}(\mathbf{r})$ are disrupted, and a pattern is formed that resembles the striped structure of magnetic domains. Figures 3a–d depict the evolution of a system of magnetic vortices produced by fluctuation noise, where the density of one of the components (m_z) is shown by changes in the shades of gray.

In the context of the present investigation it is important to note that a stage at which solitary circular configurations of the magnetic moments show up clearly does indeed exist. This intermediate stage originates in the relatively small-scale local fluctuations fixed by the relaxing field for a fairly long time.

Such a fluctuation state of the system can establish itself for an indefinitely long time if the phase transition to it is first-order and the ratio of the noise (temperature) to the height of the barrier in the energy $\mathscr{F}[\mathbf{m}]$ is such that the growth of supercritical nucleation centers becomes impossible. In computer simulations of such a system one can observe how the constantly ''flickering'' but long-lived vortices gradually order into a close-packed hexagonal lattice. These observations can be used constructively in developing a kinetic scenario that leads to stabilization of the Bogdanov–Hubert phase.

Indeed, if we lower the barrier after keeping the system for a fairly long time in the fluctuation region, the system proceeds to a low-temperature state, and both the symmetry of the distribution $\mathbf{m}(\mathbf{r})$ (on the average) and individual vortices inside the striped structure become frozen in. For instance, such patches consisting of individual vortices of both signs of m_z are clearly visible in Fig. 3d. In this context the structure of the correlation function

$$G(x,y)=\langle\mathbf{m}(x',y')\mathbf{m}(x+x',y+y')\rangle \qquad (2.7)$$

for a state consisting of vortices and stripes simultaneously is especially indicative.

Such a function for the m_z component was calculated numerically, and its typical shape is depicted in Fig. 4. Clearly visible is the sixfold symmetry for small-scale correlations, which is somewhat distorted because of the presence of stripes, which fix the twofold symmetry for intermediate-scale correlations. One can expect that for large arrays, the arbitrary nature of the orientation of the stripes makes the effect of these stripes on the symmetry of $G(x,y)$ negligible. However, for accessible moderate scales this does not happen, as Fig. 4 clearly demonstrates.

The problem of the size and shape of the computational grid (which is fourfold symmetric) is important for numerical calculations. Figure 4 indicates that fourfold symmetry begins to show up on scales greater by a factor of ten than the characteristic size of the structures described above. For specific calculations we used arrays of 256×256 cells; the characteristic size of a vortex is 8×8 cells, while the corre-

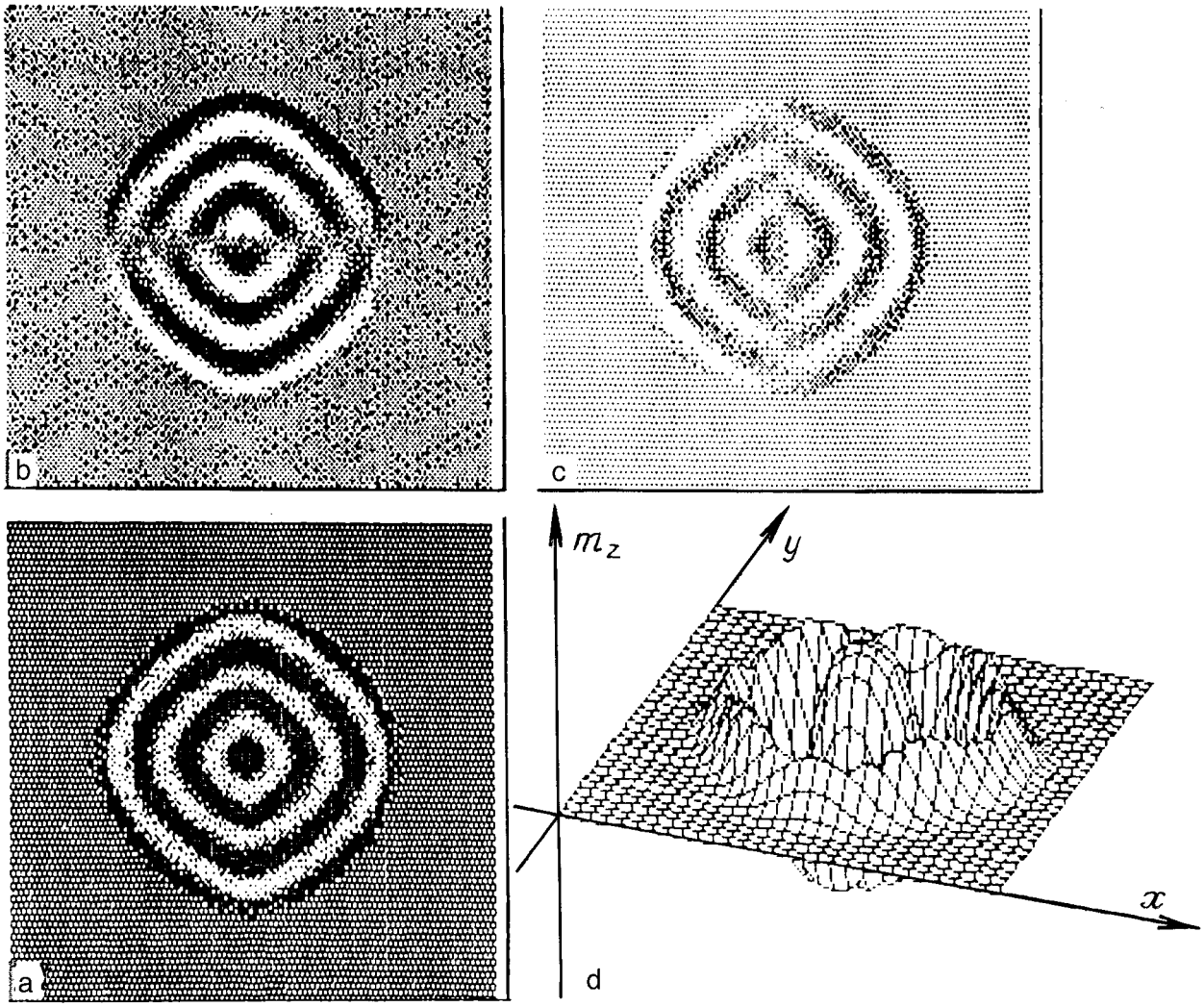

FIG. 2. Evolution of a solitary vortex. a—Projection in the xy plane of the component m_z, shown by changes in the shades of gray. Clearly visible are the central peak in m_z and the concentrically diverging waves. b—The same as in Fig. 2a for the m_x component. c—The same as in Fig. 2a for the m_y component. Comparison of Figs. 2a–c makes it possible to determine the rotation of **m** in space corresponding to the vortex described in the text. d—The relatively early stage in the formation of the vortex. The isometric system is used to depict the m_z component.

lation function depicted in Fig. 4 was calculated on an array of 64×64 cells, which amounts to one-sixteenth the area of the computational grid for $\mathbf{m}(x,y)$.

Most vortices nevertheless disappear with time. To avoid this, an external field $\mathbf{h}\neq 0$ can be used to strengthen those vortex nucleation centers that point in the same direction as the field. However, as expected,[37,38] a low-temperature phase with vortices directed along the field is unstable and degenerates, depending on the strength of the field **h**, into either a striped structure or a homogeneously ordered state. To stabilize the desired phase, we must ensure that vortices with the same sign are fairly close-packed and are directed opposite the field.

The latter requirement does not contradict the idea of strengthening vortices pointed in the same direction by a field parallel to them, and it yields a recipe for creating the desired phase. In the initial stage of the ordering process one must use a moderate field **h**, whose strength can be chosen by numerical modeling with a fixed structure $\mathscr{F}[\mathbf{m}]$. Then the field is reversed and increased in amplitude (also selected by modeling). In the process, the ordered system fixes the vortex structures formed in the previous stages.

Figure 5 depicts the typical result of a process of this type, obtained for the sake of definiteness for the local function

$$\frac{\delta\mathscr{F}[\mathbf{m}]}{\delta\mathbf{m}}=\mathbf{m}(\tau-am+bm^2) \tag{2.8}$$

with the constants $\tau=0.14$, $a=2.0$, and $b=1.0$ and for a field strength $h_0=0.35$ in the ordered phase. Naturally, these parameter values are within the stability range for the desired vortex phase, found earlier by Bogdanov and Huber.[37] The value of h used in the initial stages of the process can vary over a broad range, and usually amounts to approximately $0.2h_0$.

Figure 5 depicts the distribution of m_z with an isometric system of constant-level curves. This reveals the universal form of the majority of vortices, which are attractors of the corresponding nonlinear system (2.6). Moreover, we see that

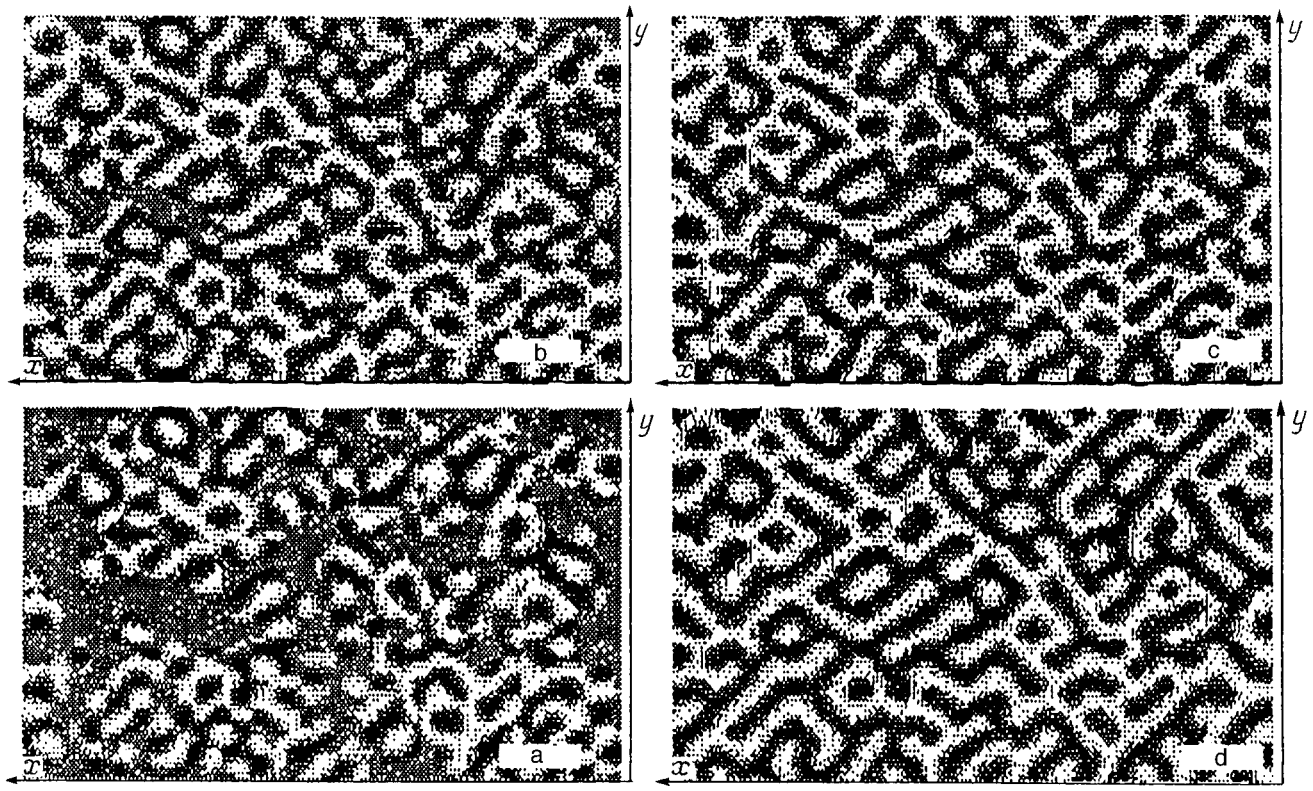

FIG. 3. Evolution of the magnetic vortices produced by fluctuation noise. The density of one of the components (m_z) is shown by changes in the shades of gray.

lattice imperfections are usually accompanied by the adhesion of two (or more) vortices, caused by the inadequate preparation of the intermediate virtual state. Physically, the adhesion of vortices can be interpreted as a consequence of kinetic competition with one-dimensional modulated ordering, which occurs spontaneously in the system if the transition process is not arranged in the special manner described above.

It would be quite natural to extend the kinetic scenario described in this paper to the preparation of such structures in real samples. Here the virtual phase may be replaced by a similar state in the fluctuation region. The way to do this is to

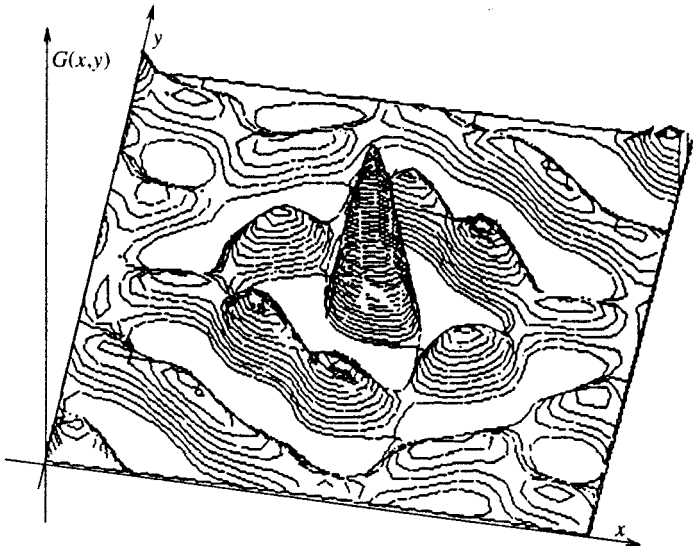

FIG. 4. A typical structure of the correlation function $G(x,y)$ of the component m_z for a state containing vortices and stripes simultaneously. Clearly visible are the sixfold symmetry for small-scale correlations, the twofold symmetry for intermediate-scale correlations, and the fourfold symmetry for correlations whose scales range from roughly one-tenth to the full size of the computational grid for $\mathbf{m}(x,y)$ (this in turn is greater by a factor of ten than the array used in calculating $G(x,y)$).

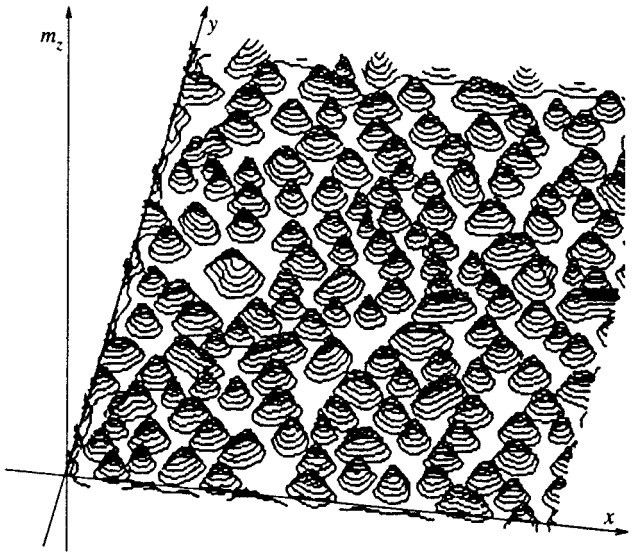

FIG. 5. Typical configuration of $m_z(x,y)$ obtained as a result of the special kinetic process described in the text.

keep the system near the transition point in a weak field. Then, after the direction of the field is reversed and its amplitude is increased, the system must be rapidly transferred to a low-temperature state.

I am deeply grateful to A. N. Bogdanov, who drew my attention to the problem of magnetic vortices and kindly provided the results of his studies prior to publication. Many thanks go to A. S. Zel'tser for the top-quality computer programs, which made all necessary computer simulation possible, and to Yu. E. Kuzovlev for writing the application programs used in processing and graphical representation of the data. The present work was sponsored by the International Science Foundation (Grant No. K58100).

[1] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3rd ed., Pergamon Press, Oxford (1980).
[2] V. L. Ginzburg and L. D. Landau, Zh. Éksp. Teor. Fiz. **20**, 1064 (1950).
[3] K. Wilson and J. Kogut, *The Renormalization Group and the ε-Expansion*, Wiley, New York (1974).
[4] S. K. Ma, *Modern Theory of Critical Phenomena*, W. A. Benjamin, Reading, Mass. (1976).
[5] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, Pergamon Press, Oxford (1979).
[6] I. F. Lyuksyutov, V. L. Pokrovskiĭ, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **69**, 1817 (1975) [Sov. Phys. JETP **42**, 923 (1975)].
[7] A. I. Sokolov and A. K. Tagantsev, Zh. Éksp. Teor. Fiz. **76**, 181 (1979) [Sov. Phys. JETP **49**, 92 (1979)].
[8] P. Bak, S. Krinski, and D. Mukamel, Phys. Rev. B **13**, 5065 (1975).
[9] P. Bak, S. Krinski, and D. Mukamel, Phys. Rev. Lett. **36**, 52 (1975).
[10] A. E. Filippov, J. Stat. Phys. **75**, 241 (1994).
[11] A. S. Zel'tser and A. É. Filippov, Zh. Éksp. Teor. Fiz. **106**, 1117 (1994) [JETP **79**, 605 (1994)].
[12] A. S. Zel'tser and A. É. Filippov, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 604 (1995) [JETP Lett. **62**, 627 (1995)].
[13] S. Semenovskaya and A. G. Khachaturyan, Phys. Rev. Lett. **67**, 2223 (1991).
[14] Long-Qing Chen and A. G. Khachaturyan, Phys. Rev. B **46**, 5889 (1992).
[15] A. G. Khachaturian, *Theory of Structural Transformations in Solids*, Wiley, New York (1983).
[16] T. M. Rogers, K. R. Elder, and R. C. Desai, Phys. Rev. B **37**, 9638 (1988).
[17] K. R. Elder and R. C. Desai, Phys. Rev. B **40**, 243 (1989).
[18] A. E. Filippov, Yu. E. Kuzovlev, and T. K. Soboleva, Phys. Lett. A **178**, 301 (1993).
[19] Yu. E. Kuzovlev, T. K. Soboleva, and A. É. Filippov, Zh. Éksp. Teor. Fiz. **103**, 1742 (1993) [JETP **76**, 858 (1993)].
[20] A. S. Zel'tser, T. K. Soboleva, and A. E. Filippov, Zh. Éksp. Teor. Fiz. **108**, 356 (1995) [JETP **81**, 193 (1995)].
[21] L. D. Landau and I. M. Khalatnikov, in *L. D. Landau: Collected Papers*, Pergamon Press, Oxford (1965), p. 626.
[22] Yu. M. Ivanchenko and A. É. Filippov, Fiz. Tverd. Tela (Leningrad) **31**, 51 (1989).
[23] A. Aharony and M. E. Fisher, Phys. Rev. B **8**, 323 (1973).
[24] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems* [in Russian], Nauka, Moscow (1987).
[25] Yu. A. Izyumov and Yu. N. Skryabin, Teoret. Mat. Fiz. **5**, 110 (1970).
[26] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966); P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).
[27] V. L. Berezinskiĭ, Zh. Éksp. Teor. Fiz. **59**, 907 (1970) [Sov. Phys. JETP **32**, 493 (1971)]; Zh. Éksp. Teor. Fiz. **61**, 1144 (1971) [Sov. Phys. JETP **34**, 610 (1972)].
[28] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973).
[29] J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974).
[30] A. A. Abrikosov, *Modern Theory of Metals*, North-Holland, Amsterdam (1990).
[31] A. E. Filippov, A. V. Radievsky, and A. S. Zeltser, Phys. Lett. A **192**, 131 (1994).
[32] I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **32**, 1547 (1957) [Sov. Phys. JETP **5**, 1259 (1957)].
[33] I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **46**, 1420 (1964) [Sov. Phys. JETP **19**, 960 (1964)].
[34] T. Morya, Phys. Rev. **120**, 91 (1960).
[35] A. N. Bogdanov and D. A. Yablonskiĭ, Zh. Éksp. Teor. Fiz. **95**, 178 (1989) [Sov. Phys. JETP **68**, 101 (1989)].
[36] A. N. Bogdanov, M. V. Kudinov, and D. A. Yablonskiĭ, Fiz. Tverd. Tela (Leningrad) **31**, No. 10, 105 (1989) [Sov. Phys. Solid State **31**, 1707 (1989)].
[37] A. N. Bogdanov and A. Hubert, J. Magn. Magn. Mater. **138**, 255 (1996).
[38] A. N. Bogdanov and A. Hubert, Phys. Status Solidi B **186**, 527 (1994).

Translated by Eugene Yankovsky

Localization and space-time dispersion of the kinetic coefficients of a two-dimensional disordered system

A. G. Groshev and S. G. Novokshonov

Physicotechnical Institute, Ural Branch of the Russian Academy of Sciences, 426001 Izhevsk, Russia
(Submitted 8 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1787–1802 (May 1997)

A generalization of the Vollhardt–Wölfle self-consistent localization theory is proposed to take into account spatial dispersion of the kinetic coefficients of a two-dimensional disordered system. It is shown that the main contribution to the singular part of the collision integral of the Bethe–Salpeter equation in the limit $\omega\to 0$ is from the diffusion pole $i\omega=(\mathbf{p}+\mathbf{p}')^2 D(|\mathbf{p}+\mathbf{p}'|,\omega)$, which provides an anomalous increase in the probability of backscattering $\mathbf{p}\to-\mathbf{p}'$. In this limit the dependence of the diffusion coefficient on $\mathbf{q}$ and ω exhibits localization behavior, $D(q,\omega)=-i\omega f(l_D q)$, where $|f(z)|\leqslant(0)=d^2$ (d is the localization length). According to the Berezinskiĭ–Gor'kov criterion, $D(q,0)=0$ for all q. Spatial dispersion of $D(q,\omega)$ is manifested on a scale $q\propto 1/l_D$, where l_D is the frequency-dependent diffusion length. In the localization state $l_D\ll l$, where l is the electron mean free path; $l_D\propto\omega$ as $\omega\to 0$, suggesting the suppression of spatial dispersion of the kinetic coefficients down to atomic scales. Under the same conditions $\sigma(q,\omega)$ exhibits a strong dependence on q on a scale $q\propto 1/d$, i.e., the nonlocality range of the electrical conductivity is of the order of the localization length d. At the microscopic level these results corroborate the main conclusions of Suslov (Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)]), which were obtained to a certain degree phenomenologically in the limit $\omega\to 0$. A major advance beyond the work of Suslov in the present study is the analysis of spatial dispersion of the kinetic coefficients at finite (rather than infinitely low) frequencies. [S1063-7761(97)01705-8]

1. INTRODUCTION

The last twenty years have witnessed enormous progress in our understanding of the Anderson localization phenomenon.[1] Today this field of research has grown into an expansive branch of physics of the condensed state (see, e.g., Refs. 2–5). One of the most productive approaches to the investigation of the problem is afforded by the Vollhardt–Wölfle self-consistent localization theory.[6–8] Results corroborating the hypothesis of total localization of charge carriers in two-dimensional disordered systems have been obtained on the basis of this theory.[6,7] The critical behavior predicted by the Vollhardt–Wölfle theory near the Anderson transition in systems of dimension $d>2$ (Ref. 8) is consistent with results obtained in the field-theory[9] and scaling[4,10] approaches to the problem. Its ideas form the basis of microscopic localization theory in the presence of an external magnetic field[11,12] and the modern theory of ''dirty'' superconductors.[5,13,14]

One of the drawbacks of self-consistent localization theory in its present formulation is the impossibility of systematically taking into account spatial dispersion of the kinetic coefficients. The problem is that the basic equation of this theory establishes an integral relation between the local diffusion coefficient $D(\omega)=D(q=0,\omega)$ and the generalized diffusion coefficient $D(q,\omega)$. Following Vollhardt and Wölfle,[6] one can circumvent this difficulty by replacing $D(q,\omega)$ everywhere by $D(\omega)$, i.e., by assuming that spatial dispersion $d(q,\omega)$ is insignificant in the localized phase. It is important to note that this problem has scarcely been studied at all to date.[5] Qualitative estimates of the dependence of the diffusion coefficient on q $D(q,\omega\to 0)$ on the basis of scaling considerations[4] lead to contradictory results and in fact destroy the structure of self-consistent localization theory.

A host of problems arise in this connection: 1) Can the spatial dispersion of the kinetic coefficients be systematically taken into account within the framework of self-consistent localization theory? 2) How does the character of such dispersion change in transition from the metallic to the insulator state? 3) What restrictions are imposed on the Vollhardt–Wölfle theory when spatial dispersion is ignored?

The first serious discussion of these problems appears to have been undertaken in a recent publication,[15] in which it is concluded that spatial dispersion of the diffusion coefficient is insignificant on scales $q\propto 1/d$ (where d is the localization length), and its presence for $q\propto 1/k_F$ (k_F is the Fermi momentum) does not influence the critical behavior near the Anderson transition. Indeed it is stated in Ref. 15 that the Vollhardt–Wölfle theory becomes asymptotically exact in a sufficiently small neighborhood of the mobility threshold (for systems of dimension $d>2$). An important result of Ref. 15 is proof of the fact that, in accordance with the Berezinskiĭ–Gor'kov[16] criterion, $D(q,\omega=0)\equiv 0$ for all q in the localized phase.

In this paper we propose an alternative approach to the solution of the above-stated problems, which represents a direct generalization of the Vollhardt–Wölfle localization theory[6] such as to facilitate investigation of the space-time dispersion of the electrical conductivity $\sigma(q,\omega)$ and the diffusion coefficient $D(q,\omega)$ of a two-dimensional disordered

system in the low-frequency limit $\omega \ll \mathscr{E}_F$ ($\mathscr{E}_F$ is the Fermi energy) and in the long-wavelength limit $q \ll k_F$. Our results agree with the main conclusions of Ref. 15 ($d>2$). In addition, our approach can be used to find explicit low-frequency and long-wavelength asymptotic forms of the kinetic coefficients in the localization state.

We close this discussion with the observation that spatial dispersion of the kinetic coefficients must be taken into account not only for the generalization or substantiation of self-consistent localization theory, but also for the solution of certain problems of practical interest, for example, the influence of carrier localization on the spectrum of long-wavelength ''acoustic'' plasmons in systems containing a quasi-two-dimensional electron gas.[17]

2. STATEMENT OF THE PROBLEM AND GENERAL EQUATIONS

We consider a two-dimensional, degenerate, ideal gas of spinless electrons subject to elastic scattering by immobile impurities, which have a concentration n_I and a Poisson distribution function in the sample. The one-electron Hamiltonian of the system has the form

$$H=\frac{p^2}{2m}+\sum_R U(\mathbf{r}-\mathbf{R}). \tag{1}$$

Here $U(\mathbf{r}-\mathbf{R})$ is the potential of an isolated impurity localized at the point R. We assume that it is short- lived and can be approximated by a delta function, $U(\mathbf{r})=U_0\delta(\mathbf{r})$, which is a good approximation under the condition $r_0 \ll \lambda_F, l$, where r_0 is the range of the potential $U(\mathbf{r})$, λ_F is the de Broglie wavelength, and l is the electron mean free path at the Fermi level. We also assume that the scattering of an electron by the isolated impurity is weak and that the first Born approximation is sufficient for calculating its amplitude.

Inasmuch as the given system is spatially homogeneous on the average, the averaged one-electron Green's function is diagonal in the momentum representation:

$$G^{\pm}_{\mathbf{p}}(\mathscr{E})=\langle\langle \mathbf{p}|R^{\pm}(\mathscr{E})|\mathbf{p}\rangle\rangle_I=[\mathscr{E}-\mathscr{E}_{\mathbf{p}}-\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})]^{-1}, \tag{2}$$

where

$$R^{\pm}(\mathscr{E})=(\mathscr{E}-H\pm i\delta)^{-1}, \quad \delta\to+0, \tag{3}$$

is the resolvent of the Hamiltonian (1), the angle brackets $\langle\rangle_I$ signify averaging over the distribution of the impurities, and $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ is the electron free-energy part, which in the language of Edwards diagrams[18] is defined by the series shown in Fig. 1a. Information on the kinetic properties of the system is contained in the two-particle Green's function

$$\varphi^{\sigma\sigma'}_{\mathbf{pp}'}(\mathbf{q},\omega)=\langle R^{\sigma}(\mathbf{p}_+,\mathbf{p}'_+;\mathscr{E}^+)R^{\sigma'}(\mathbf{p}'_-,\mathbf{p}_-;\mathscr{E}^-)\rangle_I, \tag{4}$$

where $\sigma,\sigma'=\pm$, $\mathbf{p}_\pm=\mathbf{p}\pm\mathbf{q}/2$, $\mathscr{E}^{\pm}=\mathscr{E}\pm\omega/2$, and $R^{\pm}(\mathbf{p},\mathbf{p}';\mathscr{E})$ is the matrix element of the resolvent (3) of the Hamiltonian (1). They satisfy the Bethe–Salpeter equation, which is shown in graphical form in Fig. 1b. The vertex $U^{\sigma\sigma'}_{\mathbf{pp}'}(\mathbf{q},\omega)$ in this equation is defined by the series of irreducible diagrams shown in Fig. 1c.

According to the modern theory of irreversible processes,[19] the kinetic coefficients can be expressed in terms of Kubo correlation functions. In the low-frequency, long-wavelength limit ($\omega \ll \mathscr{E}_F$, $q \ll k_F$) they satisfy the asymptotic expression ($l,m=0,1$)

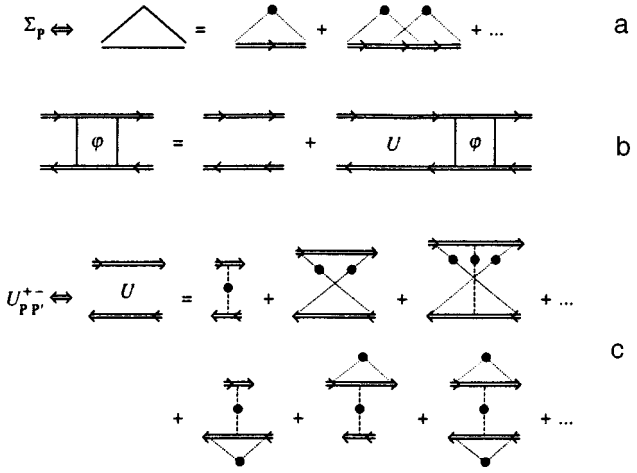

FIG. 1. a) Series of one-particle, irreducible Edwards diagrams[18] for the electron self-energy part $\Sigma_{\mathbf{p}}(\mathscr{E})$; b) graphical form of the Bethe–Salpeter equation for the two-particle Green's function (4); c) series of two-particle, irreducible diagrams for the vertex $U_{\mathbf{pp}'}(\mathbf{q},\omega)$. The conventional rules of correspondence between the analytic expressions for the terms of the perturbation series and their graphical representations are used here.

$$K_{lm}(q,\omega)=-\int d\mathscr{E}\frac{\partial f}{\partial \mathscr{E}}\langle K_{lm}(\mathbf{q},\omega)\rangle_I. \tag{5}$$

Here $f(\mathscr{E})$ is the Fermi–Dirac function,

$$\langle K_{lm}(\mathbf{q},\omega)\rangle=\frac{1}{4\pi}\sum_{\mathbf{p},\mathbf{p}'}\left(\frac{\hat{\mathbf{q}}\cdot\mathbf{p}}{m}\right)^l\left(\frac{\hat{\mathbf{q}}\cdot\mathbf{p}'}{m}\right)^m \times\{2\varphi^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)-\varphi^{++}_{\mathbf{pp}'}(\mathbf{q},\omega)-\varphi^{--}_{\mathbf{pp}'}(\mathbf{q},\omega)\}, \tag{6}$$

the volume (area) of the system is $V=1$, and $\hat{\mathbf{q}}\mathbf{q}=\mathbf{q}/q$.

The correlation functions (5) and (6) satisfy the equations of continuity derived from conservation of the number of particles

$$\begin{aligned} -\omega K_{00}(q,\omega)+qK_{01}(q,\omega)&=-in_F+O\left[\left(\frac{\omega}{\mathscr{E}_F}\right)^2\right],\\ -\omega K_{01}(q,\omega)+qK_{11}(q,\omega)&=O\left[\frac{\omega}{\mathscr{E}_F}\frac{q}{k_F}\right] \end{aligned} \tag{7}$$

(n_F is the density of states at the Fermi level) and are related by simple equations to the generalized kinetic coefficient $L(q,\omega)$ and the conductivity $\sigma(q,\omega)$:

$$L(q,\omega)=\frac{1}{n_F}K_{11}(q,\omega), \quad \sigma(q,\omega)=e^2n_F\,L(q,\omega) \tag{8}$$

and to the Green's function of the diffusion equation

$$G(q,\omega)=\frac{1}{n_F}K_{00}(q,\omega)=\frac{1}{-i\omega+q^2D(q,\omega)}. \tag{9}$$

Here $D(q,\omega)$ is the generalized diffusion coefficient[19]:

$$D(q,\omega)=\frac{L(q,\omega)}{1+(q^2/i\omega)L(q,\omega)}. \tag{10}$$

It is a well-known fact[6] that the diffusion pole $i\omega=q^2D(q,\omega)$ of the Green's function (9) is described by the first three terms in the correlation function (6) ($l=m=0$), whereas the contribution of the last two terms remains finite at the point $\omega=0$, $q=0$. Consequently, all necessary information on the low-frequency and long-wavelength asymptotic representations of our kinetic coefficients is contained in the two-particle Green's function $\varphi^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ (4).

We use the relation

$$\Delta G_{\mathbf{p}}(\mathbf{q},\omega)\Phi^{(k)}(\mathbf{p},\mathbf{q};\omega)=\sum_{\mathbf{p}'}\varphi^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)\left(\frac{\hat{\mathbf{q}}\cdot\mathbf{p}'}{m}\right)^k,$$

$$k=0,1, \tag{11}$$

to determine the relaxation functions $\Phi^{(k)}(\mathbf{p},\mathbf{q};\omega)$ of the density ($k=0$) and the current ($k=1$) in satisfaction of the transport equation

$$\left[\omega-\left(\frac{\mathbf{q}\cdot\mathbf{p}}{m}\right)+\Delta\Sigma_p(\mathbf{q},\omega)\right]\Phi^{(k)}(\mathbf{p},\mathbf{q},\omega)$$
$$=\left(\frac{\hat{\mathbf{q}}\cdot\mathbf{p}}{m}\right)^k+\sum_{\mathbf{p}'}U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)\Phi^{(k)}(\mathbf{p}',\mathbf{q},\omega), \tag{12}$$

which is easily obtained from the Bethe–Salpeter equation[6] for $\varphi^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ (see Fig. 1b). Here we have introduced the notation

$$\Delta G_{\mathbf{p}}(\mathbf{q},\omega)=G^-_{\mathbf{p}_-}(\mathscr{E}^-)-G^+_{\mathbf{p}_+}(\mathscr{E}^+),$$
$$\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)=\Sigma^-_{\mathbf{p}_-}(\mathscr{E}^-)-\Sigma^+_{\mathbf{p}_+}(\mathscr{E}^+). \tag{13}$$

The self-energy part $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ and the kernel of the integral equation (12) $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ are related by the Ward identity[6]

$$\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)=\sum_{\mathbf{p}'}U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)\Delta G_{\mathbf{p}'}(\mathbf{q},\omega). \tag{14}$$

This relation plays an important role in the ensuing calculations; in particular, it ensures conservation of the number of particles.

In contrast with the customary definition of the relaxation functions,[6,8] the delta singularity in them at $|\mathbf{p}|\simeq k_{\mathrm{F}}$ ($q\ll k_{\mathrm{F}}$) is explicitly separated out in Eq. (11). Therefore, assuming that $\Phi^{(k)}(\mathbf{p},\mathbf{q},\omega)$ is a sufficiently smooth function of $\mathbf{p}$ near the Fermi surface, we seek a solution of the transport equation (12) in the Fourier series form

$$\Phi^{(k)}(\mathbf{p},\mathbf{q},\omega)=\sum_{n=0}^{\infty}\Phi^{(k)}_n(k_{\mathrm{F}},q,\omega)\cos(n\theta),$$

$$\theta=\widehat{\mathbf{p},\mathbf{q}}. \tag{15}$$

Substituting Eq. (15) into (10), we readily obtain a system of nonlinear algebraic equations for the Fourier coefficients (for brevity we drop the subscript F from the Fermi momentum from now on):

$$\sum_{n'=0}^{\infty}\left\{(1+\delta_{n0}+\delta_{0n'}-\delta_{n0}\delta_{0n'})\left[\omega\delta_{nn'}-(\delta_{n+1,n'}+\delta_{n,n'+1})\frac{qk}{2m}\right]+M_{nn'}(q,\omega)\right\}\Phi^{(k)}_{n'}(k,q,\omega)$$
$$=\left(\frac{k}{m}\right)^k\delta_{nk}(1+\delta_{n0}). \tag{16}$$

Here

$$M_{nn'}(q,\omega)=\frac{1}{\pi^2}\int_0^{2\pi}d\theta\int_0^{2\pi}d\theta'\cos(n\theta)\cos(n'\theta')$$
$$\times\left\{\pi\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)\delta(\theta-\theta')-\sum_{\mathbf{p}'}U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)\right\} \tag{17}$$

is the matrix of memory functions. The symbol p' under the summation sign in (17) indicates that the summation is carried out only over the modulus of the vector $\mathbf{p}'$. To calculate the conductivity (8) and the diffusion coefficient (10), it is sufficient to know only the one Fourier coefficient $\Phi^{(1)}_1(k,q,\omega)$. It is therefore customary, following the work of Vollhardt and Wölfle,[6,7] to retain only the first two terms ($n=0,1$) in the expansion (15), whereupon a system of two linear equations is obtained in place of (16). It will be shown below that this approximation can be used to calculate the kinetic coefficients only if their spatial dispersion is ignored. Otherwise it is necessary to solve the complete system of equations (16).

It is readily shown with the aid of the Ward identity (14) that $M_{n0}(q,\omega)\equiv 0$ for any n. All other matrix elements (17), generally speaking, are nonvanishing and depend both on the choice of approximation for the vertex function $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ and on the series representing the impurity potential $U(\mathbf{r})$. In the next two sections we analyze the solution of the system of equations (16) and the calculation of the kinetic coefficients of a two-dimensional disordered system in the ladder approximation and in the approximation of self-consistent localization theory.

3. LADDER APPROXIMATION

Retaining only the first terms in the diagram series in Figs. 1a and 1c, we obtain the following expressions for the electron self-energy part $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ and the vertex $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$:

$$U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)=n_I|U_{\mathbf{p}-\mathbf{p}'}|^2,$$
$$U_{\mathbf{p}}=\int U(\mathbf{r})\exp(-i\mathbf{p}\cdot\mathbf{r})d\mathbf{r}, \tag{18}$$
$$\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})=n_I\sum_{p'}|U_{\mathbf{p}-\mathbf{p}'}|^2G^{\pm}_{\mathbf{p}'}(\mathscr{E}).$$

This approximation was first proposed by Edwards,[18] but so far it has been used to calculate the kinetic coefficients without regard for their spatial dispersion. Using Eq. (18), we can easily verify that all off-diagonal elements of the matrix of

memory functions (17) can be ignored in the long-wavelength limit. Indeed, to within terms of order $(q/k)^2$ we have

$$M_{nn}(q,\omega)=\frac{i}{\tau_n}=n_I\sum_{\mathbf{p}'}|U_{\mathbf{p}-\mathbf{p}'}|^2\Delta G_{\mathbf{p}'}(\mathscr{E})\times[1-\cos(n\varphi)], \tag{19}$$

where $\varphi=\theta-\theta'=\widehat{\mathbf{pp}}'$, and τ_n has the significance of the nth-order relaxation time; in particular, $\tau_1\equiv\tau$ is the transport relaxation time. To estimate the relative value of $M_{nn'}(q,\omega)$ $(n\neq n')$, we use a Gaussian model potential with the Fourier transform $U_{\mathbf{p}}=U_0\exp(-p^2r_0^2/2)$. Expanding the off-diagonal elements (17) in powers of the small parameters r_0/λ_{F}, λ_{F}/l, and q/k_{F}, we obtain

$$\left|\frac{M_{n,n\pm1}(q,\omega)}{qv_{\mathrm{F}}}\right|\propto\frac{r_0^2}{\lambda_{\mathrm{F}}^2}\frac{\lambda_{\mathrm{F}}}{l},\quad \left|\frac{M_{n,n\pm k}(q,\omega)}{M_{n,n\pm1}(q,\omega)}\right|\propto\left(\frac{q}{k_{\mathrm{F}}}\right)^{k-1}. \tag{20}$$

Consequently, within the error limits stated here, only the diagonal elements of the matrix of memory functions (17) need to be included in the system of equations (16). It is important to note that the inequalities $r_0\ll\lambda_{\mathrm{F}}\ll l$ are the conditions governing the validity of classical kinetic theory.

In the limit of the delta-function impurity scattering potential $(r_0\to0)$ all the off-diagonal elements $M_{nn'}(q,\omega)$ $(n\neq n')$ are identically zero, and all the relaxation times are equal $(\tau_n=\tau)$, because the integral of the second term over the angle φ vanishes at $U_{\mathbf{p}}=U_0=\mathrm{const}$. These simplifications permit us not only to obtain an exact solution of the system of equations (16), but also to sum the corresponding Fourier series (15)) (see Appendix A):

$$\Phi^{(0)}(\mathbf{p},\mathbf{q},\omega)=\frac{1}{1-(i/\tau)\Phi_0^0(k,q,\omega+i/\tau)}\frac{1}{\omega-\mathbf{q}\cdot\mathbf{p}/m+i/\tau}, \tag{21}$$

$$\Phi^{(1)}(\mathbf{p},\mathbf{q},\omega)=-\frac{1}{q}+\frac{\omega}{q}\Phi^{(0)}(\mathbf{p},\mathbf{q},\omega).$$

Here the symbol $\Phi_0^0(k,q,z)$ denotes the zeroth Fourier coefficient of the density relaxation function without scattering by impurities:

$$\Phi_0^0(k,q,z)=\frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{z-qv_{\mathrm{F}}\cos\theta}=(z^2-q^2v_{\mathrm{F}}^2)^{-1/2}. \tag{22}$$

The factor containing this function in $\Phi^{(0)}(\mathbf{p},\mathbf{q},\omega)$ (21) plays an important role. Because of it, the density and current scattering functions are coupled by the equation of continuity (21) and, hence, conserve the number of particles. A similar structure is found in the solution of the classical kinetic equation for a spatially nonuniform nonequilibrium distribution[20] and in the expression for the density relaxation function obtained by Götze[21] in the interacting-mode approximation.

Using the expressions (21) for the density and current relaxation functions, we readily calculate the generalized diffusion coefficient

$$D(q,\omega)=\frac{2D_0}{1-i\omega\tau+\sqrt{(1-i\omega\tau)^2+2q^2D_0\tau}},$$

$$D_0=\frac{1}{2}v_{\mathrm{F}}^2\tau \tag{23}$$

and, related to it by Eq. (10), the generalized kinetic coefficient (Lq,ω) or the electrical conductivity $\sigma(q,\omega)$. For $q=0$, from Eq. (23) we obtain the usual Drude equations for the frequency-dependent electrical conductivity and diffusion coefficient, which are related to each other by the Einstein relation:

$$\sigma(\omega)=e^2n_{\mathrm{F}}D(\omega)=\frac{\sigma_0}{1-i\omega\tau},\quad \sigma_0=\frac{ne^2\tau}{m}, \tag{24}$$

where σ_0 is the static conductivity, and n is the density of electrons. The spatial dispersion of the kinetic coefficients calculated here is of a less trivial nature. Whereas the generalized diffusion coefficient (23) as a function of ω and q is continuous at $\omega=0$ and $q=0$, the conductivity suffers a discontinuity at this point:

$$\sigma_0=\lim_{\omega\to0}\lim_{q\to0}\sigma(q,\omega)\neq\lim_{q\to0}\lim_{\omega\to0}\sigma(q,\omega)=0. \tag{25}$$

The fact that the second limit in (25) is zero follows from the low-frequency asymptotic behavior of the conductivity:

$$\sigma(q,\omega)\propto-\frac{i\omega}{q^2}e^2n_{\mathrm{F}},\quad \omega\to0,\quad q\neq0, \tag{26}$$

which is readily deduced from relations (8) and (10) with allowance for the finiteness of the static diffusion coefficient $D(q)$ (23).

Consequently, the expressions obtained here for the kinetic coefficients correctly reproduce their behavior in the vicinity of the point $\omega=0$, $q=0$. The equation $\sigma(q,0)\equiv0$ has a simple physical significance. In the inhomogeneous steady state without any forces of nonelectrical origin, the diffusion and drift currents exactly cancel one another, creating zero net current.

4. SELF-CONSISTENT LOCALIZATION THEORY APPROXIMATION

Of utmost importance in the set of diagrams contributing to the irreducible vertex $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ is the series shown in the first row of Fig. 1c (maximum crossing or ''fan'' diagrams). The sum of the series can be determined exactly in the case of scattering of electrons by impurities with a short-lived potential. When this sum is taken into account, $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ acquires a term containing a diffusion pole at the point $i\omega=(\mathbf{p}+\mathbf{p}')^2D_0(|\mathbf{p}+\mathbf{p}'|,\omega)$, where $D_0(q,\omega)$ is the classical diffusion coefficient (23). When the Bethe–Salpeter equation is solved iteratively, this term yields logarithmically divergent (as $\omega\to0$) corrections to the kinetic coefficients. It has been shown[22] that the phenomenon of weak localization in low-dimensional disordered systems is intimately related to these corrections.

This idea has been elaborated by Vollhardt and Wölfle,[6–8] who proposed a self-consistent generalization of the ''fan'' series for the irreducible vortex, augmenting it

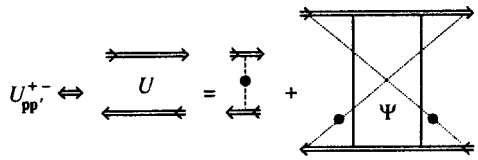

FIG. 2. Graphical form of the equation for the vertex $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ in the approximation of the Vollhardt–Wölfle self-consistent localization theory.[6,7]

with all possible diagrams containing two outwardly intersecting lines of interaction. As a result, the equation shown graphically in Fig. 2 is obtained for $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$. Here the diagram unit between the lines of interaction in the second term on the right-hand side represents the convolution over momenta $\mathbf{k}$ and $\mathbf{k}'$ of the two-particle Green's function in the Cooper channel or in the particle–particle channel (cooperon):

$$\Psi(q,\omega)=\sum_{\mathbf{k},\mathbf{k}'}\langle R^{+}(\mathbf{k}_{+},\mathbf{k}'_{+};\mathscr{E}^{+})R^{-}(-\mathbf{k}_{-},-\mathbf{k}'_{-};\mathscr{E}^{-})\rangle_{I}, \tag{27}$$

where $\mathbf{k}_{\pm}=\mathbf{k}\pm\mathbf{q}/2$ and $\mathbf{q}=\mathbf{p}+\mathbf{p}'$. In systems with time-reversal invariance, however, we have $R^{\sigma}(-\mathbf{k},-\mathbf{k}';\mathscr{E})=R^{\sigma}(\mathbf{k}',\mathbf{k};\mathscr{E})$, so that the cooperon (27) exactly coincides with the Green's function of the diffusion equation $G(q,\omega)$ (9), and the irreducible vertex acquires the form

$$U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)=W+\frac{W}{\tau}\,\frac{1}{-i\omega+(\mathbf{p}+\mathbf{p}')^{2}D(|\mathbf{p}+\mathbf{p}'|,\omega)}, \tag{28}$$

where $W=n_I U_0^2$, $\tau=(2\pi W n_{\mathrm{F}})^{-1}$ is the bare relaxation time, and $D(q,\omega)$ is the exact diffusion coefficient.

In approaching the calculation of the matrix elements of the memory functions (17), we call attention to the fact that according to the Ward identity (14), $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)$ has a term that contains the diffusion pole (28). At first glance, this property should lead to a singularity $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)\propto 1/\omega$ in the localized phase $[D(q,\omega)\propto -i\omega]$, contradicting the well-known analytic properties of the averaged one-particle Green's function (2). For $2<d<4$ this paradox has been resolved recently in Ref. 15, where it is shown that a divergence of the type $1/\omega$ on the right-hand side of the Ward identity (14) in the localized phase is precluded by the approximate (to order ω) orthogonality of the singular part of the vertex $U^{+-}_{\mathbf{pp}'}(\mathbf{q},\omega)$ relative to $\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)$. We therefore replace the first term in (17) by i/τ and, for simplicity, identify τ with the bare relaxation time present in (28), since this substitution does not qualitatively affect the main results.

Here again, as in the ladder approximation, a major role is played by the diagonal elements of the memory function matrix (17), which after simple transformations assume the form

$$M_{nn}(q,\omega)=\frac{i}{\tau}-\frac{W}{2\pi\tau}\int_0^{+\infty}k'\,dk'\times\frac{1}{2\pi}\int_0^{2\pi}\frac{\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)\cos(n\varphi)\,d\varphi}{-i\omega+(\mathbf{p}+\mathbf{p}')^{2}D(|\mathbf{p}+\mathbf{p}'|,\omega)}. \tag{29}$$

The substitution $z=\exp(i\varphi)$ reduces the integral over the polar angle in (29) to an integral over the circle of unit radius $|z|=1$ in the complex plane. We assume that the main contribution to it is from the diffusion pole (this approximation is justified in Appendix B). Then, invoking the residue theorem, we obtain the following expression for a diagonal element of the memory function matrix:

$$M_{nn}(q,\omega)=\frac{i}{\tau}-(-1)^{n}\frac{i}{\tau}\Delta,$$

$$\Delta=\frac{W}{2\pi i}\times\int_0^{+\infty}\frac{\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)k'\,dk'}{\{[-i\omega+(k-k')^{2}\widetilde{D}][-i\omega+(k+k')^{2}\widetilde{D}]\}^{1/2}}. \tag{30}$$

Here $\widetilde{D}$ is the value of the diffusion coefficient at the pole $i\omega=q^2D(q,\omega)$. The asymptotic expression (30) holds in the limit $\omega\to 0$ under the condition $\omega\ll 4k^2|\widetilde{D}|$ $(k\simeq k'\simeq k_{\mathrm{F}})$. This inequality is satisfied automatically in the metallic state, and Eq. (30) remains valid in the localized phase ($\widetilde{D}\propto -i\omega d^2$) for $d\gg\lambda_{\mathrm{F}}$, where d is the localization length. To calculate the integral with respect to k' in Δ (30), we approximately set $\mathscr{E}\simeq k^2/2m$ in $\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)$. If, in addition to the inequality $\omega\ll 4k^2|\widetilde{D}|$, the condition $k^2\gg m/\tau$ (or $l=k\tau/m\gg\lambda_{\mathrm{F}}$) also holds, the lower limit of integration with respect to $x=k'^2-k^2$ can be replaced by $-\infty$. We finally obtain

$$\Delta=\frac{\lambda_{\mathrm{F}}}{\pi^2 l}\,\frac{D_0}{\widetilde{D}}\left(1+4i\omega\frac{l^2}{\widetilde{D}}\right)^{-1/2}\ln\frac{1+(1+4i\omega l^2/\widetilde{D})^{1/2}}{1-(1+4i\omega l^2/\widetilde{D})^{1/2}}. \tag{31}$$

We therefore obtain a system of equations for the Fourier coefficients $\Phi_n^{(0)}(k,q,\omega)$, whose exact solution (see Appendix A) yields the following expression for the desired correlation function:

$$\Phi^{(0)}(\mathbf{p},\mathbf{q},\omega)=\frac{1}{1-(i/\tau)(1-\Delta)\Phi_0^0(k,q,\omega+i/\tau)}\times\frac{\omega+i/\tau+i\Delta/\tau+\mathbf{p}\cdot\mathbf{q}/m}{(\omega+i/\tau)^2+\Delta^2/\tau^2-(\mathbf{p}\cdot\mathbf{q}/m)^2}, \tag{32}$$

where

$$\Phi_0^0\left(k,q,\omega+\frac{i}{\tau}\right)=\left[\left(\omega+\frac{i}{\tau}-i\frac{\Delta}{\tau}\right)^2-q^2v_{\mathrm{F}}^2\frac{\omega+i/\tau-i\Delta/\tau}{\omega+i/\tau+i\Delta/\tau}\right]^{-1/2}. \tag{33}$$

The current relaxation function $\Phi^{(1)}(\mathbf{p},\mathbf{q},\omega)$ is related to $\Phi^{(0)}(\mathbf{p},\mathbf{q},\omega)$ by the continuity equation (21). With the help of these relations and Eqs. (8)–(10) we readily find an expression for the diffusion coefficient;

$$D(q,\omega)=\frac{D_0}{1-i\omega\tau+\Delta}\times\frac{2}{1+\{1+2D_0q^2\tau/[(1-i\omega\tau)^2-\Delta^2]\}^{1/2}}. \tag{34}$$

For $q=0$ the substitution $\widetilde{D}\to D(\omega)$ $[|i\omega l^2/D(\omega)|\ll 1]$ takes Eq. (34) into the well-known equation for the diffusion coefficient $D(\omega)$ obtained on the basis of self-consistent localization theory.[6,7] For $\Delta=0$ Eq. (34) coincides with the expression (23) for the diffusion coefficient in the ladder approximation.

5. DISCUSSION

The value of the diffusion coefficient at the pole $i\omega=q^2D(q,\omega)$ enters into the right-hand side of Eq. (34) through the parameter Δ (31). To calculate $D(q,\omega)$, therefore, it is first necessary to solve the self-consistent equation for $\widetilde{D}$ obtained from (34) for $q^2=i\omega/\widetilde{D}$. In the low-frequency limit its solution has a localized character, $\widetilde{D}\propto -i\omega d^2$. In fact, substituting this asymptotic representation into the equation for $\widetilde{D}$, we readily verify that it exists when $\omega\tau\ll l^2/d^2$, where the localization length d for $l\gg\lambda_F$ has the form

$$d=l\,\exp\left(\frac{\pi^2}{2}\,\frac{l}{\lambda_F}\right),\quad d\gg l\gg\lambda_F. \tag{35}$$

The next iteration of the equation for $\widetilde{D}$ enables us to determine the low-frequency asymptotic behavior of its real part. Thus,

$$\mathrm{Im}\,\widetilde{D}(\omega)\propto -d^2\omega,\quad \mathrm{Re}\,\widetilde{D}(\omega)\propto\pi\frac{m}{\hbar}d^4\omega^2,$$
$$\omega\tau\frac{d^2}{l^2}\ll 1. \tag{36}$$

Equations (35) and (36) do in fact reproduce the results of Vollhardt and Wölfle,[6] differing only by the numerical factor 1/2 in the exponent of (35) and the absence of an artificially introduced cutoff parameter.

Consequently, the inclusion of spatial dispersion in the self-consistent equations for the kinetic coefficient does not alter their low-frequency behavior (36) in the localization state for $q=0$; this fact concurs with one of the conclusions in Ref. 15. However, carrier localization influences the nature of the spatial dispersion of the conductivity $\sigma(q,\omega)$ and the diffusion coefficient $D(q,\omega)$ in different ways. Thus, substituting the asymptotic representations (36) into Eqs. (10) and (34), we obtain

$$\sigma(q,\omega)=e^2n_F\frac{\widetilde{D}(\omega)}{1+d^2q^2},$$
$$D(q,\omega)=\frac{2\widetilde{D}(\omega)}{1+[1+2l_D^2(\omega)q^2]^{1/2}}. \tag{37}$$

Whereas the nonlocality range, which characterizes the spatial dispersion of the diffusion coefficient (23), is equal to the mean free path l (or the classical diffusion length $l_D=l/\sqrt{2}=\sqrt{D_0\tau}$) in the ladder approximation, in the localization state it is equal to $l_D(\omega)=\sqrt{D_0\tau}|\widetilde{D}(\omega)/D_0|\ll l\ll d$. Consequently, in accordance with the conclusions of Ref. 15, the spatial dispersion of $D(q,\omega)$ is insignificant on scales $q\propto 1/d$.

The parameter l_D in the localized phase has the significance of a frequency-dependent diffusion length. According to Eq. (36), the low-frequency asymptotic representation (37) has the form $D(q,\omega)=-i\omega f(q\omega)$, which ensures the suppression of spatial dispersion of the diffusion coefficient in the localized phase ($l_D(\omega)\propto\omega\to 0$). This result is valid over the range of frequencies satisfying the condition $l_D(\omega)>\lambda_F$.[1)]

The conductivity presents a different situation. In the metallic state the nonlocality range characterizing its spatial dispersion is equal to the mean free path l in the limit $q\to 0$, $\omega\neq 0$ (23) and becomes infinite in the limit $\omega\to 0$, $q\neq 0$ (26). In the insulator phase $\sigma(q,\omega)$ (37) as a function of q and ω is continuous at the point $q=0$, $\omega=0$, and the role of the nonlocality range is taken by the localization length d.[2)] Consequently, unlike the diffusion coefficient, the conductivity exhibits appreciable spatial dispersion even on scales $q\propto 1/d$ in the localization state. We assume on this basis that the electrodynamic properties of two-dimensional disordered systems in the localization phase for $d\gg l$ must be much stronger than in the metallic state and depend on spatial dispersion effects.

It helps to understand the difference in the dependences of the diffusion coefficient and the conductivity on q (37) when we consider that they determine different fluxes. Thus, $D(q,\omega)$ as a kinetic coefficient determines the magnitude of fluxes strictly of a diffusion nature, so that the scale of its spatial dispersion is governed by the diffusion length $l_D(\omega)$. The conductivity $\sigma(q,\omega)$ determines the total magnitude of diffusion and drift fluxes, the latter type dominating in the long-wavelength limit. The spatial nonlocality range of the constitutive equation relating the current density to the electric field in the medium is equal to the mean distance over which the electrons ''remember'' their previous states. The coherent scattering of electrons by randomly distributed impurities plays a major role in the localization state, producing an anomalous increase in the backscattering probability ($\mathbf{p}'=-\mathbf{p}$) (Ref. 22) [see (28)]. Of course, the nonlocality radius in this case is determined by the coherence length, i.e., by the average size of the closed loops of self-intersecting trajectories, which is equal to the localization length d in order of magnitude.[3)]

It follows from the asymptotic expressions (36) and (37) that the static limit of the diffusion coefficient is equal to zero in the localization state for any finite wave vector, i.e., $\lim_{\omega\to 0}D(q,\omega)=D(q,0)=0$. This result agrees with the Berezinskiĭ–Gor'kov criterion,[16] which stipulates that the Green's function of the diffusion equation (9) manifests the asymptotic behavior $G(q,\omega)\propto 1/\omega$ in the limit $\omega\to 0$ in the localized state.

The present study is based on the Vollhardt–Wölfle theory, the justification of which poses a timely problem.[5] An important step toward that goal has been made in Ref. 15,

where it is shown that the singular structure of the quantity (28) is a direct consequence of time-reversal symmetry. All the same, there are still a number of problems associated with checking out the approximate procedure for establishing self-consistency.[15] The results obtained here can contribute to the solution of some of them. In particular, inequality (B6) in Appendix B defines the frequency range in which not only the Vollhardt–Wölfle procedure $D(q,\omega)=D(\omega)$, but also (subject to the additional condition $l_D(\omega)>\lambda_F$) the above-proposed approximate scheme for taking into account weak spatial dispersion of the electrical conductivity and the diffusion coefficient (37) is valid. As for the dependence of the kinetic coefficients on q, estimates of our discarded off-diagonal elements of the matrix of memory functions (17) show that the asymptotic representation (37) remains valid in the range $q \ll 1/l$.

The authors are grateful to A. K. Arzhnikov and L. I. Danilov for profitable discussions and support during the present study.

APPENDIX A: SOLUTION OF THE SYSTEM OF RECURSION RELATIONS

In the case of scattering of electrons by impurities with a delta-function potential the system of equations for the coefficients $\Phi_n^{(0)}$ in the ladder approximation has the form

$$\begin{aligned}
&\omega\Phi_0^{(0)}-\frac{qk}{2m}\Phi_1^{(0)}=1,\\
&\frac{qk}{2m}\Phi_0^{(0)}-\left(\omega+\frac{i}{\tau}\right)\Phi_1^{(0)}+\frac{qk}{2m}\Phi_2^{(0)}=0,\\
&\frac{qk}{2m}\Phi_n^{(0)}-\left(\omega+\frac{i}{\tau}\right)\Phi_{n+1}^{(0)}+\frac{qk}{2m}\Phi_{n+2}^{(0)}=0, \quad n\geqslant 1.
\end{aligned} \tag{A1}$$

This is an infinite system of homogeneous difference equations with constant coefficients. We seek a solution in the form[23]

$$\Phi_n^{(0)}=\Phi_0^{(0)}Cz^n. \tag{A2}$$

The substitution of Eq. (A2) into the last equation of the system (A1) yields a quadratic equation in z, which has two solutions, the correct choice being the one that satisfies the condition $\Phi_n^{(0)}(q=0)=0$ $(n\geqslant 1)$

$$z=-\frac{1}{qv_F}\left\{\left[\left(\omega+\frac{i}{\tau}\right)^2-q^2v_F^2\right]^{1/2}-\left(\omega+\frac{i}{\tau}\right)\right\}. \tag{A3}$$

The undetermined constant C and the zeroth Fourier coefficient $\Phi_0^{(0)}$ can be found by means of the first two equations in the system (A1), which serve as boundary conditions for the system of difference equations. The Fourier series obtained by substituting the resulting solution into (15) reduces to a geometric progression and is easily summed.

The system of equations for the coefficients $\Phi_n^{(1)}$ differs from (A1) only in the boundary conditions:

$$\begin{aligned}
&\omega\Phi_0^{(1)}-\frac{qk}{2m}\Phi_1^{(1)}=0,\\
&\frac{qk}{m}\Phi_0^{(1)}-\left(\omega+\frac{i}{\tau}\right)\Phi_2^{(1)}+\frac{qk}{2m}\Phi_2^{(1)}=-\frac{k}{m}
\end{aligned} \tag{A4}$$

and is solved analogously. The final expressions for the density and current relaxation functions in the ladder approximation have the form (21) and (22)).

According to Eq. (30), the system of equations for the Fourier coefficients $\Phi_n^{(0)}$ in the self-consistent localization theory approximation has the form

$$\begin{aligned}
&\omega\Phi_0^{(0)}-\frac{qk}{2m}\Phi_1^{(0)}=1,\\
&\frac{qk}{m}\Phi_0^{(0)}-\left(\omega+\frac{i}{\tau}+\frac{i}{\tau}\Delta\right)\Phi_1^{(0)}+\frac{qk}{2m}\Phi_2^{(0)}=0,\\
&\frac{qk}{2m}\Phi_{2k-1}^{(0)}-\left(\omega+\frac{i}{\tau}-\frac{i}{\tau}\Delta\right)\Phi_{2k}^{(0)}+\frac{qk}{2m}\Phi_{2k+1}^{(0)}=0,\\
&\frac{qk}{2m}\Phi_{2k}^{(0)}-\left(\omega+\frac{i}{\tau}+\frac{i}{\tau}\Delta\right)\Phi_{2k+1}^{(0)}+\frac{qk}{2m}\Phi_{2k+2}^{(0)}=0, \quad k\geqslant 1,
\end{aligned} \tag{A5}$$

where Δ is defined in Eq. (31). It is easily shown, using the last two equations of the system (A5), that all the Fourier coefficients of odd or even order are interrelated by the homogeneous second-order recursion relations

$$\begin{aligned}
&\left(\frac{qk}{2m}\right)^2\Phi_n^{(0)}-\left[\left(\omega+\frac{i}{\tau}\right)^2+\frac{\Delta^2}{\tau^2}-2\left(\frac{qk}{2m}\right)^2\right]\Phi_{n+2}^{(0)}\\
&\qquad+\left(\frac{qk}{2m}\right)^2\Phi_{n+4}^{(0)}=0,
\end{aligned} \tag{A6}$$

where $n=2k-1$ or $n=2k$ $(k\geqslant 1)$. In this case, by analogy with (A2), the solution of the system of equations (A6) must be of the form

$$\Phi_{2k}^{(0)}=\Phi_0^{(0)}Cz^{2k}, \quad \Phi_{2k-1}^{(0)}=\Phi_0^{(0)}Bz^{2k-1}. \tag{A7}$$

The substitution of these expressions into (A6) yields a quadratic equation in z^2. Its solution subject to the condition $\Phi_n^{(0)}(q=0)=0$ $(n\geqslant 1)$ has the form

$$z=\frac{1}{qv_F}\left\{\left[\left(\omega+\frac{i}{\tau}\right)^2+\frac{\Delta^2}{\tau^2}-q^2v_F^2\right]^{1/2}-\left[\left(\omega+\frac{i}{\tau}\right)^2+\frac{\Delta^2}{\tau^2}\right]^{1/2}\right\}. \tag{A8}$$

The constants C and B are again determined from the boundary conditions, i.e., from the first two equations in the system (A5). The system of equations for the coefficients $\Phi_n^{(1)}$ is solved analogously. The final expressions for the density and current relaxation functions in the self-consistent localization theory approximation have the form (32) and (33).

In closing, we note that to calculate the diffusion coefficient $D(\omega)$ and the conductivity $\sigma(\omega)$ without regard for spatial dispersion, it is sufficient to retain only the first pairs of equations in the systems (A1) and (A5) and to set $\Phi_n^{(0)}\equiv 0$ $(n\geqslant 2)$. In this case Eq. (A1) gives the well-known Drude limit (24)), and Eq. (A5) gives the Vollhardt–Wölfle approximation[6,7] (see relation (34) with $q=0$).

APPENDIX B: ESTIMATION OF THE INTEGRALS ALONG THE BRANCH CUTS

The explicit expression (37) for the low-frequency asymptotic behavior of the generalized diffusion coefficient $D(q,\omega)$ can be used to estimate the contribution from the previously discarded (in calculating the matrix elements of the memory functions (29)) integrals along the branch cuts. After substituting (37) into the integral over the polar angle φ (29), we set

$$q^2=(\dot{\mathbf{p}}+\mathbf{p}')^2\simeq 2k^2+k^2(z+z^{-1}),\quad z=\exp(i\varphi). \tag{B1}$$

In the z plane the integrand is symmetric under the inversion $z\to z^{-1}$ and has two branch cuts C_1 ($z_1\leqslant z\leqslant 0$) inside the unit circle $|z|=1$ and C_2 ($-\infty\leqslant z\leqslant z_2$) outside this circle. Here z_1 and z_2 are the roots of the equation

$$1+2l_D^2(\omega)k^2(2+z+z^{-1})=0, \tag{B2}$$

$z_1,z_2=1$, $|z_1|<1$ and $|z_2|>1$.

After straightforward manipulations the contribution from the branch cuts to $M_{nn}(q,\omega)$ (29) can be written in the form

$$M_{nn}^{\mathrm{cut}}\simeq \frac{1}{2\omega\tau^2}\frac{1}{2\pi i}\oint_{C_1} dz\; z^{n-1/2}\frac{\sqrt{2l_D^2(\omega)k^2(z+1)^2+z}}{d^2k^2(z+1)^2+z}. \tag{B3}$$

Only terms that branch at the points $z=0$ and $z=z_1$ are retained in the integrand of Eq. (B3). Moreover, terms small with respect to the parameter $l_D^2(\omega)/d^2\ll 1$ in the localization state are omitted in the denominator. To obtain an upper bound for the modulus of M_{nn}^{cut}, we replace the numerator of the integrand in (B3) by its maximum value at the point $z=0$ and replace the denominator by its minimum value at the point $z=z_1$. In conjunction with the inequality $|z_1|<1$ this operation yields

$$|M_{nn}^{\mathrm{cut}}|<\frac{\lambda_{\mathrm{F}}l_D(\omega)}{8\sqrt{2}\pi^2\omega\tau^2}\left(\frac{1+\sqrt{1+8l_D^2(\omega)k^2}}{d}\right)^2\frac{2}{2n+1}, \tag{B4}$$

where $n=1, 2, 3, \ldots$, and $k=2\pi/\lambda_{\mathrm{F}}$.

Consequently, over the range of frequencies satisfying the condition $l_D<\lambda_{\mathrm{F}}$ we obtain the following upper bound for the ratio of the contributions from the branch cuts and the diffusion pole to the matrix elements of the memory functions (29):

$$\frac{|M_{nn}^{\mathrm{cut}}|}{|M_{nn}^{\mathrm{pol}}|}<\frac{4}{2n+1}\left|\frac{\widetilde{D}(\omega)}{D_0}\right|^3\frac{l}{\lambda_{\mathrm{F}}}=\frac{4\sqrt{2}}{2n+1}\left|\frac{\widetilde{D}(\omega)}{D_0}\right|^2\frac{l_D(\omega)}{\lambda_{\mathrm{F}}}. \tag{B5}$$

Despite the two competing factors in (B5) [$\lambda_{\mathrm{F}}<l_D(\omega)\ll l$, $|\widetilde{D}(\omega)|\ll D_0$], there is a frequency interval in which

$$\frac{1}{\sqrt{2}}\left|\frac{\widetilde{D}(\omega)}{D_0}\right|^3\frac{l}{\lambda_{\mathrm{F}}}=\left|\frac{\widetilde{D}(\omega)}{D_0}\right|^2\frac{l_D(\omega)}{\lambda_{\mathrm{F}}}\ll 1, \tag{B6}$$

so that the principal contribution to the matrix elements of the memory function is from the diffusion pole. Consequently, the scheme developed above for the calculation of $D(q,\omega)$ works when the inequality (B6) is satisfied. The same condition defines the limits of validity of the Vollhardt–Wölfle approximation[6–8] $D(q,\omega)=D(\omega)$. Otherwise the complex analytic structure of the generalized diffusion coefficient can significantly alter the value of the integral (29).

[1)] This restriction is a consequence of the wave nature of the laws governing the motion of microscopic particles. Estimates in Appendix B (see Eq. (B6)) show that for $l\gg\lambda_{\mathrm{F}}$ there is a frequency interval in which $\lambda_{\mathrm{F}}<l_D\ll l$ and $|\widetilde{D}(\omega)|\ll D_0$ hold simultaneously.

[2)] The only effect of including spatial dispersion of the kinetic coefficients (37) in the calculation of $\sigma(q,\omega)$ is a weak renormalization of d in (37).

[3)] In our opinion, these considerations do not apply to spatial nonlocality of the diffusion coefficient, because the concentration gradient is not a force acting on electrons, and the relationship between it and the magnitude of the diffusion flux is statistical in nature.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials*, 2nd ed., Clarendon Press, Oxford (1979).
[3] Y. Nagaoka and H. Fukuyama (eds.), *Anderson Localization*, Springer-Verlag, Berlin-New York (1982).
[4] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[5] M. V. Sadovskiĭ, Sverkhprovodimost' (KIAE) **8**, 337 (1995).
[6] D. Vollhardt and P. Wölfle, Phys. Rev. B **22**, 4666 (1980).
[7] D. Vollhardt and P. Wölfle, Phys. Rev. Lett. **45**, 842 (1980).
[8] P. Wölfle and D. Vollhardt, in *Anderson Localization*, Y. Nagaoka and H. Fukuyama (eds.), Springer-Verlag, Berlin-New York (1982), p. 26.
[9] F. J. Wegner, Z. Phys. B. **25**, 327 (1976).
[10] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).
[11] D. Yoshioka, Y. Ono, and H. Fukuyama, J. Phys. Soc. Jpn. **50**, 3419 (1981).
[12] D. Yoshioka, in *Anderson Localization*, Y. Nagaoka and H. Fukuyama (eds.), Springer-Verlag, Berlin-New York (1982), p. 44.
[13] A. Theumann and M. A. Pires Idiart, J. Phys. Condens. Matter **3**, 3765 (1991).
[14] É. Z. Kuchinskiĭ and M. V. Sadovskiĭ, Sverkhprovodimost' (KIAE) **4**, 2278 (1991).
[15] I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].
[16] V. L. Berezinskiĭ and L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **77**, 2498 (1979) [Sov. Phys. JETP **50**, 1209 (1979)].
[17] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[18] S. F. Edwards, Philos. Mag. **8**, 1020 (1958).
[19] D. N. Zubarev, ''Modern methods of the statistical theory of nonequilibrium processes,'' in *Contemporary Problems in Mathematics* [in Russian], Vol. 15, VINITI AN SSSR (1979).
[20] V. P. Silin, *Introduction to the Kinetic Theory of Gases* [in Russian], Nauka, Moscow (1971).
[21] W. Götze, Philos. Mag. **43**, 219 (1981).
[22] L. P. Gor'kov, A. I. Larkin, and D. E. Khmel'nitskiĭ, JETP Lett. **30**, 228 (1979).
[23] A. A. Samarskiĭ and A. V. Gulin, *Numerical Methods* [in Russian], Nauka, Moscow (1989).

Translated by James S. Wood

Nonlocal Josephson electrodynamics of plates of finite thickness

Yu. E. Kuzovlev and A. I. Lomtev

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine

(Submitted 3 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1803–1809 (May 1997)

A nonlocal integrodifferentical equation describing the electrodynamics of a Josephson junction between superconductors of finite thickness in the direction of the magnetic field is derived. It is shown that the interaction of kinks always exhibits long-range power-law asymptotic behavior, which can strongly influence the motion of vortices in and the current–voltage characteristic of even a thick contact. The spectrum of low-amplitude excitations is studied.
 [S1063-7761(97)01805-2]

1. The nonlocal electrodynamics of Josephson junctions has undergone rapid development in recent years. Three cases have been investigated thus far: 1) a tunnel junction at a butt-joint of two ultrathin superconducting films whose thickness is much less than the London length; 2) a tunnel junction between bulk superconductors whose thickness is much greater than the London length; and, 3) a tunnel junction between superconducting layers of finite thickness in a direction perpendicular to the magnetic field. For example, in Refs. 1–8 it is shown that nonlocality effects can be large even in junctions with a large thickness d ($d \gg \lambda$, where λ is the London penetration depth) in the direction of the magnetic field (along the direction of the vortices), i.e., in cases previously studied in the local approximation. In the opposite case of junctions in thin films ($d \ll \lambda$), nonlocality becomes a decisive factor. The corresponding equations were derived and studied in Refs. 9–12.

A Josephson junction between two superconducting layers of finite thickness in a direction perpendicular to the magnetic field of the vortices was studied in Ref. 13. Nonetheless, the theory remains inadequately developed. It is necessary to go beyond the limiting cases mentioned above, since in practice junctions whose size in the direction of the orientation of the Josephson vortices is comparable to λ (such a geometry is realized in, for example, single-crystal Y–Ba–Cu–O chips with twins) are often used in practice. The present paper gives a derivation and a preliminary analysis of the equations for a Josephson junction in a plate with an arbitrary ratio d/λ.

It is shown that the relation between the jump in the phase of the order parameter at the junction and the current density always contains a strongly nonlocal component due to the long-range character of the field in free space, the amplitude of the component exhibiting only a weak (approximately linear) dependence on the parameter λ/d, and its form being completely independent of λ/d. In view of the latter circumstance, the nonlinear component of the current, albeit small in amplitude, can play an important role in several respects: first, to take account of the total transport current of the junction and the external field in the boundary conditions correctly and, second, in describing structures consisting of parallel Josephson junctions connected in series, since the nonlocal long-range coupling of the phase and current extends, of course, to the interjunction interaction. Furthermore, in sufficiently wide junctions with nonballistic transport of vortices, the indicated long-range action should be reflected (see below) in the form of the current–voltage characteristic of the contact at low voltages and currents.

We reduce the problem posed above to a problem of an Abrikosov vortex in a plane-parallel plate. Although this problem has already been studied in Refs. 14–16, for our purposes additional calculations were required. The final results concern a butt-joint contact (the plane of the joint is perpendicular to the plane of the plate), but the formulas presented below make it possible to extend the results to the case of a beveled junction and to pass to the limiting case of a lap-joint junction.

2. The magnetic field of a linear Abrikosov vortex (in the London approximation) satisfies the equation

$$\mathbf{h}-\lambda^2\Delta\mathbf{h}=\boldsymbol{\nu}(\mathbf{r}), \tag{1}$$

$$\boldsymbol{\nu}(\mathbf{r})\equiv\frac{\Phi_0}{2\pi}\,\mathrm{curl}\,\nabla\theta=\frac{\Phi_0}{2\pi}\int\delta(\mathbf{r}-\mathbf{R}(p))\,d\mathbf{R}(p)$$

in the superconductor, where θ is the phase of the order parameter and $\mathbf{R}(p)$ is the parametrically defined radius vector of the points of the vortex core. The core of the Josephson vortex spreads out along the weak-link surface S dividing the superconductor, i.e., it is a two-dimensional object. The corresponding source $\boldsymbol{\nu}(\mathbf{r})$, ''smeared'' over this surface, in Eq. (1) can be expressed, as one can easily show, by the formula

$$\boldsymbol{\nu}(\mathbf{r})=\frac{\Phi_0}{2\pi}\int\delta(\mathbf{r}-\mathbf{R}(a,b))\left[\frac{\partial\varphi}{\partial\mathbf{R}}d\mathbf{S}(a,b)\right], \tag{2}$$

where a and b are the arguments of the parametric representation of S, $\mathbf{R}(a,b)$ is the radius vector of the points on S, $d\mathbf{S}(a,b)$ is a vector element of area of S, and φ is the phase difference between the edges of the junction. Since in any case $\mathrm{div}\,\boldsymbol{\nu}=0$, this source can either be represented as a continuous sum along linear cores (and $\mathbf{h}$ can be represented as a linear combination of the fields of the Abrikosov vortices) or, conversely, it can be treated as a three-dimensional vector field. For a butt-joint junction with the magnetic flux directed

perpendicular to the plate, orienting the z axis along the thickness and the x axis along the junction, we have from Eq. (2)

$$\nu_z(\mathbf{r})=\frac{\Phi_0}{2\pi}\delta(y)\frac{\partial\varphi(x)}{\partial x},\quad \nu_x=\nu_y=0. \tag{3}$$

Here, in accordance with the condition $\mathrm{div}\boldsymbol{\nu}=0$, the phases at the edges and the phase jumps $\varphi(x)=\theta(x,+0)-\theta(x,-0)$ are independent of the coordinate z.

Ultimately, the phase jump and thereby the source will be found by solving the complete nonlinear equation for the junction. But since Eq. (1) is linear, we can write the field as $\mathbf{h}=\mathbf{H}_m+\mathbf{H}$, where $\mathbf{H}_m$ is the ''seed'' Meissner field, which is engendered by the prescribed transport supercurrent and the external magnetic field and is determined by solving the homogeneous equation (1) as if there were no weak link at all and the superconductor were continuous, and $\mathbf{H}$ is engendered by the source (it vanishes for $\boldsymbol{\nu}=0$). Taking a two-dimensional Fourier transform in the plane of a plate of thickness $2d$, $|z|<d$, we find

$$\mathbf{H}=\mathbf{H}_0+\mathbf{a}\exp(\kappa z)+\mathbf{b}\exp(-\kappa z), \tag{4}$$

$$\mathbf{H}_0=-\int_{-d}^{d}\frac{\sinh\kappa|z-z'|}{2\kappa\lambda^2}\boldsymbol{\nu}(\mathbf{k},z')dz',$$

$$\kappa=(\lambda^{-2}+k^2)^{1/2},\quad k^2=k_x^2+k_y^2, \tag{5}$$

where $\mathbf{k}$ is a two-dimensional wave vector. The vector coefficients $\mathbf{a}$ and $\mathbf{b}$ are determined, first, by the condition that the field be divergence-free and continuous at the edge of the plate and, second, by the potential nature of the tangential component of the magnetic field at the boundary of the superconductor (which means that the component of the current normal to the surface vanishes), and by the potential nature of all three components of the field in free space.

From the latter condition we obtain

$$\mathbf{H}_\parallel/H_z|_{z=\pm d}=\mp i\mathbf{k}/|\mathbf{k}|, \tag{6}$$

which accounts for the effect of free space on the field and current distributions in the superconductor. Here and below the subscript $\parallel$ denotes the x and y projections of the vectors. All conditions taken together yield

$$\begin{aligned}
a_z&=\Delta^{-1}[-(\partial H_{0z}/\partial z+kH_{0z})|_{z=d}(\kappa+k)\exp(\kappa d)\\
&\quad+(\partial H_{0z}/\partial z-kH_{0z})|_{z=-d}(\kappa-k)\exp(-\kappa d)],\\
b_z&=\Delta^{-1}[-(\partial H_{0z}/\partial z-kH_{0z})|_{z=-d}(\kappa+k)\exp(\kappa d)\\
&\quad-(\partial H_{0z}/\partial z+kH_{0z})|_{z=d}(\kappa-k)\exp(-\kappa d)],\\
\Delta&=2(\kappa^2+k^2)\sinh(2\kappa d)+4\kappa k\cosh(2\kappa d),
\end{aligned} \tag{7}$$

$$\begin{aligned}
\mathbf{a}_\parallel&=(2\sinh(2\kappa d))^{-1}\{-i\hat{\mathbf{k}}[H_z|_{z=d}\exp(\kappa d)\\
&\quad+H_z|_{z=-d}\exp(-\kappa d)]-\mathbf{H}_{0\parallel}|_{z=d}\exp(\kappa d)\\
&\quad+\mathbf{H}_{0\parallel}|_{z=-d}\exp(-\kappa d)\},\\
\mathbf{b}_\parallel&=(2\sinh(2\kappa d))^{-1}\{i\hat{\mathbf{k}}[H_z|_{z=d}\exp(-\kappa d)\\
&\quad+H_z|_{z=-d}\exp(\kappa d)]+\mathbf{H}_{0\parallel}|_{z=d}\exp(-\kappa d)\\
&\quad-\mathbf{H}_{0\parallel}|_{z=-d}\exp(\kappa d)\},
\end{aligned} \tag{8}$$

where, as a simplification, we set $\hat{\mathbf{k}}\equiv\mathbf{k}/|\mathbf{k}|$ and $k\equiv|\mathbf{k}|$. The formulas (4)–(8) make is possible to find the supercurrent and the field (inside the plate) of an arbitrary source.

Substituting (3) into Eq. (5) and transforming to the coordinate representation, we obtain for the average (over the thickness of the plate) current density through the junction from Eqs. (4)–(8) $j(x)=J_m(x)+J(x)$, where J_m is the ''seed'' Meissner current, determined by the field H_m, and

$$J(x)=\frac{c\Phi_0}{16\pi^3\lambda^2}\frac{\partial}{\partial x}\int Q(x-x')\frac{\partial}{\partial x'}\varphi(x')dx' \tag{9}$$

is the current due to the source. The nonlocal source–current coupling kernel here is given by

$$Q(x)=K_0\left(\frac{|x|}{\lambda}\right)+\frac{1}{d\lambda^2}\int_0^\infty\frac{dk\ J_0(kx)}{\kappa^3(\kappa+k\coth\kappa d)}, \tag{10}$$

where K_0 and J_0 are a modified Bessel function and a Bessel function of order zero. Here the first term corresponds to the limit of two bulk superconductors of thickness $d\gg\lambda$ and is the kernel, first obtained in Ref. 1 and extrapolated in Refs. 2–8, in the integral term of the equation. In the opposite limit of ultrathin films of thickness $d\ll\lambda$, the sum of both terms gives the kernel, first studied in Refs. 9–11, in the integral term of the equation and equal to

$$Q(x)+\frac{\lambda_{\mathrm{eff}}}{\pi}\int_0^\infty dk\frac{1}{1+2k\lambda_{\mathrm{eff}}}J_0(kx),$$

where $\lambda_{eff}=\lambda^2/2d$ is the Peierls penetration depth.

A closed equation for the phase difference at the junction can be obtained, as usual, by equating $j(x)$ to the sum of the Josephson supercurrent, the normal current, and the capacitive displacement current, all regarded as internal characteristics of the contact, and has the form, in standard notation,

$$\begin{aligned}
&\sin\varphi+\frac{\beta}{\omega_J^2}\frac{\partial\varphi}{\partial t}+\omega_J^{-2}\frac{\partial^2\varphi}{\partial t^2}\\
&\quad=\frac{J_m(x)}{j_c}+\frac{\lambda_J^2}{\pi\lambda}\frac{\partial}{\partial x}\int Q(x-x')\frac{\partial}{\partial x'}\varphi(x')dx'.
\end{aligned} \tag{11}$$

Here j_c is the Josephson current density, λ_J and ω_J are the Josephson length and frequency, and β is a dissipative parameter. The integral kernel $Q(x)$ describes excitations in a Josephson junction in a consistent manner, both in a thin film and in a sample which is thick in the direction of the magnetic field. In the general intermediate case, it consists of the sum of a well-localized and nonintegrable strongly nonlocal term (second term on the right-hand side in Eq. (10)), which originates with the slowly decreasing tangential component of the magnetic field at the surface of the plate. It is easy to see that for $|x|\gg\lambda$ the kernel has the asymptotic form

$$Q(x)\simeq\frac{\lambda^2}{d|x|}. \tag{12}$$

Accordingly, as follows from Eqs. (9) and (11), the current density of an isolated static kink and the phase of the current (near their limiting values) far from the center of the kink decrease as $|x|^{-2}$.

3. We shall now discuss the role of the first term on the right-hand side of Eq. (11). Since the total current flowing from an arbitrary source (just like the current of an individual Abrikosov vortex) through an arbitrary cross section of the superconductor equals zero, the integral of the second term in Eq. (11) over x should be zero. Therefore the first term introduces the transport current into the equation. It is clear that this holds as well for a plate of finite width (in the direction of the x axis), though the kernel in Eqs. (9) and (11) loses translational invariance and becomes a function of two arguments, $Q(x,x')$. However, our ansatz (as for other authors) that the Josephson vortices are oriented across the thickness of the plate presumes implicitly that the thickness is small compared with the width (otherwise the vortices generated by the transport current in the absence of an external field would in any case be aligned along the x and not the z axis). Then the current $J_m(x)$ is concentrated mainly at the edges of the plate, and in the local approximation (for $d \gg \lambda$), in which the second term in Eq. (10) can be dropped, the kernel as a whole is replaced by $Q(x) \to \pi\lambda\,\delta(x)$, and the right-hand side of Eq. (11) can be put into the form $\lambda_J^2 \partial^2\varphi/\partial x^2$ everywhere except at the edges; then the total transport current (the integral of the right-hand side of Eq. (11)) can be introduced in the form of boundary conditions on the gradient $\partial\varphi/\partial x$ of the phase.

In taking account of the nonlocality, in order to formulate the boundary conditions correctly, the real distribution $J_m(x)$ of the seed current, which is determined to the same extent by the geometry of the system as is the form of the kernel, must obviously also be taken into consideration. But then the edges of the plate, i.e., the finiteness of its width, should also be taken into account. This problem merits a separate analysis. Nonetheless, even without solving the problem, we note that under certain conditions, even in the case $d \gg \lambda$, nonlocality is capable of strongly influencing the current–voltage characteristic of the resistive state of a wide junction at low voltages, when the Josephson supercurrent through the junction is greater than the normal (''quasiparticle'') current.

We shall regard a junction as wide if its width is much greater than the characteristic distance traversed by an individual kink before it stops as a result of viscosity due to the dissipation of the normal current in the junction and, moreover, much greater than λ_J. For $d \gg \lambda$ we treat the nonlocality as a weak perturbation of the dynamics of the resistive state, described in the zeroth approximation by Eq. (11) with the right-hand side replaced by $\lambda_J^2 \partial^2\varphi/\partial x^2$.

As shown in Ref. 17, the viscous (nonballistic) character of the motion of the vortices gives rise to a strong nonuniformity of the vortex density (i.e., of the magnetic field and gradient of the phase); these quantities all grow toward the edges of the junction approximately as the distance from the center of the contact. At low voltages in the central region of the junction, where the density is low, the vortices are far from one another and their shape is close to that of an isolated kink. The interaction of such vortices and their corresponding viscous velocity toward the center of the junction (where they annihilate with antivortices moving from the other side), as well as the time-averaged slip rate of the phase $\partial\varphi/\partial t$ and the voltage on the junction in the indicated local approximation, are proportional to the small factor $\exp(-\Delta x/\lambda_J)$ (Δx is the characteristic distance between the vortices), i.e., they are exponentially weak. Thus, the central part of a junction plays the role of a ''bottleneck'' that determines the current–voltage characteristic of the junction in the given regime. Since Δx is approximately inversely proportional to the transport current, it is easy to see that the Josephson frequency will be an exponential function of the reverse current, and the dependence of the transport current I on the voltage U will have the qualitative form $I \simeq I_0[\ln(U_0/U)]^{-1}$ with some constants I_0 and U_0.

However, the nonlocal interaction of the vortices can qualitatively change this picture. Since in the zeroth approximation $\partial\varphi/\partial x \sim x$ far from the center of the junction (if $x=0$ corresponds to the center) and, on account of Eq. (12), at the center of the junction the contribution of the second term in Eq. (11) is determined by its periphery, the nonlocality, as is clear from Eqs. (11) and (12), effectively results in the appearance in Eq. (11) of a distributed current source that pushes the vortices toward the center of the junction. As a result, the viscous transport of flux should accelerate and, accordingly, the actual current–voltage characteristic $I(U)$ at low voltages should become smoother (compared with the local-approximation estimate). It is important that such a correction can be very substantial even for a thick plate, since the nonlocal correction, competing with the exponentially weak ''local'' interaction, decreases with increasing thickness only linearly, and the effect of the nonlocality on the form of the IVC should intensify, together with the parameter λ^2/d, with increasing temperature.

4. Let us examine the low-amplitude electromagnetic excitations propagating along the junction. The dispersion relation follows from Eq. (11):

$$\omega = \omega_J \left[1 + \frac{\lambda_J^2 q^2}{(1+\lambda^2 q^2)^{1/2}} + \frac{\lambda_J^2 q^2}{\pi\lambda} F(q) \right]^{1/2},$$
$$F(q) = \frac{1}{d\lambda^2} \int_q^\infty \frac{1}{\kappa^3} \frac{1}{[\kappa + k\coth(\kappa d)]} \frac{1}{(k^2 - q^2)^{1/2}}\, dk. \tag{13}$$

The function $F(q)$ has the asymptotic form

$$F(q) \simeq \frac{2}{3d\lambda^2 q^4} (q \to \infty), \quad F(q) \simeq 4\lambda_{\text{eff}} \ln \frac{1}{\lambda q} (q \to 0).$$

From Eq. (12) we have for the corresponding asymptotic behavior of the spectrum of electromagnetic waves

$$\omega \simeq \omega_J \lambda_J q (1+\lambda^2 q^2)^{-1/4}$$

for sufficiently large wave numbers and

$$\omega \simeq \omega_J \left\{ 1 + \lambda_J^2 q^2 \left[1 + \frac{2\lambda}{\pi d} \ln \frac{1}{\lambda q} \right] \right\}^{1/2}$$

for long-wavelength excitations.

In closing, I thank Yu. A. Genenko for helpful discussions and Yu. V. Medvedev for his interest and support.

[1] Yu. M. Aliev, V. P. Silin, and S. A. Uryupin, Sverkhprovodimost' **5**, 228 (1992) [Superconductivity **5**, 230 (1992)].
[2] A. Gurevich, Phys. Rev. B **46**, 3187 (1992).
[3] Yu. M. Aliev, V. P. Silin, and S. A. Uryupin, JETP Lett. **57**, 193 (1993).
[4] Yu. M. Aliev and V. P. Silin, Zh. Éksp. Teor. Fiz. **104**, 2526 (1993) [JETP **77**, 142 (1993)].
[5] Yu. M. Aliev and V. P. Silin, Phys. Lett. A **117**, 259 (1993).
[6] V. P. Silin, JETP Lett. **58**, 701 (1993).
[7] G. L. Alfimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **106**, 671 (1994) [JETP **79**, 369 (1994)].
[8] V. P. Silin, JETP Lett. **60**, 460 (1994).
[9] Yu. M. Ivanchenko and T. K. Soboleva, JETP Lett. **51**, 114 (1990).
[10] Yu. M. Ivanchenko and T. K. Soboleva, Phys. Lett. A **147**, 65 (1990).
[11] Yu. M. Ivanchenko and T. K. Soboleva, Fiz. Tverd. Tela (Leningrad) **32**, 2029 (1990) [Sov. Phys. Solid State **32**, 1181 (1990)].
[12] R. G. Mints and I. B. Snapiro, Phys. Rev. B **51**, 3054 (1995).
[13] I. O. Kulik and I. K. Yanson, *Josephson Effect in Superconducting Tunnel Structures* [in Russian], Nauka, Moscow (1970).
[14] Yu. M. Ivanchenko, A. I. Kozinskaya, and Yu. V. Medvedev, Fiz. Tverd. Tela (Leningrad) **21**, 3445 (1979) [Sov. Phys. Solid State **21**, 1989 (1979)].
[15] M. Fusko-Girard and F. Mancini, Physica BC **123**, 75 (1983).
[16] A. Yu. Martynovich, Zh. Éksp. Teor. Fiz. **105**, 912 (1994) [JETP **78**, 489 (1994)].
[17] A. E. Filippov, Yu. E. Kuzovlev, and T. K. Soboleva, Phys. Lett. A **183**, 123 (1993).

Translated by M. E. Alferieff

Electromagnetic excitation of ultrasound in magnetically ordered dielectrics

V. D. Buchel'nikov and Yu. A. Nikishin

Chelyabinsk State University, 454136 Chelyabinsk, Russia

A. N. Vasil'ev

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 24 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1810–1816 (May 1997)

The electromagnetic excitation of sound in magnetically ordered dielectrics—ferro- and antiferromagnets—is investigated theoretically. It is shown that sound generation in dielectrics by the Lorentz mechanism (displacement current) is much less efficient than in metals. The magnetoelastic mechanism of sound excitation in dielectrics is just as efficient as in metals. In antiferromagnets the amplitude of the excited sound depends on the relaxation parameter in the magnetic subsystem. The sound excitation efficiency increases as the orientational phase transition point or the ferromagnetic resonance frequency is approached. © *1997 American Institute of Physics.* [S1063-7761(97)01905-7]

1. INTRODUCTION

A great deal of theoretical and experimental work has been devoted to noncontact excitation of ultrasound in magnetically ordered media (see, for example, the review in Ref. 1). In this body of work, electromagnetoacoustic conversion (EMAC) processes in magnetic metals are studied. It is of interest to investigate the efficiency in magnetic dielectrics theoretically in order to study ultrasound generation mechanisms in this class of magnetically ordered substances.

The present paper is devoted to a study of EMAC processes in ferromagnetic and antiferromagnetic dielectrics.

2. FERROMAGNETIC DIELECTRIC

Let a uniform electromagnetic plane wave $h_x = h_0 \times \exp(-i\omega t + ikz)$ be normally incident on the surface of a semi-infinite ($z>0$) cubic single-domain ferromagnetic dielectric. In the ground state of the ferromagnetic dielectric, the magnetization vector $\mathbf{M}_0$ and the external constant magnetic field intensity vector $\mathbf{H}_0$ are parallel to the wave vector $\mathbf{k}$ and the $\mathbf{z}$ axis.

In studying EMAC processes in magnets, it is necessary to solve a system of coupled equations describing the propagation and interaction of electromagnetic, spin, and elastic waves in the magnet. This system includes the equation of elasticity, Maxwell's equations, and the Landau–Lifshitz equation for the magnetization vector

$$\rho\ddot{u}_i = \frac{\partial \sigma_{ik}}{\partial x_k} + f_i, \quad \text{curl } \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t},$$

$$\text{curl } \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \text{div } \mathbf{B} = 0, \tag{1}$$

$$\text{div } \mathbf{D} = 0, \quad \dot{\mathbf{M}} = g\mathbf{M}\times\mathbf{H}^{\text{eff}}.$$

Here ρ is the density of the dielectric, $\mathbf{u}$ is the displacement vector, $\sigma_{ik} = \partial F/\partial u_{ik}$ is the stress tensor, u_{ik} is the strain tensor,

$$f_i = \frac{1}{4\pi c}\frac{\partial \mathbf{D}}{\partial t}\times\mathbf{B}$$

is the analog of the Lorentz force in metals on account of the displacement current, $\mathbf{H}$ and $\mathbf{E}$ are the magnetic and electric field intensities in the magnet, c is the speed of light in vacuum,

$$\mathbf{D} = \varepsilon\mathbf{E} + \frac{1}{c}\varepsilon\dot{\mathbf{u}}\times\mathbf{B} - \frac{1}{c}\dot{\mathbf{u}}\times\mathbf{H}, \quad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M},$$

where ε is the permittivity (it is assumed that at ultrasonic frequencies the permittivity tensor is a constant scalar quantity), g is the gyromagnetic ratio,

$$\mathbf{H}^{\text{eff}} = -\frac{\partial F}{\partial \mathbf{M}} + \frac{\partial}{\partial x}\frac{\partial F}{\partial(\partial \mathbf{M}/\partial x_i)}$$

is the intensity of the effective magnetic field in the magnet, and F is the free energy density of the dielectric. This system of equations must be supplemented by the standard boundary conditions for the intensity and induction vectors of the electric and magnetic fields, and for the stress tensor and the magnetization vector.[1]

We write the free energy density of the ferromagnetic dielectric in the form

$$\begin{aligned} F = F(M^2) &+ \beta(M_x^2M_y^2 + M_x^2M_z^2 + M_y^2M_z^2) + b_1(M_x^2u_{xx} \\ &+ M_y^2u_{yy} + M_z^2u_{zz}) + b_2(M_xM_yu_{xy} + M_xM_zu_{xz} \\ &+ M_zM_yu_{zy}) + (1/2)c_{11}(u_{xx}^2 + u_{yy}^2 + u_{zz}^2) + c_{12}(u_{xx}u_{yy} \\ &+ u_{xx}u_{zz} + u_{zz}u_{yy}) + 2c_{44}(u_{xy}^2 + u_{yz}^2 + u_{zx}^2), \end{aligned} \tag{2}$$

where β is the first cubic anisotropy constant, b_i are magnetostriction constants, and c_{ik} are elastic moduli. In writing out the energy, we neglect the nonuniform exchange energy. As a result, there is no spatial dispersion of the dynamic magnetic permeability tensor of the ferromagnetic dielectric and the boundary condition for the magnetization can be neglected. This approximation holds well far from various resonances.[1,2]

The expression (2) for the free energy density does not explicitly contain a term describing the volume magnetostriction (this term is included in $F(M^2)$), since in the present geometry, where the constant and variable magnetic fields are perpendicular to one another, this term does not affect EMAC processes in ferromagnetic metals and dielectrics.

The system of equations (1), linearized near the position of equilibrium and describing the propagation of interacting electromagnetic, spin, and other waves, has the form

$$\omega^2 u_\pm + s_4^2 \frac{\partial^2 u_\pm}{\partial z^2} + \frac{H_0}{4\pi\rho}\frac{\partial h_\pm}{\partial z} + \frac{b_2 M_0}{\rho}\frac{\partial m_\pm}{\partial z} = 0,$$

$$\frac{\partial^2 h_\pm}{\partial z^2} + \frac{\varepsilon\omega^2}{c^2} h_\pm - \frac{\omega^2}{c^2}[(\varepsilon-1)H_0 + 4\pi\varepsilon M]\frac{\partial u_\pm}{\partial z} + \frac{4\pi\varepsilon\omega^2}{c^2} m_\pm = 0,$$

$$m_\pm = \chi_\pm h_\pm - b_2 M_0 \chi_\pm \frac{\partial u_\pm}{\partial z}, \tag{3}$$

where $a_\pm = a_x \pm i a_y$ $(a = u, h, m)$ are small cyclic displacement, magnetic field, and magnetization variables, $s_4^2 = c_{44}/\rho$ is the speed of transverse sound, $\chi_\pm = gM_0(\omega_{s0} \mp \omega)$ is the dynamic magnetic permeability, $\omega_{s0} = g\underline{M_0}(\beta + H_0/M_0 + b_2^2 M_0^2/c_{44})$ is the frequency of spin waves (β is the magnetostriction-renormalized cubic anisotropy constant[3]). The third term in the elasticity equation is responsible for the Lorentz mechanism of ultrasound excitation and the fourth term is responsible for the magnetoelastic mechanism.

Solving the system of equations (3) under the linearized boundary conditions gives the following results for the amplitude of ultrasound excited in the ferromagnetic dielectric. In the case of ultrasound excitation by the displacement current (Lorentz force) mechanism, the amplitude of the excited ultrasound is

$$u_\pm = \left(\frac{s_\pm}{c}\right)^2 \frac{H_0 \varepsilon \mu_\pm h_0}{2\pi\rho\omega s_\pm}\left(1 + \sqrt{\frac{\mu_\pm}{\varepsilon}}\right), \tag{4}$$

where $s_\pm^2 = s_4^2(1 - \zeta_\pm)$ is the magnetoelastic-interaction-renormalized transverse sound velocity, $\zeta_\pm = b_2 M^2 \chi_\pm / \rho s_4^2$ is the dynamic magnetoelastic interaction parameter, and $\mu_\pm = 1 + 4\pi\chi_\pm$ is the dynamic magnetic permeability.

In magnetically ordered media, a magnetoelastic EMAC mechanism that is specific to these media appears. On account of this mechanism, transverse sound with amplitude

$$u_\pm = -\frac{b_2 M_0 \chi_\pm h_0}{\rho s_4 \omega}\left[4\pi\chi_\pm\zeta_\pm + \left(1 + \sqrt{\frac{\mu_\pm}{\varepsilon}}\right)\sqrt{1-\zeta_\pm}\right]. \tag{5}$$

is excited in a ferromagnetic dielectric. Let us compare the results for the amplitude of the excited ultrasound in dielectrics with the analogous formulas for metals.[1] It should be noted that in dielectrics the skin-layer thickness δ can be formally assumed always to be greater than the wavelengths λ of the electromagnetic and ultrasonic waves. This corresponds to the case $\beta \gg 1$ in Ref. 1. Then, in the case of a Lorentz mechanism the sound excitation amplitude is approximately $(c/s_\pm)^2/\varepsilon\beta = 4\pi\sigma/\varepsilon\omega$ (σ is the conductivity of the metal) times greater in metals than in dielectrics. A similar comparison for the magnetoelastic mechanism shows that the efficiency of this mechanism of ultrasound excitation in dielectrics is of the same order of magnitude as in metals.

The amplitude of ultrasound excited by the magnetoelastic mechanism increases rapidly as the orientational phase transition point is approached ($\beta + H_0/M_0 \to 0$). This happens because the magnetic susceptibility $\chi_\pm$ and the magnetoelastic interaction parameter $\zeta_\pm$ increase near an orientational phase transition.

3. ANTIFERROMAGNETIC DIELECTRIC

EMAC processes in antiferromagnetic dielectrics are of greatest interest, since the overwhelming majority of known antiferromagnets are dielectrics. Moreover, in antiferromagnets many effects are usually intensified by uniform exchange;[4] this can also influence ultrasound transformation processes. This paper investigates EMAC processes in a semi-infinite two-sublattice antiferromagnetic dielectric.

Let a uniform electromagnetic wave $h_x = h_0 \times \exp(-i\omega t + i\omega z)$ be incident on the surface of an elastic and magnetoelastic isotropic easy-plane antiferromagnetic dielectric in the ground state, in which the antiferromagnetism vector **L** and the ferromagnetism vector **M** lie in the plane of the sample (along the **y** and **x** axes, respectively) and are perpendicular to the wave vector **k**. The external magnetic field $\mathbf{H}_0$ is directed along the **x** axis.

In the initial system of equations (1), we write the Landau–Lifshitz equations for an antiferromagnetic dielectric in the form

$$\dot{\mathbf{M}} = g\{\mathbf{M}\times\mathbf{H}_M + \mathbf{L}\times\mathbf{H}_L\} + r\mathbf{H}_M g L,$$

$$\dot{\mathbf{L}} = g\{\mathbf{M}\times\mathbf{H}_L + \mathbf{L}\times\mathbf{H}_M\} + r\mathbf{H}_L g L. \tag{6}$$

Here $\mathbf{H}_{M,L}$ are the effective magnetic fields for the ferro- and antiferromagnetism vectors, respectively, and r is the relaxation parameter of the magnetic subsystem. For brevity, the relaxation term in Eqs. (6) is written in its simplest form. The free energy density of an antiferromagnetic dielectric has the form

$$F = \frac{1}{2}A\mathbf{L}^2 + \frac{1}{4}B\mathbf{L}^4 + \frac{1}{2}D\mathbf{M}\times\mathbf{L}^2 + \frac{1}{2}D'\mathbf{M}^2\mathbf{L}^2 + \frac{1}{2}aM^2 - \mathbf{H}_0\times\mathbf{M} - \frac{1}{2}\beta L_z^2 + \frac{1}{2}b_0 L^2 u_{ll} + \frac{1}{2}bL_i L_k u_{ik} + \lambda_1 u_{ll}^2 + \lambda_2 u_{ik}^2. \tag{7}$$

Here A, a, B, D, and D' are uniform exchange constants within and between the sublattices, β is the uniaxial anisotropy constant, b_0 and b are the exchange and relativistic magnetostriction constants, and λ_i are the moduli of elasticity. Just as in the case of a ferromagnetic dielectric, the nonuniform exchange energy is neglected in Eq. (7). This makes it possible to neglect the spatial dispersion of the dynamic

magnetic susceptibility of an antiferromagnetic dielectric and ignore the boundary conditions for the ferro- and antiferromagnetism vectors.

We first investigate the case in which relaxation in the magnetic subsystem is so large that there is not enough time for the ferro- and antiferromagnetism vectors to follow the changes in the displacement and electromagnetic field vectors. This corresponds to the approximation $\omega \ll \tau\omega_E$, where $\omega_E = gM_0(a + D'L^2)$ is the exchange frequency and M_0 is the saturation magnetization of the sublattices.

The original system of Eqs. (1) and (6) linearized near the position of equilibrium can be written in the form

$$\begin{aligned}
\ddot{u}_y &= \tilde{s}_t^{\,2}\frac{\partial^2 u_y}{\partial z^2} - \frac{b_0 bLMr\omega_E}{2\rho\omega_{10}}\chi_\perp'\frac{\partial^2 u_z}{\partial z^2} - \frac{bLr\omega_E}{\rho\omega_{10}}\chi_\perp'\frac{\partial h_x}{\partial z},\\
\ddot{u}_z &= -\frac{ib_0 bLM\omega}{2\rho\omega_{10}\omega_B}\omega_D''\chi_\perp'\frac{\partial^2 u_y}{\partial z^2} + \tilde{s}_l^{\,2}\frac{\partial^2 u_z}{\partial z^2}\\
&\quad -\frac{1}{\rho}\left(\frac{1}{4\pi}B_0 + \frac{Mb_0}{\omega_B}\omega_E'\chi_\perp'\right)\frac{\partial h_x}{\partial z},\\
\frac{\partial^2 h_x}{\partial z^2} &= -\frac{i\varepsilon\omega}{c^2}\left[\frac{2i\pi\omega bL}{\omega_{10}}\chi_\perp'\frac{\partial \dot{u}_y}{\partial z}\right.\\
&\quad \left. + (B_0 + 4\pi M b_0\chi_\perp')\frac{\partial \dot{u}_z}{\partial z} + \mu_\perp'\frac{\partial h_x}{\partial t}\right],
\end{aligned} \tag{8}$$

where

$$\begin{aligned}
&\tilde{s}_t^{\,2} = \frac{\lambda_2(1-\xi_t)}{\rho}, \quad \xi_t = \frac{b^2L^2\omega_E'}{4\omega_{10}\lambda_2}\chi_\perp',\\
&\tilde{s}_l^{\,2} = \frac{(\lambda_1+2\lambda_2)(1-\xi_l)}{\rho}, \quad \xi_l = \frac{b_0^2M^2}{\omega_B}\frac{\omega_D''\chi_\perp'}{\lambda_1+2\lambda_2},\\
&\omega_E'' = \omega_E + \frac{1}{2}\omega_D',
\end{aligned}$$

$\chi_\perp' = gL/\omega_E'$ is the static perpendicular magnetic susceptibility, $\mu_\perp' = 1 + 4\pi\chi_\perp'$, $\omega_{10} = \omega_A + \omega_{me}$, $\omega_A = g\beta L$ is the anisotropy frequency, $\omega_{me} = gb^2L^3/4\lambda_2$ is the magnetoelastic frequency,

$$\begin{aligned}
&\omega_E' = \omega_E - 2M^2\omega_D'/L^2, \quad \omega_D'' = \omega_D' + \frac{1}{2}\omega_E,\\
&\omega_D' = gL^3D', \quad \omega_B = gL_0^3B
\end{aligned}$$

are the exchange frequencies, and

$$B_0 = H_0 + 4\pi M, \quad M = gH_0/\omega_E, \quad L = 2M_0$$

are the equilibrium values of the magnetic induction, the magnetization, and the modulus of the antiferromagnetism vector.

Solving (8) together with the linearized system of boundary conditions for the displacement and electromagnetic field vectors leads to the following results.

At low temperatures, when exchange magnetostriction can be neglected, and for $H_0 = 0$ $(M = 0)$, only transverse sound with amplitude

$$u_y = \frac{2ibLrh_0\omega_E\chi_\perp'}{\rho\tilde{s}_t\omega\omega_{10}}\left(1 + \sqrt{\frac{\mu_\perp'}{\varepsilon}}\right) \tag{9}$$

is excited, by the magnetoelastic mechanism, in antiferromagnetic dielectrics. At high temperatures, including near the Néel point, where relativistic magnetostriction can be neglected, and for $H_0 \neq 0$, only longitudinal ultrasound with amplitude

$$u_z = -i\left(\frac{\tilde{s}_l}{c}\right)^2\frac{B_0\varepsilon\mu_\perp' h_0}{2\pi\rho\omega\tilde{s}_l}\left(1 + \sqrt{\frac{\mu_\perp'}{\varepsilon}}\right) \tag{10}$$

is excited.

Comparing the amplitudes of transverse sound excited in ferromagnetic (5) and antiferromagnetic (9) dielectrics by the magnetoelastic mechanism shows that they differ by the factor $r\omega_E/\omega_{10}$, which need not be small if the relaxation parameter of the magnetic subsystem is not too small. However, in antiferromagnetic dielectrics the static magnetic susceptibility $\chi_\perp'$ is ordinarily small, so that the efficiency with which transverse sound is generated in an antiferromagnetic dielectric cannot be much greater than the analogous efficiency in a ferromagnetic dielectric. In the case of generation by the Lorentz force mechanism (assuming that in the ferromagnetic dielectric longitudinal sound will be excited with the same efficiency as transverse sound), the ultrasound generation efficiency is essentially the same in ferromagnetic and antiferromagnetic dielectrics.

In the opposite case, when there is enough time for the magnetic subsystem to adjust to the changes occurring in the elastic and electromagnetic subsystems $(\omega \gg r\omega_E)$, the linearized system of equations has the form

$$\begin{aligned}
\ddot{u}_y &= \tilde{s}_t^{\,2}\frac{\partial^2 u_y}{\partial z^2} + \frac{ibLMb_0\omega}{2\rho\omega_{10}}\chi\frac{\partial^2 u_z}{\partial z^2} + \frac{i\omega bL}{2\rho\omega_{10}}\chi\frac{\partial h_x}{\partial z},\\
\ddot{u}_z &= -\frac{ibb_0LM\omega}{\rho\omega_{10}}\chi\frac{\partial^2 u_y}{\partial z^2} + \tilde{s}_l^{\,2}\frac{\partial^2 u_z}{\partial z^2}\\
&\quad -\frac{1}{\rho}\left(b_0M\chi + \frac{1}{4\pi}B_0\right)\frac{\partial h_x}{\partial z},\\
\frac{\partial^2 h_x}{\partial z^2} &= -\frac{i\varepsilon\omega}{c^2}\left[\frac{2\pi ibL\omega}{\omega_{10}}\chi\frac{\partial \dot{u}_y}{\partial z}\right.\\
&\quad \left. + (B_0 + 4\pi M b_0\chi)\frac{\partial^2 \dot{u}_z}{\partial z} + \mu\frac{\partial h_x}{\partial t}\right],
\end{aligned} \tag{11}$$

where

$$\chi = \frac{gL\omega_{10}}{\omega_{1s}^2 - \omega^2}, \quad \omega_{1s}^2 = \omega_{10}\omega_E', \quad \mu = 1 + 4\pi\chi.$$

At low temperatures and for $H = 0$, $M = 0$, and $b_0 = 0$ only transverse ultrasound with amplitude

$$u_y = \frac{bL\chi h_0}{2\rho\omega_{10}\tilde{s}_t}\left(1 + \sqrt{\frac{\mu}{\varepsilon}}\right). \tag{12}$$

is once again excited. For $b=0$, $H\neq 0$, and $b_0\neq 0$ only longitudinal sound with amplitude determined by (10) with the magnetic permeability and susceptibilities from Eq. (11) is excited in antiferromagnetic dielectrics.

Comparing the results for this case with the analogous results for ferromagnetic dielectrics shows that in the case of transverse sound generation by the magnetoelastic mechanism, the amplitude of the excited waves in antiferromagnetic dielectrics differs from (5) by the factor ω_{10}/ω, provided that the magnetic susceptibility in the magnets studied is identical. In the case of the Lorentz mechanism, the amplitudes of the excited longitudinal sound are once again of the same order of magnitude.

Note that according to (10) and (12) the amplitude of the excited ultrasound increases near the orientational phase transition of easy-plane antiferromagnetic dielectrics to easy-axis antiferromagnetic dielectrics (at the transition point, the frequency ω_{10} decreases to the value of the magnetoelastic gap), as well as near the antiferromagnetic resonance frequency ω_{1s}.

In all other cases both transverse and longitudinal sound are excited simultaneously. Transverse ultrasound is excited mainly by the magnetoelastic mechanism and longitudinal ultrasound is excited by the induction mechanism.

Note that in the present geometry, a quasiantiferromagnetic (high-frequency) spin branch appears in the final formulas for ultrasound excitation in an antiferromagnetic dielectric. The frequency of this branch is ordinarily several orders of magnitude higher than that of the quasiferromagnetic (low-frequency) spin branch. In other geometries of the problem (for example, with $\mathbf{k}\|\mathbf{M}$), where the frequency of the quasiferromagnetic branch of the spin waves will appear in the final formulas, it may happen that the ultrasound generation efficiency will be much higher in an antiferromagnetic dielectric.

This work was supported in part by Russian Fund for Fundamental Research, Grants No. 96-02-19474 and No. IP-1095 ISSEP.

[1] V. D. Buchel'nikov and A. N. Vasil'ev, Usp. Fis. Nauk **162**, 89 (1992) [Sov. Phys. Usp. **35**, 192 (1992)].
[2] V. D. Buchel'nikov and V. G. Shavrov, Fiz. Tverd. Tela (St. Petersburg) **33**, 3284 (1993) [Phys. Solid State **33**, 1853 (1993)].
[3] V. D. Buchel'nikov and V. G. Shavrov, Fiz. Tverd. Tela (Leningrad) **23**, 1296 (1981) [Sov. Phys. Solid State **23**, 760 (1981)].
[4] E. A. Turov and V. G. Shavrov, Usp. Fiz. Nauk **140**, 429 (1983) [Sov. Phys. Usp. **26**, 593 (1983)].

Translated by M. E. Alferieff

Influence of disorder in crystal structure on ferroelectric phase transitions

A. A. Bokov

Physics Research Institute, Rostov State University, 344090 Rostov on Don, Russia
(Submitted 22 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1817–1832 (May 1997)

A new model of ferroelectric phase transitions in disordered crystals is developed. The model takes into account the nonequivalence of the structural environment of identical ions, which alters the balance of forces governing ferroelectric structural instability. In contrast with its predecessors, the new model can be used for qualitative and, in many cases, quantitative predictions of the temperature range of the diffuse ferroelectric phase transition and the value of the disorder-induced shift of the average transition temperature as a function of the degree of disorder and chemical composition. This conclusion is confirmed by a comparison of the theoretical results with an abundance of known experimental facts. [S1063-7761(97)02005-2]

1. INTRODUCTION

The influence of disorder in the structure of crystals on the ferroelectric phase transition can be manifested both as a change in the transition temperature and as broadening of the transition temperature (diffuse transition).[1–3] In contrast with normal ferroelectric phase transitions, where the structure and properties of the crystal change sharply at a certain temperature (or at a definite value of some other external parameter), such a point cannot be found in the case of diffuse phase transitions. Anomalies of the physical characteristics of a substance as a result of diffuse phase transitions are observed over a broad temperature range spanning tens and sometimes hundreds of kelvins.[1,2] One of the fundamental problems attending the investigation of diffuse transitions is to establish which factors govern the width of such an interval. This problem is interesting not only from a fundamental point of view, but also from a practical standpoint, since many potential applications of ferroelectric materials are directly associated with the diffuse transition effect.

Factors that can contribute to broadening of the ferroelectric transition in real samples are large-scale inhomogeneities of the chemical composition and distribution of defects,[4] size effects (e.g., in thin films and intergranular ceramics),[5,6] and nonuniform internal and external stresses.[4] In macroscopically homogeneous samples ferroelectric phase transitions are smeared by point defects if their density is sufficiently high.[7] These factors are directly related to the technology of preparation and processing of real samples, and their role can be minimized (at least in principle) by technological measures. However, there are more fundamental causes responsible for the broadening of phase transitions, including disorder in the distribution of ions of different species among identical crystallographic sites (compositional disorder).[1,3]

Despite the enormous abundance of experimental data gathered in studies of diffuse phase transitions, to date there is a lack of clarity in regard to the influence of compositional disorder on ferroelectric transitions. The broadening effect is usually identified with spatial fluctuations of the Curie temperature, which lead to the formation—at temperatures much higher than the average Curie point—of ''new phase'' microdomains: randomly oriented polar clusters, whose number and size increases as the temperature is lowered. The emergence of such polar clusters of diameter ~ 10 nm is corroborated by a great many experimental factors (see, e.g., Refs. 8–13).

The standard models of the evolution of ferroelectric transitions differ primarily in causes postulated as responsible for fluctuations of the local Curie temperature $T_0^{\rm loc}$. Historically the first and most universally recognized is the Isupov–Smolenskiĭ model,[1,14,15] which attributes the fluctuations of $T_0^{\rm loc}$ to fluctuations of composition. Spatial fluctuations of composition (e.g., fluctuations of the densities of B′ and B″ cations in solid solutions $A(B'_{1-x}B''_x)O_3$ with a perovskite-type structure) must occur as a result of compositional disorder (in the given example, disorder in the positions of the B′ and B″ ions). If the Curie temperature depends on x, phase transitions take place at different temperatures in microdomains with differing values of x.

Critical surveys of other transition broadening models are given in Refs. 1, 16, and 17. More or less successfully, each explains distinct phenomena in individual substances, but cannot be used to calculate *a priori* the width of the broadening temperature interval (degree of broadening) in a substance having a certain composition, or at least to make some kind of qualitative prediction along these lines. We have proposed[17,18] a new model of ferroelectric phase transitions in compositionally disordered crystals, where spatial fluctuations of $T_0^{\rm loc}$ are associated with random static displacements of ions from lattice sites as a result of differences in the sizes of the disordered ions. Here, on the basis of these notions, we explain the known characteristics of the influence of disorder on ferroelectric transitions in specific compounds and solid solutions having the most thoroughly investigated perovskite structure of all ferroelectrics. We show, in particular, that the simple model equations can be used to calculate *a priori* the degree of broadening.

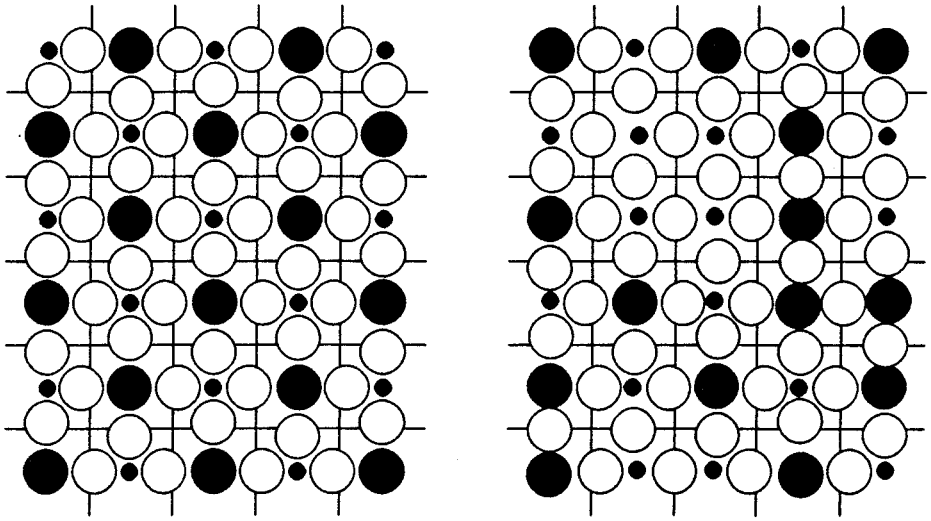

FIG. 1. Arrangement of B′ cations (small dark circles), B″ cations (large dark circles), and oxygen ions (light circles) in the (200) plane of an oxide $A(B'_{0.5}B''_{0.5})O_3$ with a perovskite structure in the compositionally ordered (a) and disordered (b) states. The line segments represent projections of the basal planes, at whose intersections A cations are situated.

2. NATURE OF THE BROADENING OF PHASE TRANSITIONS IN COMPOSITIONALLY DISORDERED FERROELECTRICS

To investigate the causes of spatial fluctuations of the Curie temperature, which result in the creation of polar clusters in compositionally disordered crystals, we proceed on the assumption that the nature of the spontaneous polarization is exactly the same in ordered ferroelectrics and in ferroelectrics with diffuse phase transitions. This hypothesis is entirely reasonable in view of the well-known fact that if disorder is introduced in an ordered crystal (e.g., by annealing a complex compound[19] or by varying the concentration of a solid solution[20]), the variation of the properties of the substance in the vicinity of the ferroelectric transition as the degree of disorder varies is generally a gradual process. We can therefore assume that a cluster acquires a dipole moment through the loss of balance between the forces tending to displace ions from sites occupied by them in the paraelectric phase (usually long-range Coulomb dipole–dipole interaction forces) and forces tending to stabilize the highly symmetrical structure (short-range repulsive forces).

All ions in an ordered crystal have the same structural environment, but disorder leads to distortion of the configuration of neighboring ions. We illustrate this process for oxides of the perovskite family, which have the general formula ABO_3 and in the paraelectric phase are endowed with a cubic structure[21] formed by a shell of oxygen octahedra joined at the vertices. The B cations are situated at the centers of the octahedra, and the positions between octahedra are occupied by A cations. The perovskite structure is shown schematically in Fig. 1 for an $A(B'_{0.5}B''_{0.5})O_3$ crystal, in which the centers of the oxygen octahedra are occupied by different cations B′ and B″. If an octahedron containing a B″ cation is in the neighborhood of an octahedron containing a B′ cation, the oxygen ion situated between these cations is displaced from the basal plane formed by A cations, moving toward a smaller-radius B cation.[21,22] In the ordered state (Fig. 1a) the arrangement of the ions is identical in all unit cells. In the disordered state (Fig. 1b), B cations of one species are located in certain neighboring unit cells, and the oxygen ions situated between them are not (in the first approximation) displaced from the basal A planes. The size and shape of different oxygen octahedra with identical B cations therefore differ, and for this reason the interionic distances also differ, as do the forces acting on the ions, which are the short-range as well as long-range forces responsible for the ferroelectric transition. In addition, the random configuration of disordered ions in neighboring and more distant cells will necessarily create differences in the Coulomb forces applied to identical ions (even when the dimensions of the disordered ions are identical and the oxygen octahedra are not distorted). The local Curie temperature in any microdomain of the crystal will be determined by the forces acting on the ions located inside it, and because these forces are of a random nature, the local Curie temperature will also be random.

3. DESCRIPTION OF THE EFFECTS OF BROADENING OF FERROELECTRIC PHASE TRANSITIONS IN PEROVSKITES OF THE TYPE $A(B'_{1-x}B''_x)O_3$

To analyze the qualitative notions set forth in the preceding section, we use the model of coupled anharmonic oscillators for displacive ferroelectrics with one soft mode.[23,24] In this model the effective Hamiltonian in the mean field approximation is the sum of the Hamiltonians of all unit cells of the crystal:

$$H=\sum_l H_l, \tag{1}$$

and the Hamiltonian of the lth cell has the form

$$H_l=\frac{1}{2}\Pi_l^2+\frac{A}{2}\xi_l^2+\frac{B}{4}\xi_l^4-\sum_{l'} v_{ll'}\xi_l\langle\xi_{l'}\rangle,$$

where Π_l and ξ_l are the generalized momentum and coordinate of the displacement for the soft mode, $A>0$ and $B>0$ are parameters of the one-particle potential, which are determined by interactions (mainly short-range) between ions of one unit cell, and $v_{ll'}$ are parameters characterizing the interaction (long-range dipole–dipole) between different cells.

The transition temperature can be expressed as follows in this model[23]:

$$kT_0 \approx 0.5\frac{V(V-A)}{B}, \quad V=\sum_{l'} v_{ll'}. \tag{2}$$

Let us assume that the soft mode in an $AB'O_3$ perovskite crystal with a sharp ferroelectric transition described by the Hamiltonian (1) is associated with displacements of B' ions relative to the oxygen octahedra (i.e., B' is a ferroelectrically active ion). When impurity B'' ions are added, the $A(B'_{1-x}B''_x)O_3$ crystal becomes disordered and, as mentioned, the dipole–dipole interaction forces and the short-range restoring forces responsible for the ferroelectric transition become different in different unit cells. This means that the Hamiltonian H_l for each cell must be characterized by different values of the parameters.

We refer to a cell containing a B' cation and having among its nearest neighbors m cells containing a B'' cation as an m-type cell ($m=0,1,2,\ldots,Z$); Z is the coordination number (in our case $Z=6$).

It is known from experimental work (see above) that the phase transition in a disordered perovskite is diffuse, i.e., local phase transitions take place at different temperatures in separate microdomains (in other words, polar clusters appear). In writing the Hamiltonian $H_{\rm loc}$ of such a microdomain, we assume for simplicity that the number of unit cells is the same and equal to N in all unit cells, the parameters of the Hamiltonian are also identical for all cells of one type (i.e., for all cells having the same m), and the critical coordinates ξ_l do not depend on the type of cell. We then obtain

$$H_{\rm loc}=N\sum_m c_m\left(\frac{1}{2}\Pi_{lm}^2(x)+\frac{A_m(x)}{2}\xi_l^2(x)+\frac{B_m(x)}{4}\xi_l^4(x) - \sum_{l'=1}^{N(1-x)} v_{ll'lm}(x)\xi_l(x)\langle\xi_{l'}(x)\rangle\right)+H_\rho+H_E.$$

Here c_m is the density of m-type cells in the microdomain, Π_{lm}, A_m, B_m, and $v_{ll'm}$ are the parameters of the Hamiltonian for an m-type cell, the summation over l' encompasses only interactions with ferroelectrically active ions belonging to the same microdomain, H_ρ is the total Hamiltonian of cells containing impurity ions, which are nonferroelectrically active by assumption (H_ρ does not depend on the critical coordinates), and H_E is the contribution associated with long-range interactions of ferroelectrically active ions of the microdomain with ions situated outside the microdomain.

We know[25] that Coulomb interactions with distant domains can be taken into account by introducing a macroscopic electric field that does not change within the confines of the unit cell (and in the case of a homogeneous crystal depends on the conditions at its boundaries). The Hamiltonian describing interaction with the macroscopic field is a constant, which can be discarded in microscopic analysis and then taken into account in corrections to the macroscopic characteristics of the system. The dependence of the parameters of $H_{\rm loc}$ on x is determined by the type of impurity B'' ions.

Multiplying the numerator and denominator in the expression for $H_{\rm loc}$ by $\Sigma_m c_m=1-x$ and discarding constant terms H_ρ and H_E, we obtain

$$H_{\rm loc}=N(1-x)\left(\frac{1}{2}(\Pi_l^2)_{\rm loc}+\frac{A_{\rm loc}}{2}\xi_l^2(x)+\frac{B_{\rm loc}}{2}\xi_l^4(x) - \sum_{l'=1}^{N(1-x)} v_{ll'}^{\rm loc}\xi_l(x)\langle\xi_{l'}(x)\rangle\right),$$

where

$$A_{\rm loc}=\frac{\sum_m c_m A_m(x)}{\sum_m c_m}, \quad B_{\rm loc}=\frac{\sum_m c_m B_m(x)}{\sum_m c_m},$$

$$V_{\rm loc}=\frac{\sum_m c_m V_m(x)}{\sum_m c_m}, \quad v_{ll'}^{\rm loc}=\frac{\sum_m c_m v_{ll'm}}{\sum_m c_m}$$

are parameters averaged over all ferroelectrically active cells of the given microdomain. Consequently, the Hamiltonian $H_{\rm loc}$ of the microdomain represents a sum of identical Hamiltonians, i.e., is formally the same as the Hamiltonian of an ordered macroscopic crystal. We therefore obtain the local transition temperature (i.e., the transition temperature in the microdomain) in the form

$$kT_0^{\rm loc}\approx 0.5\,\frac{V_{\rm loc}(V_{\rm loc}-A_{\rm loc})}{B_{\rm loc}}. \tag{3}$$

According to Ref. 18, the fluctuations of the density of m-type cells in the investigated microdomains $\Delta c_m=c_m-\langle c_m\rangle$ (angle brackets are used everywhere to denote averaging over the macroscopic volume) obey a normal distribution or a Poisson distribution with variance

$$\langle(\Delta c_m)^2\rangle=(1-x)(1-p_m)p_m/N, \tag{4}$$

where

$$p_m=\frac{Z!}{m!(Z-m)!}[(1-\alpha)x]^m[1-(1-\alpha)x]^{Z-m}, \tag{5}$$

and α is the compositional short-range order parameter.

Inasmuch as $T_0^{\rm loc}$ is a function of c_m, the same distribution can be used to approximate the probability $F_N(T_0^{\rm loc})$ of fluctuations of the local Curie temperature $T_0^{\rm loc}$ (subject to the condition that the nonlinearity of the function $T_0^{\rm loc}(c_m)$ is not too pronounced):

$$F_N(T_0^{\rm loc})=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(\Delta T_0^{\rm loc})^2}{2\sigma^2}\right], \tag{6}$$

where we have introduced the notation $\Delta T_0^{\rm loc}=T_0^{\rm loc}-\langle T_0\rangle$, and $\sigma^2\equiv\langle(\Delta T_0^{\rm loc})^2\rangle$ (σ is called the broadening parameter).

We note that the quantity $T_0^{\rm loc}$ defined by Eq. (3) is the thermodynamic phase equilibrium temperature, at which the densities of volume free energies of the two phases are equal. To determine the temperature at which a new-phase cluster emerges in some microvolume, as mentioned above, we need to take into account the energy of the macroscopic (depolarizing) field, along with the surface and elastic energies of the cluster. Allowance for these factors within the framework of the macroscopic approach has shown[26] that for all clusters their onset temperature differs from $T_0^{\rm loc}$ by the same amount, so that the distribution of the cluster-onset temperatures has the same form as $F_N(T_0^{\rm loc})$ and the same value of σ.

Equation (3) can be used to obtain[18] an approximate expression for the broadening parameter,

$$\sigma^2=\left(\frac{\langle T_0\rangle}{1-x}\right)^2\sum_m Q_m^2\langle(\Delta c_m)^2\rangle+\sigma_x^2, \tag{7}$$

where

$$Q_m\approx\frac{\nu_m\langle V\rangle-\mu_m\langle A\rangle}{\langle V\rangle-\langle A\rangle},$$

$\mu_m=A_m/\langle A\rangle$, $\nu_m=V_m/\langle V\rangle$, and σ_x^2 is the contribution associated with composition fluctuations, which according to Ref. 18 are usually so small as to be negligible.

From Eq. (7), making use of (4) and assuming for simplicity that $A=V$ (in fact $A\approx V$; Ref. 25), we obtain the final expression

$$\sigma^2=\frac{0.25\langle A\rangle^4}{(1-x)Nk^2\langle B\rangle^2}\sum_m(\nu_m-\mu_m)^2p_m(1-p_m). \tag{8}$$

To make the model useful for quantitative estimates, we attempt to express the parameters μ_m and ν_m in terms of structural characteristics. Since the parameters A and B of the one-particle potential describe short-range repulsive forces, we can assume that the latter act between the ferroelectrically active B′ ion and neighboring oxygen ions forming the octahedron about B′. The parameters A and B then vary according to a power law as functions of the distance B′–O inside the unit cell,[27] and the relations between the parameters associated with 0-type and 6-type cells can be written in the form

$$A_6=A_0\left(\frac{a}{a-2b}\right)^f,\quad B_6=B_0\left(\frac{a}{a-2b}\right)^g, \tag{9}$$

where a is the average unit cell parameter, b is the displacement of oxygen ions from the basal planes formed by A ions, and $f\approx f\gg1$. In other types of unit cells the oxygen octahedra are distorted, and the distances B′–O differ even within the same cell. Presumably the parameters of the Hamiltonian in this case assume values intermediate between the values of the corresponding parameters of the 0-type and 6-type cells. By analogy with (9) we assume that

$$A_m=A_0\left(\frac{a}{2I_m}\right)^f,\quad B_m=B_0\left(\frac{a}{2I_m}\right)^g, \tag{10}$$

where $I_m=a/2-mb/Z$ is the average distance between B′ and O ions in an m-type unit cell (of course, the actual relationship between the parameters will necessarily be more complicated; in particular, it must depend on the configuration of B″ ions in neighboring cells).

Using Eqs. (10), we obtain[18]

$$\mu_m=\frac{1+2fmb/aZ}{1+2f(1-\alpha)xb/a}. \tag{11}$$

Expanding μ_m in powers of the small ratio b/a and retaining only the linear term, we obtain

$$\mu_m=1+2f\left[\frac{m}{Z}-(1-\alpha)x\right]\frac{b}{a}. \tag{12}$$

To estimate ν_m, in general, interactions between ions of different cells must be taken into account — by no means a simple undertaking. We therefore assume, as in the case of short-range interactions, that the relative variations of dipole–dipole forces acting on a ferroelectrically active ion are determined entirely by the average size of the oxygen octahedron surrounding that ion, and do not depend on the composition of nearest-neighbor and more distant unit cells (the assumption is far from obvious but, as will be shown below, works quite well for the qualitative description of experimental data). By analogy with (12) we then readily obtain

$$\nu_m=1+2h\left[\frac{m}{Z}-(1-\alpha)x\right]\frac{b}{a}, \tag{13}$$

where $h\ll f$ (the short-range forces vary more rapidly with distance than the dipole–dipole forces). Substituting Eqs. (12) and (13) into (8), we obtain

$$\sigma=\delta D\frac{b}{a}, \tag{14}$$

where

$$\delta=\frac{\langle A\rangle^2|f-h|}{k\langle B\rangle\sqrt{N}},$$

$$D=\sqrt{\frac{1}{1-x}\sum_m\left[\frac{m}{Z}-(1-\alpha)x\right]^2p_m(1-p_m)}. \tag{15}$$

The ratio a/b can be determined by a method proposed by Sakhnenko et al.,[22] who expressed a and b in an ordered $A(B'_{0.5}B''_{0.5})O_3$ perovskite in terms of the known parameters of the ions. Using the same approach, we find

$$\frac{b}{a}=\left|\frac{(1/\sqrt{2})L_A\lambda_A[\lambda_{B'}-\lambda_{B''}]+L_{B''}\lambda_{B''}[\lambda_A+\lambda_{B'}]-L_{B'}\lambda_{B'}(\lambda_A+\lambda_{B''})}{(\lambda_{B'}+\lambda_{B''})[\sqrt{2}\lambda_AL_A+2(1-x)\lambda_{B'}L_{B'}+2x\lambda_{B''}L_{B''}]}\right|, \tag{16}$$

TABLE I. Broadening parameters of ferroelectric phase transitions and structural characteristics of compounds and solid solutions with $A(B'_{1-x}B''_x)O_3$ perovskite structure.

No.	Composition	s	L_A, Å	$L_{B'}$, Å	$L_{B''}$, Å	$(b/a)\cdot 10^4$	$D\cdot 10^3$	σ_{exp}, K	σ_{cal}, K	Refs.
1	$PbNb_{0.5}Fe_{0.5}O_3$	0	2.83	2.01	2.00	13	270	8	4	32, 33
2	$PbTa_{0.5}Fe_{0.5}O_3$	0	2.83	2.02	2.00	26	270	11	8	33, 34
3	$PbNb_{0.5}Sc_{0.5}O_3$	0	2.83	2.01	2.11	106	270	26	31	35
4	$PbTa_{0.5}Sc_{0.5}O_3$	0	2.83	2.02	2.11	93	270	25	27	36
5	$PbTa_{0.5}Sc_{0.5}O_3$	0.40	2.83	2.02	2.11	93	267	25	27	37, 38
6	$PbTa_{0.5}Sc_{0.5}O_3$	0.52	2.83	2.02	2.11	93	259	20	26	36
7	$PbTa_{0.5}Sc_{0.5}O_3$	0.74	2.83	2.02	2.11	93	221	17	23	36
8	$PbTa_{0.5}Sc_{0.5}O_3$	0.85	2.83	2.02	2.11	93	176	11	18	37, 38
9	$PbTa_{0.5}Sc_{0.5}O_3$	0.94	2.83	2.02	2.11	93	115	8	12	36
10	$PbNb_{0.5}In_{0.5}O_3$	0	2.83	2.01	2.15	148	270	49	44	33
11	$PbNb_{0.5}Lu_{0.5}O_3$	0	2.83	2.01	2.20	200	270	26	59	33
12	$PbNb_{0.5}Yb_{0.5}O_3$	0	2.83	2.01	2.21	211	270	46	63	33
13	$PbNb_{2/3}Mg_{1/3}O_3$	0	2.83	2.01	2.14	102	219	30	25	1, 35
14	$PbNb_{2/3}Zn_{1/3}O_3$	0	2.83	2.01	2.15	112	219	22	27	39
15	$PbFe_{2/3}W_{1/3}O_3$	0	2.83	2.00	1.89	156	219	37	38	40
16	$BaTi_{0.95}Sn_{0.05}O_3$	0	2.91	1.97	2.06	112	78	10	10	41
17	$BaTi_{0.90}Sn_{0.10}O_3$	0	2.91	1.97	2.06	112	111	12	14	41
18	$BaTi_{0.88}Sn_{0.12}O_3$	0	2.91	1.97	2.06	112	123	14	15	41
19	$BaTi_{0.85}Sn_{0.15}O_3$	0	2.91	1.97	2.06	112	140	18	17	41
20	$BaTi_{0.95}Zr_{0.05}O_3$	0	2.91	1.97	2.11	174	77	13	14	41
21	$BaTi_{0.90}Zr_{0.10}O_3$	0	2.91	1.97	2.11	174	111	19	21	41
22	$BaTi_{0.80}Zr_{0.20}O_3$	0	2.91	1.97	2.11	173	166	30	32	41
23	$BaTi_{0.70}Zr_{0.30}O_3$	0	2.91	1.97	2.11	172	207	42	39	41

Notes: In Ref. 36 the broadening temperature interval is characterized by the width of the permittivity curve at 3/4 of the maximum; we have calculated σ_{exp} by dividing this width by 1.633; the broadening parameter used in Ref. 38 is, by definition, twice the value of our parameter σ.

where λ_A, $\lambda_{B'}$, and $\lambda_{B''}$ are the valences of A, B′, and B″ ions, respectively, and L_A, $L_{B'}$, and $L_{B''}$ are the ''lengths of unstressed cation–anion bonds'' which characterize the size of the ions and are defined in Refs. 21 and 22. (The relations $L_A \approx R_O + R_A$, $L_{B'} \approx R_O + R_{B'}$, and $L_{B''} \approx R_O + R_{B''}$, where R_O, R_A, $R_{B'}$, and $R_{B''}$ are the ionic radii of the oxygen, A, B′, and B″ ions, can be used for approximate calculations.)

4. VARIATIONS OF THE AVERAGE CURIE TEMPERATURE IN THE PRESENCE OF COMPOSITIONAL DISORDER

Normally in experimental work the degree of long-range compositional disorder s is to be determined[19,28,29]; we therefore analyze the dependence of the average ferroelectric transition temperature $\langle T_0 \rangle$ on s. For ordering of the NaCl type (as in Fig. 1a), which is observed in $A(B'_{0.5}B''_{0.5})O_3$ perovskites and which we propose to consider below, $\alpha = -s^2$ for large enough s.[30] Using this relation and Eq. (12), we find the average of the parameter A over the volume of the crystal:

$$\langle A(s) \rangle = \langle A(0) \rangle \left(1 + 2fxs^2 \frac{b}{a} \right), \tag{17}$$

where $\langle A(0) \rangle$ is the average parameter A in the completely ordered crystal. We similarly obtain

$$\langle B(s) \rangle = \langle B(0) \rangle \left(1 + 2gxs^2 \frac{b}{a} \right), \tag{18}$$

$$\langle V(s) \rangle = \langle V(0) \rangle \left(1 + 2hxs^2 \frac{b}{a} \right), \tag{19}$$

where, as in the analysis of the influence of disorder on the degree of broadening of the phase transition, for simplicity we ignore effects associated with disorder-induced changes in the configuration of cations in neighboring and more distant unit cells. Substituting (17)–(19) into Eq. (2), we obtain

$$\langle T_0(s) \rangle = \langle T_0(0) \rangle \left(1 + 2xs^2 \frac{a}{b} \frac{hV_0 - fA_0}{V_0 - A_0} \right), \tag{20}$$

5. COMPARISON WITH EXPERIMENTAL DATA FOR PEROVSKITES OF THE TYPE $A(B'_{1-x}B''_x)O_3$

In the vicinity of a diffuse ferroelectric phase transition the temperature dependence of the dielectric permittivity ε is described by the well-known expression[1]

$$\frac{1}{\varepsilon} = \frac{1}{\varepsilon_m} + K(T - T_m)^2, \tag{21}$$

where T_m is the temperature at which ε attains its maximum value ε_m, and K is a parameter that differs in different substances. It has been shown[1,31] that for polar microdomains having a normal distribution in the vicinity of a diffuse phase transition with respect to their local Curie temperatures (and indeed such is the distribution considered in our model), the parameter K can be related to the variance σ^2 of the normal distribution:

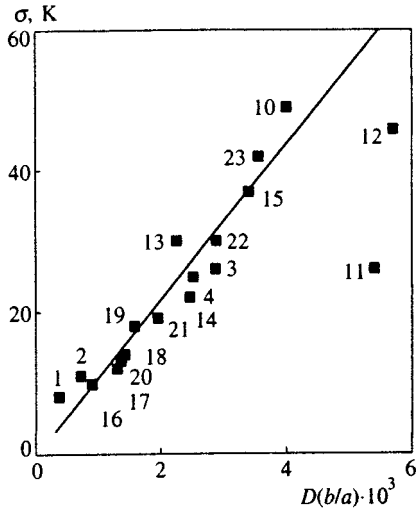

FIG. 2. Broadening parameter σ versus Db/a for compounds and solid solutions $A(B'_{1-x}B''_x)O_3$ having a perovskite structure and ferroelectric cations in the B sublattice. The points represent experimental values of σ, and the straight line corresponds to Eq. (14) for $\delta = 1.1 \cdot 10^4$ K. The points are numbered in accordance with the numbering of the compositions in Table I.

$$K = \frac{1}{2\varepsilon_m \sigma^2}. \tag{22}$$

Table I summarizes all the results of experimental measurements of σ (using Eqs. (21) and (22)) that we have been able to find in the literature for perovskites of the type $A(B'_{1-x}B''_x)O_3$. Where different authors have given different values of σ for one compound, we have chosen the smallest value (higher values are most likely found in samples where defects other than compositional disorder also contribute to broadening of the transition).

Figure 2 shows σ as a function of the parameter Db/a, calculated for the corresponding compounds according to Eqs. (15) and (16). In accordance with Ref. 42, we assume that the Ti, Nb, or Ta ion is ferroelectrically active. To calculate the ratio b/a, we use the values given in Ref. 43 for the lengths of the unstressed bonds, which are more accurate than those in Refs. 22 and 23. The resulting graph is found to be linear, so that, according to (14), the model parameter δ is approximately the same for all compositions. Calculated by the least-squares method from the slope of the line $\sigma = F(Db/a)$, it is found to be $1.1 \cdot 10^4$ K.

The broadening parameter σ can be determined with Eq. (14) for any perovskite of the type $A(B'_{1-x}B''_x)O_3$ with ferroelectrically active cations in the B sublattice by setting $\delta = 1.1 \cdot 10^4$ K. The results of such calculations are given in Table I and exhibit good agreement with experiment in most cases.

We now analyze the reasons why such a crude model can be used to calculate the broadening parameter in real compounds. The crux of the matter, in our opinion, is that the primary factor governing broadening is not the absolute strengths of interatomic interactions (e.g., the Curie temperature), but their relative variations induced by variations of the interatomic distances in the locally deformed disordered structure. In our model these relative variations are described by the parameters μ_m and ν_m, which in turn depend on the local strains of the structure and on the rate of change of the interatomic interactions with distance. The latter rate is dictated by fundamental parameters that are the same in different compounds of identical structure (e.g., the power exponent in the expression for the energy of interatomic repulsion in the Born–Landé form B/r^n (Ref. 44), which is often used for perovskites). With regard to local strains, we have stimulated them in such a way as to have μ_m and ν_m described by relations (12) and (13). Of course, this procedure introduces errors, which are not very likely to be small. However, the local strains are similar in all structurally similar compounds, so the relative errors are equal. As a result, the quantity δ in the expression (14) for σ can be regarded simply as a fitting parameter to be evaluated from the experimental data.

We now consider the influence of the degree of disorder on the average ferroelectric transition temperature. The model predicts (see (20)) a quadratic dependence $T_0(s)$, which is consistent with the results of a thermodynamic analysis[45] and with experimental results.[3,19] The influence of s on the transition temperature should be stronger for higher values of the ratio a/b, i.e., for a greater difference between the sizes of the ordered ions, and this too is consistent with experimental results.[46]

This behavior makes sense at the qualitative level when we consider the fact that disorder changes the average diameter of the oxygen octahedra surrounding the ferroelectrically active B cations, causing the average Hamiltonian parameters governing the transition temperature to change as well. In the known compositionally ordered $Pb(B'_{0.5}B''_{0.5})O_3$ perovskites (Refs. 3 and 19), the B″ cations are larger than the ferroelectrically active B′ cations; in the presence of ordering, the average diameter of the oxygen octahedra around the latter cations decreases, the short-range restoring forces increase, and the ferroelectric transition temperature is therefore lowered. (For a similar reason, the ferroelectric transition temperature drops in perovskites under the influence of hydrostatic pressure.[24]) In contrast, the antiferroelectric–paraelectric transition temperature must increase as s increases (in fact, hydrostatic pressure raises this temperature in perovskites[24]).

Experimental results[3,19] concur with these conclusions. Special consideration must be given to lead scandium titanate $Pb(Ta_{0.5}Sc_{0.5})O_3$ (PST), in which the phase transition temperature, in contrast with other known ferroelectrics, rises in the presence of ordering. This is probably attributable to the fact, demonstrated in Ref. 47, that as the temperature is lowered, the crystal initially enters a modulated antiferroelectric (incommensurate) phase and, from there, makes the transition to the ferroelectric phase. But the Curie–Weiss temperature (which has the significance of the threshold of stability of the paraelectric phase and whose dependence on s, strictly speaking, is treated in the simple models of this study) diminishes when ordering takes place in PST,[36] as in other ferroelectrics.

6. GENERAL LAWS IN THE BROADENING OF FERROELECTRIC TRANSITIONS IN PEROVSKITES

We now summarize qualitative trends in the broadening of the phase transition, which are valid for disordered perovskites, and we determine the extent to which the theoretical conclusions agree with the available experimental data.

In the case of $A(B'_{1-x}B''_x)O_3$ perovskites with ferroelectrically active cations in the B sublattice, an analysis of Eq. (16) shows as the difference in the length of the unstressed cation–anion bonds $\Delta L_B = |L_{B'} - L_{B''}|$ (or the difference in the radii) of the B′ and B″ cations increases, the relative displacement of the oxygen ions from the basal A planes (b/a) increases, so that, according to (14), the broadening of the transition increases as well.

As a result of substitution in the A sublattice (compounds or solid solutions of the type $(A'_{1-x}A''_{1-x})BO_3$), oxygen ions are displaced in the A planes without ever leaving them. It is clear from geometrical considerations that since the displacements of the oxygen ions are perpendicular to the B–O segment, the length of this segment changes only slightly (in comparison with the case of substitution in the B sublattice), so that comparatively very little difference exists among the B–O distances. Consequently, if the B sublattice is ferroelectrically active, the ferroelectric transition should not broaden appreciably (even though its broadening increases with an increase in $\Delta L_A = |L_{A'} - L_{A''}|$).

If the A ions are ferroelectrically active, then for similar reasons substitution in the B sublattice should not induce significant broadening of the phase transition, whereas substitution in the A sublattice, resulting in a relatively marked variation of the A–O distances, should induce broadening.

Consequently, in different binary solid solutions with identical concentrations the broadening is greater, the greater the difference in the sizes of the disordered ions; it should be relatively large if ferroelectrically active ions are replaced, and it should be small if nonferroelectrically active ions are replaced.

Good examples of this kind of behavior are solid solutions based on $BaTiO_3$, in which the replacement of the ferroelectrically active (according to Ref. 42) Ti cation by Sn, Zr (Ref. 41), Hf, and Th (Ref. 48) ions diffuses the transition substantially, whereas the replacement of the nonferroelectrically active Ba cation by Ca ions leaves the transition sharp,[49,50] and its replacement by Sr ions produces slight broadening of the transition.[41,51] In $PbTiO_3$ (in which the Pb ion is ferroelectrically active[42]) the transition becomes diffuse when Pb is replaced by Ca and Sr (Ref. 51) and remains sharp when the nonferroelectrically active Ti ion is replaced by Sn (Refs. 52 and 53) and Zr (Refs. 53 and 54). Here the transition in (Pb, Sr)TiO_3 broadens only slightly in comparison with the transition in (Pb, Ca)TiO_3, because the difference in the sizes of the cations in the A sublattice is smaller in the first case. In $(K_{0.5}Bi_{0.5})TiO_3$ and $(Na_{0.5}Bi_{0.5})TiO_3$, in which the disordered cations differ markedly in size and are situated in a ferroelectrically active sublattice,[42] the transition is substantially broadened.[55–57]

Different cases can be considered as the concentration of binary solid solutions is varied.

1) If a ferroelectrically active ion does not change when the impurity concentration is increased (as should be expected in solid solutions of a ferroelectric and a nonferroelectric or two ferroelectrics with identical ferroelectrically active ions), then:

a) replacement of the ferroelectrically active cation should cause the degree of broadening to increase monotonically with the concentration (such an increase is described by Eq. (14)), and indeed this kind of behavior has been observed experimentally, for example, when the ferroelectrically active Ti cation in $BaTiO_3$ was replaced by Zr and Sn (see the table in Ref. 41);

b) replacement of the nonferroelectrically active cation should lead to slight broadening with a maximum at the concentration $x \approx 0.5$ (obviously, the oxygen environment of the ferroelectrically active cations is deformed to the maximum at this concentration), and such a maximum of the degree of broadening has been observed in Pb(Zr, Ti)O_3 and (Ba, Sr)TiO_3 (Ref. 58).

2) If a ferroelectrically active cation changes when the concentration is increased (as should be expected at a morphotropic phase boundary in solid solutions of two ferroelectrics with different ferroelectrically active cations), then:

a) The replacement in the ferroelectrically active sublattice can be described by equations of the type (14), in which different cations must be interpreted as ferroelectrically active at low and high concentrations x. Thus, the broadening parameter should tend to zero in the limits $x \to 0$ and $x \to 1$, and a maximum should be observed at the concentration at which replacement of the ferroelectrically active cation takes place. This kind of behavior has been observed in solid solutions (Pb, Sr)TiO_3 (Ref. 51) (as in $PbTiO_3$, the A sublattice is ferroelectrically active in $SrTiO_3$; (Ref. 42));

b) if not only the ferroelectrically active cation, but also the ferroelectrically active sublattice (bearing in mind the A and B sublattices) is replaced at a certain concentration, sudden jumps in the degree of broadening can be observed at that concentration. This behavior is inherent in solid solutions $(1-x)BaTiO_3$–$xSrTiO_3$. In this case the transition broadens slightly at $x < 0.8$ (Refs. 51 and 59), because the A sublattice, in which replacement occurs, is not ferroelectrically active (detailed studies[58] have disclosed a maximum of the degree of broadening at $x \sim 0.5$, which should be observed in accordance with case 1b above). At $x \gtrsim 0.8$ the Sr ion probably becomes ferroelectrically active (as in pure $SrTiO_3$), as a result of which the transition broadens considerably (the possibility of a morphotropic phase boundary at $x \approx 0.8$ is suggested in Ref. 59). The broadening diminishes with a further increase in x (Refs. 51 and 59), a trend that can be attributed to an increase in the concentration of the newly ferroelectrically active Sr ion.

If ions of three or more different species are present in a ferroelectrically active sublattice, an increase in the impurity concentration can lead to a monotonic decrease in the degree of broadening. Indeed, impurity ions whose size differs little from that of the ferroelectrically active ion contribute little to broadening; consequently, if their concentration increases at the expense of the concentration of ions with distinctly different sizes, the degree of broadening must decrease. Good examples are the solid solutions

$(1-x)Pb(Mg1/3Nb2/3)O3-xPbTiO_3$

and

$(1-x)Pb(Mg1/3Zn2/3)O3-xPbTiO_3$,

which display[60–62] a decrease in the broadening as x is increased, where the concentration threshold separating solutions that exhibit behavior typical of diffuse phase transition from solutions with a sharp phase transition coincides with the morphotropic phase boundary ($x=0.35$ for the first solution, and $x=0.1$ for the second).

Within the framework of our model the decrease in the degree of broadening with increasing x is explained by a decrease in the concentration of Mg ($L=2.14$ Å) or Zn ($L=2.15$ Å) ions, which differ sharply in size from the ferroelectrically active Nb ions ($L=2.01$ Å). The contribution of Ti ions ($L=1.97$ Å) to broadening of the transition is insignificant, because their size is close to that of Nb. After the morphotropic boundary the transition becomes sharp as a result of the change in the type of ferroelectrically active sublattice: In pure lead titanate the A sublattice is ferroelectrically active,[42] and it is reasonable to assume that the same sublattice will remain ferroelectrically active in the solid solution up to the morphotropic boundary. In solid solutions $(1-x)\ Pb(Fe_{1/2}Nb_{1/2})O_3-xPbTiO_3$, on the other hand, the Ti impurity increases the degree of broadening,[63] because the difference between the sizes of Ti and Nb may be small, but it is still greater than the difference between Nb and Fe ($L=2.00$ Å).

According to the given model, strong local structural distortions, and hence marked broadening of the phase transition, are produced (at high concentrations) by vacancies, interstitial atoms, dislocations, and other extended defects. A large increase in the degree of broadening due to the appearance of vacancies has been observed experimentally, in particular, in $Pb(Sc_{1/2}Nb_{1/2})O_3$ and $Pb(Sc_{1/2}Ta_{1/2})O_3$ ceramics.[7,64] It has been shown[65,66] that dislocations are the main cause of broadening in $BaTiO_3$, (Ba, Sr)TiO_3, and Pb(Zr, Ti)O_3 thin films (compositional disorder, as mentioned above, should not contribute much to broadening of the transition in these solid solutions). A significant contribution from dislocations and other structural defects to broadening of the ferroelectric transition has been observed in $Pb(Sc_{1/2}Nb_{1/2})O_3$ crystals.[67]

7. CONCLUSION

To summarize, all the fundamental laws underlying the influence of disorder in crystal structure on ferroelectric phase transitions have been successfully explained, where, in contrast to the work of other authors, different types of disorder (point and extended defects and compositional disorder) have been treated within the context of a single approach. A distinctive feature of this approach is that it takes into account disorder-induced differences in the structural environment of identical ions, such differences altering the balance of forces governing ferroelectric structural instability. According to the model developed here, the cause of spatial fluctuations of the local transition temperature, which lead to the formation of polar microclusters and broadening of the transition, are random static distortions of the configuration of ions in the units cells due to the difference in the sizes of the compositionally disordered ions or to the existence of vacancies, interstitial ions, dislocations, and other defects. The influence of the degree of compositional disorder on the average transition temperature can be attributed to the disorder-related variation of the average diameter of the oxygen octahedra surrounding ferroelectric ions in the perovskite structure. The degree of broadening of the ferroelectric phase transition can be determined for perovskites having the general formula $A(B'_{1-x}B''_x)O_3$, using only data on the chemical composition, degree of compositional disorder, and ionic radii in the calculations.

The work described in this publication has been made possible by a grant from the International Science Foundation and the Russian government, No. J35100.

[1] G. A. Smolenskiĭ, V. A. Bokov, V. A. Isupov *et al.*, *Physics of Ferroelectric Phenomena* [in Russian], Nauka, Leningrad (1985).
[2] L. E. Cross, Ferroelectrics **76**, 241 (1987).
[3] A. A. Bokov, Ferroelectrics **183**, 65 (1996).
[4] B. N. Rolov and V. É. Yurkevich, *Physics of Diffuse Phase Transitions* [in Russian], Izd. Rostov. Univ., Rostov-on-Don (1983).
[5] V. P. Dudkevich, and E. G. Fesenko, *Physics of Ferroelectric Films* [in Russian], Izv. Rostov. Univ., Rostov-on-Don (1979).
[6] G. Arlt, Ferroelectrics **104**, 217 (1990).
[7] F. Chu, I. M. Reaney, and N. Setter, Ferroelectrics **151**, 343 (1994).
[8] C. A. Randall, D. J. Barber, P. Groves *et al.*, J. Mater. Sci. **23**, 3678 (1988).
[9] C. A. Randall, D. J. Barber, and R. W. Whatmore, J. Meteorol. **145**, 275 (1987).
[10] S. B. Vakchrushev, B. E. Kvyatkovsky, A. A. Nabereznov *et al.*, Physica B **156-157**, 90 (1989).
[11] S. B. Vakchrushev, B. E. Kvyatkovsky, A. A. Nabereznov *et al.*, Ferroelectrics **90**, 173 (1989).
[12] C. N. W. Darlington, J. Phys. C **21**, 3851 (1988).
[13] C. N. W. Darlington, Phys. Status Solidi A **113**, 63 (1989).
[14] G. A. Smolenskiĭ and V. A. Isupov, Zh. Tekh. Fiz. **24**, 1375 (1954).
[15] V. A. Isupov, Ferroelectrics **90**, 113 (1989).
[16] Z.-G. Ye, Ferroelectrics **184**, 193 (1996).
[17] A. A. Bokov, Ferroelectrics **131**, 49 (1992).
[18] A. A. Bokov, Solid State Commun. **90**, 687 (1994).
[19] A. A. Bokov, Izv. Ross. Akad. Nauk Ser. Fiz. **57**, 25 (1993).
[20] V. É. Yurkevich, *Physics of Phase Transitions in Ferroelectrically Active Solid Solutions* [in Russian] Izv. Rostov. Univ. Rostov-on-Don (1988).
[21] E. G. Fesenko, *The Perovskite Family and Ferroelectricity* [in Russian] Atomizdat, Moscow (1972).
[22] V. P. Sakhnenko, E. G. Fesenko, A. T. Shuvaev *et al.*, Kristallografiya **17**, 316 (1972) [Sov. Phys. Crystallogr. **17**, 268 (1972)].
[23] M. E. Lines, Phys. Rev. **177**, 797 (1969).
[24] M. E. Lines and A. M. Glass, *Principles and Applications of Ferroelectrics and Related Materials*, Clarendon Press, Oxford (1977).
[25] V. G. Vaks, *Introduction to the Microscopic Theory of Ferroelectrics* [in Russian], Nauka, Moscow (1973).
[26] A. A. Bokov, Fiz. Tverd. Tela (St. Petersburg) **36**, 36 (1994) [Phys. Solid State **36**, 19 (1994)].
[27] H. Braeter, N. M. Plakida, and W. Windsch, Solid State Commun. **67**, 1219 (1988).
[28] C. G. F. Stenger, F. L. Scholten, and F. J. Burggraaf, Solid State Commun. **32**, 989 (1979).
[29] N. Setter and L. E. Cross, J. Appl. Phys. **51**, 4356 (1980).
[30] G. S. Zhdanov, *Physics of the Solid State* [in Russian], Izd. MGU, Moscow (1962), p. 349.
[31] V. V. Kirilov and V. A. Isupov, Izv. Akad. Nauk Ser. Fiz. **35**, 2602 (1971).
[32] N. Yasuda and Y. Ueda, Phys. Lett. **134**, 501 (1989).
[33] V. Yu. Shonov, *Dissertation for the Degree of Candidate of Physicomathematical Sciences* [in Russian], Rostov. Gos. Univ. Rostov-on-Don (1992).
[34] N. Yasuda, S. Fujimoto, and H. Terasawa, Trans. IEEE Ultrason. Ferroelectr. Freq. Control **UFFC-36**, 402 (1989).

[35] M. Yu. Leshchenko, *Author's Abstract of Dissertation for the Degree of Candidate of Physicomathematical Sciences* [in Russian], Uzhgorod. Univ. Uzhgorod (1993).
[36] C. G. F. Stenger, and A. J. Burggraaf, Phys. Status Solidi A **61**, 653 (1980).
[37] N. Setter and L. E. Cross, Phys. Status Solidi A **61**, K71 (1980).
[38] P. Groves, Ferroelectrics **76**, 81 (1987).
[39] S. Nomura, J. Kuwata, S. J. Jang *et al.*, Mater. Res. Bull. **14**, 769 (1979).
[40] N. Yasuda, S. Fujimoto, and K. J. Tanaka, J. Phys. D **18**, 1909 (1985).
[41] V. A. Isupov, Fiz. Tverd. Tela (Leningrad) **28**, 2235 (1986) [Sov. Phys. Solid State **28**, 1253 (1986)].
[42] Yu. N. Venevtsev, E. D. Politova, and S. A. Ivanov, *Ferroelectrics and Antiferroelectrics of the Barium Titanate Family* [in Russian], Khimiya, Moscow (1985), p. 114.
[43] G. A. Geguzina, V. P. Sakhnenko, E. G. Fesenko *et al.*, All-Union Institute of Scientific and Technical Information (VINITI) No. 3049-76, Moscow (1976).
[44] J. Slater, *Insulators, Semiconductors, and Solids*, McGraw-Hill, New York (1967).
[45] A. A. Bokov, I. P. Raevskiĭ, and V. P. Smotrakov, Fiz. Tverd. Tela (Leningrad) **25**, 2025 (1983) [Sov. Phys. Solid State **25**, 1168 (1983)].
[46] A. A. Bokov and I. P. Raevskiĭ, Pis'ma Zh. Tekh. Fiz. **16**(17), 44 (1990) [Sov. Tech. Phys. Lett. **16**, 660 (1990)].
[47] K. Z. Baba-Kishi and D. J. Batber, J. Appl. Crystallogr. **23**, 43 (1990).
[48] O. I. Prokopalo and E. G. Fesenko, in *Ferroelectrics* [in Russian], Izd. Rostov. Gos. Univ., Rostov-on-Don (1961), p. 123.
[49] Z. A. Takare, in *Ferroelectric Phase Transitions* [in Russian], Izd. Latv. Univ., Riga (1978), p. 149.
[50] K. Ya. Borman, in *Ferroelectric Phase Transitions* [in Russian], Izd. Latv. Univ., Riga (1978), p. 162.
[51] D. M. Kazarnovskiĭ, *Ferroceramic Capacitors* [in Russian], Gosénergoizdat, Moscow–Leningrad (1956).
[52] G. A. Smolesnkiĭ, A. I. Agranovskaya, and A. M. Kalinina, Zh. Tekh. Fiz. **25**, 2134 (1955).
[53] V. Ya. Fritsberg, in *Phase Transitions in Ferroelectric Solid Solutions* [in Russian], Izd. Latv. Univ., Riga (1976), p. 5.
[54] K. Tsuzuki, K. Sakat, and M. Wada, Ferroelectrics **8**, 501 (1974).
[55] G. A. Smolenskiĭ, V. A. Bokov, V. A. Isupov *et al.*, *Ferroelectrics and Antiferroelectrics* [in Russian], Nauka, Leningrad (1971).
[56] I. P. Pronin, P. P. Syrnikov, V. A. Isupov *et al.*, Pis'ma Zh. Tekh. Fiz. **5**, 705 (1979) [Sov. Tech. Phys. Lett. **5**, 289 (1979)].
[57] S. M. Emel'yanov, I. P. Raevskiĭ, and O. I. Prokopalo, Fiz. Tverd. Tela (Leningrad) **25**, 1542 (1983) [Sov. Phys. Solid State **25**, 889 (1983)].
[58] V. Ya. Gritsberg and K. Ya. Borman, in *Phase Transitions in Ferroelectrics with a Perovskite Structure* [in Russian], Uch. Zap. Latv. Gos. Univ. **189**, 99 (1974).
[59] V. V. Lemanov, E. P. Smirnova, and E. A. Tarakanov, Fiz. Tverd. Tela (St. Petersburg) **37**, 1854 (1995) [Phys. Solid State **37**, 1010, (1995)].
[60] S. W. Choi, T. R. Shrout, S. J. Jang *et al.*, Mater. Lett. **8**, 253 (1989).
[61] C. A. Randall, A. S. Bhalla, T. R. Shrout *et al.*, J. Mater. Res. **5**, 829 (1990).
[62] J. Kuwata, K. Uchino, and S. Nomura, Jpn. J. Appl. Phys. **21**, 1298 (1982).
[63] A. A. Bokov, L. E. Shpak, and I. P. Rayevsky, J. Phys. Chem. Solids **54**, 495 (1983).
[64] F. Chu, I. M. Reaney, and N. Setter, J. Am. Ceram. Soc. **78**, 1947 (1995).
[65] Z. Surowiak, E. Nogas, A. M. Margolin *et al.*, Ferroelectrics **115**, 21 (1991).
[66] Z. Surowiak, D. Czekaj, A. A. Bakirov *et al.*, Integr. Ferroelectr. **8**, 267 (1995).
[67] K. G. Abdulvakhidov and M. F. Kupriyanov, Izv. Ross. Akad. Nauk Ser. Fiz. **59**, 73 (1995).

Translated by James S. Wood