JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Generalization of the effective mass method for semiconductor structures with atomically sharp heterojunctions

É. E. Takhtamirov*) and V. A. Volkov†)

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

(Submitted 28 June 1999)
Zh. Éksp. Teor. Fiz. **116**, 1843–1870 (November 1999)

The Kohn–Luttinger method of envelope functions is generalized to the case of heterostructures with atomically sharp heterojunctions based on lattice-matched layers of related semiconductors with zinc-blende symmetry. For electron states near the Γ point in (001) heterostructures the single-band effective-mass equation is derived, taking into account both the spatial dependence of the effective mass and effects associated with the atomically sharp heterojunctions. A small parameter is identified, in powers of which it is possible to classify the various contributions to this equation. For hole states only the main contributions to the effective Hamiltonian, due to the sharpness of the heterojunctions, are taken into account. An expression is derived for the parameter governing mixing of states of heavy and light holes at the center of the 2D Brillouin zone. *© 1999 American Institute of Physics.* [S1063-7761(99)02511-1]

1. INTRODUCTION

The Kohn–Luttinger effective-mass method[1,2] is widely used to describe electron states in external fields varying smoothly over scales of the lattice constant a. Although the original method, based on the formalism of envelope functions, is applicable only to homogeneous semiconductors, various modifications of it have been used to described the electron states in semiconductor heterostructures. In recent years there has been a revival of discussion on the applicability of the effective-mass method to describe electron and hole states in real nanostructures.[3–22] Many different modifications of the effective-mass method have been proposed, which apply to the case of a spatially varying effective mass $m(\mathbf{r})$. There are two ways of constructing the effective-mass approximation for heterostructures. 1) Derivation of the effective Hamiltonian for the envelope functions, defined over all space. By integrating the effective-mass equation (which contains this Hamiltonian) near the heteroboundary it is possible to obtain boundary conditions on the envelope functions (if needed). 2) Derivation or, as is done much more often, postulation of phenomenological boundary conditions on the envelope functions at the heteroboundary. This approach makes use of symmetry arguments, continuity of the probability flux density, etc. (By the way, these arguments, as a rule, are insufficient to uniquely determine the boundary conditions.) The second approach is applicable in the case of sharp heterojunctions (all the models in which such boundary conditions were obtained dealt with mathematically sharp heterojunctions between the left-hand and right-hand materials). It is implicitly assumed that the envelope function on the left side (on the right side) of the heteroboundary satisfies the same equation as in the bulk case for the left-hand (right-hand) material. In this case the very delicate problem of increasing the accuracy arises (which, by the way, has not been discussed to this date): the boundary conditions should hold with the same accuracy as the equations for the envelope functions hold.

Below we will follow the first approach, in which it is possible to rigorously treat the problem of accuracy (see Sec. 3).

It is well known that two main problems arise along the path of constructing a common equation for the envelope functions. The first is the problem of ordering the momentum operators in the kinetic-energy operator [due to the noncommutativity of the momentum operator and the function $m(\mathbf{r})$], on the form of which the derived effective-mass equations can depend substantially.[18] The second problem is that the effective potential near a heteroboundary, as a rule, is not a smooth function on scales of the order of a. This calls into question the validity of using differential equations in the method of envelope functions. Let us discuss these problems in the indicated order.

1.1. Account of the spatial dependence of the effective mass

A necessary condition for the applicability of one-band equations for the envelope functions (one equation is understood here, valid near the bottom of the nondegenerate conduction band, or a system of equations for the degenerate valence band) used in the effective-mass method is ‘‘shallowness’’ of these states: their energy, measured from the band edge, should be small in comparison with the interband energy. Therefore, bearing in mind the one-band version of the effective-mass method, we restrict the discussion to heterostructures consisting of related materials, where the band discontinuities are small in comparison with the characteristic widths of the band gaps; this means, as a rule, that other band parameters of the semiconductors differ only slightly. Let us consider the first problem, which arises even for heterostructures whose chemical composition varies smoothly on scales of the order of a. As the zeroth-order potential we choose the potential of the crystalline lattice, continued to all space, of one of the materials of the structure (this is not a unique choice, see Ref. 7), and we treat the difference between the potentials of the lattices of the remaining semiconductors and the basis potential as a small perturbation. Following the approach of Luttinger and Kohn and deriving the many-band $\mathbf{k}\cdot\mathbf{p}$ system of equations (see, e.g., Ref. 3), we can next attempt to solve the problem of the correct order of noncommuting operators in the kinetic-energy operator for the one-band equations. But here yet another problem arises.

Reduction of the many-band system of equations to a one-band effective-mass equation is achieved by eliminating the small envelope functions from the many-band $\mathbf{k}\cdot\mathbf{p}$ system in favor of the large ones by means of some procedure. We make a small departure here and make use of a formal analogy between the relativistic Dirac equation and the many-band $\mathbf{k}\cdot\mathbf{p}$ system of equations for the envelope functions,[23] which is most simply seen in the two-band approximation (the conduction band and the nondegenerate valence band). In the relativistic theory there are two approaches to deriving an equation for shallow electron states. One of them consists in eliminating the small positron component of the wave function by the method of substitution. In this case, we obtain either an exact equation for the electron component, which is not an eigenvalue equation (Ref. 24, Ch. 20, Sec. 28) or an approximate equation whose Hermitian nature must be checked separately.[25] The second approach is a Foldy–Wouthuysen transformation, an approximate unitary transformation of the Dirac equation (Ref. 24, Ch. 20, Sec. 33).

In our case the first approach is comparatively simple to realize only in the two-band approximation (see, e.g., Ref. 26). In a treatment of the contribution of distant bands (and this is necessary, in particular, for a valid description of the contribution of the heavy holes) a number of problems arise. Thus, the authors of Refs. 14 and 20 were able to take into account only a few of the first-order corrections to the ''standard'' Kohn–Luttinger equation with position-independent effective mass (the small parameter here is the ratio of the characteristic band discontinuity to the characteristic interband energy). However, treating the expression obtained, for example, in Ref. 14 for the position-dependent effective mass, it can be shown that the effective mass of the edge of the conduction band of one of the non-basis semiconductors does not contain interband matrix elements of the perturbation potential, obtained using Bloch functions of the band edge of the basis crystal (see Sec. 4.1 below). It can be easily seen that this is equivalent to the poorly justified approach of neglecting the difference between the interband matrix elements of the momentum operator or, what is equivalent, the difference between the Bloch functions for the materials making up the structure.

Hence it follows that we should give special attention to the problem of taking distant bands into account. Efforts at solving it by direct elimination of the small envelope functions by the method of substitution, in addition to its laboriousness, lead finally to a non-Hermitian equation, whose solution is still in need of a valid interpretation.

Below we will follow the second approach, i.e., we will apply the unitary transformation eliminating the small envelope functions.[1,27] Since we are considering heterostructures consisting of related materials, the standard effective-mass method will play the role of a first approximation. An account of the spatial dependence of the effective mass necessitates treating corrections to the standard theory, where it is necessary to take into account all corrections of the same order without increasing the accuracy.

In order to understand what corrections should be taken into account, let us turn to the relativistic analogy with the hypothetical Dirac equation containing the inhomogeneous gap $2m(\mathbf{r})c^2$, where c is the speed of light in vacuum (see Appendix A). The ordinary one-band effective-mass equation is an analog of the nonrelativistic Schrödinger equation. It is important, however, that the effective mass in the two-band approximation is proportional to the local band gap $E_g(\mathbf{r})$ (this is valid if the effective mass is formed mainly by the $\mathbf{k}\cdot\mathbf{p}$ interaction), and its relative variation $\delta m/m \simeq \delta E_g/E_g$. Since the correction to the kinetic energy describing the spatial dependence of the effective mass will have a ''relativistic'' character, the desired equations for the heterostructures will be analogous to the Schrödinger equation with all relativistic corrections—both the usual ones (the contribution of nonparabolicity of the dispersion law, proportional to $\mathbf{p}^4$, where $\mathbf{p}$ is the momentum operator; the contribution of the spin–orbital interaction; and the so-called Darwin term, proportional to the second derivative of the potential energy) and a new pseudorelativistic correction describing $\delta m(\mathbf{r})$. Of course, the present arguments are valid for describing states whose energies, reckoned from the band edge of any of the materials making up the structure, are of the order of the band discontinuity. The case of a very small band discontinuity, where the discontinuity is small in comparison with the energies of the states is quite trivial: depending on the energies of the states under consideration an account of the spatial dependence of the effective mass can require treating terms with higher and higher powers of the momentum operator. We will not consider such a situation. In this sense, introducing a term proportional to the fourth power of the momentum operator into the effective-mass equation is a necessary condition for a consideration of its spatial dependence. Note that for homogeneous semiconductors an effective-mass equation analogous to the Schrödinger equation with first relativistic corrections was discussed already in Ref. 27 (Sec. 27).

A typical shortcoming of previous works dedicated to a generalization of the effective-mass method to electron states in heterostructures is that they take account within the framework of perturbation theory of only some of the terms of a given order. Thus, Refs. 3–22 take account of the spatial dependence of the effective-mass parameters, but neglect corrections $\sim\mathbf{p}^4$. In Refs. 3, 7, 10, and 14, dedicated to deriving the one-band equations for the envelope functions directly from a many-band $\mathbf{k}\cdot\mathbf{p}$ system, the main error, leading to an invalid result, is an incorrect estimate, according to which the contribution of the $\mathbf{k}\cdot\mathbf{p}$ interaction terms (i.e., the terms $\hbar\mathbf{k}\mathbf{p}_{nn'}/m_0$, where m_0 is the free electron mass and $\mathbf{p}_{nn'}$ is the interband matrix element of the momentum) is of order the contribution of the potential energy terms (by potential energy here we mean the difference in the periodic potentials of the semiconductors making up the structure, also treated in a perturbation-theory context). In the case of a smooth heterojunction the correct procedure for deriving the one-band effective-mass equation near the bottom of the conduction band with all the above-indicated contributions taken into account was followed in Ref. 28.

1.2. Account of atomically sharp heterojunctions

The second problem requiring careful study is the non-smooth nature of actual heterojunctions, where the transition from one material to the other occurs over scales of the order of a. In this case, first, Leibler's many-band $\mathbf{k}\cdot\mathbf{p}$ system,[3] where smoothness of the potential was a necessary condition for validity of the system, is in need of refinement, and second, the problem of transforming to $\mathbf{r}$ space from the region of $\mathbf{k}$ space bounded by the first Brillouin zone is more complicated.[1] It is also necessary to analyze the consequences of the unitary transformation eliminating distant bands. It is important to estimate the error that enters at each step. An estimate of this error either gives us confidence in the absence of an excess of accuracy or it challenges the validity of the effective-mass approximation. In the works known to us which treat sharp heterojunctions, such an estimate is lacking. References 14 and 20, for example, only point to its smallness, and Ref. 19 made some approximations whose accuracy were not even estimated.

Thus, we can formulate the following steps in the construction of an effective-mass approximation for heterostructures: a) obtaining a many-band $\mathbf{k}\cdot\mathbf{p}$ system of equations for the envelope functions taking proper account of possible sharpness of the heterojunction; b) reducing this system to one-band equations with the help of the unitary transformation to $\mathbf{k}$ space, transforming to $\mathbf{r}$ space and transforming the resulting equation to differential form; c) estimating the accuracy of these transformations. Following this scheme, step (a) is realized in Sec. 2. The equations include contributions associated with the non-smoothness of the heterojunction on scales of order a which are treated within the framework of an approach similar to that used in Ref. 27 to describe the short-range part of the impurity potential. Section 3 shows that a common differential equation over all space for the envelope functions for sharp heterojunctions exists, and its accuracy is determined by the procedure of transforming to one-band equations in $\mathbf{r}$ space. One-band equations are considered in Sec. 4.1 (the conduction band) and in Sec. 4.2 (the valence band). Section 4.3 is devoted to heterostructures with superthin layers. It is shown that additional contributions to the ''standard'' effective-mass equation can be classified by powers of the small parameter $\overline{k}a_m$, where $\overline{k}$ is a characteristic value of the quasimomentum of the state and a_m is on the order of the lattice constant. Section 5 constructs a hierarchical scheme of effective-mass equations, the nth level of which corresponds to taking account of these additional contributions up to $(\overline{k}a_m)^n$. The zeroth order of the hierarchy ($n=0$) corresponds to the ''standard'' effective-mass equation with position-independent parameters. At the first level of the hierarchy each heteroboundary gives an additional δ-function contribution to the potential energy. Only at the second level of the hierarchy does spatial dependence of the effective mass appear, along with corrections associated with weak nonparabolicity of the spectrum and hetero-surface terms of the spin–orbit interaction. At higher levels of the hierarchy nonlocal contributions arise, and the one-band differential effective-mass equations do not exist. Results are compared with the conclusions of other authors. Brief reports on the results obtained have been published elsewhere.[29–31]

2. MANY-BAND k·p SYSTEM OF EQUATIONS FOR ENVELOPE FUNCTIONS IN THE CASE OF A SHARP (001) HETEROJUNCTION

Let us consider a heterojunction formed from related, lattice-matched semiconductors with zinc-blende structure. The Schrödinger equation without relativistic corrections (which will be taken into account below) and in the absence of external potentials has the usual form

$$\left(\frac{\mathbf{p}^2}{2m_0}+U(\mathbf{r})\right)\Psi(\mathbf{r})=\epsilon\Psi(\mathbf{r}).$$

Here $U(\mathbf{r})\equiv U$ is the crystal potential of the heterostructure. To start with, we will use the following model of this potential:

$$U=U_1+G(z)[U_2-U_1]\equiv U_1+G(z)\,\delta U, \tag{1}$$

where $U_1\equiv U_1(\mathbf{r})$ and $U_2\equiv U_2(\mathbf{r})$ are periodic (with the same period) potentials, continued through the entire structure, of the left-hand and right-hand materials, respectively, the z axis is directed perpendicular to the plane of the heterojunction, $G(z)$ is the form factor of the heterojunction

$$G(z)|_{z<-d}=0, \quad G(z)|_{z>d}=1;$$

and the width of the transitional region of the heterojunction is $2d$ (non-one-dimensionality of $G(z)$ will be taken into account below).

It is natural to treat the potential $G(z)\,\delta U$ as a perturbation. As the basis for expanding the wave function we use the complete orthonormal set of Kohn–Luttinger functions $\{u_{n0}e^{i\mathbf{k}\cdot\mathbf{r}}\}$

$$\Psi(\mathbf{r})=\sum_{n'}\int \mathscr{F}_{n'}(\mathbf{k}')e^{i\mathbf{k}'\cdot\mathbf{r}}u_{n'0}d^3k', \tag{2}$$

where $u_{n0}\equiv u_{n0}(\mathbf{r})$ is the periodic Bloch amplitude for the edge ϵ_{n0} of the nth band of the left-hand crystal at the Γ point of the Brillouin zone (in the nonrelativistic limit),

$$\left(\frac{\mathbf{p}^2}{2m_0}+U_1\right)u_{n0}=\epsilon_{n0}u_{n0}.$$

The sum in Eq. (2) is over all bands, and the integral, unless otherwise stated, is over the Brillouin zone here and in what follows; $\mathscr{F}_n(\mathbf{k})$ is the envelope function for the nth band in $\mathbf{k}$ space. Following the standard procedure,[1] we obtain a system of $\mathbf{k}\cdot\mathbf{p}$ equations:[3]

$$\left(\epsilon_{n0}+\frac{\hbar^2\mathbf{k}^2}{2m_0}\right)\mathcal{F}_n(\mathbf{k})+\sum_{n'}\frac{\hbar\mathbf{p}_{nn'}\cdot\mathbf{k}}{m_0}\mathcal{F}_{n'}(\mathbf{k})$$

$$+\sum_{n'}\int\mathcal{M}_{nn'}(\mathbf{k},\mathbf{k}')\mathcal{F}_{n'}(\mathbf{k}')d^3k'=\epsilon\mathcal{F}_n(\mathbf{k});$$

$$\mathcal{M}_{nn'}(\mathbf{k},\mathbf{k}')=\sum_j C_j^{nn'}\mathcal{G}(k_z-k'_z+K_{zj})\times\delta(\mathbf{k}_\parallel-\mathbf{k}'_\parallel+\mathbf{K}_{\parallel j}). \qquad (3)$$

Here $\mathbf{p}_{nn'}=\langle n|\mathbf{p}|n'\rangle$ and $C_j^{nn'}=\langle n|\delta U\epsilon^{i\mathbf{K}_j\cdot\mathbf{r}}|n'\rangle$, and the matrix elements of the periodic operator $\mathbf{f}$ are defined as follows:

$$\langle n|\mathbf{f}|n'\rangle\frac{(2\pi)^3}{\Omega}\int_{\text{cell}}u^*_{n0}\mathbf{f}u_{n'0}d^3r,$$

where Ω is the volume of the unit cell; $\mathbf{k}_\parallel=(k_x,k_y,0)$, and K_{zj} and $\mathbf{K}_{\parallel j}$ are the components of the vector $\mathbf{K}_j$ of the inverse lattice perpendicular and parallel to the plane of the heteroboundary, respectively; $\mathcal{G}(k_z)$ is the Fourier transform of $G(z)$. Let us analyze the expression for the matrix elements $\mathcal{M}_{nn'}(\mathbf{k},\mathbf{k}')$:

$$\mathcal{M}_{nn'}(\mathbf{k},\mathbf{k}')=\sum_{j(\mathbf{K}_{\parallel j}=0)}C_j^{nn'}\mathcal{G}(k_z-k'_z+K_{zj})\,\delta(\mathbf{k}_\parallel-\mathbf{k}'_\parallel)$$

$$+\sum_{j(\mathbf{K}_{\parallel j}\neq 0)}C_j^{nn'}\mathcal{G}(k_z-k'_z+K_{zj})\times\delta(\mathbf{k}_\parallel-\mathbf{k}'_\parallel+\mathbf{K}_{\parallel j}). \qquad (4)$$

The second sum in Eq. (4) describes transfer processes in the two-dimensional Brillouin zone, when the projections of any pair of vectors $\mathbf{k}$ and $\mathbf{k}'$ from the bulk Brillouin zone onto the plane of the heterojunction satisfy the condition $\mathbf{k}'_\parallel-\mathbf{k}_\parallel=\mathbf{K}_{\parallel j}\neq 0$. For a heterojunction of arbitrary orientation such transfer processes exist. However, for the orientation of interest to us—the (001) orientation—their contribution to the desired equations for the envelope functions disappears (see Appendix B):

$$\mathcal{M}_{nn'}(\mathbf{k},\mathbf{k}')=\delta(\mathbf{k}_\parallel-\mathbf{k}'_\parallel)\left[\mathcal{G}(k_z-k'_z)\delta U_{nn'}+\sum_{j\neq 0}C_j^{nn'}\mathcal{G}(k_z-k'_z+K_j)\right], \qquad (5)$$

where we have introduced the notation $K_j=(4\pi/a)j$, $j=\pm1,\pm2,\dots$, and $\delta U_{nn'}=C_0^{nn'}$. If $G(z)$ is a sufficiently smooth function, $a\ll d$, and we are interested in states with $\overline{k_z}\ll 2\pi/a$, where $\overline{k_z}$ is a characteristic value of the quasimomentum of the state, we can neglect the second term inside the brackets in Eq. (5) and as a result obtain the well-known set of equations for the envelope functions.[3,28] In the case of an atomically sharp heterojunction, on the other hand, it is possible to proceed in the spirit of the method used in Ref. 27 to describe a short-range impurity potential. We introduce the function $G'(z)\equiv dG(z)/dz$, localized on the heteroboundary, $|z|\leqslant d$. Then, for $j\neq 0$ we have

$$\mathcal{G}(k_z-k'_z+K_j)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}G(z)\exp(-i(k_z-k'_z+K_j)z)dz$$

$$=\frac{1}{2\pi i}\frac{1}{k_z-k'_z+K_j}\int_{-\infty}^{+\infty}G'(z)\times\exp(-i(k_z-k'_z+K_j)z)dz$$

$$=\frac{1}{2\pi iK_j}\left(1-\frac{k_z-k'_z}{K_j}+\dots\right)\times\int_{-d}^{d}G'(z)\exp(-iK_jz)\times[1-i(k_z-k'_z)z+\dots]dz, \qquad (6)$$

and we can write the sum in Eq. (5) in the form of an expansion in powers of $k_z-k'_z$:

$$\sum_{j\neq 0}C_j^{nn'}\mathcal{G}(k_z-k'_z+K_j)=\sum_{s=0,1,2,\dots}\frac{(k_z-k'_z)^s}{2\pi}D_{snn'}. \qquad (7)$$

The constants in the expansion (7) have the form

$$D_{0nn'}=\sum_{j\neq 0}C_j^{nn'}\frac{1}{iK_j}\int_{-d}^{d}G'(z)\exp(-iK_jz)dz,$$

$$D_{1nn'}=\sum_{j\neq 0}C_j^{nn'}\frac{1}{iK_j}\int_{-d}^{d}G'(z)\exp(-iK_jz)\times\left(-\frac{1}{K_j}-iz\right)dz,\dots$$

The present approach fundamentally allows one to treat even mathematically sharp heterojunctions, since the necessary convergence of the coefficients $D_{lnn'}$ is ensured by the property $C_j^{nn'}\to 0$ as $K_j\to\infty$ (by the way, a physically realizable heterojunction cannot be mathematically discontinuous).

Let us consider the corrections associated with sharpness of the heterojunction. Simple estimates show that terms proportional to $D_{0nn'}$, $D_{1nn'},\dots$ can give corrections not greater in order of magnitude than $a\overline{k_z}$, $(a\overline{k_z})^2,\dots$ respectively. Our goal is to obtain one-band equations with position-dependent effective-mass parameters, which is achieved by taking account of corrections of order $(\lambda\overline{k_z})^2$ to the standard approximation. Here we have introduced a characteristic "two-band" length $\lambda=\hbar(2mE_g)^{-1/2}$. For GaAs, for example, $\lambda\approx 6$ Å. We will make use of the smallness of the parameter $\overline{k_z}d$, which will allow us to write down the final equation in quite simple form (see below). Thus, three quantities having the dimensions of length (a, d, and λ) in combination with $\overline{k_z}$ form three parameters whose smallness is employed in the present method. In our view, the situation $a\lesssim d\lesssim\lambda$ is the most realistic, being realized in semiconductor heterostrutures with sharp heteroboundaries. Thus, the parameter $\lambda\overline{k_z}$ can be taken as the main small parameter of the problem and the sum in Eq. (7) can be restricted to terms with $s=0$ and $s=1$.

As a result, the many-band system of $\mathbf{k}\cdot\mathbf{p}$ equations (3) takes the form

$$0=\left(\epsilon_{n0}-\epsilon+\frac{\hbar^2\mathbf{k}^2}{2m_0}\right)\mathscr{F}_n(\mathbf{k})+\sum_{n'}\frac{\hbar\mathbf{p}_{nn'}\cdot\mathbf{k}}{m_0}\mathscr{F}_{n'}(\mathbf{k})$$
$$+\sum_{n'}\int\left[\mathscr{G}(k_z-k_z')\,\delta U_{nn'}+\frac{1}{2\pi}D_{0nn'}\right.$$
$$\left.+\frac{k_z-k_z'}{2\pi}D_{1nn'}\right]\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'. \tag{8}$$

Here we can distinguish different contributions of the perturbation potential: the contribution of the smooth part is represented by the first term inside the brackets, and the contribution of the discontinuous part is represented by the second and third terms.

2.1. Account of the $3D$ character of the form factor

Let us now consider instead of Eq. (1) a more realistic form of the heteropotential:

$$U=U_1+g(\mathbf{r}_\parallel,z)\,\delta U, \tag{9}$$

where $\mathbf{r}_\parallel=(x,y,0)$. By definition, $g(\mathbf{r}_\parallel,z)|_{z<-d}=0$ and $g(\mathbf{r}_\parallel,z)|_{z>d}=1$, and the function $g(\mathbf{r}_\parallel,z)$ is periodic in $\mathbf{r}_\parallel$. For our case of a (001) heterostructure the unit translation vectors in the plane of the heteropotential are $\mathbf{a}_1=(1,1,0)a/2$ and $\mathbf{a}_2=(-1,1,0)a/2$. The sites of the two-dimensional inverse lattice with basis vectors $\mathbf{b}_1=(1,1,0)2\pi/a$ and $\mathbf{b}_2=(-1,1,0)2\pi/a$ are projections of the sites of the three-dimensional inverse lattice onto the (001) plane.

We expand $g(\mathbf{r}_\parallel,z)$ in a $2D$ Fourier series:

$$g(\mathbf{r}_\parallel,z)=\sum_l G_l(z)\exp(i\mathbf{K}_l\cdot\mathbf{r}_\parallel),$$

where the summation index l is defined so that the vectors $\mathbf{K}_l$ determine the sites of the indicated two-dimensional inverse lattice, and

$$G_l(z)=\frac{1}{\Omega_\parallel}\int_{\substack{2D\\ \text{cell}}} g(\mathbf{r}_\parallel,z)\exp(-i\mathbf{K}_l\cdot\mathbf{r}_\parallel)d^2r_\parallel.$$

The integration is over a unit cell of the $2D$ lattice with area $\Omega_\parallel a^2/2$. Denoting $G(z)$ as $G_0(z)$, we obtain for the perturbation potential

$$g(\mathbf{r}_\parallel,z)\,\delta U=G(z)\,\delta U+\delta U\sum_{l\neq0}G_l(z)\exp(i\mathbf{K}_l\cdot\mathbf{r}_\parallel). \tag{10}$$

It can be seen that the simple model (1) takes into account the first term in the sum (10). The functions $G_l(z)$ for $l\neq0$ are nonzero only for $z\in[-d,d]$. Therefore, the left-hand side of Eq. (3) will include an additional sum of interface contributions:

$$\sum_{n'}\int\mathscr{M}^{\parallel}_{nn'}(\mathbf{k},\mathbf{k}')\mathscr{F}_{n'}(\mathbf{k}')d^3k',$$

where

$$\mathscr{M}^{\parallel}_{nn'}(\mathbf{k},\mathbf{k}')=\sum_{l\neq0;j}\frac{C_j^{nn'}}{(2\pi)^3}\int_{\substack{\text{all}\\ \text{space}}}G_l(z)\exp(i\mathbf{K}_l\cdot\mathbf{r}_\parallel)$$
$$\times\exp(-i(\mathbf{k}-\mathbf{k}'+\mathbf{K}_j)\cdot\mathbf{r})d^3r.$$

For states with $|k_x|+|k_y|<\pi/a$ (see Appendix B) we obtain

$$\mathscr{M}^{\parallel}_{nn'}(\mathbf{k},\mathbf{k}')=\delta(\mathbf{k}_\parallel-\mathbf{k}_\parallel')$$
$$\times\sum_{l\neq0;j}\frac{C_j^{nn'}}{2\pi}\delta_{\mathbf{K}_l,\mathbf{K}_{\parallel j}}\int_{-d}^{d}G_l(z)$$
$$\times\exp(-i(k_z-k_z'+K_{zj})z)dz$$
$$=\delta(\mathbf{k}_\parallel-\mathbf{k}_\parallel')\sum_{j(\mathbf{K}_{\parallel j}\neq0)}\frac{C_j^{nn'}}{2\pi\Omega_\parallel}$$
$$\times\int_{-d}^{d}dz\int_{\substack{2D\\ \text{cell}}}d^2r_\parallel g(\mathbf{r}_\parallel,z)$$
$$\times\exp(-i\mathbf{K}_{\parallel j}\cdot\mathbf{r}_\parallel)$$
$$\times\exp(-i(k_z-k_z'+K_{zj})z).$$

For a smooth heterojunction it is necessary to retain only terms with $K_{zj}=0$ in the sum and develop the expression in the standard way^3 since the equations for the envelope functions in the $\mathbf{r}$ representation will include an additional interface potential. Note that while the smooth part $G(z)\,\delta U$ ensures mixing of states of the same crystal symmetry (i.e., the local symmetry of the smooth part of the perturbation potential $G(z)\,\delta U$ coincides with the symmetry of the bulk crystal), this additional potential also ensures mixing of states of different symmetry. We are interested in the case $\overline{k_z}d\ll1$, and this obviates the necessity of separating effects of this additional potential into contributions of smooth and discontinuous parts and allows us to use the expansion

$$\mathscr{M}^{\parallel}_{nn'}(\mathbf{k},\mathbf{k}')=\delta(\mathbf{k}_\parallel-\mathbf{k}_\parallel')\sum_{s=0,1,2,\ldots}\frac{(k_z-k_z')^s}{2\pi}D^{\parallel}_{snn'}, \tag{11}$$

where

$$D^{\parallel}_{0nn'}=\sum_{j(\mathbf{K}_{\parallel j}\neq0)}\frac{C_j^{nn'}}{\Omega_\parallel}\int_{-d}^{d}dz\int_{\substack{2D\\ \text{cell}}}d^2r_\parallel g(\mathbf{r}_\parallel,z)$$
$$\times\exp(-i\mathbf{K}_j\cdot\mathbf{r}),$$

$$D^{\parallel}_{1nn'}=\sum_{j(\mathbf{K}_{\parallel j}\neq0)}\frac{C_j^{nn'}}{\Omega_\parallel}\int_{-d}^{d}dz\int_{\substack{2D\\ \text{cell}}}d^2r_\parallel g(\mathbf{r}_\parallel,z)$$
$$\times\exp(-i\mathbf{K}_j\cdot\mathbf{r})(-iz),\ldots$$

In the expansion (11) it is necessary to keep only the first two terms; the terms proportional to $D^{\parallel}_{0nn'}$ and $D^{\parallel}_{1nn'}$ can give contributions of order $\overline{k_z}d$ and $(\overline{k_z}d)^2$, respectively.

We have shown that taking the three-dimensionality of the form factor into account [see Eq. (9)] causes no special difficulty for analysis, and we now make an important observation which will allow us to use the simple model (1). The function $g(\mathbf{r}_\parallel,z)$ has lower symmetry than $G(z)$. Specifically, it is invariant under symmetry transformations from the point group C_{2v}. The complete perturbation potential in both models, (1) and (9), on the other hand, possesses the same symmetry, both its point-group symmetry (C_{2v}) and translational symmetry in the plane of the heterojunction. Information about C_{2v} symmetry will be preserved, however, only if the contribution of the discontinuous part of the potential $G(z)\delta U$ in model (1) is taken into account. Therefore, using model (9) instead of model (1) does not give anything qualitatively new, and only leads to renormalization of some parameters, namely those that are negligibly small for the smooth heterojunction in model (1). The expressions for these parameters are very complicated, so in what follows we will stick with model (1).

2.2. Account of relativistic corrections

Let us now consider relativistic effects. We restrict the discussion to the spin–orbital interaction. The remaining relativistic contributions only influence the values of the constants that we will obtain. We assume that within the framework of perturbation theory the characteristic parameter of the spin–orbital interaction and also the difference of this parameter for the left-hand and right-hand crystals are less than or of the same order as the characteristic band discontinuity. The expansion of the total wave function, as before, is given by expression (2). Omitting intermediate manipulations, we give the resulting, quite lengthy $\mathbf{k}\cdot\mathbf{p}$ system of equations which take into account the spin–orbit interaction

$$\begin{aligned}
&\left(\epsilon_{n0}-\epsilon+\frac{\hbar^2\mathbf{k}^2}{2m_0}\right)\mathscr{F}_n(\mathbf{k})+\sum_{n'}\frac{\hbar\mathbf{p}_{nn'}\cdot\mathbf{k}}{m_0}\mathscr{F}_{n'}(\mathbf{k})\\
&+\sum_{n'}\delta U_{nn'}\int\mathscr{G}(k_z-k_z')\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'\\
&+\sum_{n'}\frac{\hbar\langle n|[\nabla U_1\times\mathbf{p}]|n'\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2}\mathscr{F}_{n'}(\mathbf{k})\\
&+\sum_{n'}\frac{\hbar\langle n|[\nabla\,\delta U\times\mathbf{p}]|n'\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2}\\
&\times\int\mathscr{G}(k_z-k_z')\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'\\
&+\sum_{n'}\frac{\hbar\langle n|[\mathbf{n}\delta U\times\mathbf{p}]|n'\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2}\int i(k_z-k_z')\\
&\times\mathscr{G}(k_z-k_z')\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'\\
&+\sum_{n'}\int\frac{\hbar^2\langle n|[\nabla\,\delta U\times\mathbf{k}']|n'\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2}\delta(\mathbf{k}_\parallel-\mathbf{k}_\parallel')\\
&\times\mathscr{G}(k_z-k_z')\mathscr{F}_{n'}(\mathbf{k}')d^3k'+\sum_{n'}\int\left(\frac{1}{2\pi}D_{0nn'}\right.\\
&\left.+\frac{k_z-k_z'}{2\pi}D_{1nn'}\right)\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'+\sum_{n'}\int\left(\frac{1}{2\pi}\mathbf{S}_{0nn'}\right.\\
&\left.+\frac{k_z-k_z'}{2\pi}\mathbf{S}_{1nn'}\right)\cdot\boldsymbol{\sigma}\mathscr{F}_{n'}(k_z',\mathbf{k}_\parallel)dk_z'+\sum_{n'}\int\frac{1}{2\pi}\delta(\mathbf{k}_\parallel\\
&-\mathbf{k}_\parallel')(\hbar\mathbf{k}'\times\boldsymbol{\sigma})\cdot\mathbf{B}_{0nn'}\mathscr{F}_{n'}(\mathbf{k}')d^3k'0. \qquad (12)
\end{aligned}$$

The vectors $\mathbf{S}_{0nn'}$, $\mathbf{S}_{1nn'}$, and $\mathbf{B}_{0nn'}$ have the following form:

$$\mathbf{S}_{0nn'}=\sum_{j\neq 0}\frac{\hbar\langle n|[\nabla(e^{iK_jz}\delta U)\times\mathbf{p}]|n'\rangle}{4iK_jm_0^2c^2}\int_{-d}^{d}G'(z)e^{-iK_jz}dz,$$

$$\begin{aligned}
\mathbf{S}_{1nn'}=&-\sum_{j\neq 0}\frac{\hbar\langle n|[\nabla(e^{iK_jz}\delta U)\times\mathbf{p}]|n'\rangle}{4K_jm_0^2c^2}\\
&\times\int_{-d}^{d}G'(z)e^{-iK_jz}z\,dz\\
&-\sum_{j\neq 0}\frac{\hbar\langle n|e^{iK_jz}[\nabla\,\delta U\times\mathbf{p}]|n'\rangle}{4iK_j^2m_0^2c^2}\int_{-d}^{d}G'(z)e^{-iK_jz}dz,
\end{aligned}$$

$$\mathbf{B}_{0nn'}=\sum_{j\neq 0}\frac{\hbar\langle n|\nabla(e^{iK_jz}\delta U)|n'\rangle}{4iK_jm_0^2c^2}\int_{-d}^{d}G'(z)e^{-iK_jz}dz.$$

Here $\mathbf{n}$ is the unit vector along the z axis, $\mathbf{n}G'(z)\equiv\nabla G(z)$, and $\boldsymbol{\sigma}$ is the Pauli matrix. On the left-hand side of Eq. (12) the fourth term describes the spin–orbital interaction in the potential of the basis semiconductor; the fifth, sixth, and seventh terms are due to the smooth part of the perturbation potential. The terms proportional to $\mathbf{S}_{0nn'}$, $\mathbf{S}_{1nn'}$, and $\mathbf{B}_{0nn'}$ are due to the sharpness of the potential. In Ref. 28, in a consideration of the state of the conduction band in heterostructures with smooth heteroboundaries, we neglected the sixth and seventh terms on the left-hand side of Eq. (12) as small. We noted that in second-order perturbation theory they, together with $\hbar\mathbf{k}\cdot\mathbf{p}_{nn'}/m_0$, give a correction only of order $(\lambda\overline{k_z})^2m/m_0$, which in this case is of order the effective mass is small in comparison with m_0. For the hole states, on the other hand, m/m_0 is able not to be a small parameter.

We do not consider $\mathbf{k}$-linear contributions of the spin–orbital interaction due to the potential U_1. They give corrections of order $(\lambda\overline{k_z})^3$ (third-order corrections, along with two terms of the form $\hbar\mathbf{k}\cdot\mathbf{p}_{nn'}/m_0$) similar to the contribution responsible for removing the spin degeneracy in the conduction band of the bulk semiconductor (we neglect terms of this order), and for the valence band it is well known that to first order the contribution of these terms is small, and to second order, along with $\hbar\mathbf{k}\cdot\mathbf{p}_{nn'}/m_0$ they only renormalize the effective-mass parameters.

As for the contributions from the sharpness of the heterojunction potential to the spin–orbit interaction, the terms proportional to $\mathbf{S}_{0nn'}$ can give corrections of order $a\overline{k_z}$, while the terms $\mathbf{S}_{1nn'}$ and $\mathbf{B}_{0nn'}$ can give terms of order $(a\overline{k_z})^2$.

It is trivial to generalize to the case of many heterojunctions. In this case it is convenient to choose the coordinates of the heteroboundaries so that the distances between them be integer multiples of $a/2$, so that the phase factor of each of the expansions of the type (7) are equal to unity.

3. PROBLEM OF TRANSFORMATION OF THE EFFECTIVE-MASS EQUATIONS FROM QUASIMOMENTUM SPACE TO COORDINATE SPACE

In the following section we obtain the one-band effective-mass equations for the conduction band and the valence band. But first of all, we must discuss a problem arising in the method of envelope functions and associated with the boundedness of **k** space. Let us consider the following one-band equation for the envelope functions $f(k_z)$ in **k** space:

$$\int H(k_z,k_z')f(k_z')dk_z'=\epsilon f(k_z), \tag{13}$$

where k_z and k_z' are bounded by the Brillouin zone. Using the transformation (13) to go over to the coordinate representation we obtain, generally speaking, an integral equation. The problem consists in the accuracy with which it is possible to obtain a differential equation in **r** space. Let us consider an equation similar to Eq. (13), but in which k_z and k_z' belong to the entire inverse space:

$$\int_{-\infty}^{+\infty} H(k_z,k_z')g(k_z')dk_z'=\epsilon g(k_z). \tag{14}$$

The Fourier transform of Eq. (14) with the system of equations (12) taken into account gives a differential equation in the **r** representation. If the function $g(k_z)$ vanishes for k_z not in the Brillouin zone, they are also solutions of Eq. (13) and we have solved our problem exactly. In general this is not so. But in order for Eqs. (13) and (14) to be approximately equivalent, it is necessary that $g(k_z)$ be small for k_z not in the Brillouin zone. In the theory of smooth perturbations this smallness is ensured by exponentially decaying envelope functions in the **k** representation; however, in the case of discontinuous perturbations the envelope functions are decreasing functions of k_z with only a power-law falloff. Thus, if the envelope function possesses one discontinuity, its Fourier transform satisfies $g(k_z)\propto(\delta\overline{g}/\overline{g})(k_z)^{-1}$ for large k_z (where the exponential contributions associated with the effects of smooth fields have decayed); here $(\delta\overline{g}/\overline{g})$ is a typical relative discontinuity of the function in **r** space. If we consider, for example, the standard effective-mass equations[1] with discontinuous potentials, then the second derivatives of the corresponding envelope functions will be discontinuous with characteristic relative discontinuity of the order of unity (again, for states whose energies, measured from the band edge of the left-hand or right-hand material, is of the order of the band discontinuity), and the error incurred by using differential equations will be of order $(\overline{k_z}/K)^3$, where K is the radius of the Brillouin zone along the k_z axis.

In the case of a quantum well of width L it is possible to treat two cases: $\overline{k_z}L\gtrsim 1$ and $\overline{k_z}L\ll 1$. In the first case the error is of the same order as for a single heterojunction; in the second case it can be of the order $(\overline{k_z}L)^{-1}(\overline{k_z}/K)^3$. This is an upper estimate. For a symmetric quantum well in the conduction band, for example, the error depends on the sign of the product of the values of the envelope function on the heteroboundaries, and for states of the second $2D$ subband it is more. In the limiting case of a narrow quantum well, $L\lesssim 1/K$, the potential can be replaced by a δ function. We obtain an envelope function with a discontinuous derivative and error of order $(\overline{k_z}/K)^2$.

Above, in Eq. (13), we tacitly assumed that the Hamiltonian $H(k_z,k_z')$ is defined for all k_z and k_z' belonging to the Brillouin zone. Since our goal is to obtain the one-band equations, we must take one more circumstance into account. In **k** space near the Γ point there exists a region Λ_1 in which the spectrum of states of the conduction band of the bulk material can be written in the form of a series in powers of the quasimomentum (for states of a degenerate band the spectrum is determined by a matrix whose elements are such series). The series converges for $|\mathbf{k}|<1/2\lambda$, as follows from the two-band approximation (such an estimate is valid if the effective mass is formed mainly by the $\mathbf{k}\cdot\mathbf{p}$ interaction). This is the region described by the Hamiltonian of the one-band equation. There is also a region Λ_2, where the interaction of states of isolated bands with distant bands is not described by this series. In our case of sharp heterojunctions the envelope functions in the **k** representation fall off according to a power law; therefore we should also provide a valid description of the region Λ_2, which will be done elsewhere in connection with the problem of intervalley mixing of states in heterostructures. Here we only mention that if the ratio of the characteristic band discontinuity to the energy gap between the interesting states in the region Λ_1 and the states in the region Λ_2 is a small parameter $\omega\ll 1$, then the error incurred by neglecting the region Λ_2 will be less than or of order $\omega(a\overline{k_z})^2$. Thus, the effective radius in **k** space determining the accuracy of reducing the integral equation to a differential equation is in fact not determined by the size of the Brillouin zone along the k_z axis, but depends on which bands are taken to be distant and are ''eliminated'' by the unitary transformation. In our case, this radius is of the order of $1/\lambda$.

Below we will obtain an equation for the conduction band, leaving out details of the unitary transformation to **k** space and at once carrying out the transformation to the **r** representation. Formally, the final equation is a fourth-order differential equation, and the envelope function satisfying it in the case of a discontinuous potential, the most unfavorable case for accuracy, has a discontinuous second derivative with characteristic discontinuity of the order of the second derivative itself. It is possible to proceed otherwise. Reducing the fourth-order differential equation to a physically equivalent second-order equation,[28] we obtain a discontinuous envelope function with characteristic relative discontinuity of order $(\overline{k_z}\lambda)^2$. This means that for a single heterojunction or a wide quantum well the accuracy of the obtained effective-mass equation is limited as a result of having to take account of all terms up to order $(\overline{k_z}\lambda)^3$. In the case of a narrow quantum well, on the other hand, even for $L\sim\lambda$ the effective-mass equation should include only first-order corrections associated with effects of sharpness of the heterojunction, and considering all remaining orders, including those leading to spatial dependence of the effective-mass parameters, will yield an excess of accuracy. In such a case the short-range potential formalism, already used above to obtain the expansion (11), is convenient. This will be taken up in Sec. 4.3.

4. ONE-BAND EQUATIONS

4.1. Conduction band

4.1.1. Smooth heterojunction

The transformation from a many-band system of equations for the envelope functions to a single-band equation is realized in the standard way.[1,27] For smooth heterojunctions the one-band equation for the envelope functions for the *c* band (conduction band) was derived in Ref. 28. In the **r** representation it has the form

$$\begin{aligned}&\epsilon_{c0}F_c(\mathbf{r})+\frac{1}{2}m^{\alpha_1}(z)\mathbf{p}m^{\beta_1}(z)\mathbf{p}m^{\alpha_1}(z)F_c(\mathbf{r})\\&\quad+\Gamma(z)\Delta U_cF_c(\mathbf{r})+\alpha_0\mathbf{p}^4F_c(\mathbf{r})+\beta_0(\mathbf{p}_\parallel^2p_z^2\\&\quad+p_x^2p_y^2)F_c(\mathbf{r})+\eta(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}\Gamma'(z)F_c(\mathbf{r})=\epsilon F_c(\mathbf{r}).\end{aligned}\tag{15}$$

The product of the discontinuity in the conduction band ΔU_c and the modified form factor of the heterojunction $\Gamma(z)$ is given by

$$\Gamma(z)\Delta U_c=G(z)\delta U_{cc}+\sum_n{}'\frac{|\delta U_{cn}|^2}{\epsilon_{c0}-\epsilon_{n0}}G^2(z),$$

so that in all small corrections $G(z)$ can be replaced by $\Gamma(z)$. The position-dependent effective mass is given by

$$m(z)=m_1[1+m_1(\mu_1-\mu_2)\Gamma(z)],$$

and m_1 is the effective mass of the edge of the conduction band of the left-hand material, and if m_2 is the effective mass of the conduction band of the right-hand material we have

$$1/m_2=1/m_1+\mu_2-\mu_1.$$

The parameters α_1 and β_1 are defined as follows:

$$\alpha_1=\frac{\mu_1}{2(\mu_2-\mu_1)},\quad 2\alpha_1+\beta_1=-1.$$

The parameters μ_1 and μ_2 are given by

$$\begin{aligned}\mu_1=&\sum_n{}'\frac{2|\langle c|p_x|n\rangle|^2\delta U_{cc}}{m_0^2(\epsilon_{c0}-\epsilon_{n0})^2}\\&-\sum_{n,l}{}'\frac{4\langle c|p_x|n\rangle\langle n|p_x|l\rangle\delta U_{lc}}{m_0^2(\epsilon_{c0}-\epsilon_{n0})(\epsilon_{c0}-\epsilon_{l0})},\end{aligned}\tag{16}$$

$$\mu_2=\sum_{n,l}{}'\frac{2\langle c|p_x|n\rangle\delta U_{nl}\langle l|p_x|c\rangle}{m_0^2(\epsilon_{c0}-\epsilon_{n0})(\epsilon_{c0}-\epsilon_{l0})}.\tag{17}$$

In Eq. (15) α_0 and β_0 are the nonparabolicity parameters of the bulk spectrum. Finally, the last parameter entering into the equation is

$$\eta=\sum_{n,l}{}'\frac{\hbar^2\langle c|p_z|n\rangle\langle n|[\nabla\delta U\times\mathbf{p}]_x|l\rangle\langle l|p_y|c\rangle}{4im_0^4c^2(\epsilon_{c0}-\epsilon_{n0})(\epsilon_{c0}-\epsilon_{l0})}.$$

In the Introduction it was pointed out that an invalid expression for the effective mass of the edge of the conduction band of a non-basis semiconductor was obtained in Ref. 14 by direct elimination of the small envelope functions. This corresponds to the absence in expression (16) of the second sum, and in expression (17), of the terms with $n\neq l$. Thus it is necessary to be careful when using the method of direct elimination of small envelope functions.

The relation of the envelope function of the conduction band $F_c(\mathbf{r})$ with the total wave function is given by

$$\begin{aligned}\Psi(\mathbf{r})=&u_{c0}\{1+2Rm_1(\Gamma(z)\Delta U_c+\epsilon_{c0}-\epsilon)\}F_c(\mathbf{r})\\&+\sum_n{}'\frac{u_{n0}}{(\epsilon_{c0}-\epsilon_{n0})}\left[\frac{\hbar\langle n|\mathbf{p}|c\rangle\cdot\nabla}{im_0}+\delta U_{nc}\Gamma(z)\right.\\&\left.+\sum_{l,\alpha,\beta}{}'\frac{\langle n|p_\alpha|l\rangle\langle l|p_\beta|c\rangle}{m_0^2(\epsilon_{c0}-\epsilon_{l0})}p_\alpha p_\beta\right]F_c(\mathbf{r}).\end{aligned}\tag{18}$$

Here

$$R=\frac{1}{2}\sum_n{}'\frac{|\langle c|p_x|n\rangle|^2}{m_0^2(\epsilon_{c0}-\epsilon_{n0})^2},$$

and the term $2Rm_1(\Gamma(z)\Delta U_c+\epsilon_{c0}-\epsilon)$ inside the braces in expression (18) comes from the term $-R\mathbf{p}^2$ treated as a perturbation using the standard effective-mass equation. In the brackets in expression (18) we neglected the term

$$\frac{\hbar\langle n|[\nabla U_1\times\mathbf{p}]|c\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2}+\Gamma(z)\frac{\hbar\langle n|[\nabla\delta U\times\mathbf{p}]|c\rangle\cdot\boldsymbol{\sigma}}{4m_0^2c^2},\tag{19}$$

since the largest contribution to the matrix elements (19) comes from the region of the potential near the atomic nuclei in which the spin–orbital interaction operator can be written in the form of a product of operators of the electron spin and the orbital angular momentum, and the function u_{c0} is spherically symmetric (the orbital momentum is zero).

In Ref. 28 it was shown that for $\overline{k_z}d\ll 1$ Eq. (15) can be replaced by an equivalent equation where the Heaviside step function $\theta(z-z_0)$ replaces the function $\Gamma(z)$, and the coordinate z_0 assigning the position of the mathematically sharp heterojunction can be chosen arbitrarily within the limits $-d\leqslant z_0\leqslant d$. The method used in Ref. 28 for this transformation is not the most convenient. There is a simpler way of transforming to a mathematically sharp heterojunction based on the following chain of identities, valid for operators acting on smooth functions:

$$\begin{aligned}\Gamma(z)=&\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk_ze^{ik_zz}\int_{-\infty}^{+\infty}\Gamma(z')e^{-ik_zz'}dz'\\&=\theta(z-z_0)+\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk_ze^{ik_z(z-z_0)}\int_{-d}^{d}(\Gamma(z')\\&-\theta(z'-z_0))e^{-ik_z(z'-z_0)}dz'\approx\theta(z-z_0)\\&+\delta(z-z_0)\left[\int_{-d}^{d}\Gamma(z)dz-(d-z_0)\right]+\delta'(z-z_0)\\&\times\left[\int_{-d}^{d}\Gamma(z)(z_0-z)dz+\frac{(d-z_0)^2}{2}\right].\end{aligned}$$

Setting $z_0=0$, instead of Eq. (15) we obtain a more convenient form of the effective-mass equation for a smooth heterojunction:

$$[\epsilon_{c0}-\epsilon+\theta(z)\Delta U_c+\Delta U_c\rho_0\delta(z)]F_c(\mathbf{r})$$
$$+\frac{1}{2}m^{\alpha_2}(z)\mathbf{p}m^{\beta_2}(z)\cdot\mathbf{p}m^{\alpha_2}(z)F_c(\mathbf{r})+\alpha_0\mathbf{p}^4F_c(\mathbf{r})$$
$$+\beta_0(\mathbf{p}_\parallel^2p_z^2+p_x^2p_y^2)F_c(\mathbf{r})+\eta(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}\delta(z)F_c(\mathbf{r})=0. \tag{20}$$

Here

$$m(z)=m_1[1+m_1(\mu_1-\mu_2)\theta(z)],$$

and the parameter α_2 not only depends on the materials of the heterojunction but also takes into account its finite width through its dependence on $\Gamma(z)$:

$$\alpha_2=\frac{\mu_1+2\Delta U_c\hbar^{-2}[d^2-\int_{-d}^{d}2\Gamma(z)zdz]}{2(\mu_2-\mu_1)}.$$

The relation $2\alpha_2+\beta_2=-1$ is preserved, and

$$\rho_0=\int_{-d}^{d}\Gamma(z)dz-d.$$

The transformation to a mathematically sharp heterojunction described above is dictated only by arguments of convenience, and the explicit form of the function $\Gamma(z)$ appears only in integral form in the expressions for the two parameters α_2 and ρ_0.

4.1.2. Sharp heterojunction

To include corrections associated with the nonsmoothness of the heterojunction in Eq. (20), it is necessary in the standard procedure[1,27] to take into account the contributions of the terms $D_{0nn'}$, $D_{1nn'}$, and $\mathbf{B}_{0nn'}$ in first-order perturbation theory ($\mathbf{S}_{0cc}=\mathbf{S}_{1cc}=0$, as follows from zinc-blende symmetry), and the contributions of the terms $D_{0nn'}$ and $\mathbf{S}_{0nn'}$ should still be treated in the second-order theory along with the terms $\hbar\mathbf{k}\cdot\mathbf{p}_{nn'}/m_0$. Utilizing symmetry properties, it is not hard to obtain the following additional term H_{abr} to the Hamiltonian of Eq. (15):

$$H_{abr}=D_{0cc}\delta(z)+\rho\delta'(z)+\tilde{\eta}(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}\delta(z). \tag{21}$$

Here

$$D_{0cc}=-\sum_{j\neq0}\frac{\langle c|\delta U\cos(K_jz)|c\rangle}{K_j}\int_{-d}^{d}G'(z)\sin(K_jz)dz,$$

$$\rho=\sum_{n;j\neq0}{}'\frac{\hbar\langle c|p_z|n\rangle\langle n|\delta U\sin(K_jz)|c\rangle}{iK_jm_0(\epsilon_{c0}-\epsilon_{n0})}$$
$$\times\int_{-d}^{d}G'(z)\cos(K_jz)dz+\sum_{j\neq0}\frac{\langle c|\delta U\cos(K_jz)|c\rangle}{K_j}$$
$$\times\int_{-d}^{d}G'(z)\left(\frac{\cos(K_jz)}{K_j}+z\sin(K_jz)\right)dz,$$

$$\tilde{\eta}=\sum_{j\neq0}\left(\sum_n{}'\frac{\hbar\langle c|[\nabla(\sin(K_jz)\delta U)\times\mathbf{p}]_x|n\rangle\langle n|p_y|c\rangle}{2K_jm_0^3c^2(\epsilon_{c0}-\epsilon_{n0})}\right.$$
$$\left.-\frac{\hbar\langle c|\nabla_z(\sin(K_jz)\delta U)|c\rangle}{4K_jm_0^2c^2}\right)\int_{-d}^{d}G'(z)\cos(K_jz)dz.$$

Comparing expression (21) with the Hamiltonian of Eq. (20), we see that for the conduction band taking the sharpness of the heterojunction into account does not alter the form of the one-band equation and simply renormalizes the constants used in it. If we take Eqs. (20) and (21) into account, the desired equation for the conduction band takes the form

$$(\epsilon_{c0}-\epsilon+\theta(z)\Delta U_c+d_1\delta(z))F_c(\mathbf{r})+\frac{1}{2}m^{\alpha}(z)\mathbf{p}m^{\beta}(z)$$
$$\times\mathbf{p}m^{\alpha}(z)F_c(\mathbf{r})+\alpha_0\mathbf{p}^4F_c(\mathbf{r})+\beta_0(\mathbf{p}_\parallel^2p_z^2$$
$$+p_x^2p_y^2)F_c(\mathbf{r})+d_2(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}\delta(z)F_c(\mathbf{r})=0. \tag{22}$$

The term $\rho\delta'(z)$ in expression (21) was absorbed by the kinetic-energy operator; as a result, for α we obtain

$$\alpha=\frac{\mu_1+2\Delta U_c\hbar^{-2}[d^2-\int_{-d}^{d}2\Gamma(z)zdz]+4\rho\hbar^{-2}}{2(\mu_2-\mu_1)}$$

(here the error incurred in Ref. 29 has been corrected), and $2\alpha+\beta=-1$, with $d_1=D_{0cc}+\Delta U_c\rho_0$ and $d_2=\eta+\tilde{\eta}$.

Let us discuss the Hamiltonian of Eq. (22). The first term represents the potential energy of an electron in the c band. The possible existence of a hetero-interface term proportional to d_1 was discussed in Ref. 5; it is clear that this contribution disappears for a mathematically discontinuous heterojunction (in this rather unrealistic case models (1) and (9) are identical). The second term is the position-dependent kinetic-energy operator, which is quadratic in the momentum; such a form was proposed in Ref. 6 more generally. Note that the parameter α is not a universal constant but depends both on the materials and on the shape of the transitional region of the heterojunction, as seems natural even intuitively (see also Ref. 8). If it happens that $m(z)=\mathrm{const}$, i.e., $\mu_1=\mu_2$, indeterminate expressions of the form 1^∞ arise in the form of the kinetic-energy operator used here which are easily evaluated. Such indeterminate forms do not arise if we use a different, equivalent form of this operator where we separate out a term analogous to the relativistic Darwin term (see Ref. 28):

$$T_2=\mathbf{p}\frac{1}{2m(z)}\cdot\mathbf{p}+\left(\frac{\mu_1\hbar^2}{4}+\Delta U_c\right.$$
$$\left.\times\left[\frac{d^2}{2}-\int_{-d}^{d}\Gamma(z)zdz\right]+\rho\right)\delta'(z).$$

The third and fourth terms in the Hamiltonian of Eq. (22) describe corrections to the weak nonparabolicity and depend only on the bulk parameters. The fifth term describes the interface spin–orbital interaction (see, e.g., Ref. 32), whose strength (d_2) depends not only on the materials of the heterojunction but also on the shape of the transitional region. The relation between the total wave function and the envelope functions of the conduction band is given by Eq. (18), where we can set $\Gamma(z)=\theta(z)$.

4.1.3. Boundary conditions on the envelope functions

From Eq. (22) it is not hard to obtain boundary conditions imposed on the envelope functions at the heterojunction; to this end it is necessary to reduce Eq. (22) to a second-order differential equation by employing the smallness of the contribution proportional to $\mathbf{p}^4$ (Ref. 28). We present only the result:

$$\left.\begin{pmatrix} F_c(\mathbf{r}) \\ F_c'(\mathbf{r}) \end{pmatrix}\right|_{z=+0} = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix} \left.\begin{pmatrix} F_c(\mathbf{r}) \\ F_c'(\mathbf{r}) \end{pmatrix}\right|_{z=-0},$$

where

$$d_{11} = 1 + 4m_1^2\alpha_0\Delta U_c + m_1\alpha\left(\frac{1}{m_2} - \frac{1}{m_1}\right),$$

$$d_{21} = \frac{2m_1}{\hbar^2}(d_1 + d_2(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}),$$

and

$$d_{22} = \frac{m_2}{m_1} + 4m_1^2\alpha_0\Delta U_c - m_1\alpha\left(\frac{1}{m_2} - \frac{1}{m_1}\right).$$

In the approximation of the standard effective-mass method we have a position-independent effective mass and a discontinuous (step-function) potential. Corrections which are first order in the small parameter of the problem are taken into account by including in the standard equation a δ-function potential proportional to d_1 (formally, this is a correction of order $\overline{k_z}d$). The complete equation (22) also includes all corrections of order $(\lambda\overline{k_z})^2$. It is not possible to take into account smaller contributions in the one-band version of the method of envelope functions because the unavoidable error arising in the transformation from the many-band system of integral equations to the one-band differential equation is of the same order.

It is not hard to generalize the above results to the case of an arbitrary heterostructure. It is easy to do this proceeding from symmetry arguments for a symmetric quantum well with two equivalent heterojunctions with coordinates $z=0$ and $z=L$, where $L=\mathscr{N}a/2$, where $\mathscr{N}$ is an integer. The effective Hamiltonian is

$$\begin{aligned} H^{el} = {} & \epsilon_{c0} + \Delta U_c\{\theta(z) - \theta(z-L)\} + \frac{1}{2}m^{\alpha}(z)\mathbf{p}m^{\beta}(z) \\ & \times \mathbf{p}m^{\alpha}(z) + \alpha_0\mathbf{p}^4 + \beta_0(\mathbf{p}_{\parallel}^2p_z^2 + p_x^2p_y^2) + d_1\{\delta(z) \\ & + \delta(z-L)\} + d_2\{\delta(z) - \delta(z-L)\}(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}. \end{aligned}$$

4.2. Valence band

The distinction between the effective-mass method for the valence band and the effective-mass method for the conduction band consists, in principle, simply of the necessity of considering more elaborate equations in the case of the valence band. The main points of the problem of deriving the equation for the hole states with position-dependent effective-mass parameters were noted in Ref. 29. The equation for the c band already contains three new parameters which depend on the bulk properties of the materials of the heterostructure and on the properties of the heteroboundary. For the valence band, there is a larger number of such parameters, which may be seen as rendering such an equation of little practical value. Therefore we limit ourselves here to a derivation of first-order corrections to the standard equation in the above small parameter.

In the basis $\{|J,j_z\rangle\}$ of eigenfunctions of the total angular momentum J and its projection j_z, with these eigenfunctions being combinations of Bloch functions $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{Z}$ of the top of the valence band (transforming according to the representation Γ_{15}) and the spin, the matrix of the effective Hamiltonian for the valence band $\mathbf{H}$ in this approximation is a sum of the 6×6 matrix of the standard kinetic-energy operator $\mathbf{T}$ (we neglect the small contribution of $\mathbf{k}$-linear bulk terms from the spin–orbit interaction) and the 6×6 matrix of the potential-energy operator $\mathbf{V}$:

$$\mathbf{H} = \mathbf{T} + \mathbf{V}.$$

Of course, in the standard approximation $\mathbf{V}$ contains only diagonal discontinuous (step-function) potentials. Additional terms appear in the approximation that follows which are diagonal and nondiagonal δ-function potentials.

It is convenient first to find the elements of the potential-energy matrix $\tilde{\mathbf{V}}$ in the basis $\{\mathscr{X},\mathscr{Y},\mathscr{Z}\}$, and then compose from them the necessary linear combinations and transform to $\mathbf{V}$:

$$\begin{aligned} \tilde{V}_{\mathscr{Z}\mathscr{Z}} = \tilde{V}_{\mathscr{X}\mathscr{X}} & = \epsilon_{\mathscr{X}0} + \delta U_{\mathscr{X}\mathscr{X}}G(z) + D_{0\mathscr{X}\mathscr{X}}\delta(z) \\ & \approx \epsilon_{\mathscr{X}0} + \delta U_{\mathscr{X}\mathscr{X}}\theta(z) + (D_{0\mathscr{X}\mathscr{X}} + \rho_0\delta U_{\mathscr{X}\mathscr{X}})\delta(z), \end{aligned}$$

$$\begin{aligned} \tilde{V}_{\mathscr{X}\mathscr{Y}} & = \frac{1}{3i}(\Delta + \delta\Delta G(z))\sigma_z + D_{0\mathscr{X}\mathscr{Y}}\delta(z) + S^z_{0\mathscr{X}\mathscr{Y}}\delta(z)\sigma_z \\ & \approx \frac{1}{3i}(\Delta + \delta\Delta\,\theta(z))\sigma_z + D_{0\mathscr{X}\mathscr{Y}}\delta(z) \\ & \quad + \left(S^z_{0\mathscr{X}\mathscr{Y}} + \frac{\rho_0\delta\Delta}{3i}\right)\delta(z)\sigma_z, \end{aligned}$$

where

$$\Delta = \frac{3\hbar i\langle\mathscr{X}|[\nabla U_1\times\mathbf{p}]_z|\mathscr{Y}\rangle}{4m_0^2c^2},$$

$$\delta\Delta = \frac{3\hbar i\langle\mathscr{X}|[\nabla\,\delta U\times\mathbf{p}]_z|\mathscr{Y}\rangle}{4m_0^2c^2}.$$

Analogously,

$$\begin{aligned} \tilde{V}_{\mathscr{X}\mathscr{Z}} & = \frac{i}{3}(\Delta + \delta\Delta G(z))\sigma_y + S^y_{0\mathscr{X}\mathscr{Z}}\delta(z)\sigma_y \\ & \approx \frac{i}{3}(\Delta + \delta\Delta\,\theta(z))\sigma_y + \left(S^y_{0\mathscr{X}\mathscr{Z}} + \frac{i\rho_0\delta\Delta}{3}\right) \\ & \quad \times\delta(z)\sigma_y. \end{aligned}$$

The remaining elements of $\tilde{\mathbf{V}}$ can be obtained from those shown above by cyclic permutation of the indices.

The contribution of the sixth and seventh terms on the left-hand side of Eq. (12), as can be seen, is not included in $\tilde{\mathbf{V}}$ since it is negligibly small for the following reason:

$$\langle\mathscr{X}|\nabla_z U_1|\mathscr{Y}\rangle = 0$$

(see Ref. 27, Sec. 21). This means that the matrix element

$$\langle \mathscr{X}|\nabla_z \delta U|\mathscr{Y}\rangle = \langle \mathscr{X}|\nabla_z U_2|\mathscr{Y}\rangle.$$

For the Bloch functions of the nth band at the Γ point of the right-hand crystal in the nonrelativistic limit ($\tilde{u}_{n0}$) we now have

$$\tilde{u}_{n0} = u_{n0} + {\sum_{n'}}' \frac{\delta U_{n'c} u_{n'0}}{\epsilon_{c0} - \epsilon_{n'0}}, \tag{23}$$

and for the corresponding functions of the edge of the valence band of the right-hand crystal $\tilde{\mathscr{X}}$ and $\tilde{\mathscr{Y}}$ the relation $\langle \tilde{\mathscr{X}}|\nabla_z U_2|\tilde{\mathscr{Y}}\rangle = 0$ holds. That is, the seventh term on the left-hand side of Eq. (12) gives corrections only of order $(\lambda \overline{k_z})^3$, and not $\lambda \overline{k_z}$. In our approximation, on the other hand, it is necessary in general to neglect the difference between the Bloch functions $\tilde{u}_{n0}$ and u_{n0}. Hence it follows from invariance of the equation under time reversal that the contribution from the sixth term on the left-hand side of Eq. (12) is negligibly small.

It is not difficult now to obtain the elements of **V**. We choose the phases the same as was done in Ref. 1 and write

$$|1\rangle \equiv \left|\frac{3}{2}, \frac{3}{2}\right\rangle, \quad |2\rangle \equiv \left|\frac{3}{2}, -\frac{3}{2}\right\rangle, \quad |3\rangle \equiv \left|\frac{3}{2}, \frac{1}{2}\right\rangle,$$

$$|4\rangle \equiv \left|\frac{3}{2}, -\frac{1}{2}\right\rangle, \quad |5\rangle \equiv \left|\frac{1}{2}, \frac{1}{2}\right\rangle, \quad |6\rangle \equiv \left|\frac{1}{2}, -\frac{1}{2}\right\rangle$$

(pairs of states of heavy holes, light holes, and states of the split-off band). Thus, the desired potential-energy matrix takes the form

$$\mathbf{V} = \begin{pmatrix} V_{\Gamma_8}\mathbf{1} & V_0\sigma_y & -i\sqrt{2}V_0\sigma_y \\ V_0^{\dagger}\sigma_y & V_{\Gamma_8}\mathbf{1} & \mathbf{0} \\ i\sqrt{2}V_0^{\dagger}\sigma_y & \mathbf{0} & V_{\Gamma_7}\mathbf{1} \end{pmatrix}, \tag{24}$$

where

$$V_{\Gamma_8} = E_{\Gamma_8} + \Delta U_{\Gamma_8}\theta(z) + \chi_1\delta(z),$$

$$V_{\Gamma_7} = E_{\Gamma_7} + \Delta U_{\Gamma_7}\theta(z) + \chi_2\delta(z),$$

$$V_0 = \frac{D_{0\mathscr{X}\mathscr{Y}}}{\sqrt{3}}\,\delta(z).$$

Here we have introduced the notation

$$E_{\Gamma_8} = \epsilon_{\mathscr{X}0} + \frac{1}{3}\Delta, \quad \Delta U_{\Gamma_8} = \delta U_{\mathscr{X}\mathscr{X}} + \frac{1}{3}\delta\Delta,$$

$$E_{\Gamma_7} = \epsilon_{\mathscr{X}0} - \frac{2}{3}\Delta, \quad \Delta U_{\Gamma_7} = \delta U_{\mathscr{X}\mathscr{X}} - \frac{2}{3}\delta\Delta,$$

$$\chi_1 = D_{0\mathscr{X}\mathscr{X}} + \rho_0\delta U_{\mathscr{X}\mathscr{X}} + iS^z_{0\mathscr{X}\mathscr{Y}} + \frac{1}{3}\rho_0\delta\Delta,$$

$$\chi_2 = D_{0\mathscr{X}\mathscr{X}} + \rho_0\delta U_{\mathscr{X}\mathscr{X}} - 2iS^z_{0\mathscr{X}\mathscr{Y}} - \frac{2}{3}\rho_0\delta\Delta.$$

The expressions for $D_{0\mathscr{X}\mathscr{X}}$, $D_{0\mathscr{X}\mathscr{Y}}$, and $S^z_{0\mathscr{X}\mathscr{Y}}$ have the following form:

$$D_{0\mathscr{X}\mathscr{X}} = -\sum_{j\neq 0} \frac{\langle \mathscr{X}|\delta U \cos(K_j z)|\mathscr{X}\rangle}{K_j} \times \int_{-d}^{d} G'(z)\sin(K_j z)\,dz,$$

$$D_{0\mathscr{X}\mathscr{Y}} = \sum_{j\neq 0} \frac{\langle \mathscr{X}|\delta U \sin(K_j z)|\mathscr{Y}\rangle}{K_j} \times \int_{-d}^{d} G'(z)\cos(K_j z)\,dz, \tag{25}$$

$$S^z_{0\mathscr{X}\mathscr{Y}} = -\sum_{j\neq 0} \frac{\hbar\langle \mathscr{X}|[\nabla(\cos(K_j z)\,\delta U)\times\mathbf{p}]_z|\mathscr{Y}\rangle}{4K_j m_0^2 c^2} \times \int_{-d}^{d} G'(z)\sin(K_j z)\,dz.$$

Thus, within the framework of the $\mathbf{k}\cdot\mathbf{p}$ method we have shown that in (001) III–V heterostructures mixing of heavy (hh) and light (lh) holes takes place at the center of the $2D$ Brillouin zone (see Ref. 33 and the references cited therein) which bears no relation to the k_z-linear bulk terms from the spin–orbit interaction. This mixing is governed by the parameter $D_{0\mathscr{X}\mathscr{Y}}$, which was estimated in Ref. 33 on the basis of experimental data for GaAs/AlAs heterostructures: $D_{0\mathscr{X}\mathscr{Y}} \simeq 500\,\text{meV}\cdot\text{Å}$.

In Ref. 29 it was concluded that the strength of the mixing of the heavy and light holes at the center of the $2D$ Brillouin zone is greater for sharp heterojunctions than for heterojunctions with smoothly varying chemical composition. But this is valid only in model (1). Generally speaking, one can draw conclusionsonly about the dependence of the strength of this mixing on the structure of the transitional region of the heterojunctions.

For a symmetric quantum well with boundaries at $z=0$ and $z=L$ the elements of the potential-energy matrix can be easily obtained from symmetry arguments:

$$V_{\Gamma_8} = E_{\Gamma_8} + \Delta U_{\Gamma_8}(\theta(z) - \theta(z-L)) + \chi_1(\delta(z) + \delta(z-L)),$$

$$V_{\Gamma_7} = E_{\Gamma_7} + \Delta U_{\Gamma_7}(\theta(z) - \theta(z-L)) + \chi_2(\delta(z) + \delta(z-L)),$$

$$V_0 = \frac{D_{0\mathscr{X}\mathscr{Y}}}{\sqrt{3}}(\delta(z) - \delta(z-L)).$$

4.3. Equation for the envelope functions for a narrow well in the conduction band

We will devote separate attention to the problem of electron states in narrow quantum wells because at present it is widely held that the effective-mass method is inapplicable in such cases. Here we treat only states of the c band in a (001) III–V heterostructure consisting of related isolated quantum well (or narrow barrier) semiconductors for the case when its width satisfies $L \lesssim \lambda$. Hole states are easily treated in an analogous way. Now the contributions from the steepness of the potential begin to play a much greater role than in the case of wide quantum wells. In fact, the estimate $\overline{k_z^2}/2m \sim \delta U_{cc}$ is valid only when the width of the quantum well is greater than or of the order of the characteristic localization length of the states. For states in a narrow quantum well, on the other hand, of course $\overline{k_z^2}/2m \sim \delta U_{cc}\overline{k_z}L$, and the contribution to the energy eigenvalue from the terms associated with steepness of the potential can be estimated to first order as $\delta U_{cc}\overline{k_z}a$, which implies that they must be taken into account even in the zeroth approximation.

Thus, the potential of the heterostructure under consideration can be written as

$$U = U_1 + P[U_2 - U_1] \equiv U_1 + P(z)\,\delta U,$$

where U_1 and U_2 are periodic crystal potentials of the ''barrier'' (basis) and ''well'' semiconductor, respectively, and $P(z)$ is the form factor of the heterostructure. We choose L so that $P(z<-L/2) = P(z>L/2) = 0$. It is natural to consider $P(z)$ as a local function on the scale of variation of the envelope function of the conduction band since $L\overline{k_z} \ll 1$. The one-band equation has the following form:

$$\left[\epsilon_{c0} - \epsilon + \frac{\mathbf{p}^2}{2m} + b_1\delta(z) + b_2\delta'(z) + b_3\delta(z)(\mathbf{p}\times\mathbf{n})\cdot\boldsymbol{\sigma}\right]F_c(\mathbf{r}) = 0, \tag{26}$$

where m is the position-independent effective mass, and we have the following expressions for the three parameters b_i ($i = 1,2,3$):

$$\begin{aligned} b_1 &= \sum_j \langle c|\delta U\cos(K_j z)|c\rangle \int_{-L/2}^{L/2} P(z)\cos(K_j z)\,dz, \\ b_2 &= \sum_{n;j}{}' \frac{\hbar\langle c|p_z|n\rangle\langle n|\delta U\sin(K_j z)|c\rangle}{i m_0(\epsilon_{c0}-\epsilon_{n0})} \\ &\quad\times \int_{-L/2}^{L/2} P(z)\sin(K_j z)\,dz - \sum_j \langle c|\delta U\cos(K_j z)|c\rangle \\ &\quad\times \int_{-L/2}^{L/2} P(z)\cos(K_j z)\,z\,dz, \\ b_3 &= \sum_j \left(\sum_n{}' \frac{\hbar\langle c|[\nabla(\sin(K_j z)\,\delta U)\times\mathbf{p}]_x|n\rangle\langle n|p_y c\rangle}{2m_0^3c^2(\epsilon_{c0}-\epsilon_{n0})} \right. \\ &\quad \left. - \frac{\hbar\langle c|\nabla_z(\sin(K_j z)\,\delta U)|c\rangle}{4m_0^2c^2}\right)\int_{-L/2}^{L/2} P(z)\sin(K_j z)\,dz, \end{aligned} \tag{27}$$

in which the summation index j includes zero (the terms with $j=0$ represent the contribution of the smooth part of the potential). The term in the potential energy proportional to b_1 gives the main contribution, and the two remaining terms are corrections of order $\lambda\overline{k_z}$. For a symmetric structure, $P(z) = P(-z)$, the equation simplifies:

$$b_2 = b_3 = 0.$$

Strictly speaking, Eq. (26) is invalid from a mathematical point of view (it does not have a rigorous nontrivial solution); therefore, it needs to be put into a different form using the smallness of the term proportional to b_2 and discarding corrections of order $(\lambda\overline{k_z})^2$. It is necessary to invoke the approximate relation

$$b_1\delta(z) + b_2\delta'(z) \approx b_1\delta\left(z + \frac{b_2}{b_1}\right), \tag{28}$$

and in the term proportional to b_3, $\delta(z)$ can be replaced by $\delta(z + b_2/b_1)$ for simplicity. We obtain the valid equation

$$\left[\epsilon_{c0} - \epsilon + \frac{\mathbf{p}^2}{2m} + b_1\delta\left(z+\frac{b_2}{b_1}\right) + b_3\delta\left(z+\frac{b_2}{b_1}\right)[\mathbf{p}\times\mathbf{n}]\cdot\boldsymbol{\sigma}\right]F_c(\mathbf{r}) = 0. \tag{29}$$

Note that for states of one band (if we are not interested, for example, in interband transitions) in a heterostructure with one narrow layer the value of b_2 does not play a role: in Eq. (26) we can shift the origin $z' = z + b_2/b_1$.

We assume that we are dealing with a structure containing two narrow layers lying near one another (such that the distance between them is of the order of λ). In this case, an upper estimate on the error arising from the transformation to a one-band differential equation is $\lambda\overline{k_z}$ (this is valid, in particular, in the case when the constant b_1 describing the potential of the first layer is equal to the constant describing the potential of the second layer taken with opposite sign). Then we should remove from consideration those terms containing b_2 and b_3 in Eq. (26). We could probably treat both layers as one local perturbation, thereby decreasing the error, and obtain Eq. (26) with one set of parameters of the local perturbation. In the situation for which we obtained an upper estimate of the error, however, we could not then guarantee the smallness of the parameter $\overline{k_z}\tilde{b}_2/\tilde{b}_1$ on which Eq. (28) is based. If it is not small, then we could not say that Eq. (28) is mathematically correct, which would imply the inapplicability of such an approach.

Thus, taking the above modification into account, we can also apply the effective-mass method to electron states in heterostructures with superthin layers. In this regard the situation can arise in which the potential of a thick layer of some semiconductor plays the role of a barrier for the electron states while a thin layer of the same material will couple states and *vice versa*, depending on the sign of the parameter b_1. The sign of this parameter, as can be seen from Eqs. (27), can be different from the sign of the parameter $\langle c|\delta U|c\rangle$, which defines the discontinuity in the conduction band at the heterojunction.

5. HIERARCHY OF EFFECTIVE-MASS EQUATIONS AND DISCUSSION OF RESULTS

We have derived a many-band $\mathbf{k}\cdot\mathbf{p}$ system of integral equations (12) which can be used to describe electron states in heterostructures with atomically sharp variation of their chemical composition. The system contains contributions in the form of converging power series in $k_z - k_z'$ which are due to the sharpness of the heterojunction. For example, such terms were discarded in Refs. 14, 20, and 19 so that the effects of a discontinuous change in the crystal potential of the structure near a heterojunction were in fact neglected. It is specifically the presence of these terms that distinguishes the many-band system of $\mathbf{k}\cdot\mathbf{p}$ equations derived here from the system obtained by Leibler[3] for heterostructures with smooth heterojunctions. All of the papers known to us which use one-band or many-band schemes of the method of envelope functions to describe electron states in heterostructures apply the $\mathbf{k}\cdot\mathbf{p}$ system of Leibler. Very often the distinction between the Bloch functions for the component semiconductors of the heterostructure is also neglected, which gives, in particular, $\delta U_{nn'}=0$ for $n\neq n'$. Taking into account the terms due to sharpness of the heterojunction, the $\mathbf{k}\cdot\mathbf{p}$ method can also be used to describe intervalley mixing of states in heterostructures, including the problem of $\Gamma-X_z$ mixing of states in (001) heterostructures.

The most important limitation on the accuracy of the method of envelope functions employing differential equations is the procedure of transforming from **k** to **r** space. The one-band differential equation (of fourth order) with position-dependent effective mass is valid for structures with characteristic width of layers (much) greater than λ, where the length λ was defined in Sec. 2.

Above we considered in detail how taking account of contributions of the sharpness of the interface potential modifies the equation for states of the conduction band in (001) heterostructures of related, lattice-matched III–V semiconductors, derived in Ref. 28. Formally, the resulting equation for a sharp heterojunction differs from that for a smooth heterojunction only by renormalization of the parameters entering into it. In the case of the valence band, on the other hand, taking account of the sharpness of the heterojunction leads to qualitatively new effects (mixing of heavy and light holes for $\mathbf{k}_\parallel=0$).

For heterostructures with wide layers it is possible to construct a hierarchy of approximations of the one-band method of envelope functions according to the parameter $\lambda\overline{k_z}$, where $\overline{k_z}$ is the characteristic value of the quasimomentum of the state. For example, for an isolated heterojunction we have the following.

0) Zeroth level of the hierarchy for electrons.

In the effective-mass approximation in which small corrections in the order parameter have been neglected, we have the usual equation with position-independent effective mass and a discontinuous (step-function) potential:

$$\left(\epsilon_{c0}+\theta(z)\Delta U_c+\frac{\mathbf{p}^2}{2m}\right)F_c(\mathbf{r})=\epsilon F_c(\mathbf{r}). \tag{30}$$

1) First level of the hierarchy.

First-order corrections are taken into account (here, on the contrary, the small parameter $\overline{k_z}d$ plays a role; $2d$ here is the width of the transitional region of the heterojunction) by including a δ-function potential in Eq. (30), which is localized at the heteroboundary:

$$\left(\epsilon_{c0}+\theta(z)\Delta U_c+d_1\delta(z)+\frac{\mathbf{p}^2}{2m}\right)F_c(\mathbf{r})=\epsilon F_c(\mathbf{r}),$$

where d_1 is given by the complicated expression in Sec. 4.1.

2) Second level of the hierarchy.

Equation (22) includes all corrections of order $(\lambda\overline{k_z})^2$. Smaller contributions, of third order and higher, cannot be taken into account correctly in a one-band version of the method.

For hole states we obtained [see Eq. (24)] the first-order corrections to the standard equation (we obtained the first step of the hierarchy of effective-mass equations for the holes) and showed that for (001) heterostructures mixing of heavy and light holes at the center of the $2D$ Brillouin zone does indeed take place and that contributions from the sharpness of the heterojunction potential determine the strength of this mixing so that it depends on the microscopic structure of the heteroboundaries. In Ref. 20 it is asserted, in particular, that such mixing of heavy and light holes is caused by a difference in the Bloch functions for the component semiconductors of the heterostructure and it is absent, if one neglects such a difference, or, what should be equivalent, if the Bloch functions of all the bulk semiconductors comprising the structure are the same set of functions. If this is indeed the case, then the contribution of these terms (even without taking symmetry arguments into account) would be only of order $(\lambda\overline{k_z})^2$, as can be seen from Eq. (23). In fact, however, the difference in the Bloch functions does not play a substantial role. To prove this, consider the idealized situation of a (001) homojunction—the problem of hole states in a weak but not smooth external potential, say, $W(z)=G(z)W_0$, where W_0 is a constant assigning the jump of the potential, small in comparison with the band gap. In this case the point symmetry of the structure (C_{2v}) also admits the existence of mixing of heavy and light holes at the center of the $2D$ Brillouin zone,[33] and for the coefficient governing this mixing, $D_{0\mathscr{X}\mathscr{Y}}$, instead of formula (25) we have

$$D_{0\mathscr{X}\mathscr{Y}}=\sum_{j\neq0}\frac{W_0\langle\mathscr{X}|\sin(K_jz)|\mathscr{Y}\rangle}{K_j}\times\int_{-d}^{d}G'(z)\cos(K_jz)\,dz.$$

This is direct proof of our assertion.

The independent parameters α, d_1, and d_2 introduced in the present work and appearing in Eq. (22), and also χ_1, χ_2, and $D_{0\mathscr{X}\mathscr{Y}}$ entering into the effective potential energy operator for the valence-band states, depend not only on the bulk properties of the materials of the heterojunction, but also on its microscopic structure. All these parameters determine the hetero-interface contribution to the potential energy. At the same time, it is well known[34] that as a consequence of the possible appearance of an electric dipole moment at the interface of two materials the magnitude of the potential jump at the interface can also depend on the microscopic structure of the boundary. This is not described in our model of a heterojunction because we do not take into account the effect of such a dipole. Including the corresponding discontinuous electrostatic potential in Eq. (1) and developing it according to the scheme laid out above also yields the desired effect.

The electron states in heterostructures consisting of thin layers whose thickness is less than or on the order of λ can be treated only in the approximation quadratic in the momentum operator. In this case an account of terms due to the sharpness of the potential becomes necessary already in the zeroth approximation. This is clear from Eqs. (27). In this regard, the following situation is possible: as the width of the quantum well is decreased, the bound state can disappear or conversely, a thin layer—nominally a ''barrier'' layer—of some semiconductor can create an attractive potential and form a bound state.

It is possible that just such a situation was observed in Ref. 35 and then modeled in Ref. 36.

The authors are grateful to E. L. Ivchenko and A. A. Gorbatsevich for helpful discussions of a number of results of this work. This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grant No. 99-02-17592) and INTAS–RFBR (Grant No. 97-11475), and also the Russian Ministry of Science under the auspices of the programs ''Physics of Solid-State Nanostructures'' (Grant No. 99-1124) and ''Atomic Surface Structures'' (Grant No. 3.1.99).

APPENDIX A: THE DIRAC EQUATION WITH POSITION-VARIANT GAP

Let us consider a model Dirac equation with a position-dependent gap $2m(\mathbf{r})c^2$

$$\begin{bmatrix} m(\mathbf{r})c^2 & c\boldsymbol{\sigma}\cdot\mathbf{p} \\ c\boldsymbol{\sigma}\cdot\mathbf{p} & -m(\mathbf{r})c^2 \end{bmatrix}\begin{pmatrix}\varphi_e\\ \varphi_p\end{pmatrix}=\epsilon\begin{pmatrix}\varphi_e\\ \varphi_p\end{pmatrix},$$

where φ_e and φ_p are the electron and positron components of the wave function, respectively. Let $m(\mathbf{r})$ vary weakly in space, i.e., $m(\mathbf{r})=\tilde{m}+\delta m(\mathbf{r})$ so that $\delta m(\mathbf{r})/\tilde{m}\ll 1$. With the help of a Foldy–Wouthuysen unitary transformation it is quite simple, for example, following the scheme laid out in Ch. 20, Sec. 33 or Ref. 24, to obtain a one-band equation describing the states of the electron. Thus, the equation in which all small terms have been discarded is the ordinary Schrödinger equation

$$\left(m(\mathbf{r})c^2+\frac{\mathbf{p}^2}{2\tilde{m}}\right)\tilde{\varphi}_e=\epsilon\tilde{\varphi}_2,$$

where $\tilde{\varphi}_e$ is the transformed electron wave function. The equation, on the other hand, in which all terms of higher order than $\delta m(\mathbf{r})/\tilde{m}$ have been neglected (the following approximation) has the form

$$H\tilde{\varphi}_e=\epsilon\tilde{\varphi}_e,$$

$$H=m(\mathbf{r})c^2+\frac{\mathbf{p}^2}{2\tilde{m}}-\frac{\mathbf{p}\delta m(\mathbf{r})\cdot\mathbf{p}}{2\tilde{m}^2}+\frac{\hbar^2\nabla^2\delta m(\mathbf{r})}{8\tilde{m}^2}$$
$$-\frac{\mathbf{p}^4}{8\tilde{m}^3c^2}-\frac{\hbar[\nabla\delta m(\mathbf{r})\times\mathbf{p}]\cdot\boldsymbol{\sigma}}{4\tilde{m}^2}. \tag{A1}$$

All terms in Eq. (A1) with the exception of the third which describes the position-dependent mass can be taken as ''ordinary.'' The second, third, and fourth terms (the fourth is the Darwin term) can either be written in the following form:

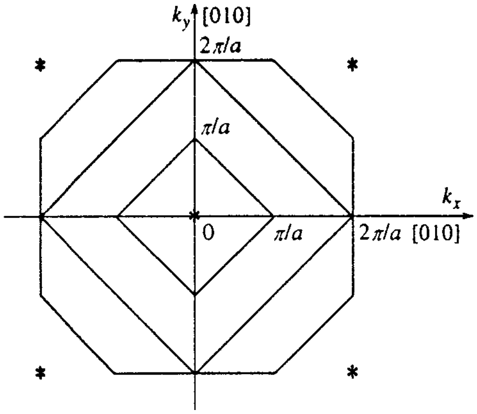

FIG. 1. Projection of the bulk Brillouin zone (the region bounded by the octagon) and sites of the inverse lattice (denoted by asterisks) onto the (001) plane. The square with diagonal $2\pi/a$ bounds the region where there are no $2D$ transfer processes.

$$T_2=\frac{1}{2}\mathbf{p}\cdot\frac{1}{m(\mathbf{r})}\mathbf{p}+\frac{\hbar^2\nabla^2\delta m(\mathbf{r})}{8\tilde{m}^2},$$

or combined into one (quadratic in the momentum) kinetic-energy operator:

$$T_2=\frac{1}{2}\frac{1}{\sqrt[4]{m(\mathbf{r})}}\mathbf{p}\frac{1}{\sqrt{m(\mathbf{r})}}\cdot\mathbf{p}\frac{1}{\sqrt[4]{m(\mathbf{r})}},$$

or some other equivalent form can be used. For example, Ref. 21 uses the following form for T_2:

$$T_2=\frac{1}{4}\left[\mathbf{p}\frac{1}{\sqrt{m(\mathbf{r})}}\cdot\mathbf{p}\frac{1}{\sqrt{m(\mathbf{r})}}+\frac{1}{\sqrt{m(\mathbf{r})}}\mathbf{p}\cdot\frac{1}{\sqrt{m(\mathbf{r})}}\mathbf{p}\right].$$

Thus, in the model Dirac equation with position-dependent gap the concept of a position-dependent effective mass shows up only within the context of the nonrelativistic (quadratic) approximation.

APPENDIX B: REGARDING TRANSFER PROCESSES IN THE TWO-DIMENSIONAL BRILLOUIN ZONE FOR A (001) HETEROSTRUCTURE

Let us consider the second sum in Eq. (4), describing transfer processes in the two-dimensional Brillouin zone and prove that it does not contribute in the case of interest to us of states near the Brillouin zone center in (001) heterostructures. Since the function $\mathscr{G}(q)$ isonzero for any q, there also exist nonzero vectors of the inverse lattice $\mathbf{K}_j$ for which $\mathbf{k}_\parallel'-\mathbf{k}_\parallel=\mathbf{K}_{\parallel j}$. Here $\mathbf{k}_\parallel'$ and $\mathbf{k}_\parallel$ are components of vectors belonging to the Brillouin zone. There exits a finite number of vectors $\mathbf{K}_j$ possessing this property. Therefore, in general we should also retain the second sum in expression (4). Let us now consider the interesting case of a (001) heterostructure. The octagon in Fig. 1 represents the projection of the bulk Brillouin zone onto the (001) plane, and projections of sites of the lattice are denoted by asterisks. The function $\mathscr{F}_n'(k_z',\mathbf{k}_\parallel')$ on which the operator $\mathscr{M}_{nn'}(\mathbf{k},\mathbf{k}')$ acts in Eq. (3) is defined only for $\mathbf{k}_\parallel'$ belonging to the projection of the bulk Brillouin zone onto the (001) plane. But since all sites $\mathbf{K}_{\parallel j}$ for $j\neq 0$ lie outside this projection, there exists a region of $\mathbf{k}_\parallel$ for which $|\mathbf{k}'_\parallel-\mathbf{k}_\parallel|<\mathbf{K}_{\parallel\mathbf{j}}$, $j\neq 0$. This region is defined by the inequality $|k_x|+|k_y|<\pi/a$ and is indicated in the figure by the square with diagonal $2\pi/a$. The area of this region is one-fourth that of the area of the first $2D$ Brillouin zone (the square with diagonal $4\pi/a$). In the region $|k_x|+|k_y|<\pi/a$ the second sum in expression (4) does not contribute to the equations for the envelope functions. Its larger dimensions ensure satisfaction of the conditions of applicability of the derived equations for the envelope functions describing states near the Γ point.

*⁾E-mail: gibbon@royal.net
†⁾E-mail: VoVA@mail.cplire.ru

---

[1] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[2] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).
[3] L. Leibler, Phys. Rev. B **12**, 4443 (1975).
[4] G. Bastard, Phys. Rev. B **24**, 5693 (1981).
[5] Q. G. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).
[6] R. A. Morrow and K. R. Brownstein, Phys. Rev. B **30**, 678 (1984).
[7] D. L. Smith and C. Mailhiot, Phys. Rev. B **33**, 8345 (1986).
[8] W. Trzeciakowski, Phys. Rev. B **38**, 12493 (1988).
[9] T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11609 (1989).
[10] K. Young, Phys. Rev. B **39**, 13434 (1989).
[11] B. Laikhtman, Phys. Rev. B **46**, 4769 (1992).
[12] J. P. Cuypers and W. van Haeringen, J. Phys.: Condens. Matter **4**, 2587 (1992).
[13] B. A. Foreman, Phys. Rev. B **48**, 4964 (1993).
[14] M. G. Burt, J. Phys.: Condens. Matter **4**, 6651 (1994).
[15] G. T. Einevoll and L. J. Sham, Phys. Rev. B **49**, 10533 (1994).
[16] T. Yamanaka, H. Kamada, Y. Yoshikuni, W. W. Lui, S. Seki, and K. Yokoyama, J. Appl. Phys. **76**, 2347 (1994).
[17] W. Trzeciakowski, Semicond. Sci. Technol. **10**, 768 (1995).
[18] A. Brezini, M. Sebbani, and S. Marouf, Phys. Status Solidi B **189**, 389 (1995).
[19] B. A. Foreman, Phys. Rev. B **52**, 12241 (1995).
[20] B. A. Foreman, Phys. Rev. B **54**, 1909 (1996).
[21] F. S. A. Cavalcante, R. N. Costa Filho, J. Ribeiro Filho, C. A. S. de Almeida, and V. N. Freire, Phys. Rev. B **55**, 1326 (1997).
[22] M. V. Kisin, B. L. Gelmont, and S. Luryi, Phys. Rev. B **58**, 4605 (1998).
[23] V. Keldysh, Zh. Éksp. Teor. Fiz. **45**, 364 (1963) [Sov. Phys. JETP **18**, 253 (1964)].
[24] A. Messiah, *Quantum Mechanics* (Interscience, New York, 1961, 1962), Vol. 2.
[25] V. B. Berestetskiĭ, E. M. Lifshitz, L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. (Pergamon Press, Oxford, 1982).
[26] R. A. Suris, Fiz. Tekh. Poluprovodn. **20**, 2008 (1986) [Sov. Phys. Semicond. **20**, 1258 (1986)].
[27] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1975).
[28] E. E. Takhtamirov and V. A. Volkov, Semicond. Sci. Technol. **12**, 77 (1997).
[29] V. A. Volkov, É. E. Takhtamirov, Usp. Fiz. Nauk **167**, 1123 (1997).
[30] E. E. Takhtamirov and V. A. Volkov, in *Proceedings of the International Symposium ''Nanostructures: Physics and Technology,''* Ioffe Institute, St. Petersburg, 1997, p. 134.
[31] E. E. Takhtamirov and V. A. Volkov, in *Proceedings of the International Symposium ''Nanostructures: Physics and Technology,''* Ioffe Institute, St. Petersburg, 1998, p. 332.
[32] F. T. Vas'ko, JETP Lett. **30**, 541 (1979).
[33] E. L. Ivchenko, A. Yu. Kaminski, and U. M. Rössler, Phys. Rev. B **54**, 5852 (1996).
[34] R. G. Dandrea, C. B. Duke, and A. Zunger, J. Vac. Sci. Technol. B **10**, 1744 (1992).
[35] R. Schwabe, F. Pietag, W. Faulkner, and S. Lassen, J. Appl. Phys. **77**, 6295 (1995).
[36] M. Di Ventra and A. Baldereschi, in *Proceedings of the 23rd International Conference on the Physics of Semiconductors*, Berlin, Germany, 1996 (World Scientific, Singapore, 1996), p. 1719.

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Complex dynamics and chaos in the parametric coupling of counter-propagating waves

T. V. Dmitrieva and N. M. Ryskin*)

Higher College of Applied Sciences, Saratov State University, 410036 Saratov, Russia

(Submitted 5 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1871–1881 (November 1999)

We study the dynamics of a distributed self-oscillating system of three parametrically coupled waves, one of which is propagating counter to the other two. We show that an infinite number of natural modes are self-excited as the bifurcation parameter, which has the meaning of the pump amplitude, increases without bound. Exact solutions describing steady-state oscillation regimes are found. We present the results of computer simulation, which show that for moderate pump amplitudes the transient process terminates when a stationary state corresponding to the fundamental mode sets in. As supercriticality increases, the oscillations become chaotic, with the transition to chaos being rapid. We note an analogy that exists between the dynamics of such a system and the dynamics of a Lorentz system. © *1999 American Institute of Physics.* [S1063-7761(99)02611-6]

1. INTRODUCTION

Investigations into the complex dynamics of distributed self-oscillating systems are an important part of the theory of oscillations and waves. The relationship between such investigations and the onset of turbulence in the formation of dissipative structures is quite obvious.[1,2] However, direct computer simulation of complex and, especially, chaotic oscillation regimes in specific distributed self-oscillating systems, such as devices of vacuum and quantum electronics and hydrodynamic flow, pose many difficulties for the researcher. Hence the search of fairly simple distributed model systems with complex dynamics becomes crucial, since such models can be studied in detail by numerical and, possibly, analytical methods.

Among such models we should like to mention the equations of parametric coupling of three wave packets propagating in a quadratically nonlinear medium, equations that describe parametric (decay) instability in hydrodynamic problems, nonlinear optics, plasma physics, etc., andare among the most important standard equations of nonlinear wave theory.[2,3] If some kind of feedback is present, the system becomes self-oscillatory and exhibits complex dynamical behavior. For instance, Burlak and Ishkabulov[4] detected dynamical chaos in the case where one of the waves is reflected at the boundaries.

In the present paper we study the problem of three-wave coupling in which two wave propagate in opposite directions, which guarantees the existence of internal distributed feedback. We believe that this situation is common and can be realized, in particular, when low- and high-frequency waves interact,[2] e.g., when optical waves interact with acoustic waves or ion-acoustic waves interact with Langmuir waves. In electronics, a system of three parametrically coupled waves one of which is counter-propagating can be taken as the simplest model of a free-electron laser, a generator of a counter-propagating wave, in which the electron beam interacts with the fields of two electromagnetic waves. It was Ginzburg and Sergeev[5] who pointed out that the dynamics can be complex in this case.

In Sec. 2 we discuss the main equations and find the conditions for self-excitation of self-oscillations. Since the system being studied is distributed, it is characterized by an infinite number of natural modes, which are self-excited as the bifurcation parameter increases (here the amplitude of the pump wave acts as the bifurcation parameter). The exact time-dependent nonlinear solutions corresponding to the fundamental and higher modes are analyzed in Sec. 3. Section 4 is devoted to describing the results of computer simulation. We study in detail a scenario for the transition to chaos. We also discuss the analogy between sequences of bifurcations in the present system and those in some finite-dimensional systems, e.g., in the Lorentz system, which is one of the best known and most thoroughly studied models of nonlinear dynamics with a small number of degrees of freedom.

2. THE STARTING EQUATIONS: CONDITIONS FOR SELF-EXCITATION OF SELF-OSCILLATIONS

For the staring equations we take the equation of three-wave parametric interaction,[2,3] which in our case assume the form

$$\begin{aligned}
&\frac{\partial A_1}{\partial t}+v_1\frac{\partial A_1}{\partial x}=\sigma_1 A_2 A_3^*,\\
&\frac{\partial A_2}{\partial t}+v_2\frac{\partial A_2}{\partial x}=-\sigma_2 A_1 A_3, \qquad (1)\\
&\frac{\partial A_3}{\partial t}-v_3\frac{\partial A_3}{\partial x}=\sigma_3 A_2 A_1^*.
\end{aligned}$$

Here $A_{1,2,3}(x,t)$ are the complex-valued amplitude of the idler wave, the pump wave, and the signal wave, respectively, $v_{1,2,3}$ are the group velocities of these waves, and $\sigma_{1,2,3}$ are the real nonlinear-coupling constants. The ''minus'' in front of v_3 is an indication that the signal wave is propagating counter to the idler and pump waves. We examine the situation where an intense pump wave with an amplitude $\alpha=\text{const}$ is input into the system at the point $x=0$ and there is no idler wave. We assume that at the right boundary of the system, in the cross section $x=l$, there is no signal wave. Thus, the boundary conditions are

$$A_1(x=0;t)=0,\quad A_2(x=0;t)=\alpha,\quad A_3(x=l;t)=0, \tag{2}$$

which corresponds to a distributed parametric generator of a counter wave. Since Eqs. (1) are invariant under the transformations

$$A_1\to A_1\exp(i\varphi_1),\quad A_2\to A_2\exp(i\varphi_2),$$

$$A_3\to A_3\exp[i(\varphi_2-\varphi_1)], \tag{3}$$

where $\varphi_{1,2}=\text{const}$, we can assume, without loss of generality, that α is real. We select the initial conditions in the form

$$A_2(x;t=0)=\alpha,\quad |A_{1,3}(x;t=0)|\ll\alpha \tag{4}$$

and are interested in the conditions for self-excitation of the system by small fluctuations of the idler and signal waves.

We introduce the dimensionless coordinate $\xi=x/l$ and time

$$\tau=\frac{2v_1v_3t-(v_3-v_1)x}{l(v_3+v_1)}$$

and the new dependent variables

$$A_i'=\frac{A_il}{\alpha}\sqrt{\frac{\sigma_j\sigma_k}{v_jv_k}},$$

where $i,j,k=1,2,3$, with $i\neq j\neq k$. Substituting all this in Eqs. (1) and dropping the primes, we get

$$\frac{\partial A_1}{\partial\tau}+\frac{\partial A_1}{\partial\xi}=\alpha A_2A_3^*,$$

$$(1+u)\frac{\partial A_2}{\partial\tau}+\frac{\partial A_2}{\partial\xi}=-\alpha A_1A_3, \tag{5}$$

$$\frac{\partial A_3}{\partial\tau}-\frac{\partial A_3}{\partial\xi}=\alpha A_2A_1^*,$$

where

$$u=\frac{2v_3(v_1-v_2)}{v_2(v_1+v_3)}$$

is the dimensionless parameter characterizing the detuning of the group velocities of the co-propagating waves. The boundary conditions (2) and the initial conditions (4) become

$$A_1(\xi=0;\tau)=0,\quad A_2(\xi=0;\tau)=1,\quad A_3(\xi=1;\tau)=0, \tag{6}$$

$$A_2(\xi;\tau=0)=1,\quad |A_{1,3}(\xi;\tau=0)|\ll 1. \tag{7}$$

In order to derive the conditions for self-excitation of self-oscillations we start with the initial stage in the development of an instability, when the amplitudes of the signal and idler waves are small and depletion of the pump wave still has no effect. Putting $A_2=1$ in Eqs. (5), we obtain

$$\frac{\partial A_1}{\partial\tau}+\frac{\partial A_1}{\partial\xi}=\alpha A_3^*,$$

$$\frac{\partial A_3}{\partial\tau}-\frac{\partial A_3}{\partial\xi}=\alpha A_1^*. \tag{8}$$

Equations (8) with the boundary conditions

$$A_1(\xi=0;\tau)=A_3(\xi=1;\tau)=0$$

coincide with the equations that describe the time-dependent linear coupling of counter-propagating waves with positive and negative energies. The theory of such coupling has been thoroughly described in the literature (see, e.g., the review by Trubetskov and Chetverikov[6]). Separating the variables in (8), we find the general solution as a superposition of natural modes:

$$A_1=\sum_n C_n\exp(k_nt)\sin[\sqrt{\alpha^2-k_n^2}\xi],$$

$$A_3=\sum_n D_n\exp(k_nt)\sin[\sqrt{\alpha^2-k_n^2}(\xi-1)],$$

where the amplitudes C_n and D_n can be found from the initial conditions, and the k_n are the solutions of the transcendental equation

$$\sqrt{\alpha^2-k^2}\cos\sqrt{\alpha^2-k^2}+k\sin\sqrt{\alpha^2-k^2}=0. \tag{9}$$

This equation has an infinite number of complex-valued solutions, which is a reflection of the fact that the distributed system contains an infinite number of natural modes, i.e., degrees of freedom.

The condition for the loss of stability of a node with number n can be found by putting $k=0$ in Eq. (9), which yields $\alpha_n=\pi n+\pi/2$. Thus, when the parameter α exceeds the value α_n, self-excitation of the nth mode is possible. The mode with $n=0$ has the smallest value $\alpha_0=\pi/2$, i.e., the smallest pumping amplitude.

3. STEADY-STATE OSCILLATIONS REGIMES

Let us now find the solutions that describe steady-state self-oscillating regimes. To this end we put $\partial/\partial\tau\equiv 0$ in Eqs. (5). This yields a system of ordinary differential equations:

$$\frac{dA_1}{d\xi}=\alpha A_2A_3^*,\quad\frac{dA_2}{d\xi}=-\alpha A_1A_3,\quad\frac{dA_3}{d\xi}=-\alpha A_2A_1^* \tag{10}$$

with the boundary conditions

$$A_1(0)=0,\quad A_2(0)=1,\quad A_3(1)=0. \tag{11}$$

Note that these equations do not contain the parameter u, i.e., stationary solutions do not depend on the detunings of the group velocities (in the normalization of the variables adopted here).

Assuming that $A_j=a_j\exp(i\varphi_j)$, where a_j and φ_j are the real-valued amplitude and phase, and separating the real and imaginary parts in (10), we arrive at the system of equations

$$\frac{da_1}{d\xi}=\alpha a_2a_3\cos\Phi, \tag{12}$$

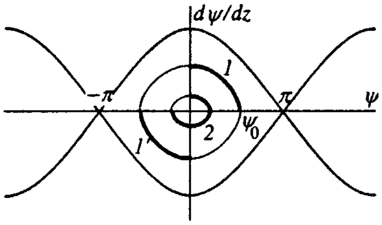

FIG. 1. Phase portrait of the nonlinear oscillator (16). Paths *1* and *1′* correspond to stationary solutions for the fundamental mode, and the path *2*, for the second mode.

$$\frac{da_2}{d\xi}=-\alpha a_1 a_3\cos\Phi, \tag{13}$$

$$\frac{da_3}{d\xi}=-\alpha a_2 a_1\cos\Phi, \tag{14}$$

$$\frac{d\Phi}{d\xi}=\alpha\left(\frac{a_1a_2}{a_3}+\frac{a_1a_3}{a_2}-\frac{a_2a_3}{a_1}\right)\sin\Phi, \tag{15}$$

where $\Phi=\varphi_2-\varphi_1-\varphi_3$. The exact solutions of the system of equations (12)–(15) are well known (see, e.g., Ref. 2). The unique property of the problem is that the boundary conditions (11) are set at the different ends of the system. We write Eq. (15) as follows:

$$\frac{d\Phi}{d\xi}=-\frac{d\ln(a_1a_2a_3)}{d\xi}\tan\Phi.$$

Integration of this equation with allowance for the boundary conditions yields $a_1a_2a_3\sin\Phi=0$. Since this equality is valid for all values of ξ, nontrivial solutions correspond to $\sin\Phi=0$. Thus, in a steady-state regime the phases of the interacting waves are found to be synchronized.

From Eqs. (12) and (13) we can obtain one more constant of motion:

$$a_1^2(\xi)+a_2^2(\xi)=a_1^2(0)+a_2^2(0)=1.$$

Assuming that $a_1=\sin(\psi/2)$ and $a_2=\cos(\psi/2)$ and substituting these quantities in (14), we arrive at the pendulum equation

$$\frac{d^2\psi}{dz^2}+\sin\psi=0, \tag{16}$$

where $z=\alpha\xi$, with the boundary conditions

$$\psi(z=0)=d\psi(z=\alpha)/dz=0.$$

The phase portrait of Eq. (16) is depicted in Fig. 1. The solutions we are interested in are portions of the phase paths that begin at the vertical axis and end at the horizontal axis. Thus, we must find the solution in the form of periodic oscillations with a period $T=4\alpha$ (the path *1* in Fig. 1). Allowing for the expression for the period of pendulum oscillations (see, e.g., Ref. 7), we find that $\alpha=K[\sin^2(\psi_0/2)]$, where K is the complete elliptic integral of the first kind, with $\psi_0=\psi(\alpha)$. Since $K>\pi/2$, the solution exists for $\alpha>\pi/2$, which agrees with conditions for self-excitation established in Sec. 2.

When $\alpha>3\pi/2$, there appears an additional solution, according to which the pendulum performs three-fourths of a complete revolution (path *2* in Fig. 1). The solution corresponds to the second natural mode. As α increases, new stationary solutions corresponding to higher modes appear. The conditions for self-excitation of such modes were found in Sec. 2. These solutions can be written explicitly:

$$\psi=2\arcsin(m\,\mathrm{sn}(z;m)),\quad a_1=m\,\mathrm{sn}(z;m),$$
$$a_2=\mathrm{dn}(z;m),\quad a_3=m\,\mathrm{cn}(z;m), \tag{17}$$

where $\mathrm{sn}(\cdots)$, $\mathrm{cn}(\cdots)$, and $\mathrm{dn}(\cdots)$ are Jacobi elliptic functions, and $m=\sin^2(\psi_n/2)$, $n=0,1,2,\ldots$, with ψ_n determined by the equation $\alpha=(2n+1)K[\sin^2(\psi_n/2)]$. Obviously, in view of the invariance of Eq. (6) under the transformation $\psi\to-\psi$, to each solution in (17) there corresponds a solution of opposite polarity. The phase paths in Fig. 1 that originate in the lower half-plane (e.g., the path *1′*) correspond to these new solutions.

4. RESULTS OF COMPUTER SIMULATION AND DISCUSSION

The analysis done in Sec. 3 makes it possible to find the exact solutions describing steady-state oscillation modes. However, it is still unclear which of these regimes actually sets in as a result of the transient process (and whether such regimes set in at all). To answer these questions we are forced to turn to direct numerical integration of the time-dependent equations.

So solve Eqs. (5) with the boundary and initial conditions (6) and (7), we use the second-order Lax–Wendorff method,[8] with the time and space step sizes related by the formula $\Delta\tau=\Delta\xi/2$. Calculations show that the selection of $N\sim 100$, where $N=1/\Delta\xi$ is the number of steps in the difference scheme in coordinate, ensures good accuracy of the results (except for the range of parameters for which $u\ll 1$; for more details see the text below). Note that any finite-difference approximation actually amounts to passing from the distributed system to a chain of N elements, which has a large but finite number of degrees of freedom.

For the sake of simplicity we limit ourselves to the case where the amplitudes of the interacting waves are purely real. Here we have $\sin\Phi=0$, i.e., the phase path belongs to a manifold on which stationary solutions are realized (see Sec. 3). It can be shown that if this condition was met initially, it will remain valid at any subsequent moment in time, i.e., the manifold $\sin\Phi=0$ is invariant.

The system in question has two simplifying parameters: the parameter α, which measures the extent to which the system departs from equilibrium, and the parameter u, which is the detuning of the group velocities. First we turn to the analysis of one-parameter dynamics. We describe the sequence of bifurcations observed as α increases with a constant value $u=0.1$.

For $\alpha<\pi/2$ the only stable state of the system is given by the trivial solution $A_{1,3}=0$ and $A_2=1$. In phase space this corresponds to an equilibrium state in the form of a stable

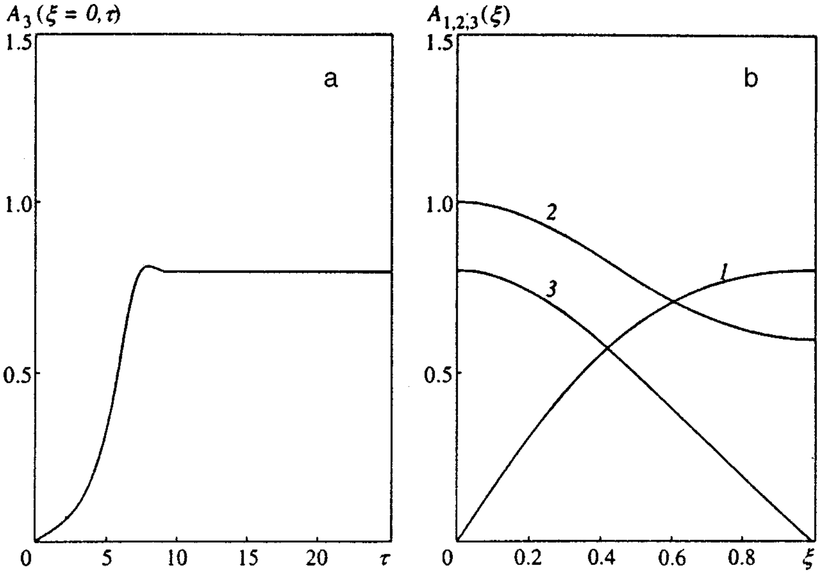

FIG. 2. Time dependence of the output-signal amplitude $A_3(\xi=0)$ in the process of the establishment of oscillations (a), and the steady-state distributions of the wave amplitudes (b) at $u=0.1$ and $\alpha=2.0$.

node, which we denote by O. Self-excitation of oscillations occurs at $\alpha=\pi/2$, which agrees with the results of Sec. 2. Here the equilibrium state O becomes a saddle point and there appear two nontrivial equilibrium states $C^{\pm}$ that correspond to the stationary solutions for the fundamental mode found in Sec. 3. These equilibrium states are symmetric under the transformations $A_1 \rightarrow -A_1$ and $A_3 \rightarrow -A_3$. Figure 2 depicts the time dependence of the output-signal amplitude $A_3(\xi=0)$ in the establishment of oscillations (a) spatial distributions of the amplitudes in the steady-state regime (b), when α is only slightly larger than $\pi/2$. Obviously, $C^{\pm}$ are equilibrium states of the stable-node type. The saddle-point equilibrium state O has a stable manifold W_s and a pair of symmetric unstable manifolds $W_u^{\pm}$, with W_u^{+} attracted to C^{+} and W_u^{-} to C^{-}. Thus, the initial perturbations determine which of the two stable states is realized after the transient process terminates.

As the control parameter α increases, the transient process acquires an oscillatory character and its duration increases (Fig. 3). The oscillations are due to the retarding nature of the internal distributed feedback, their period is close to the time lag $l/v_1 + l/v_3$, which in the normalization of variables adopted here is equal to two. In the process, the equilibrium states $C^{\pm}$ change from nodes to foci. For large enough values of α the stationary distributions of the amplitudes are close to the solutions that correspond to the motion of the pendulum along the separatrix (see Sec. 3): $A_1 = \tanh \alpha\xi$ and $A_{2,3} = \operatorname{sech} \alpha\xi$. At $\alpha \approx 5.6$ the manifolds $W_u^{\pm}$ close on W_s. As a result of this bifurcation, W_u^{+} begins to be attracted by C^{-} and W_u^{-} by C^{+}. The curves corresponding to this case are depicted in Fig. 4. As the bifurcation parameter increases, the transient process becomes more complicated, which is due to the distributed nature of the system, i.e., the occurrence of an ever-increasing number of natural modes. However, the stationary state corresponding to the higher modes are found to be unstable, and as a result of competition only the fundamental mode survives.

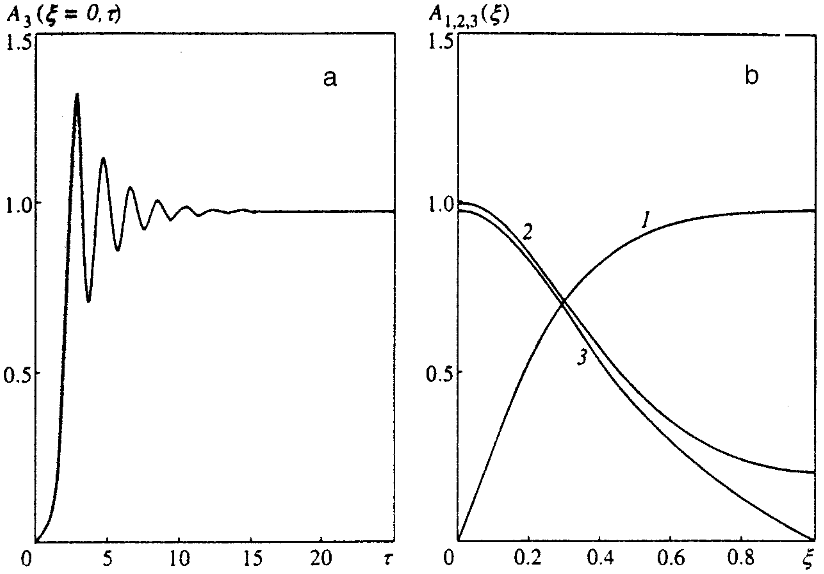

FIG. 3. Time dependence of the output-signal amplitude $A_3(\xi=0)$ in the process of the establishment of oscillations (a), and the steady-state distributions of the wave amplitudes (b) at $u=0.1$ and $\alpha=4.0$.

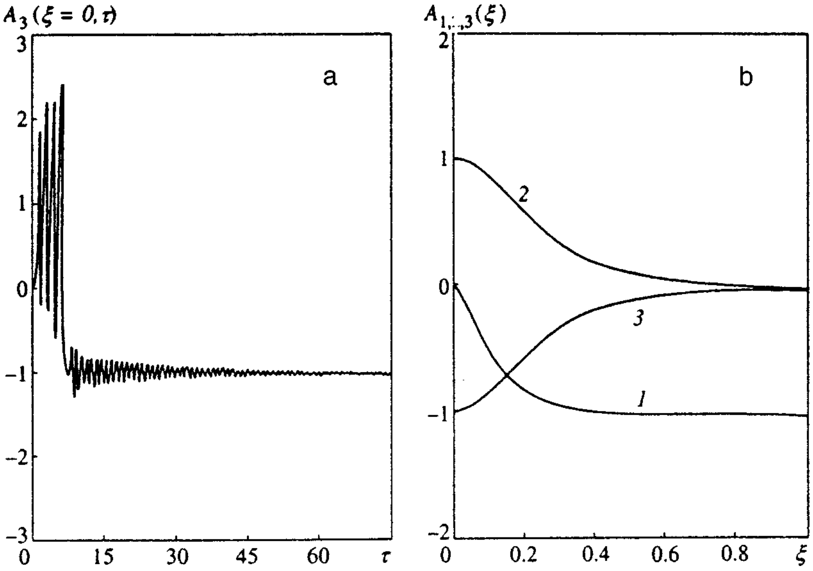

FIG. 4. Time dependence of the output-signal amplitude $A_3(\xi=0)$ in the process of the establishment of oscillations (a), and the steady-state distributions of the wave amplitudes (b) at $u=0.1$ and $\alpha=6.0$, after which bifurcation of the formation of a separatrix loop occurred.

Finally, at $\alpha \approx 8.19$, the motion becomes chaotic. Figure 5 depicts the temporal realization of the process and the projection of the strange attractor restored by the Packard–Takens method. Note that the attractor and the sequence of bifurcation preceding the appearance of the attractor demonstrate an analogy with some well-known finite-dimensional systems, in particular, with the Lorentz system[1,2]

$$\dot{X}=\sigma(Y-X), \quad \dot{Y}=rX-Y-XZ, \quad \dot{Z}=-bZ+XY, \tag{18}$$

where r, σ, b are parameters, and with equations describing the decay limitation of parametric instability:[9]

$$\begin{aligned} \dot{A}_1 &= hA_2^* - \nu_1 A_1 - A_2 A_3, \\ \dot{A}_2 &= hA_1^* - \nu_2 A_2 + A_1 A_3^*, \\ \dot{A}_3 &= -A_3 + A_1 A_2^*, \end{aligned} \tag{19}$$

which differ from (18) only in the fact that the second equation has an additional nonlinearity and that the wave amplitudes are, in general, complex-valued. In Eqs. (19) the waves $A_{1,2}$ are excited due to a parametric process, with the parameter h proportional to the amplitude of the pump wave, which is assumed fixed. The coefficients $\nu_{1,2}$ determine the linear decay of the waves. At the same time, the wave participate in another resonant interaction, in the process of which a quantum of wave A_1 decays into quanta of the waves $A_{2,3}$. This

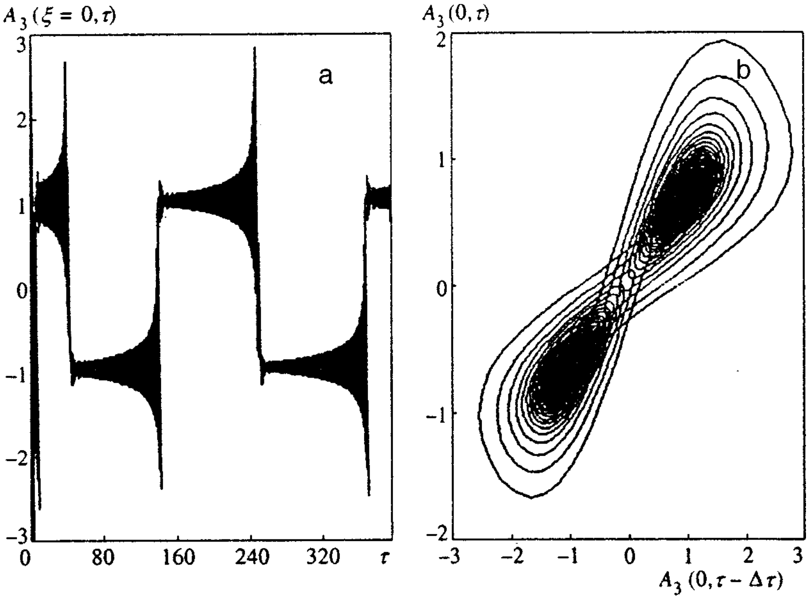

FIG. 5. Realization and projection of the phase portrait corresponding to the strange attractor ay $u=0.1$ and $\alpha=10.0$.

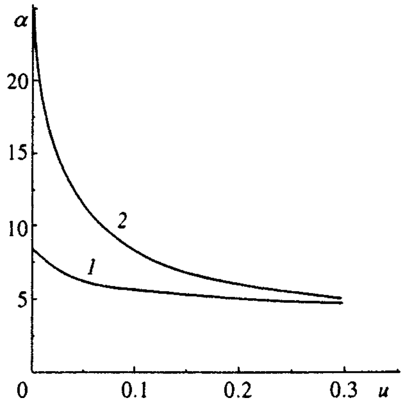

FIG. 6. Line of formation of a separatrix loop (curve *1*) and the boundary of transition to chaos (curve *2*) in the α,u plane.

decay process leads to saturation of the parametric instability. The application of the system of equations (19) to a number of problems in plasma physics has been discussed by Pikovskiĭ *et al.*[9]

As in the finite-dimensional systems mentioned earlier, a strange attractor appears before the equilibrium states $C^{\pm}$ lose their stability. Thus, within a narrow range of values of the parameter α, stable stationary states and chaotic motion coexist, and when the value of α increases smoothly the transition to chaos is accompanied by hysteresis. This can easily be verified by fixing the initial conditions in a form close to a stationary solution. For instance, with $u=0.1$ a strange attractor is born at $\alpha\approx 8.19$, and $C^{\pm}$ lose their stability at $\alpha\approx 8.27$. For the Lorentz system, in the most thoroughly studied case of $\sigma=10$ and $b=8/3$, a strange attractor is born at $r\approx 24.06$, and the stationary state loses its stability at $r\approx 24.74$ (the closure of the separatrix loops occurs at $r\approx 13.92$).

For other positive values of *u*, the dynamics of the system is similar, although the bifurcation values of the parameters change. Figure 6 depicts the boundary of the transition to chaos and the line of formation of a separatrix loop in the plane of parameters α and *u*. However, as $u\to 0$, the value of α at which there is a transition to chaos tends to infinity. A similar situation takes place in the Lorentz system when σ tends to $b+1$. Thus, chaotic oscillations are possible only if the group velocity of the pump wave is smaller than that of the idler wave. Note that in the region where $u\ll 1$ holds one is forced to increase the number of steps in the difference scheme substantially to obtain satisfactory accuracy of the calculations.

Studies of the dynamics of the system in the general case where the A_i are assumed to be complex-valued have shown that for values of α that are not very large the phase path is attracted, in the transient process, to the manifold $\sin\Phi=0$, i.e., the phases of the coupled waves become synchronized. However, in chaotic regimes, where the wave amplitudes undergo strong oscillations, states with rapidly varying phases arise and the synchronization of the phases is violated. Here the phase path jumps off the manifold $\sin\Phi=0$ and then jumps on. Note that the system (19) behaves in a similar manner, as noted by Pikovskiĭ.[10]

5. CONCLUSION

Our results prove without any doubt that in a system of three parametrically coupled waves one of which propagates counter to the other two, chaotic self-oscillations of a deterministic nature can develop. Of special interest here is the fact that although the system under investigation is of distributed type and is characterized by an infinite number of degrees of freedom, its dynamics is close to that of some finite-dimensional systems, in particular, the Lorentz system, which is a standard model of nonlinear dynamics. The reason is that three unstable stationary states play the main role in the organization of complex dynamics: the zero state *O* and the nonzero states $C^{\pm}$, which correspond to the lowest natural mode and are symmetric to each other. The higher natural modes affect only the initial stage in the transient process, and the corresponding stationary states are not realized.

Note that the rapidity of the transition to chaos observed in the given case is not typical of distributed self-oscillating systems with counter-propagating waves,[6] where a different scenario is more common, namely, a gradual loss of stability by the steady-state regime through self-modulation due delayed internal feedback.

The work was sponsored by the Program of Basic Research at Russian Universities.

*⁾E-mail: ryskin.np.sgu@oda.ssu.runnet.ru

---

[1] Yu. N. Neĭmark and P. S. Landa, *Stochastic and Chaotic Oscillations*, Kluwer Academic, Dordrecht (1992).

[2] M. I. Rabinovich and D. I. Trubetskov, *Oscillations and Waves in Linear and Nonlinear Systems*, Kluwer Academic, Dordrecht (1989).

[3] M. B. Vinogradova, O. V. Rudenko, A. P. Sukhorukov, *Theory of Waves* [in Russian], Nauka, Moscow (1990).

[4] G. N. Burlak and K. Ishkabulov, Zh. Éksp. Teor. Fiz. **109**, 774 (1996) [JETP **82**, 416 (1996)].

[5] N. S. Ginzburg and A. S. Sergeev, Radiotekh. Elektron. (Moscow) **33**, 580 (1988).

[6] D. I. Trubetskov and A. P. Chetverikov, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelin. Mekh. **2**(5), 9 (1994).

[7] R. Z. Sagdeev, D. A. Usilov, and G. M. Zaslavskiĭ, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos*, Harwood Academic, New York (1988).

[8] D. Anderson, J. Tannehill, and R. Pletcher, *Computational Fluid Mechanics and Heat Transfer*, Hemisphere, New York (1984).

[9] A. S. Pikovskiĭ, M. I. Rabinovich, and V. Yu. Trakhtengerts, Zh. Éksp. Teor. Fiz. **74**, 1366 (1978) [Sov. Phys. JETP **47**, 715 (1978)].

[10] A. S. Pikovskiĭ, Cand. Sci. (Phys.–Math.) thesis, Institute of Applied Physics, USSR Academy of Sciences, Gorkiĭ (1982).

Translated by Eugene Yankovsky

ERRATA

Erratum: Nonmonotonic behavior of the superconducting transition temperature in bimetallic ferromagnet–superconductor structures [JETP 86, 930–942 (May 1998)]

Yu. N. Proshin and M. G. Khusainov

Kazan State University, Kazan, Russia

Zh. Éksp. Teor. Fiz. **116**, 1882 (November 1999)

[S1063-7761(99)02711-0]

Regrettably, the authors of this paper erred in the numerical analysis of Eqs. (36), (37), and (41) for *F/S* junctions. Nevertheless, the qualitatively different variants of the behavior of the curves $T_c(d_f)$ shown in Fig. 1 are obtained but for somewhat different values of the parameters of the theory.

An accurate numerical solution shows that the values of the parameters in Fig. 1 should be a) $\sigma_s=0.5$, $2I\tau_f=0.5$, $l_s=200\ \text{Å}$, $N_s v_s=N_f v_f$; b) $\sigma_s=10$, $2I\tau_f=0.7$, $l_s=300\ \text{Å}$, $N_s v_s=0.2N_f v_f$; c) $\sigma_s=4$, $2I\tau_f=3$, $l_s=200\ \text{Å}$, $N_s v_s=N_f v_f$; d) $\sigma_s=18$, $2I\tau_f=5$, $l_s=300\ \text{Å}$, $N_s v_s=0.5N_f v_f$. In addition, $d_s=500\ \text{Å}$ and $\xi_{s0}=400\ \text{Å}$ everywhere.

Accordingly, for *F/S* superlattices phase diagrams $T_c(d_f)$ similar to those in Fig. 2 occur for the following values of the parameters of the theory: a) $\sigma_s=0.7$, $2I\tau_f=0.5$, $l_s=200\ \text{Å}$; b) $\sigma_s=10$, $2I\tau_f=0.5$, $l_s=200\ \text{Å}$; c) $\sigma_s=0.5$, $2I\tau_f=5$, $l_s=200\ \text{Å}$; d) $\sigma_s=2$, $2I\tau_f=5$, $l_s=300\ \text{Å}$. In addition, $d_s=500\ \text{Å}$, $\xi_{s0}=400\ \text{Å}$, and $N_s v_s=N_f v_f$.

These corrections alter neither the main text of the paper nor our results and conclusions.

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Investigation of the antineutrino angular distribution in experiments on the β decay of polarized neutrons

G. G. Bunatyan*)

Joint Institute of Nuclear Research, 141980 Dubna, Moscow Region, Russia

(Submitted 25 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1505–1522 (November 1999)

Since the emission of γ rays unavoidably accompanies β decay, the final state after the β decay of a neutron includes a photon along with a proton, an electron, and an antineutrino, i.e., four particles, rather than three. Therefore, when only the electron and proton momenta are detected and the γ-ray momentum is not detected in an experiment, the antineutrino momentum cannot be uniquely reconstructed, and only its mean value over a γ-ray momentum distribution determined from corresponding calculations can be considered. The γ rays are significant for finding the asymmetry parameter B of the antineutrino angular distribution from experiments on the β decay of polarized neutrons, where the electron momentum $\mathbf{p}$ directed along the $\mathbf{x}$ axis and the projection of the proton momentum P_x onto the $\mathbf{x}$ axis are detected, and the neutron polarization vector $\boldsymbol{\xi}$ is parallel or antiparallel to $\mathbf{x}$. Since the γ rays are not detected in such experiments, the antineutrino kinematics are not uniquely specified by the observables $\mathbf{p}$ and P_x and can be reconstructed only on the average, so that the antineutrino momentum distribution averaged over a γ-ray momentum distribution is considered. Thus, the exact value of B cannot be obtained from these experiments, but the true value of B can be estimated on the average by considering the mean (most likely) value $\langle B \rangle$ and the dispersion (rms deviation) ΔB. The unavoidable uncertainty in the estimate of B amounts to several percent and is thus significant for present-day experiments, which are intended to obtain the value of B to a very high accuracy of $\sim(0.1-1)\%$. If electromagnetic interactions are taken into account, measurements of the electron and proton momentum distributions can also be used to obtain g_A, i.e., the axial β-decay amplitude, to high accuracy. [S1063-7761(99)00111-0]

1. INTRODUCTION

There has recently been growing interest in achieving highly accurate knowledge of the characteristics of neutron β decay, above all the neutron lifetime τ (Ref. 1) and, in the case of polarized neutrons, the asymmetry parameter of the electron angular distribution A (Ref. 2) and the corresponding parameter for antineutrinos B (Ref. 3) relative to the neutron polarization vector $\boldsymbol{\xi}$. The value of A can be obtained in a simple manner from the experimentally observed electron angular distribution, while obtaining the value of B is a far trickier matter, since there is no way to directly measure the antineutrino angular distribution. The idea behind the experiment in Ref. 3 for obtaining the value of B originated in the distant past. It was back in the nineteen-sixties when a method was proposed[4] for determining the value of B by treating experimental data on the electron and proton angular distributions. It was assumed that the antineutrino kinematics can be uniquely reconstructed if the electron and proton momenta are given. However, this claim would only be rigorously true if there were no emission of γ rays, which, as we know (see, for example, Refs. 5–8), unavoidably accompanies β decay. Of course, the need to take into account the γ rays is only a question of the accuracy which is required to obtain the value of B by processing the experimental data. However, inasmuch as we are dealing with knowledge of the characteristics of β decay to within $\sim(0.1-1)\%$,[1–3] there is no basis to assume that it would be reasonable to neglect the γ rays in obtaining the value of B.[3]

Our professed goal is specifically to represent the influence of the electromagnetic interactions on the electron, proton, and antineutrino distributions studied in Ref. 3 and to establish the accuracy that can be achieved in describing the antineutrino angular distribution in the experiments in Ref. 3. An accurate description of semileptonic processes requires the unequivocal establishment of the characteristics of neutron β decay, particularly B, to high accuracy with proper allowance for electromagnetic interactions.

2. DESCRIPTION OF THE β DECAY OF POLARIZED NEUTRONS WITH ALLOWANCE FOR ELECTROMAGNETIC INTERACTIONS

It is presently perfectly clear that exact knowledge of the characteristics of semileptonic decays is of great importance for the theory of elementary particles, which imposes a series of rigorous constraints on the parameters entering into semileptonic interactions. Of course, the validity of the relations following from the theory must be carefully verified in order to assess with complete confidence the limits and the accuracy within which the basic principles of the present-day theory are valid. In this sense, in fact, we have no choice but to compare experimental data obtained to high accuracy with the results of systematic theoretical calculations based on the description of weak interactions following from general field theory.

The effective Lagrangian describing the β decay of baryons with allowance for electromagnetic interactions can be represented (see, for example, Refs. 9–11) in the form

$$L_{\mathrm{int}}=L_{BfBiw}+L_{e\gamma}+L_{B\gamma}, \tag{1}$$

where

$$\begin{aligned}L_{BfBiw}(x)=&\frac{G_{if}}{\sqrt{2}}(\overline{\psi}_e(x)\gamma^\alpha(1+\gamma^5)\psi_\nu(x))\overline{\Psi}_{Bf}(x)\\&\times[\gamma_\alpha g_V^B(q)+g_{WM}^B\sigma_{\alpha\nu}q_\nu+\gamma^5(\gamma_\alpha g_A^B(q)\\&+g_{IP}^Bq_\alpha)]\Psi_{Bi}(x)\end{aligned} \tag{2}$$

is the baryon-lepton $(V-A)$ weak interaction, and q is the momentum transferred in the β-decay process. The expression

$$L_{e\gamma}(x)=-e\overline{\psi}_e(x)\gamma^\mu\psi_e(x)A_\mu(x) \tag{3}$$

describes the interaction of an electromagnetic field with leptons, and similarly $L_{B\gamma}$ describes the interaction with baryons. The notation in (2) and (3) corresponds to the notation adopted in Ref. 9, but here the index B specifies the type of baryon, and we choose units with $h=c=1$. In addition, $\Psi_{Bi}(x)$ and $\Psi_{Bf}^+(x)$ are the baryon fields in the initial and final states, and ψ_e, ψ_ν, and A_μ denote the electron (positron), (anti)neutrino, and electromagnetic fields, respectively.

As for $g_V^B(0)$, by analogy with Ref. 10, we set $g_V^n(0)=1$ for neutron decay [and $g_V^\Sigma(0)=0$ for the strangeness-conserving decay process $\Sigma^\pm\to\Lambda^0+e^\pm+\nu(\overline{\nu})+\gamma$]. Then, as we know (see, for example, Refs. 9–11), the amplitudes G_{if} in (2) for various semiweak decays associated with definite $i\to f$ quark transitions $(u\to d,\ s\to d,\ b\to d)$ can be represented in the form

$$G_{if}=G_F|V_{if}|. \tag{4}$$

Here $G_F=1.16639(2)\times10^{-5}\,\mathrm{GeV}^{-2}$ can be determined from the muon lifetime,[12] and the Cabibbo–Kobayashi–Maskawa (CKM) matrix elements V_{if} (Ref. 13), which mix different quark states, satisfy the unitarity relation

$$|V_{ud}|^2+|V_{sd}|^2+|V_{bd}|^2=1, \tag{5}$$

which should hold exactly within the standard model.[10,11] Thus, any deviation from the identity (5) following from an experiment, no matter how small, would be of fundamental significance and would clearly indicate a definite contradiction in the basic principles of the present-day theory of elementary particles. The main, decisive contribution to the left-hand side of Eq. (5) stems specifically from the strangeness-conserving $u\to d$ transition: $|V_{ud}|\approx0.9744\pm0.0010$, as can be corroborated, for example, by Refs. 1 and 12. In fact, the accuracy to which the value of $|V_{ud}|$ can be established guarantees the accuracy to which the identity (5) is valid. It can be assumed that a thorough investigation of neutron β decay will permit the reliable, highly accurate determination of the CKM matrix element $|V_{ud}|$, as well as the other parameters appearing in the semiweak interaction (2). Below we shall consider only neutron β decay.

Thus, if there are experimental data that were obtained to high accuracy, $\sim0.1\%$, regarding the neutron lifetime[1] and the momentum distribution of the particles in the final state,[2,3] our task is to systematically calculate these characteristics according to (1)–(3) and then to compare the results of the calculations with the experimental data and thereby obtain reliable, highly accurate values of $|V_{ud}|$ and the amplitudes $g(q)$ in (2). When the nucleon mass is assumed to be infinite $(M\to\infty)$, $g_V(0)$ and $g_A(0)$ are known to be the main factors determining the probability of β decay, while consideration of other g amplitudes, the finite nature of M, the q dependences of $g_V(q)$ and $g_A(q)$, and the electromagnetic interactions (3) causes small corrections to it. Of course, the accuracy to which the physical quantity must be obtained determines the importance of taking into account the various corrections. In this study we consider only the effect of the electromagnetic interactions on the momentum distributions of the particles, especially the antineutrinos, in the final state after the decay of polarized neutrons, ignoring all the other corrections, which may also be important. We recall that corrections due to the finite value of the nucleon mass were thoroughly investigated in Ref. 14.

When electromagnetic interactions are properly taken into account,[8] the probability of the β decay of polarized neutrons, which is accompanied by the emission of an electron with energy-momentum $(\varepsilon,\mathbf{p})$, an antineutrino in the $\mathbf{n}_\nu$ direction, and a γ photon with energy $\omega\equiv k$ less than a certain given value k_m $(k\leqslant k_m\leqslant\Delta-\varepsilon)$, can be represented in the form

$$\begin{aligned}&d\mathbf{W}(\varepsilon,\mathbf{p},\mathbf{n}_\nu,k_m,\boldsymbol{\xi})\\&\quad=d\mathbf{w}\,\frac{d\mathbf{n}_\nu}{4\pi}\exp[\mathscr{B}(\varepsilon,k_m)]\\&\quad\times\{W_0(\varepsilon,\mathbf{p},k_m,g_V,g_A)\\&\quad+(\mathbf{v}\cdot\boldsymbol{\xi})W_{v\xi}(\varepsilon,\mathbf{p},k_m,g_V,g_A)\\&\quad+(\mathbf{n}_\nu\cdot\boldsymbol{\xi})W_{\nu\xi}(\varepsilon,\mathbf{p},k_m,g_V,g_A)\\&\quad+(\mathbf{n}_\nu\cdot\mathbf{v})W_{v\nu}(\varepsilon,\mathbf{p},k_m,g_V,g_A)\}.\end{aligned} \tag{6}$$

Here $\boldsymbol{\xi}$ is the neutron polarization vector, and the following notation has been introduced:

$$\mathscr{B}=\frac{2\alpha}{\pi}\mathscr{L}\ln\frac{2k_m}{m},\quad \mathscr{L}=\frac{1}{v}\ln\frac{p+\varepsilon}{m}-1,$$

$$W_a=w_a^0[1+\widetilde{C}_a(\varepsilon,\mathbf{p},k_m)]+C_a(g_V,g_A,\varepsilon,\mathbf{p}),$$

$$a\equiv0,\ v\xi,\ \nu\xi,\ v\nu,$$

$$w_0^0=g_V^2+3g_A^2,\quad w_{v\xi}^0=2g_A(g_V-g_A),$$

$$w_{\nu\xi}^0=2g_A(g_V+g_A),\quad w_{v\nu}^0=g_V^2-g_A^2,$$

$$d\mathbf{w}=\frac{G_{ud}^2}{2\pi^3}\varepsilon p\omega_{\nu0}^2d\varepsilon\,\frac{d\mathbf{n}_e}{4\pi},\quad \omega_{\nu0}=\Delta-\varepsilon,$$

$$\Delta=M_n-M_p,\quad \mathbf{n}_e=\frac{\mathbf{p}}{p},\quad \mathbf{v}=\frac{\mathbf{p}}{\varepsilon},\quad \mathbf{n}_\nu=\frac{\mathbf{p}_\nu}{\omega_{\nu0}}. \tag{7}$$

The exponential function $\exp[\mathscr{B}(\varepsilon,k_m)]$ in (6) corrects the infrared behavior of the decay probability: $d\mathbf{W}(\varepsilon,\mathbf{p},\mathbf{n}_\nu,k_m,\boldsymbol{\xi})\to 0$ as the cutoff γ-ray energy $k_m\to 0$. According to the general theory,[9,15] this means that β decay is impossible without the emission of infrared (soft) γ rays. For the purposes of the present work there is no need to represent the expressions for $\widetilde{C}$ and C,[8] which are multiples of the fine structure constant α, in explicit form.

After plugging in the maximum value $k_m=\Delta-\varepsilon$, the expression (6) gives the corresponding decay probability, which takes into account all the γ rays that are possible for some given value of ε. Clearly, the quantities

$$A=\frac{W_{\nu\xi}(\varepsilon,\mathbf{p},k_m,g_V,g_A)}{W_0(\varepsilon,\mathbf{p},k_m,g_V,g_A)},$$

$$B=\frac{W_{\nu\xi}(\varepsilon,\mathbf{p},k_m,g_V,g_A)}{W_0(\varepsilon,\mathbf{p},k_m,g_V,g_A)},$$

$$a=\frac{W_{\nu v}(\varepsilon,\mathbf{p},k_m,g_V,g_A)}{W_0(\varepsilon,\mathbf{p},k_m,g_V,g_A)} \tag{8}$$

describe the asymmetry of the electron (A) and antineutrino (B) angular distributions, as well as the electron-antineutrino angular correlations (a). Note that if the proton mass were not assumed to tend to infinity, none of the resultant effects—proton recoil, the q dependences of g_V and g_A, and the contribution of the terms with g_{WM} and g_{IP} in (2)—would alter the general form of the expression (6); the corresponding corrections to the quantities (8) were calculated in Ref. 14. Of course, when all the corrections just cited are omitted, the quantities (8) take the usual well-known forms

$$A_0=\frac{2g_A(g_V-g_A)}{g_V^2+3g_A^2},$$

$$B_0=\frac{2g_A(g_V+g_A)}{g_V^2+3g_A^2},\qquad a_0=\frac{g_V^2-g_A^2}{g_V^2+3g_A^2}, \tag{9}$$

and the expression (6) becomes

$$d\mathbf{W}(\varepsilon,\mathbf{p},\mathbf{n}_\nu,\boldsymbol{\xi})=d\mathbf{w}\,\frac{d\mathbf{n}_\nu}{4\pi}(g_V^2+3g_A^2)\{1+(\mathbf{v}\cdot\boldsymbol{\xi})A_0+B_0(\mathbf{n}_\nu\cdot\boldsymbol{\xi})+a_0(\mathbf{n}_\nu\cdot\mathbf{v})\}. \tag{10}$$

After the substitution $k_m=\Delta-\varepsilon$ and the integration of (6) over $d\mathbf{p}\,d\mathbf{n}_\nu$, we obtain the total probability of neutron β decay, $W=1/\tau$. Then the first relation for determining the quantities appearing in (2), particularly $|V_{ud}|$, follows from the requirement of equality between the calculated lifetime τ and the experimentally measured value.[1] The experimental value of A (Ref. 2) is obtained by studying the electron momentum distribution, which corresponds to the expressions (6) and (10) integrated over the antineutrino escape direction $d\mathbf{n}_\nu$. In that case, of course, the terms with the coefficients B and a in (6) and (10) vanish. The relation defining g_A is then obtained by equating the calculated and experimentally observed values of A (8).[2]

No matter how accurate the experimental measurements of τ and A (Refs. 1 and 2) and their theoretical calculation, even with inclusion of all the corrections discussed above, it is still highly desirable to include other characteristics of β decay along with τ and A, especially the antineutrino angular distribution relative to the neutron polarization vector $\boldsymbol{\xi}$, along with the angular correlations between the electron and antineutrino escape directions, which appear in (6) and (10) as the coefficients B and a of $(\mathbf{n}_\nu\cdot\boldsymbol{\xi})=\cos\theta_{\nu\xi}$ and $(\mathbf{n}_\nu\cdot\mathbf{v})=v\cos\theta_{v\nu}$, respectively. The inclusion of these characteristics in the treatment is useful both in rechecking the accuracy to which the quantities G_{ud}, g_A, g_{IP}, and g_{WM} in (2) are determined (see Ref. 16), and in ascertaining just how rigorous the general expression (2) is. As stated in Ref. 17, according to the ideas in Ref. 18, if the experimental value of B is known to within $\sim 0.1\%$, it is useful for testing whether the weak interactions can be left–right symmetric on the Lagrangian level and whether parity breaking results exclusively from the spontaneous breaking of this symmetry. As stated in Refs. 17 and 18, if the value of B were determined to such high accuracy, it would be possible to find the magnitude of an admixture to the Lagrangian (2) with the same transformation properties as (2) but differing from (2) by the replacement $\gamma^5\to-\gamma^5$. Thus, the problem is to verify the possibility of a contribution of right-handed currents to the semiweak interactions and to estimate the bound on the mass of the right-handed gauge boson associated with them.[19] Of course, this whole idea only makes sense when the value of B is extracted from experimental data to an accuracy of $\sim 0.1\%$, for which consideration of the electromagnetic interactions is absolutely necessary.[8]

Obviously, establishment of the value of B in (6) and (10) requires determination of the antineutrino angular distribution, which corresponds to the expression (6) integrated over the electron momentum $d\mathbf{p}$, in which the terms with the coefficients A and a vanish. However, it is totally impossible to observe the antineutrino angular distribution directly in an experiment, because there is no way to detect antineutrinos.

An extremely clever and pretty way to successfully circumvent this problem was apparently found a long time ago in Ref. 4.

3. ELECTRON AND PROTON MOMENTUM DISTRIBUTIONS AND THE ANTINEUTRINO ANGULAR DISTRIBUTION

Usually, any time we think about the β-decay process, we have in mind its description specifically by Eqs. (6)–(10), but these expressions by themselves are not applicable to obtaining B from experimental data. To describe the electron and antineutrino distributions in the final state of the β decay of polarized neutrons, the relations (6)–(10) are clearly obtained from the total decay probability by integrating over the proton and γ-ray momenta. Such distributions would correspond to an experiment in which the protons and γ rays emitted are not detected at all, i.e., we would be dealing with a decay probability which includes protons and γ rays with all the momenta that are feasible for a given $d\mathbf{p}\,d\mathbf{n}_\nu$. In experiments in which the value of A is obtained,[2] we observe only the electron momentum distribution without detecting the antineutrinos, i.e., taking all feasible antineutrino momenta into account. This observable distribution is described by the expression (6) integrated over $d\mathbf{n}_\nu$, which should then

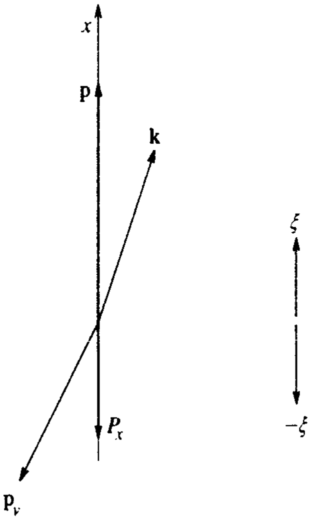

FIG. 1.

contain only the term proportional to $(\mathbf{v}\cdot\boldsymbol{\xi})A$. To obtain the value of B, i.e., the coefficient of $(\mathbf{n}_\nu\cdot\boldsymbol{\xi})$ in (6) and (10), an experiment in which, conversely, the antineutrino angular distribution is measured without regard to the proton, electron, and γ ray momenta would have to be set up. Thus, after integrating the expression (6) over $d\mathbf{p}$, the terms with A and a vanish, and only the term proportional to $(\mathbf{n}_\nu\cdot\boldsymbol{\xi})B$ remains. However, it is known that this desirable experiment is infeasible, since it is impossible to detect antineutrinos, and formulas (6) and (10) are consequently useless for directly obtaining the value of B from experimental data.

In order to have the antineutrino angular distribution at our disposal without detecting the antineutrinos themselves, we need the electron, proton, and γ-ray angular distributions; the original neutrons are assumed to be at rest. The electron momentum can be determined in present-day experiments directly to high accuracy,[2] while a simple method for measuring the proton and γ-ray momenta is presently impossible in practical terms. Thus, the situation can appear to be hopeless when we try to determine B from experimental data. However, despite how complicated this problem seems, the investigations in Ref. 4 suggest a way to solve it. A method was proposed in those papers and subsequently in Ref. 3, and a corresponding experiment was carefully developed to reconstruct the antineutrino kinematics and to consequently obtain the value of B (8) in (6) from measurements of the electron momentum distribution and the distribution of the values of the projection of the proton momentum P_x onto the **x** axis, to which the neutron polarization vector is parallel or antiparallel (see Fig. 1). The setup of a real experiment was carefully spelled out in Ref. 3, and we merely recall here that in its idealized scheme (shown in Fig. 1), which is perfectly adequate for our purposes, the detected electron momentum **p** is directed strictly along the **x** axis, the polarization vector $\boldsymbol{\xi}$ of the stationary neutron is also strictly parallel or antiparallel to the **x** axis, and the projection of the proton momentum P_x onto the **x** axis is detected in coincidence with the electron momentum **p**, while the components of the proton momentum **P** perpendicular to **x**, as well as all the γ rays, are not observed at all. When the energy ε is fixed, P_x varies in the range

$$|\mathbf{p}|-(\Delta-\varepsilon)\leqslant P_x\leqslant|\mathbf{p}|+(\Delta-\varepsilon). \tag{11}$$

For clarity, the quantity P_x in (11) and below represents the x component of the proton momentum with the opposite sign, as in Ref. 3, i.e., the quantity $P_x>0$ appears in Fig. 1. If we momentarily ignore the γ rays and neglect the kinetic energy of the proton in view of its large mass, then the antineutrino energy $\omega_{\nu 0}$ and the cosine of the angle between the **x** axis and the antineutrino escape direction (see Fig. 1) are clearly given by

$$\omega_{\nu 0}=\Delta-\varepsilon,\quad y_0\equiv\cos\theta_{\nu x}=\frac{P_x-|\mathbf{p}|}{\omega_{\nu 0}}. \tag{12}$$

In the real experiment in Ref. 3 the measurements consisted of counting the number of events with given values of **p** and P_x and a neutron polarization vector $\boldsymbol{\xi}$ parallel or antiparallel to the **x** axis. What was measured in this experiment is the probability of β decay of a polarized neutron when the electron momentum equals **p** and the projection of the proton momentum onto the **x** axis equals P_x regardless of the antineutrino and γ ray momenta, as well as the projections of the proton momentum perpendicular to the **x** axis. We can thus obtain the electron momentum distribution together with the distribution of the values of P_x:

$$d\mathbf{W}^z_{\exp}(P_x,\mathbf{p})=w^z_{\exp}(P_x,\mathbf{p})\,d\mathbf{p}\,dP_x, \tag{13}$$

the contributions of antineutrinos, γ rays, and protons with all feasible momenta being included therein. In Eq. (13) and below $z=+$ for neutron polarization parallel to the **x**, and $z=-$ for polarization in the opposite direction.

The overall expression for the probability of β decay of a polarized neutron when the electron carries momentum **p** and the x component of the momentum of the proton equals P_x, and the event is accompanied by the emission of γ rays with all energies $\omega=k=|\mathbf{k}|$ less than a certain given value k_m, $k\leqslant k_m\leqslant\Delta-\varepsilon$, can be derived using the Lagrangian (1)–(3) in the same manner as (6) was obtained in Ref. 8:

$$\begin{aligned}
d\mathbf{W}^z(P_x,\mathbf{p},k_m)&=dP_x\,\frac{d\mathbf{w}}{2\omega_{\nu 0}}\,w^z(P_x,\mathbf{p},k_m),\\
w^z(P_x,\mathbf{p},k_m)&=\exp[\mathscr{B}(\varepsilon,k_m)]\{w^0_0[1+\tilde{C}_0(\mathbf{p},k_m)]\\
&\quad+C_0(P_x,\mathbf{p},g_V,g_A)+(\mathbf{v}\boldsymbol{\xi})\{w^0_{v\xi}[1\\
&\quad+\tilde{C}_{v\xi}(\mathbf{p},k_m)]+C_{v\xi}(P_x,\mathbf{p},g_V,g_A)\}\\
&\quad+w^0_{\xi\nu}zy_0[1+\tilde{C}_{\xi\nu}(P_x,\mathbf{p},k_m)]\\
&\quad+vw^0_{v\nu}zy_0[1+\tilde{C}_{v\nu}(P_x,\mathbf{p},k_m)]\\
&\quad+zC(P_x,\mathbf{p},g_V,g_A)\}.
\end{aligned} \tag{14}$$

For the purposes of the present work there is no need to write out the detailed explicit expressions for the functions $\tilde{C}$ and

C, which are multiples of the fine structure constant α. If $k_m=\Delta-\varepsilon$, the expression (14) represents the experimental distribution (13). Instead of using (13) directly, it is more convenient to process the experimental data[2,3] using the quantity

$$X=\frac{w^{+}_{\text{exp}}(P_x,\mathbf{p})-w^{-}_{\text{exp}}(P_x,\mathbf{p})}{w^{+}_{\text{exp}}(P_x,\mathbf{p})+w^{-}_{\text{exp}}(P_x,\mathbf{p})}. \tag{15}$$

With consideration of (14), the quantity defined by (15) can be represented in the form

$$X=\frac{v[w^0_{\nu\xi}(1+\widetilde{C}_{\nu\xi})+C_{\nu\xi}]+w^0_{\xi\nu}y_0(1+\widetilde{C}_{\xi\nu})+vw^0_{\nu\nu}y_0(1+\widetilde{C}_{\nu\nu})+C}{w^0_0[1+\widetilde{C}_0]+C_0}. \tag{16}$$

Thus, comparing (14) with the corresponding experimental data for (13) and (16) (Ref. 3), we obtain an equation (in addition to the ones used in Refs. 1 and 2) for G_{ud}, $g_A,\ldots$ in (2). Thus, the value of g_A obtained in the experiments in Ref. 2 can be verified once again.[16]

However, in this way the value of B, i.e., the coefficient of $(\mathbf{n}_\nu\cdot\boldsymbol{\xi})$ in (6) and (10), which we wish to find, cannot enter into the discussion at all, because integration over all feasible antineutrino momenta was carried out during the derivation of (14), and accordingly, the distribution (13) observed in the experiment in Ref. 3 includes a contribution from antineutrinos with all feasible momenta. Thus, we must find another way that does not employ (14), but properly takes into account the influence of the γ rays in the process of extracting the value of B from the experimental distribution (13) observed in Ref. 3.

4. ESTIMATION OF *B* FROM THE ELECTRON AND PROTON MOMENTUM DISTRIBUTIONS

Clearly, in a hypothetical simplified case in which the relations (12) are valid, the antineutrino kinematics would be uniquely specified by the values of $\mathbf{p}$ and P_x detected in the experiment in Ref. 3. Accordingly, without consideration of the γ rays and under the assumption that the proton mass is infinite, i.e., $M\rightarrow\infty$, the distribution (14) transforms into

$$dW^z(P_x,\mathbf{p})=dP_x\frac{d\mathbf{w}}{2\omega_{\nu0}}w^z(P_x,\mathbf{p}),$$

$$w^z(P_x,\mathbf{p})=w^0_0[1+Azv+B_0y_0z+ay_0v], \tag{17}$$

while the antineutrino angular distribution takes on the form (10). Thus, in this simplified case there would be a one-to-one correspondence between the distribution (13), which describes the experiment in Ref. 3, and the antineutrino angular distribution (10), the quantity $(\boldsymbol{\xi}\cdot\mathbf{n}_\nu)$ in (10) taking the value of zy_0 in (17), and the quantity $dP_x/2\omega_{\nu0}$ in (17) replacing $d\mathbf{n}_\nu/4\pi$ in (10). Then a comparison of (17) with the experimental distribution (13) (Ref. 3) would yield the equation

$$w^z_{\text{exp}}(P_x,\mathbf{p})=f_0(\omega_{\nu0})(1+zAv)$$

$$+f_0(\omega_{\nu0})y_0(zB_0+av). \tag{18}$$

Accordingly, the coefficient of $(\boldsymbol{\xi}\cdot\mathbf{n}_\nu)$ in (10) would be expressed directly in terms of w^z_{exp} (13):

$$B_0=\frac{1}{zy_0f_0}[w^z_{\text{exp}}-f_0(1+zAv)-f_0avy_0],$$

$$f_0\frac{G^2_{ud}\omega_{\nu0}}{16\pi^4}w^0_0, \tag{19}$$

or, as in Ref. 3, in terms of X (15):

$$B_0=[X(1+avy_0)-Av]/y_0. \tag{20}$$

The value $B_0=0.9821\pm0.0040$ was given in Ref. 3. Of course, we ignored all the uncertainties that inevitably arise in a real experiment because of the sparse statistics, imperfect equiipment geometry, etc. This ultimately does not mean that we must regard the corresponding corrections as being negligible, but our purpose here is to thoroughly investigate only the influence of the γ rays on obtaining B by processing the experimental data in Ref. 3. In (17)–(20) we added the subscript 0 to B to stress that this value would be obtained with exclusion of the γ rays, as in (9) and (10). As we see, B_0 turns out to be expressed in terms of f_0, y_0f_0, and $\omega_{\nu0}$ [(12) and (19)], which would be known exactly in this case for each event with the values of P_x and $\mathbf{p}$ that were detected in the experiment in Ref. 3.

However, since neutron β decay, as has been well known for a long time, is accompanied by the emission of γ rays, the probability of β decay was measured in the real experiment in Ref. 3 with given values of P_x and $\mathbf{p}$ and with inclusion of γ rays with all feasible values of the momentum $\mathbf{k}$. In describing each individual event the expressions for y_0 and $\omega_{\nu0}$ in (12) undergo the following replacements:

$$y_0\rightarrow y(\omega)=\cos\theta_{\nu x}=\frac{P_x-|\mathbf{p}|-x\omega}{\omega_\nu},\quad x=\cos\theta_{\gamma x},$$

$$f_0\rightarrow f(\omega)=\frac{G^2_{ud}\omega_\nu}{16\pi^4}w^0_0,\quad \omega_{\nu0}\rightarrow\omega_\nu(\omega)=\Delta-\varepsilon-\omega, \tag{21}$$

where $\omega=|\mathbf{k}|$ is the γ-ray energy and $\theta_{\gamma x}$ is the angle defining the direction of the γ rays relative to the $\mathbf{x}$ axis (see Fig. 1). Thus, to uniquely reconstruct the antineutrino kinematics, we would have to know the values of ω and x associated with the γ rays accompanying each individual β-decay event with the values of $\mathbf{p}$ and P_x that were detected in the experiment in Ref. 3.

Of course, for real β decay, which is accompanied by the emission of γ rays, the required coefficient B cannot be expressed according to (19) and (20) in terms of f_0, f_0y_0, and $\omega_{\nu 0}$, which themselves, strictly speaking, do not have rigorous physical meaning. It is then natural to estimate the coefficient B in (6) by introducing into the discussion the expected values $\langle f\rangle$ and $\langle yf\rangle$ of $f(\omega)$ and $f(\omega)y(\omega,x)$ (21), which replace f_0 and f_0y_0 in a state with definite values of P_x, $\mathbf{p}$, and z. The values of $\langle f\rangle$ and $\langle yf\rangle$ must be calculated by averaging $f(\omega)$ and $f(\omega)y(\omega,x)$ over the momentum distribution $W_\gamma^z(P_x,\mathbf{p},\mathbf{k})$ of the γ rays accompanying decay with given values of P_x, $\mathbf{p}$, and z. To estimate the required value of B [given by (8)] in (6), we must introduce the expected value $\langle B\rangle$ in terms of $\langle yf\rangle$ and $\langle f\rangle$.

Recent experiments[3] have detected only the total number of β-decay events with definite values of $\mathbf{p}$ and P_x accompanied by the emission of γ rays with all kinematically feasible values of ω and x, i.e., the integral

$$w^z_{\mathrm{exp}}(\mathbf{p},P_x)=\int d\mathbf{k}\, w^z_{\mathrm{exp}}(\mathbf{p},P_x,\mathbf{k}) \qquad (22)$$

over all feasible $\mathbf{k}$. The problem is to be able to extract accurate information regarding the coefficient B (8) in (6) from the experimentally observed distribution $w^z_{\mathrm{exp}}(\mathbf{p},P_x)d\mathbf{p}\,dP_x$ (13) (Ref. 3). Each individual decay event with a given value of $\mathbf{k}$ appears in the experimental quantity $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ with its own individual weight and with a probability $W^z_\gamma(\mathbf{p},P_x,\mathbf{k})d\mathbf{k}$ of the emission of γ rays with a given momentum $\mathbf{k}$ accompanying β decay with given values of $\mathbf{p}$ and P_x. Thus, (18) can be replaced by a new relation, in which the experimentally observed quantity $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ is equated to the β-decay probability averaged with the weight $W^z_\gamma(P_x,\mathbf{p},\mathbf{k})$:

$$w^z_{\mathrm{exp}}(\mathbf{p},P_x)=\frac{\int d\mathbf{k}\,W^z_\gamma(P_x,\mathbf{p},\mathbf{k})f(\omega)[1+zAv+z\langle B\rangle^z y(\omega,x)+avy(\omega,x)]}{\int d\mathbf{k}\,W^z_\gamma(\mathbf{p},P_x,\mathbf{k})}=\langle f\rangle^z(1+zAv)+\langle yf\rangle^z(z\langle B\rangle^z+av), \qquad (23)$$

where we have introduced the ordinary averaging notation:

$$\langle F\rangle^z(P_x,\mathbf{p})=\frac{\int d\mathbf{k}\,W^z_\gamma(P_x,\mathbf{p},\mathbf{k})F(\mathbf{k})}{\int d\mathbf{k}\,W^z_\gamma(P_x,\mathbf{p},\mathbf{k})}=\frac{\int_0^{\Delta-\varepsilon}\omega^2\,d\omega\int_{x_1}^{x_2}dx\,F(P_x,\mathbf{p},\omega,x)\int_0^{2\pi}d\phi\,W^z_\gamma(P_x,\mathbf{p},\omega,x,\phi)}{\int_0^{\Delta-\varepsilon}\omega^2\,d\omega\int_{x_1}^{x_2}dx\int_0^{2\pi}d\phi\,W^z_\gamma(P_x,\mathbf{p},\omega,x,\phi)}. \qquad (24)$$

Here the limits

$$x_1=1-(\Delta+|\mathbf{p}|-\varepsilon-P_x)/\omega\geqslant -1,$$

$$x_2=-1+(\Delta-|\mathbf{p}|-\varepsilon+P_x)/\omega\leqslant 1$$

follow simply from the kinematics of the process under consideration, the averaged values of $f(\omega)$ and $f(\omega)y(\omega,x)$ being independent of the azimuth ϕ of the γ rays. The value of P_x varies in the range (11) for a given electron energy ε. We discuss the form of the distribution $W^z_\gamma(P_x,\mathbf{p},\mathbf{k})$ and its main properties somewhat later on. Since $W^z_\gamma(P_x,\mathbf{p},\mathbf{k})$ depends on the neutron polarization ($z=\pm$), all the mean values (24) and (23), in turn, also depend on z, being different for the different polarizations $z=+$ and $z=-$.

Thus, we have (23), which replaces the former relation (18). Since B_0 in (18) is the coefficient of zy_0f_0, it should be equal specifically to the asymmetry parameter of the antineutrino angular distribution B_0 (9) in (10), as explained above. Since we are dealing with the description of a real experiment, the expression (23) defines $\langle B\rangle^z$, which is the coefficient of the mean (expected) value $\langle yf\rangle^z$, which replaces f_0y_0 in (18). Comparing the distribution (23), which contains $\langle B\rangle^z$, and the antineutrino angular distribution (6), which contains B (8), we see that $\langle B\rangle^z$ can be used to estimate the value of B (8) in (6) in the mean, which is now our goal. According to (23), $\langle B\rangle^z(\mathbf{p},P_x)$ can be expressed in terms of the experimental values of $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ (22) (Ref. 3). Of course, it would be desirable to calculate $\langle B\rangle^z(\mathbf{p},P_x)$ directly with the values of $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$, from which the value of B_0 [(19) and (20)] was obtained without consideration of the γ rays in Ref. 3. However, we are unable to calculate $\langle B\rangle^z(\mathbf{p},P_x)$ from $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$, because these measurement results were not explicitly given in Ref. 3. Therefore, we obtain the expected value $\langle B\rangle^z(\mathbf{p},P_x)$ [and then its dispersion $\Delta B^z(\mathbf{p},P_x)$] from B_0 [(19) and (20)], rather than from the values of $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ themselves. The expression for $\langle B\rangle^z$ in terms of B_0 can be obtained directly from (17)–(19) and (23):

$$\langle B\rangle^z=z[(1+zAv)(f_0-\langle f\rangle^z)+y_0f_0(av+zB_0)]/\langle yf\rangle^z-zav. \qquad (25)$$

As we see, we have obtained two different values of $\langle B\rangle^z$ to describe the antineutrino angular distribution for the two different neutron polarizations $z=\pm$. Obviously, (25) reduces to $\langle B\rangle^z=B_0$ when $\langle f\rangle^z=f_0$ and $\langle yf\rangle^z=f_0y_0$. Otherwise, $\langle B\rangle^z$ turns out to be represented in the form of a function of the expected (mean) values of $f(\omega)$ and $f(\omega)y(\omega,x)$. Thus, in order to assess the accuracy and even the very possibility of such an estimate of B using the mean value $\langle B\rangle^z$ with complete confidence, we must ascertain the form that the distributions of $f(\omega)$ and $f(\omega)y(\omega,x)$ have in the vicinity of the most likely (expected) values $\langle f\rangle$ and $\langle yf\rangle$, i.e., we must investigate the dispersion of $f(\omega)$ and

$f(\omega)y(\omega,x)$. Thus, along with the expected values $\langle f\rangle$ and $\langle yf\rangle$ themselves, we must also calculate the rms deviations of $f(\omega)$ and $f(\omega)y(\omega,x)$ from their expected values $\langle f\rangle$ and $\langle yf\rangle$:

$$(\Delta f)^z=\sqrt{\langle(\Delta f)^2\rangle^z}=\sqrt{\langle f^2\rangle^z-(\langle f\rangle^z)^2},$$
$$(\Delta yf)^z=\sqrt{\langle(\Delta(yf))^2\rangle^z}=\sqrt{\langle(yf)^2\rangle^z-(\langle yf\rangle^z)^2}, \tag{26}$$
$$\langle\Delta(f\cdot yf)\rangle^z=\langle f\cdot yf\rangle^z-\langle f\rangle^z\langle yf\rangle^z.$$

Thus, the dispersion

$$\Delta B^z\equiv\sqrt{\langle(\Delta B^z)^2\rangle}=\sqrt{\langle B^2\rangle^z-(\langle B\rangle^z)^2}, \tag{27}$$

which specifies the accuracy achieved in the estimate (25) of B, can be expressed in the usual manner (see, for example, Ref. 20) in terms of the quantities in (26) and the derivatives

$$\frac{\partial\langle B\rangle^z}{\partial\langle f\rangle^z},\quad \frac{\partial\langle B\rangle^z}{\partial\langle yf\rangle^z}. \tag{28}$$

Of course, in (26) the expected values $\langle f^2\rangle^z$, $\langle(fy)^2\rangle^z$, and $\langle f^2y\rangle^z$ are obtained by averaging $f^2(\omega)$, $[f(\omega)y(\omega,x)]^2$, and $f^2(\omega)y(\omega,x)$ in accordance with (24) with the weight $W_\gamma^z(P_x,\mathbf{p},\mathbf{k})$.

Thus, the uncertainties in estimating the true value of B in terms of $\langle B\rangle^\pm$ stem from the difference between $\langle B\rangle^+$ and $\langle B\rangle^-$ themselves and from the appearance of the dispersion $\Delta B^\pm$.

The expected value $\langle B\rangle^z$ is suitable for estimating the true value of B (8) in (6) when the distributions of the values of $f(\omega)$ and $f(\omega)y(\omega,x)$ are ''sharp,'' i.e., when the ratios $\Delta f/\langle f\rangle$ and $\Delta(fy)/\langle fy\rangle$ and, consequently, $\Delta B/\langle B\rangle$ for given values of P_x and $\mathbf{p}$ are very small (essentially negligible) compared to the errors entailed in determining B to a desirable accuracy in Ref. 3. The value of the ratio $\Delta B/\langle B\rangle$ essentially sets the limit of the accuracy of obtaining the value of B (8) in (6) by processing the experimental data for (13) in Ref. 3. If the distributions of $f(\omega)$ and $f(\omega)y(\omega,x)$ in the vicinity of $\langle f\rangle$ and $\langle fy\rangle$ for certain values of P_x and $\mathbf{p}$ are so spread out that $\Delta f/\langle f\rangle\sim 1$ and $\Delta(fy)/\langle fy\rangle\sim 1$ and, accordingly, $\Delta B/\langle B\rangle\sim 1$, then there will not be any basis for estimating the value of B (8) in (6) in terms of $\langle B\rangle^z$. In such a case the antineutrino kinematics and the antineutrino angular distribution (6) cannot be reconstructed from the experimentally observed[3] distribution (13) even in the mean. For such circumstances and for such values of P_x and $\mathbf{p}$ there is no reasonable way to accurately estimate the value of B (8) in (6) using the experimental data in Ref. 3.

Thus, we must clearly understand to what limits and to what accuracy we are capable of reconstructing the antineutrino kinematics, having at our disposal only experimentally observed[3] values of $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ (13) for given values of $\mathbf{p}$ and P_x without detecting the γ rays. The problem facing us is therefore to calculate the uncertainties in estimating B in terms of $\langle B\rangle^z$ and, in essence, to ascertain the physical meaning of introducing $\langle B\rangle^\pm$ for given values of $\mathbf{p}$ and P_x. Of course, if the differences $B^\pm-B_0$ between the calculated values of $B^\pm$ (25) and B_0 [(19) and (20)] found in Ref. 3, like the values of $\Delta B^\pm$ (27), are actually negligible for any $\mathbf{p}$ and P_x in comparison to the measurement error of B stated in Ref. 3, $\Delta B\approx 0.4\%$, then consideration of the γ rays would be superfluous; however, this is not the case in reality.

5. CALCULATION OF $\langle B\rangle^z$ AND ΔB^z; DISCUSSION OF RESULTS

The distribution $W^z_\gamma(P_x,\mathbf{p},\mathbf{k})$ in (23) and (24) can be obtained to first order in α according to Ref. 8 for the β decay of polarized neutrons directly from (1)–(3) in the same way that the probability $W_\gamma(\omega)$ of the emission of γ rays with an energy ω was calculated back in Ref. 5 for all feasible values of P_x, $\mathbf{p}$, and $\mathbf{k}/k$ in the case of the β decay of unpolarized neutrons. Then, after integrating the γ-ray distribution $W^z_\gamma(P_x,\mathbf{p},\omega,x,\phi)$ over $d\phi$, which is contained in (24), calculating the expected values $\langle f\rangle^z$ and $\langle fy\rangle^z$ in (23)–(28) reduces to averaging corresponding values of $f(\omega)$ and $f(\omega)y(\omega,x)$ with the function

$$\begin{aligned}&\omega^2 d\omega\, dx\, d\mathbf{p}\, dP_x\int_0^{2\pi}d\phi\, W^z_\gamma(P_x,\mathbf{p},\omega,x,\phi)\\&=\left(\frac{eG_{ud}}{2\sqrt{2}}\right)^2\frac{8}{(2\pi)^7}\frac{1}{4\varepsilon^2}\frac{\omega_\nu}{[1-xv]^2}\frac{1}{\omega}\,dx\,d\omega\,dP_x d\mathbf{p}\\&\times\{(1-x^2)\varepsilon v[v(\varepsilon+\omega)(g_V^2+3g_A^2)+y(\omega+v^2\varepsilon)\\&\times(g_V^2-g_A^2)]+\omega^2[(g_V^2+3g_A^2)+yx(g_V^2-g_A^2)]\\&\times(1-vx)+2zg_A[(1-x^2)\varepsilon v[(g_V-g_A)\\&\times(v^2\varepsilon+\omega)+(g_V+g_A)vy(\varepsilon+\omega)]+\omega^2\\&\times(1-vx)[(g_V-g_A)x+(g_V+g_A)y]]\}.\end{aligned} \tag{29}$$

However, this expression, in and of itself, is still not applicable directly to the calculation of (23)–(28) because of the explicit nonintegrable singularity at $\omega\to 0$, which leads to logarithmic divergence when (24) is integrated over $d\omega$. Since the averaged values of $f(\omega)$ and $f(\omega)y(\omega,x)$ are not dependent on x when $\omega\to 0$, we are concerned essentially only with the correct behavior of the distribution $W^z_\gamma(\mathbf{p},P_x,\omega,x,\phi)$ integrated over $dx\,d\phi$ at $\omega\to 0$. To systematically describe the γ rays at small ω, where $\alpha\ln(\Delta/\omega)\geqslant 1$, i.e., infrared radiation, the processes which involve an arbitrary number of ''soft'' photons must be taken into account.[15] In accordance with the method described in Refs. 9 and 15, the correct behavior of the γ-ray distribution at $\omega\to 0$ can be obtained, according to Ref. 8, by making the following replacement in (29) at $\omega\to 0$:

$$\frac{1}{\omega}\to\frac{1}{m}\left(\frac{m}{\omega}\right)^{1-o},\quad o=\frac{2\alpha}{\pi}\left[\frac{1}{v}\ln\left(\frac{\varepsilon+|\mathbf{p}|}{m}\right)-1\right]. \tag{30}$$

It might be appropriate to recall that at such small values of $\omega\to 0$, where $\alpha\ln(\Delta/\omega)\geqslant 1$, we are dealing with the isotropic emission of an unspecified, infinite number of ''soft'' photons,[9,15] i.e., the classical emission of γ rays occurs under such conditions. Systematic consideration of the emission of infrared γ rays (''soft photons'') leads, according to Ref. 8, to the multiplier $\exp[\mathscr{B}(\varepsilon,k_m)]$ in formulas (6) and (14), and

TABLE I. Values of $(\langle B\rangle^{\pm}-B_0)/B_0$ (in percent) as a function of ε and y_0 (see the text).

ε, MeV	y_0 -0.80	-0.40	-0.20	-0.10	-0.05	0.05	0.10	0.20	0.40	0.80
					$(\langle B\rangle^{+}-B_0)/B_0$					
0.55	0.00	-0.04	-0.12	-0.28	-0.59	0.66	0.34	0.19	0.10	0.03
0.70	-0.01	-0.09	-0.25	-0.57	-1.20	1.34	0.70	0.38	0.21	0.08
0.80	-0.02	-0.12	-0.33	-0.74	-1.57	1.74	0.91	0.49	0.27	0.11
0.90	-0.02	-0.15	-0.40	-0.90	-1.90	2.11	1.10	0.59	0.33	0.14
1.00	-0.03	-0.17	-0.47	-1.05	-2.20	2.45	1.28	0.69	0.39	0.17
1.25	-0.04	-0.23	-0.62	-1.37	-2.86	3.20	1.66	0.89	0.50	0.24
					$(\langle B\rangle^{-}-B_0)/B_0$					
0.55	0.02	0.08	0.17	0.34	0.67	-0.65	-0.32	-0.15	-0.06	-0.01
0.70	0.03	0.13	0.29	0.58	1.17	-1.18	-0.58	-0.28	-0.12	-0.02
0.80	0.03	0.15	0.34	0.70	1.40	-1.43	-0.71	-0.34	-0.15	-0.03
0.90	0.04	0.17	0.38	0.79	1.59	-1.64	-0.81	-0.40	-0.18	-0.04
1.00	0.04	0.18	0.42	0.86	1.74	-1.81	-0.90	-0.44	-0.20	-0.05
1.25	0.05	0.21	0.49	1.01	2.03	-2.14	-1.07	-0.52	-0.24	-0.07

then (29) and (30) can also be derived directly by differentiating (14) with respect to k_m and then plugging in $k_m=\Delta-\varepsilon$ (Ref. 8).

With the distribution $W_{\gamma}^{z}(P_x,\mathbf{p},\mathbf{k})$ [(29) and (30)] the integrals (24) do not diverge, and the values of $\langle B\rangle^{z}(P_x,\mathbf{p})$ and $\Delta B^{z}(P_x,\mathbf{p})$ can be calculated directly. It is noteworthy that as the fine structure constant $\alpha\to 0$, all the values of (24) calculated with $W_{\gamma}^{z}(P_x,\omega,\mathbf{p},x,\phi)$ [(29) and (30)] become equal to the values of $F(0)$ found by averaging $F(\omega)$ at $\omega=0$. In fact, when $\alpha=0$, the value of o in (30) also vanishes, and the normalization integral over $d\omega$ in the denominator in (24) diverges logarithmically at the lower limit $\omega\to 0$. Accordingly, if the numerator in (24) had a finite value, the entire expression (24) would vanish. Expanding the averaged function in ω, $F(\omega)\approx F(0)+\omega F'(0)+\dots$, we see that all terms containing ω make finite contributions to the integral in the numerator in (24), while the term with $F(0)$ is exactly a multiple of the normalization integral in the denominator. This normalization integral cancels out, and each value of (24) simply reduces to $F(0)$. Thus, at $\alpha\to 0$, we would have

$$\langle f(\omega)\rangle^{z}\to f(0)\equiv f_0,\quad \langle y(\omega)\rangle^{z}\to y(0)\equiv y_0,$$

$$\langle f(\omega)y(\omega)\rangle^{z}\to f(0)y(0)\equiv f_0y_0, \tag{31}$$

$$\langle B\rangle^{z}=B_0$$

and all the uncertainties would vanish:

$$\Delta f=\Delta(yf)=\Delta B=0.$$

Thus, according to its physical meaning, formula (25) would transform into (19) if α were to vanish, and we would obtain the result corresponding to the idealized case in which γ rays do not enter into the calculation, as assumed in Ref. 3.

The results of the calculation of $\langle B\rangle^{z}$ and ΔB^{z} are presented in Tables I and II. Table III shows which value of P_x corresponds to specific values of y_0 and ε. The first row in each table contains the values of y_0 (12), and the electron energies (in Mev) are given in the first column. Thus, each value of the parameters in the tables corresponds to a definite pair of values of y_0 and ε. In these results, y_0 varies from -0.8 to $+0.8$, as in Ref. 3. Table I presents the values of

$$\frac{\langle B\rangle^{\pm}-B_0}{B_0}\cdot 100, \tag{32}$$

TABLE II. Same as in Table I, but for $\Delta B^{\pm}/\langle B\rangle^{\pm}$.

ε, MeV	y_0 -0.80	-0.40	-0.20	-0.10	-0.05	0.05	0.10	0.20	0.40	0.80
					$\Delta B^{+}/\langle B\rangle^{+}$					
0.55	0.39	1.53	4.48	10.34	22.01	24.53	12.85	6.96	3.86	1.68
0.70	0.37	2.14	5.92	13.45	28.44	31.82	16.62	9.00	5.05	2.28
0.80	0.41	2.45	6.67	15.05	31.70	35.58	18.54	10.02	5.64	2.64
0.90	0.45	2.70	7.29	16.37	34.39	38.76	20.15	10.87	6.14	2.96
1.00	0.48	2.92	7.82	17.50	36.67	41.51	21.54	11.60	6.56	3.24
1.25	0.57	3.37	8.97	19.74	41.18	47.09	24.34	12.98	7.38	3.82
					$\Delta B^{-}/\langle B\rangle^{-}$					
0.55	1.19	3.29	6.47	12.47	24.31	22.99	11.05	5.04	1.96	0.56
0.70	1.01	3.37	7.07	14.04	27.75	27.43	13.38	6.32	2.66	0.65
0.80	0.96	3.40	7.27	14.56	28.85	29.02	14.20	6.77	2.92	0.71
0.90	0.94	3.43	7.40	14.89	29.55	30.13	14.77	7.08	3.10	0.77
1.00	0.93	3.45	7.50	15.13	30.02	30.96	15.18	7.31	3.24	0.83
1.25	0.94	3.51	7.72	15.48	30.68	32.38	15.87	7.61	3.46	0.95

TABLE III. Same as in Table I, but for P_x (in MeV).

ε, MeV	y_0									
	−0.80	−0.40	−0.20	−0.10	−0.05	0.05	0.10	0.20	0.40	0.80
0.55	−0.39	−0.09	0.05	0.13	0.17	0.24	0.28	0.35	0.50	0.80
0.70	0.00	0.24	0.36	0.42	0.45	0.51	0.54	0.60	0.72	0.95
0.80	0.22	0.42	0.52	0.57	0.59	0.64	0.66	0.71	0.81	1.01
0.90	0.43	0.58	0.66	0.70	0.72	0.76	0.78	0.82	0.90	1.06
1.00	0.62	0.74	0.80	0.83	0.84	0.87	0.89	0.92	0.98	1.10
1.25	1.11	1.12	1.13	1.14	1.14	1.14	1.15	1.15	1.16	1.18

i.e., the deviations of $\langle B\rangle^z$ (25) from B_0 (19) in percent. For example, we conclude from Tables I and II that for $y_0=0.2$ and $\varepsilon=1$ Mev, B^+ exceeds B_0 by 0.69%, while for the corresponding values of ε and P_x, B^- is 0.44% less than B_0. The differences in Table I increase significantly with decreasing $|y_0|$. It becomes clear from these tables that the differences $\langle B\rangle^{\pm}-B_0$ and $\langle B\rangle^{+}-\langle B\rangle^{-}$ cannot be considered negligible so long as we are dealing with an accuracy of 1% or better in the determination of B. Thus, the results in Table I are enough convince us that there is no basis to assume that an accuracy of ~0.4% is achieved in determining B by processing the experimental data in Ref. 3, in which γ rays are not taken into account.

To continue, the values of

$$\frac{\Delta B^{\pm}}{\langle B\rangle^{\pm}}\cdot 100, \tag{33}$$

which represent the width ΔB^z (dispersion) of the possible distribution of values of B about the expected (mean) value $\langle B\rangle^z$, are given in Table II. Of course, having only the observable $w^z_{\mathrm{exp}}(\mathbf{p},P_x)$ (13) (Ref. 3) at our disposal, we are not able to estimate the true value of B to an accuracy better than $\Delta B^z(y_0,\varepsilon)$, as explained above. However, as can be seen from Table III, even the smallest uncertainties ΔB for $|y_0|=0.8$ reach ~1%, and the values of ΔB rise very quickly with decreasing $|y_0|$, as do the values in Table I. From the data in the tables we conclude that the uncertainties in the estimation of B due to the large dispersion $\Delta B^{\pm}$ are far more significant than those due to the difference between $\langle B\rangle^{+}$ and $\langle B\rangle^{-}$. Thus, it is unreasonable to determine the value of B (8) in (6) by processing the experimental data in Ref. 3 for values of $\mathbf{p}$ and P_x corresponding to small values of $|y_0|$.

Of course, there is nothing surprising in the fact that the quantities in Table I and especially in Table II increase very rapidly as $y_0\to 0$. The physical reason for such behavior of $\langle B\rangle$ and ΔB is perfectly clear. In fact, when $y_0\approx 0$, i.e., when $|\mathbf{p}|\approx P_x$, the presence of the term $x\omega$ in $y(\omega)$ (21) leads to significant values of the ratios $(y-y_0)/y_0$ and $(\Delta y f)/\langle y f\rangle$ for all ω, no matter how small. In this case the presence of γ rays completely alters the antineutrino kinematics in comparison to the kinematics that would be observed in the absence of electromagnetic interactions. The values of $(\langle B\rangle^{\pm}-B_0)/B_0$ and $\Delta B^{\pm}/\langle B\rangle^{\pm}$, in turn, increase significantly and can even become arbitrarily large as $|y_0|\to 0$. Of course, under these conditions we can say nothing definite about the expected (mean) values of these parameters.

This result leads us to a clear understanding that there is no significant physical basis to use the experimental data for small values of $|y_0|\sim 0.1$ to obtain a credible value of the asymmetry parameter B of the antineutrino angular distribution. Having at our disposal only the probability of the β decay of polarized neutrons for given values of $\mathbf{p}$ and P_x without detecting the accompanying γ rays, we are able to reconstruct the antineutrino angular distribution only to an approximation, on the average, and accordingly to estimate the value of B based on the results in the tables to the same approximation. One can then conclude that the determination of B becomes invalid and essentially devoid of any credibility when events corresponding to $|y_0|\leqslant 0.1$ are included in the analysis without assigning an appropriate weight to them. Processing the experimental data at other than these small values of $|y_0|$, we can claim to obtain a semiquantitative estimate of B to an accuracy of roughly several percent. In the best case it can be assumed that an accuracy better than 1% can be achieved in reconstructing B from the electron and proton momentum distributions in Ref. 3 by taking into account only events with $|y_0|\approx 0.8-1.0$.

6. CONCLUDING REMARKS

We thus have no basis for neglecting the emission of γ rays in finding the value of B, or confirming the very high accuracy of ≈0.4% in the measurement of B claimed in Ref. 3.

It should be clear and should be specially stressed that the uncertainties introduced by the undetected γ rays in the process of obtaining B, i.e., the differences $|B^{+}-B^{-}|$ and $|B^{\pm}-B_0|$ and the dispersion ΔB^z, cannot be eliminated or even diminished by improving the statistics in the experiment in Ref. 3 or by skillfully improving any of the existing experimental equipment.[3] The values listed in the tables were obtained during our analysis, because the antineutrino kinematics can be reconstructed from the measurements of $\mathbf{p}$ and P_x in Ref. 3 only in the mean, since the γ rays were not detected. In any decay event the antineutrino kinematics could be uniquely reconstructed if the γ-ray momentum were detected along with the values of $\mathbf{p}$ and P_x. Accordingly, an experiment should be set up to detect triple coincidences in the decay of polarized neutrons, i.e., events in which ω and x (21) of the accompanying γ rays are detected simultaneously with $\mathbf{p}$ and P_x. There would then be a one-to-one correspondence between such an observed triple distribution and the antineutrino angular distribution, in which the parameter B of interest to us would appear. Despite the difficulty of carrying out such an experiment, there is no reason to doubt that it is feasible if it actually turns out to be necessary.

The conclusion to be drawn from our calculations (see Table I) is that the uncertainties in B due to γ rays could have been reduced to $\leqslant 1\%$ if the value of B had been extracted in Ref. 3 by processing experimental data only for $|y_0| \geqslant 0.8$. Thus, it might be reasonable to reprocess the data in Ref. 3 with consideration of this limitation, and to directly calculate the expected value $\langle B\rangle^z(\mathbf{p},P_x)$ and its dispersion $\Delta B^z(\mathbf{p},P_x)$ according to (23) from the corresponding measured values of $w^z_{\exp}(\mathbf{p},P_x)$ (22) (Ref. 3).

We also recall that, as noted in Sec. 3 [see (14)–(16)], the experimental data obtained in Ref. 3 might be used directly to find the value of g_A in (2). Of course, explicit, complete expressions for $\tilde{C}(P_x,\mathbf{p},g_V,g_A)$ and $C(P_x,\mathbf{p},g_V,g_A)$ in (14) must be obtained. This can essentially be accomplished in exactly the same way that all the coefficients in (6) were previously obtained.[8] Conversely, having the value of g_A obtained in Ref. 2 at our disposal, we can substitute it into (14) to see whether the experimental data in Ref. 3 are thereby reconstructed. A discussion of the values of g_A obtained from different experiments is given in Ref. 16.

In light of the foregoing (see Sec. 2) regarding the importance of establishing the values of both B itself and of g_A to the required high accuracy, it becomes perfectly clear that the investigations performed in Refs. 3 and 4 should be developed as vigorously as possible.

We thank V. A. Vtyurin, V. G. Nikolenko, Yu. N. Pokotilovskiĭ, and Yu. V. Khol'nov for some very useful discussions and for their support.

This work was supported by the Russian Fund for Fundamental Research (Project 96-02-188826).

*)E-mail: bunat@cv.jinr.dubna.su

---

[1] A. Pichlmeier, P. Geltenbort, V. Nesvizhevsky *et al.*, in *Proceedings of ISINN-6, E3-98-202*, JINR, Dubna, Russia, 15–18 May, 1998, p. 220; S. Arzumanov, L. Bondarenko, S. Chernyavski *et al.*, in *Proceedings of ISINN-5, E3-97-213*, JINR, Dubna, Russia, 14–17 May, 1997, p. 53; K. Schreckenbach and W. Mampe, J. Phys. G **12**, 1 (1992); W. Mampe, P. Ageron, J. C. Bates *et al.*, Nucl. Instrum. Methods Phys. Res. A **284**, 111, (1989); W. Mampe, L. N. Bondarenko, V. I. Morozov *et al.*, JETP Lett. **57**, 82 (1993); J. Byrne, P. G. Dawber, C. G. Habeck *et al.*, Europhys. Lett. **33**, 187 (1996).

[2] J. Reich, H. Abele, M. A. Hoffman *et al.*, in *Proceedings of ISINN-6, E3-98-209*, JINR, Dubna, Russia, 15–18 May, 1998, p. 226; H. Abele, S. Bäßler D. Dubbers *et al.*, Phys. Lett. B **407**, 212 (1997); P. Liaud, K. Schreckenbach, R. Kossakowski *et al.*, Nucl. Phys. A **612**, 53 (1997); K. Schreckenbach, P. Liaud, R. Kossakowski *et al.*, Phys. Lett. B **349**, 427 (1995).

[3] I. A. Kuznetsov, A. P. Serebrov, I. V. Stepanenko *et al.*, Phys. Rev. Lett. **75**, 794 (1995); I. A. Kuznetsov, A. P. Serebrov, I. V. Stepanenko *et al.*, JETP Lett. **60**, 315 (1994); A. P. Serebrov, I. A. Kuznetsov, I. V. Stepanenko *et al.*, Zh. Éksp. Teor. Fiz. **113**, 1963 (1998) [JETP **86**, 1074 (1998)].

[4] B. G. Erozolimskiĭ, Usp. Fiz. Nauk **116**, 145 (1975) [Sov. Phys. Usp. **18**, 377 (1975)]; B. G. Erozolimskiĭ, L. N. Bondarenko, Yu. A. Mostovoĭ *et al.*, Yad. Fiz. **8**, 176 (1968) [Sov. J. Nucl. Phys. **8**, 98 (1969)]; Yad. Fiz. **12**, 323 (1970) [Sov. J. Nucl. Phys. **12**, 178 (1971)].

[5] R. R. Levis and G. W. Ford, Phys. Rev. **107**, 756 (1957); S. Barshay and R. Behrends, Phys. Rev. **114**, 931 (1959); J. K. Knipp and G. E. Uhlenbeck, Physica (Amsterdam) **3**, 425 (1936); F. Bloch, Phys. Rev. **50**, 272 (1936); I. S. Batkin and Yu. G. Smirnov, Fiz. Élem. Chastits At. Yadra **11**, 1421 (1980) [Sov. J. Part. Nucl. **11**, 564 (1980)].

[6] A. Sirlin, Nucl. Phys. B **71**, 29 (1974); **100**, 291 (1975); **196**, 83 (1982); Rev. Mod. Phys. **50**, 573 (1978); Phys. Rev. D **22**, 971 (1980).

[7] F. Glück and K. Toth, Phys. Rev. D **46**, 2090 (1992); A. Garcia and W. Juarez, Phys. Rev. D **22**, 1132 (1980); A. Garcia, Phys. Rev. D **25**, 1348 (1982); **35**, 232 (1987); K. Toth, K. Szegö, and A. Margaritis, Phys. Rev. D **33**, 3306 (1986); F. Glück and K. Toth, Phys. Rev. D **41**, 2160 (1990); F. Glück, Phys. Rev. D **47**, 2840 (1993).

[8] G. G. Bunatian, in *Proceedings of ISINN-5, E3-97-213*, JINR, Dubna, Russia, 15–18 May, 1997, p. 102; G. G. Bunatyan, Yad. Fiz. **62**, 697 (1999) [Phys. At. Nucl. **62**, 1648 (1999)].

[9] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Relativistic Quantum Theory*, Parts 1 and 2, Pergamon Press, Oxford (1971–1974).

[10] E. D. Commins and P. H. Bucksbaum, *Weak Interactions of Leptons and Quarks*, Cambridge Univ. Press, Cambridge (1983); E. D. Commins, *Weak Interactions*, McGraw-Hill, New York (1973); P. H. Frampton and W. K. Tung, Phys. Rev. D **3**, 1114 (1971).

[11] C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, McGraw-Hill, New York (1981).

[12] *Review of Particle Properties*, Phys. Rev. D **50** 1177 (1994).

[13] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963); M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 625 (1973).

[14] S. M. Bilen'kiĭ, R. M. Ryndin, Ya. A. Smorodinskiĭ, and Ho Tso-Hsiu, Zh. Éksp. Teor. Fiz. **37**, 1758 (1959) [Sov. Phys. JETP **10**, 1241 (1960)].

[15] A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **30**, 96 (1956) [Sov. Phys. JETP **3**, 71 (1956)]; B. B. Sudakov, Zh. Éksp. Teor. Fiz. **30**, 87 (1956) [Sov. Phys. JETP **3**, 65 (1956)]; D. R. Yennie, S. C. Frautschi, and H. Suura, Ann. Phys. **13**, 379 (1961); N. Meister and D. R. Yennie, Phys. Rev. **130**, 1210 (1963); F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).

[16] Yu. A. Mostovoy, Yad. Fiz. **59**, 1013 (1996) [Phys. At. Nucl. **59**, 968 (1996)]; Yu. A. Mostovoy, Kurchatov Institute Preprint IAE-6040/2, Moscow (1997).

[17] A. P. Serebrov and N. V. Romanenko, JETP Lett. **55**, 503 (1992).

[18] B. R. Hostein and S. B. Treiman, Phys. Rev. D **16**, 2369 (1977); M. A. Beg, R. V. Budny, R. Mohapatra *et al.*, Phys. Rev. Lett. **38**, 1252 (1977).

[19] R. N. Mohapatra and D. P. Sidhu, Phys. Rev. Lett. **38**, 667 (1977); R. N. Mohapatra and J. C. Pati, Phys. Rev. D **11**, 566 (1975).

[20] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 1st ed., Pergamon Press, London; Addison-Wesley, Reading, MA (1958)

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Common vacuum conservation amplitude in the theory of the radiation of mirrors in two-dimensional space-time and of charges in four-dimensional space-time

V. I. Ritus*)

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 7 June 1999)
Zh. Éksp. Teor. Fiz. **116**, 1523–1538 (November 1999)

The changes in the action (and thus the vacuum conservation amplitudes) in the proper-time representation are found for an accelerated mirror interacting with scalar and spinor vacuum fields in $1+1$ space. They are shown to coincide to within a factor of e^2 with changes in the action of electric and scalar charges accelerated in $3+1$ space. This coincidence is attributed to the fact that the Bose and Fermi pairs emitted by a mirror have the same spins 1 and 0 as do the photons and scalar quanta emitted by charges. It is shown that the propagation of virtual pairs in $1+1$ space can be described by the causal Green's function $\Delta_f(z,\mu)$ of the wave equation for $3+1$ space. This is because the pairs can have any positive mass and their propagation function is represented by an integral of the causal propagation function of a massive particle in $1+1$ space over mass which coincides with $\Delta_f(z,\mu)$. In this integral the lower limit μ is chosen small, but nonzero, to eliminate the infrared divergence. It is shown that the real and imaginary parts of the change in the action are related by dispersion relations, in which a mass parameter serves as the dispersion variable. They are a consequence of the same relations for $\Delta_f(z,\mu)$. Therefore, the emergence of a real part in the change in the action is a direct consequence of causality, according to which $\mathrm{Re}\,\Delta_f(z,\mu)\neq 0$ only for timelike and lightlike intervals. © *1999 American Institute of Physics.* [S1063-7761(99)00211-5]

1. INTRODUCTION

An intriguing symmetry between the creation of particle pairs by an accelerated mirror in $1+1$ space and the emission of single quanta by a charge accelerated as a mirror in $3+1$ space was discovered in Refs. 1–3. This symmetry is confined to coincidence of the spectra of the Bose and Fermi pairs created by the mirror with the spectra of the photons and scalar quanta emitted by electric and scalar charges, if twice the frequencies ω and ω' of the quanta in a pair created by the mirror are identified with the components $k_\pm=k^0\pm k^1$ of the 4-wave vector k^α of the quantum emitted by the charge:

$$2\omega=k_+,\quad 2\omega'=k_-. \tag{1}$$

It was shown in Ref. 3 that the Bogolyubov coefficients $\beta^B_{\omega'\omega}$ and $\beta^F_{\omega'\omega}$, which describe the spectra of Bose and Fermi radiation of a mirror, are related to the Fourier transforms of the 4-current density $j_\alpha(k_+,k_-)$ and the scalar charge density $\rho(k_+,k_-)$, which describe the spectra of photons and scalar quanta emitted by charges:[1)]

$$\beta^{B*}_{\omega'\omega}=-\sqrt{\frac{k_+}{k_-}}\,\frac{j_-}{e}=\sqrt{\frac{k_-}{k_+}}\,\frac{j_+}{e}, \tag{2}$$

$$\beta^{F*}_{\omega'\omega}=\frac{1}{e}\rho(k_+,k_-). \tag{3}$$

It was also shown that $\beta^*_{\omega'\omega}$ is the source amplitude of a pair of particles which are only potentially emitted to the right and to the left with the frequencies ω and ω'. In other words, it is the virtual-pair creation amplitude. The pair becomes real when one of its particles undergoes internal reflection with a frequency change and both particles move in the same direction, i.e., to the right in the case of a right-sided mirror and to the left in the case of a left-sided mirror. Therefore, for a right-sided mirror, for example, the emission amplitude $\langle out\;\omega''\omega|in\rangle$ of a real pair of particles with the frequencies ω and ω'' is related to the virtual-pair creation amplitude $\beta^*_{\omega'\omega}$ by

$$\langle out\;\omega''\omega|in\rangle=-\sum_{\omega'}\langle out\;\omega''|\omega'\;in\rangle\beta^*_{\omega'\omega}, \tag{4}$$

where $\langle out\;\omega''|\omega'\;in\rangle$ is the amplitude of single-particle scattering on the mirror. The energy and momentum of this real pair equal $\omega+\omega''$ and $\omega+\omega''$, i.e., the pair does not have mass, nor do its components.

A virtual pair is another matter. According to (1), the zeroth and first components of the 4-momentum k^α of a quantum emitted by a charge are equal to the energy and momentum of a virtual pair of massless particles created by a mirror:

$$k^0=\omega+\omega',\quad k^1=\omega-\omega', \tag{5}$$

and form the timelike 2-momentum of the pair in $1+1$ space. Clearly, the quantity

$$m=\sqrt{k_+k_-}=2\sqrt{\omega\omega'}, \tag{6}$$

being an invariant of Lorentz transformations along axis 1, is the mass of the virtual pair, and at the same time it equals the transverse momentum $k_\perp=\sqrt{k_2^2+k_3^2}$ of the massless real quantum emitted by the charge.

The fact that the source amplitude $\beta^{B}_{\omega'\omega}$ of a virtual pair of bosons is specified by the current $j^{\alpha}(k_+,k_-)$, while the source amplitude $\beta^{F}_{\omega'\omega}$ of a virtual pair of fermions is specified by the scalar $\rho(k_+,k_-)$, means that the spin of a boson pair equals 1, while the spin of a fermion pair equals 0. Thus, the coincidence between the emission spectra of a mirror in $1+1$ space and charges in $3+1$ space can be attributed to the coincidence between the angular momentum of a pair emitted by the mirror and the spin of a particle emitted by the charge.[3]

The relation (2) can be written in the manifestly invariant form

$$e\beta^{B*}_{\omega'\omega}=\varepsilon_{\alpha\beta}k^{\alpha}j^{\beta}/\sqrt{k_+k_-}, \tag{7}$$

and, more specifically, in the form of the scalar product of the 2-vector current j^{β} and the 2-pseudovector polarization a_{β} of a Bose pair

$$a_{\beta}=\frac{\varepsilon_{\alpha\beta}k^{\alpha}}{\sqrt{k_+k_-}},\quad a_0=-\frac{k^1}{\sqrt{k_+k_-}},\quad a_1=\frac{k^0}{\sqrt{k_+k_-}}. \tag{8}$$

The spacelike pseudovector a_{β} is constructed from the zeroth and first components of the 4-momentum k^{α} of the quantum emitted by the mirror. It is orthogonal to the 2-momentum of the pair, has length 1, and can be represented in the comoving frame of the pair by just a spatial component, like the current vector j^{α}.

In this paper we find the vacuum conservation amplitude for acceleration of a mirror, which is defined by the change ΔW in the self-interaction of the mirror due to its acceleration. In essence, the problem here is finding $\mathrm{Re}\,\Delta W$ from the previously found quantity $\mathrm{Im}\,\Delta W$, twice the value of which coincides in a certain approximation (see below) with the mean number of real pairs formed by the mirror. Three different methods are used for this purpose.

The first (and principal) method is considered in Sec. 2, and involves transforming the original space-time representation for the mean number of pairs into an proper-time representation, whose kernel turns out to be the relativistically invariant singular even solution $(1/2)D^1(z)$ of the wave equation in $3+1$ space. Then, the function $D^1(z)$ in the expression obtained for the number of pairs is replaced by the even solution $\Delta^1(z,\mu)$ of the Klein–Gordon equation in order to invariantly and symmetrically eliminate the infrared divergence in the integral for the number of pairs using the small mass parameter μ instead of the large trajectory-length parameter L used in the original expression. The parameters μ, $L^{-1}\ll\kappa$, if κ is the characteristic acceleration on the trajectory. Finally, by treating the function $(1/2)\Delta^1(z,\mu)$ as the imaginary part of the kernel defining ΔW, we can reconstruct a relativistically invariant kernel which is even in z and coincides with the causal Green's function $\Delta_f(z,\mu)$ specific to $3+1$ space by finding the analytic continuation with respect to z^2. The resultant changes in the action of a mirror and a charge differ only by a factor of e^2, and the interactions are described by the same causal propagation function. Thus, the difference in dimensionality between the spaces is compensated by the difference in the mechanism of interaction transfer: it is mediated by pairs in $1+1$ space and by individual particles in $3+1$ space.

Section 3 presents a direct calculation of the changes $\Delta W^{B,F}_f$ in the self-interaction for a concrete, but fairly general mirror trajectory. The invariant functions of the relative velocity of the trajectory ends obtained for $\Delta W^{B,F}_f$ are consistent with the results of Sec. 2.

In Sec. 4 $\mathrm{Re}\,\Delta W_f$ is reconstructed from $\mathrm{Im}\,\Delta W_f$ using dispersion relations, in which μ appears as the dispersion variable. It is shown that the dispersion relations for ΔW_f are a consequence of the same relations for $\Delta_f(z,\mu)$ with a timelike z as the variable. As a consequence of causality only for such z, the values of $\mathrm{Re}\,\Delta_f$ and $\mathrm{Re}\,\Delta W_f$ are nonzero and are related to $\mathrm{Im}\,\Delta_f$ and $\mathrm{Im}\,\Delta W_f$, respectively, by the dispersion relations.

The fifth section examines other analytic continuations of $i\Delta^1/2$ onto the real z^2 axis that lead to kernels for ΔW whose real parts are not even in z.

A physical interpretation of the results is presented in the sixth, concluding section. The emergence of a causal propagation function characteristic of four-dimensional space-time in two-dimensional space-time can be attributed to mediation of the interaction by pairs of differing mass.

2. PROPER-TIME REPRESENTATION OF THE CHANGE IN THE ACTION

The following representations were obtained in Ref. 2 for the mean numbers of radiated Bose and Fermi particles:

$$N^{B,F}=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}du\,K^{B,F}(u), \tag{9}$$

$$K^{B}(u)=\fint_{-\infty}^{\infty}\frac{dv}{v-f(u)}\left[\frac{1}{g(v)-u}-\frac{f'(u)}{v-f(u)}\right], \tag{10}$$

$$K^{F}(u)=-\sqrt{f'(u)}\int_{-\infty}^{\infty}\frac{dv}{v-f(u)}\left[\frac{\sqrt{g'(v)}}{g(v)-u}-\frac{\sqrt{f'(u)}}{v-f(u)}\right]. \tag{11}$$

It follows from these representations for trajectories with the asymptotically constant velocities β_1 and β_2 at the ends and a nonzero Lorentz-invariant relative velocity

$$\beta_{21}=\frac{\beta_2-\beta_1}{1-\beta_2\beta_1},\quad \theta=\tanh^{-1}\beta_{21}, \tag{12}$$

that the mean number of massless quanta emitted is infinite (there is infrared divergence). In fact, in this case it follows from formulas (10) and (11) for $u\rightarrow\pm\infty$ (more precisely, for $|u|\gg\kappa^{-1}$, i.e., outside the region where the mirror experiences the characteristic acceleration κ) that the functions $K^{B,F}(u)$ exhibit universal behavior, which depends only on β_{21}:

$$K^{B}(u)\approx\pm\frac{1}{u}\left(\frac{\theta e^{\mp\theta}}{\sinh\theta}-1\right)=\pm\frac{1}{u}\left(\frac{\theta}{\tanh\theta}-1\right)-\frac{\theta}{u}, \tag{13}$$

$$K^{F}(u)\approx\pm\frac{1}{u}\left(1-\frac{\theta}{\sinh\theta}\right). \tag{14}$$

The coefficients of u^{-1} are relativistically invariant on the portions of the trajectories with asymptotically constant velocities. As a result, the mean number of quanta emitted on a portion of the trajectory covering the acceleration region grows logarithmically as the length $2L$ of that portion is increased:

$$N^B=\frac{1}{2\pi^2}\left(\frac{\theta}{\tanh\theta}-1\right)\ln(L\kappa)+2b^B(\theta),\tag{15}$$

$$N^F=\frac{1}{2\pi^2}\left(1-\frac{\theta}{\sinh\theta}\right)\ln(L\kappa)+2b^F(\theta),\quad L\kappa\gg 1.\tag{16}$$

Let us focus our attention on the fact that the odd (with respect to both u and θ) term in the asymptote of $K^B(u)$ does not contribute to the integral defining NB. The terms $2b^{B,F}$ do not depend on L if $L\kappa\gg 1$, but they can depend on the specific form of the trajectories.

We note that there are representations for $N^{B,F}$ which differ from (9)–(11) by mirror symmetry, i.e., by the replacements $u\rightleftarrows v$ and $f(u)\rightleftarrows g(v)$. The integrands $K^{B,F}(v)$ defining them differ from $K^{B,F}(u)$, but are denoted below by the same letter, since they are values of the same functional taken for a mirror-symmetrical pair of trajectories: $K(u)\equiv K[u;g]$ and $K(v)\equiv K[v;f]$. As $v\rightarrow\pm\infty$, $K^{B,F}(v)$ have asymptotes which differ from (13) and (14) by the replacements $u\rightarrow v$ and $\theta\rightarrow-\theta$.

The vacuum conservation amplitude of an accelerated mirror is specified by the change in the action $\Delta W=W|_0^F$ (i.e., the difference between the actions for the accelerated and unaccelerated mirror) and has the form $\exp(i\Delta W)$, where $2\,\mathrm{Im}\,\Delta W=N$, if the interference effects in the creation of two or more pairs are neglected. Here we consider a particle and an antiparticle to be nonidentical; otherwise, in the same approximation $2\,\mathrm{Im}\,\Delta W=(1/2)N$ (see Ref. 3).

Now the main task is to find $\mathrm{Re}\,\Delta W$. For this purpose, we obtain a suitable representation for $\mathrm{Im}\,\Delta W$ and utilize relativistic invariance and causality arguments.

Let us consider the space-time representation for N which was the direct ‘‘parent’’ of the representation (9)–(11) [see Ref. 2]. In this representation

$$N^B=\int_{-\infty}^{\infty}\int du\,dv\,S(u,v)|_0^F,$$

$$S(u,v)=\frac{1}{8\pi^2}\left[\frac{1}{(v-f(u)-i\varepsilon)(g(v)-u-i\delta)}+\text{c.c.}\right].\tag{17}$$

We go over from the independent characteristic variables u and v to the moments in proper time τ and τ' of two points on the world trajectory of the mirror $x^\alpha(\tau)$:

$$u=x^0(\tau)-x^1(\tau)=x_-(\tau),$$

$$v=x^0(\tau')+x^1(\tau')=x_+(\tau').\tag{18}$$

Then

$$f(u)=x^0(\tau)+x^1(\tau)=x_+(\tau),$$

$$g(v)=x^0(\tau')-x^1(\tau')=x_-(\tau'),\tag{19}$$

and $S(u,v)$ becomes a relativistically invariant function of the two-dimensional vector $z^\alpha=x^\alpha(\tau)-x^\alpha(\tau')\equiv(x-x')^\alpha$ joining the points $x^\alpha=x^\alpha(\tau)$ and $x'^\alpha=x^\alpha(\tau')$ on the mirror trajectory:

$$\begin{aligned}S(z)&=\frac{1}{8\pi^2}\left[\frac{1}{(x'_+-x_+-i\varepsilon)(x'_--x_--i\delta)}+\text{c.c.}\right]\\&=\frac{1}{8\pi^2}\left[\frac{1}{z_+z_-+i\varepsilon\,\mathrm{sgn}\,z^0}+\text{c.c.}\right]\\&=\frac{1}{8\pi^2}\left[\frac{1}{-z^2+i\varepsilon\,\mathrm{sgn}\,z^0}+\text{c.c.}\right]=-P\frac{1}{4\pi^2z^2}.\end{aligned}\tag{20}$$

The individual terms in (20) and their sum are well-known relativistically invariant singular functions in quantum electrodynamics (we use Thirring's notation,[4] but our D^1 and Δ^1 lack the factor i):

$$\begin{aligned}D^{\pm}(z)&=\frac{\pm i}{4\pi^2(z^2\pm i\varepsilon\,\mathrm{sgn}\,z^0)}\\&=\frac{1}{4\pi^2}\left[\pi\varepsilon(z^0)\,\delta(z^2)\pm\frac{i}{z^2}\right],\end{aligned}\tag{21}$$

$$D^1(z)=\frac{1}{2\pi^2z^2},$$

so that

$$S(z)=-\frac{i}{2}[D^-(z)-D^+(z)]=-\frac{1}{2}D^1(z).\tag{22}$$

We stress that these functions are singular solutions of the wave equation in $3+1$ space, if z^α is construed to be a four-dimensional, rather than a two-dimensional, vector. Here the presence of these functions, which depend on the 2-vector z^α, is the result of a deep symmetry between the creation of a pair by a mirror in $1+1$ space and the emission of single quanta by a charge in $3+1$ space.

Using

$$\begin{aligned}du\,dv&=d\tau\,d\tau'\dot{x}_-\dot{x}'_+=d\tau\,d\tau'\left[\frac{1}{2}(\dot{x}_-\dot{x}'_++\dot{x}_+\dot{x}'_-)\right.\\&\quad\left.+\frac{1}{2}(\dot{x}_-\dot{x}'_+-\dot{x}_+\dot{x}'_-)\right]\\&=d\tau\,d\tau'(-\dot{x}_\alpha\dot{x}'^\alpha+\varepsilon_{\alpha\beta}\dot{x}^\alpha\dot{x}'^\beta)\end{aligned}\tag{23}$$

in the form of a sum of terms which are even and odd with respect to the interchange $\tau\rightleftarrows\tau'$ (a dot denotes differentiation with respect to the proper time), we obtain

$$N^B=\int_{-\infty}^{\infty}\int d\tau\,d\tau'(\dot{x}_\alpha\dot{x}'^\alpha-\varepsilon_{\alpha\beta}\dot{x}^\alpha\dot{x}'^\beta)\frac{1}{2}D^1(z)\Big|_0^F.\tag{24}$$

It is natural to use a manifestly relativistic method that preserves the symmetry under the interchange $\tau\rightleftarrows\tau'$ to eliminate the infrared divergence in (24). It consists of replacing $D^1(z)$ by $\Delta^1(z,\mu)$, which is also even in z and has the small mass parameter $\mu\ll\kappa$, where κ is the characteristic acceleration of the mirror.

This function

$$\frac{1}{2}\Delta^1(z,\mu)=\frac{\mu}{8\pi s}N_1(\mu s)$$

$$=-\frac{1}{4\pi^2 s^2}-\frac{\mu}{4\pi^2 s}J_1(\mu s)\ln\frac{2}{\mu s}+R \tag{25}$$

(where J_1 and N_1 are Bessel and Neumann functions, and R is a regular function of s) is a singular solution of the wave equation in $3+1$ space, which depends only on the interval $s=\sqrt{-z^2}$ between the two points and preserves all the features with respect to s at $s=0$. It is called the Hadamard elementary function or the fundamental solution.[5] The coefficient of the logarithm, which is called the Riemann function,[5] is a regular function of s, which satisfies the same equation as Δ^1. Just these two functions define the imaginary and real parts of the change in the action.

Thus,

$$N^B=\int_{-\infty}^{\infty}\int d\tau\, d\tau'(\dot{x}_\alpha \dot{x}'^\alpha-\varepsilon_{\alpha\beta}\dot{x}^\alpha\dot{x}'^\beta)\frac{1}{2}\Delta^1(z,\mu)\Bigg|_0^F. \tag{26}$$

In the expressions for N^B the odd term is insignificant, since $D^1(z)$ and $\Delta^1(z,\mu)$ are even under the replacement $z\to -z$.

Now regarding N as the imaginary part of twice the action, we naturally consider $(1/2)\Delta^1(z,\mu)$ to be the imaginary part of some function $F(z^2)$, which is taken on the real z^2 axis and which is analytic in the z^2 complex plane with a cut along the $z^2\leqslant 0$ semiaxis, where Lorentz invariance allows it to still depend on the sign of z^0, and coincides with $(i/2)\Delta^1(z,\mu)$ at $z^2>0$. Then the transition from iN^B to $2W$ is equivalent to the analytic continuation of $F(z^2)$ onto the real semiaxis $z^2\leqslant 0$. It is generally known[6] that the boundary value of such a function, which does not depend on the sign of z^0, and is therefore even, is the limit from above ($\varepsilon\to +0$), which is called a causal function:

$$\Delta_f(z,\mu)=F(z^2+i\varepsilon)=\frac{\mu}{4\pi^2 s}K_1(i\mu s)$$

$$=\frac{1}{4\pi}\delta(s^2)-\frac{\mu}{8\pi s}$$

$$\times[J_1(\mu s)-iN_1(\mu s)]. \tag{27}$$

Here K_1 is the modified Bessel function of the second kind, and $s=\sqrt{-z^2-i\varepsilon}$. The latter equality was written for $z^2\leqslant 0$, where $s\geqslant 0$ and Δ_f has a real part, which coincides with the Riemann function multiplied by $\pi/2$. If $z^2>0$, then $s=-i\sqrt{z^2}$, Δ_f is purely imaginary, and its imaginary part is positive.

Thus, for ΔW_f^B we obtain

$$\Delta W_f^B=\frac{1}{2}\int\int d\tau\, d\tau'\dot{x}_\alpha(\tau)\dot{x}^\alpha(\tau')\Delta_f(z,\mu)\Bigg|_0^F. \tag{28}$$

As shown in Ref. 2, the space-time representation for N^F differs from the representation (17) for N^B by the additional factor $-\sqrt{f'(u)g'(v)}$ under the integral. Therefore, after the replacement of variables (18), instead of (23) we have

$$-du\, dv\sqrt{f'(u)g'(v)}=-d\tau\, d\tau'. \tag{29}$$

Then

$$N^F=\frac{1}{2}\int\int d\tau\, d\tau'\Delta^1(z,\mu)\Bigg|_0^F, \tag{30}$$

and the change in the action is

$$\Delta W_f^F=\frac{1}{2}\int\int d\tau\, d\tau'\Delta_f(z,\mu)\Bigg|_0^F. \tag{31}$$

The proper-time representations obtained for $\Delta W_f^{B,F}$ differ from the changes ΔW_1 and ΔW_0 in the self-interactions of electric and scalar charges moving along the same trajectories as the mirror, but in $3+1$ space, only by the absence of the factor e^2.

As $\mu\to 0$, the coefficients of $\ln\mu^{-1}$ in the imaginary parts of the proper-time integrals (28) and (31) should coincide with the coefficients of $\ln L$ in the corresponding expressions for N^B and N^F [see (15) and (16)], since these coefficients cannot depend on the method used to eliminate the infrared divergence in the different representations for N^B and for N^F.

Since for an interval between two points on the timelike trajectory

$$\operatorname{Re}\Delta_f(z,\mu)=-\frac{\mu}{8\pi s}J_1(\mu s), \tag{32}$$

which differs from the coefficient of the logarithm in $\operatorname{Im}\Delta_f$ only by the factor $\pi/2$ [see (27) and (25)], $\operatorname{Re}\Delta W_f$ also differs by the same factor from the coefficient of $\ln\mu^{-1}$ in $\operatorname{Im}\Delta W_f$. Thus, to within terms which vanish at $\mu\to 0$, we have

$$\Delta W_f=\pi a(\theta)+i\left[a(\theta)\ln\frac{\kappa^2}{\mu^2}+b(\theta)\right], \tag{33}$$

$$a^B(\theta)=\frac{1}{8\pi^2}\left(\frac{\theta}{\tanh\theta}-1\right),$$

$$a^F(\theta)=\frac{1}{8\pi^2}\left(1-\frac{\theta}{\sinh\theta}\right). \tag{34}$$

The function $b(\theta)$ can depend on other dimensionless parameters, for example, the velocity changes on portions of the trajectory containing other extreme values of the self-acceleration.

It is significant that $\operatorname{Re}\Delta W_f=\pi a$ has a finite positive limit at $\mu\to 0$, which is positive for $\theta\neq 0$.

To conclude this section we recall that $\operatorname{Re}\Delta W_f$ is the acceleration-induced self-energy shift of the source integrated over the proper time and that $2\operatorname{Im}\Delta W_f$ is the mean number of pairs emitted (or particles emitted in the case of nonidentity to the antiparticles). More precisely, $\exp(-2\operatorname{Im}\Delta W_f)$ is the probability of the noncreation of pairs during all the time of acceleration.

3. CHANGE IN THE ACTION IN THE CASE OF QUASIHYPERBOLIC MIRROR MOTION

It would be interesting to directly calculate $\Delta W_f^{B,F}$ for the special, but very important mirror trajectory

$$x=\xi(t)=v_\infty\sqrt{\frac{v_\infty^2}{\kappa^2}+t^2}, \tag{35}$$

which can be called quasihyperbolic. Here $\pm v_\infty$ are the velocities of the mirror at $t\to\pm\infty$, and κ is its acceleration at the turning point $(t=0)$. This motion is remarkable in that as $v_\infty\to 1$, it becomes increasingly close to uniformly accelerated (hyperbolic) motion over the increasingly longer time interval

$$|t|\lesssim t_1=\frac{v_\infty}{\kappa}(1-v_\infty^2)^{-1/2},$$

smoothly going over to uniform motion outside this interval. This can be seen from the expression for the magnitude of the acceleration in the comoving frame

$$a=\kappa\left(1+\frac{t^2}{t_1^2}\right)^{-3/2}.$$

The spectrum and total radiated energy were found for an electric charge moving along the trajectory (35) in Ref. 7.

To calculate ΔW^B, in (28) instead of t we use the variable u, which is defined by the formula

$$t=\frac{v_\infty}{\kappa}\sinh u.$$

Then

$$d\tau\, d\tau'\, \dot{x}_\alpha(\tau)\dot{x}^\alpha(\tau')=-dx\, dy\,\frac{v_\infty^2}{\kappa^2}\left(\frac{1+v_\infty^2}{2}\cosh x+\frac{1-v_\infty^2}{2}\cosh 2y\right), \tag{36}$$

$$(x-x')^2=-2\frac{v_\infty^2}{\kappa^2}(\cosh x-1)[v_\infty^2+(1-v_\infty^2)\cosh^2 y],$$

$$x=u-u',\quad y=\frac{u+u'}{2},$$

and ΔW^B can be expressed in terms of an integral of modified Bessel functions

$$\Delta W^B=-\frac{1}{32\pi^2}\int_0^\infty\frac{d\xi}{\xi^2}e^{-i\mu^2\xi}\int_{-\infty}^\infty dy\,\frac{v_\infty^2}{\kappa^2}e^{iz}[(1+v_\infty^2)\times K_1(iz)+(1-v_\infty^2)K_0(iz)\cosh 2y]_0^F, \tag{37}$$

if we use the representation

$$\Delta_f(x-x',\mu)=\frac{1}{16\pi^2}\int_0^\infty\frac{d\xi}{\xi^2}\exp\left[i\frac{(x-x')^2}{4\xi}-i\mu^2\xi\right] \tag{38}$$

for the causal function and introduce the notation

$$z=\frac{v_\infty^2}{2\kappa^2\xi}[v_\infty^2+(1-v_\infty^2)\cosh^2 y]. \tag{39}$$

Now switching from the integration variable ξ to z in (37), we obtain

$$\Delta W^B=-\frac{1}{8\pi^2}\int_{-\infty}^\infty dy\left\{\frac{1+v_\infty^2}{2Q}[S_1(\Lambda)+S_0(\Lambda)]-S_0(\Lambda)\right\}, \tag{40}$$

where

$$\Lambda=\lambda v_\infty^2 Q,\quad \lambda=\frac{\mu^2}{\kappa^2},\quad Q=v_\infty^2+(1-v_\infty^2)\cosh^2 y, \tag{41}$$

$$S_n(\Lambda)=(-1)^{n+1}\int_0^\infty dz\, e^{-i\Lambda/2z}\left[e^{iz}K_n(iz)-\sqrt{\frac{\pi}{2iz}}\right]. \tag{42}$$

The subtraction $|_0^F$ in (40) reduced to subtraction of the asymptote $(\pi/2iz)^{1/2}$ of the integrand in (42) for $S_n(\Lambda)$. As shown in Refs. 8 and 9, the functions $S_n(\Lambda)$ can be expressed in terms of the product of the modified Bessel functions $I_n(\sqrt{\Lambda})$ and $K_n(\sqrt{\Lambda})$. We also turn attention to the more compact expressions for the derivatives

$$S_n'(\Lambda)=(-1)^n\pi\left[I_n(x)K_n(x)-\frac{1}{2x}\right]+iK_n^2(x),$$

$$x=\sqrt{\Lambda}. \tag{43}$$

It can be seen from formulas (40)–(42) that ΔW^B depends on two dimensionless parameters, namely λ and $v_\infty=\tanh(\theta/2)$.

To calculate the asymptote of the integral (40) at $\lambda\to 0$ we note that the values $\Lambda\to 0$ will be effective in the first term in this case, and therefore

$$S_1(\Lambda)+S_0(\Lambda)\approx-\pi-i\ln\frac{4}{\gamma^2\Lambda},\quad \gamma=1.781\ldots, \tag{44}$$

and that in the second term the integral can be reduced to the expression

$$\int_{-\infty}^\infty dy\, S_0(\Lambda)\approx-\int_0^\infty d\Lambda\, S_0'(\Lambda)\ln\Lambda+S_0(0)\ln\frac{4}{\lambda v_\infty^2(1-v_\infty^2)}=-\pi-i\left[\ln\frac{16}{\gamma^2 v_\infty^2(1-v_\infty^2)\lambda}-2\right]. \tag{45}$$

As a result, to within terms that vanish at $\lambda\to 0$ we obtain

$$\Delta W^B\approx\frac{1}{8\pi^2}\left\{\pi\left(\frac{\theta}{\tanh\theta}-2\right)+i\left[\left(\frac{\theta}{\tanh\theta}-1\right)\times\ln\left[\frac{8(\cosh\theta+1)^2}{\gamma^2\lambda(\cosh\theta-1)}\right]+2-\frac{L_2(1-e^{-2\theta})+\theta^2}{\tanh\theta}\right]\right\}. \tag{46}$$

Here $\theta=\tanh^{-1}\beta_{21}=2\tanh^{-1}v_\infty$, and $L_2(x)$ is the Euler dilogarithm.[10,11]

For a quasiuniformly accelerated mirror interacting with a spinor field, instead of (40) we obtain

$$\Delta W^F = \frac{1}{8\pi^2}\int_{-\infty}^{\infty} dy \left\{ \int_0^{\infty} dz \times \exp\left(-\frac{i\Lambda}{2z} + iz \right) R(iz) - S_0(\Lambda) \right\}, \tag{47}$$

where

$$R(iz) = \int_0^{\infty} dx \left(\sqrt{\frac{(\cosh x + c)^2 - s^2}{(1+c)^2 - s^2}} - 1 \right) \times \exp(-iz\cosh x), \tag{48}$$

$$c = \cosh\theta\cosh 2y, \quad s = \sinh\theta\sinh 2y,$$

$$\theta = 2\tanh^{-1} v_\infty,$$

and the remaining notation is the same as in (40). It is seen that ΔW^F depends on the two dimensionless parameters λ and θ.

When $\lambda \to 0$, the expression in the large parentheses in (48) can be replaced by

$$\frac{\cosh x - 1}{\sqrt{(1+c)^2 - s^2}}.$$

This approximation holds for $\cosh x \gg 1$ and has the correct (zero) value at $x = 0$. Then

$$\Delta W^F \approx \frac{1}{8\pi^2}\int_{-\infty}^{\infty} dy \left\{ \frac{1}{\sqrt{(1+c)^2 - s^2}} [S_1(\Lambda) + S_0(\Lambda)] - S_0(\Lambda) \right\}, \tag{49}$$

and using (44) and (45), we obtain

$$\Delta W^F \approx \frac{1}{8\pi^2}\left\{ \pi\left(1 - \frac{\theta}{\sinh\theta}\right) + i\left[\left(1 - \frac{\theta}{\sinh\theta}\right) \times \ln\left[\frac{8(\cosh\theta + 1)^2}{\gamma^2\lambda(\cosh\theta - 1)} \right] - 2 + \frac{L_2(1 - e^{-2\theta}) + \theta^2}{\sinh\theta} \right] \right\} \tag{50}$$

to within terms that vanish at $\lambda \to 0$.

The formulas obtained for $\Delta W^{B,F}$ not only have the structure (33), but also contain explicit expressions for $b^{B,F}(\theta)$. It can also be seen that $\Delta W^{B,F}$ do not depend on the sign of θ or β_{21}, if we take into account that $L_2(1 - e^{-2\theta}) + \theta^2$ is an odd function of θ [see Landen's formula (1.12) in Ref. 11]. We note in this connection that for small values of θ

$$L_2(1 - e^{-2\theta}) + \theta^2 = 2\theta + \frac{2}{9}\theta^3 - \frac{2}{225}\theta^5 + \ldots, \tag{51}$$

and that as $\theta \to \pm\infty$, to within exponentially small terms we have

$$L_2(1 - e^{-2\theta}) + \theta^2 = \pm\left(\theta^2 + \frac{\pi^2}{6} \right) + \ldots. \tag{52}$$

The imaginary and real parts of $\Delta W^{B,F}$ in (46) and (50) are positive owing to unitarity and causality. When $\theta = 0$, ΔW vanishes, since the quasihyperbolic trajectory becomes a straight line.

The point $\theta = \infty$ for $\Delta W_f^{B,F}(\theta,\lambda)$ is an essential singularity. It physically corresponds to a purely hyperbolic trajectory for which $\beta_{21} = 1$ or -1, according to the sign of κ. At a fixed value of λ and $\theta \to \pm\infty$, from (40) and (47) we obtain

$$\Delta W_f^{B,F}(\theta,\lambda) = \mp\theta\frac{1}{8\pi^2} S_{1,0}(\lambda). \tag{53}$$

Here $\pm\theta = |\kappa|(\tau_2 - \tau_1) \gg 1$, and when the interval (τ_1, τ_2) of proper time approaches infinite length, the relative velocity β_{21} approaches $+1$ or -1. Formula (53) was obtained for uniformly accelerated charges in $3+1$ space in Ref. 12 and was discussed in detail in Refs. 8 and 9. In those studies it defined the classical mass shift of a uniformly accelerated charge:

$$\Delta m_{1,0} = -\frac{\partial \Delta W_{1,0}}{\partial \tau_2} = \frac{\alpha}{2\pi}|\kappa| S_{1,0}(\lambda). \tag{54}$$

In accordance with unitarity and causality of this function, the imaginary and real parts of Δm are negative. At $\kappa = 0$ the function $\Delta m(\kappa)$ is nonanalytic, and therefore cannot be reproduced by perturbation theory with respect to κ or with respect to the field accelerating the charge.

4. DISPERSION RELATIONS FOR ΔW AND THEIR ORIGIN

It was shown in Ref. 9 that the changes $\Delta W_s(\mu^2)$ in the action of point charges moving along timelike trajectories, as functions of the square of the mass of quanta of their self-field with spin $s = 1,0$, are analytic in the μ^2 complex plane with a cut along the positive μ^2 semiaxis, on whose edges the imaginary parts of each of the functions coincide, while the real parts differ in sign. Such functions satisfy the dispersion representations ($\mathrm{Im}\,\mu < 0$)

$$\Delta W(\mu^2) = \frac{2i}{\pi}\int_0^{\infty} \frac{dx\, x\, \mathrm{Re}\,\Delta W(x^2)}{x^2 - \mu^2} = -\frac{2\mu}{\pi}\int_0^{\infty} \frac{dx\, \mathrm{Im}\,\Delta W(x^2)}{x^2 - \mu^2}, \tag{55}$$

which reconstruct the function $\Delta W_s(\mu^2)$ in the μ^2 complex plane from its real or imaginary part on the lower edge of the cut. When $\mu = i\kappa$ and $\kappa > 0$, these relations yield the important equalities

$$\frac{2}{\pi}\int_0^{\infty} \frac{dx\, x\, \mathrm{Re}\,\Delta W(x^2)}{x^2 + \kappa^2} = \frac{2\kappa}{\pi}\int_0^{\infty} \frac{dx\, \mathrm{Im}\,\Delta W(x^2)}{x^2 + \kappa^2} = \mathrm{Im}\,\Delta W(-\kappa^2) > 0, \tag{56}$$

$$\mathrm{Re}\,\Delta W(-\kappa^2) = 0. \tag{57}$$

As a consequence of unitarity, $\mathrm{Im}\,\Delta W(\mu^2)$ is positive on the real semiaxis $\mu^2 > 0$. Then, according to the second of the representations (55), $\operatorname{Im}\Delta W(\mu^2)$ is positive definite over the entire μ^2 complex plane (or in the lower μ half-plane).

Here we show that the dispersion relations for $\Delta W(\mu^2)$ are due to the analytic properties of the causal Green's function $\Delta_f(z,\mu)$, which, as we see, specifies not only $\Delta W_s(\mu^2)$ for the vacuum amplitude of accelerated charges in $3+1$ space, but also $\Delta W^{B,F}(\mu^2)$ for the vacuum amplitude of an accelerated mirror in $1+1$ space.

We can show that the causal function $\Delta_f(z,\mu)$ for a timelike z satisfies the dispersion relations presented. According to formulas (2.12.4.28) and (2.13.3.20) from Ref. 13

$$\int_0^\infty \frac{dx\, x^2 J_1(sx)}{x^2+\kappa^2} = -\kappa \int_0^\infty \frac{dx\, x N_1(sx)}{x^2+\kappa^2} = \kappa K_1(s\kappa), \tag{58}$$

where s, $\operatorname{Re}\kappa>0$. After analytic continuation in κ to the point $\kappa=i\mu+\varepsilon$ where $\mu>0$ and $\varepsilon\to+0$, these relations transform into

$$\int_0^\infty \frac{dx\, x^2 J_1(sx)}{x^2-\mu^2+i\varepsilon} = -i\mu \int_0^\infty \frac{dx\, x N_1(sx)}{x^2-\mu^2+i\varepsilon} = i\mu K_1(i\mu s)$$

$$= -\frac{i\pi\mu}{2}[J_1(\mu s)-iN_1(\mu s)]. \tag{59}$$

After multiplying by $-i/4\pi^2 s$, they form the first pair of dispersion relations (55), which, instead of $\Delta W(\mu^2)$, contain the causal function (27) with a timelike vector z^α, for which $s=\sqrt{-z^2}>0$. For spacelike z^α the interval $s=-i\sqrt{z^2}$, and $\Delta_f(z,\mu)$ is purely imaginary.

After multiplication by $-1/4\pi^2 s$, the original formulas (58) coincide with the second pair of the relations (56) with the replacement of $\Delta W(\mu^2)$ by $\Delta_f(z,\mu)$. The function appearing on the right-hand side of these relations

$$-\frac{\kappa}{4\pi^2 s}K_1(\kappa s) = \operatorname{Im}\Delta_f(z,-i\kappa), \tag{60}$$

unlike $\operatorname{Im}\Delta W(-\kappa^2)$, is negative. In addition,

$$\operatorname{Re}\Delta_f(z,-i\kappa)=0, \tag{61}$$

as can be seen from (27). This property is a consequence of causality, according to which $\operatorname{Re}\Delta_f(z,\mu)=0$ outside the light cone, i.e., for spacelike z^α. In this case the argument of the Bessel function in (27) is real and positive. When we go over to timelike z^α and a purely imaginary negative $\mu=-i\kappa$, this argument remains real and positive, whence follows (61).

While satisfying the dispersion relations (55) and (56) with respect to the "dispersion" variable μ, the function $\Delta_f(z,\mu)$, unlike $\Delta W(\mu^2)$, still depends on the parameter s, which equals the invariant interval between the two points chosen on the mirror trajectory with the proper times τ and τ', i.e., on $s=s(\tau,\tau')$. Integrating the dispersion relations for Δ_f over τ, τ' with the weight $(1/2)\dot{x}_\alpha(\tau)\dot{x}^\alpha(\tau')$ or 1/2 and performing the subtraction procedure, we obtain the dispersion relations for ΔW^B or ΔW^F, if, of course, the familiar conditions for changing the order of integration over x and τ, τ' are satisfied.

Thus, the dispersion relations for $\Delta W(\mu^2)$ are a consequence of the dispersion relations for $\Delta_f(z,\mu)$.

It follows from (56) that if $\operatorname{Im}\Delta W(\mu^2)$ is bounded at zero, then $\operatorname{Re}\Delta W(\mu^2)$ must vanish as $\mu\to+0$. If, on the other hand, $\operatorname{Im}\Delta W(\mu^2)$ logarithmically tends to infinity as $\mu\to0$ according to the relation

$$\operatorname{Im}\Delta W(\mu^2) = a\ln\mu^{-2}+b(\mu^2) \tag{62}$$

[$a>0$, and $b(\mu^2)$ is bounded at zero], it follows from (56) that $\operatorname{Re}\Delta W(\mu^2)$ tends to the positive value $\operatorname{Re}\Delta W(0)=\pi a$ as $\mu\to0$. According to (57), this means that the function $\operatorname{Re}\Delta W(\mu^2)$ has a discontinuity equal to πa on the real μ^2 axis at $\mu^2=0$.

5. INFLUENCE OF THE BOUNDARY CONDITIONS ON Re ΔW

Let us now consider the other boundary values of $F(z^2)$, which is analytic in the z^2 complex plane with a cut along the $z^2\leqslant0$ semiaxis and coincides with $(i/2)\Delta^1(z,\mu)$ on the $z^2>0$ semiaxis.

The limit $F(z^2-i\varepsilon)$ from below on the real axis is distinguished from the limit (27) from above by the opposite sign of the real part. According to this function, free fields would transport negative energy in $3+1$ space; therefore, this boundary condition is not considered here.

The other boundary values of $F(z^2)$, which already depend on the sign of z^0, may be the limits $F(z^2\pm i\varepsilon\,\operatorname{sgn} z^0)$, $\varepsilon\to+0$. They are positive- and negative-frequency functions, or more precisely, $\pm\Delta^\pm(z,\mu)$ (Ref. 4):

$$\pm\Delta^\pm(z,\mu) = \pm\varepsilon(z^0)\operatorname{Re}\Delta_f + i\operatorname{Im}\Delta_f. \tag{63}$$

Such functions naturally vary only the real part of the action obtained for Δ_f, so that

$$\operatorname{Re}\Delta W^B_\pm = \pm\frac{1}{2}\int\int d\tau\, d\tau' (\dot{x}^\alpha \dot{x}'^\alpha - \varepsilon_{\alpha\beta}\dot{x}^\alpha \dot{x}'^\beta)\operatorname{Re}\Delta^\pm(z,\mu)\Big|_0^F \tag{64}$$

differ from $\operatorname{Re}\Delta W^B_f$ and are given in the limit $\mu\to0$ by the expressions

$$\operatorname{Re}\Delta W^B_\pm = \mp\frac{1}{8\pi}\int\int d\tau\, d\tau'\, \varepsilon_{\alpha\beta}\dot{x}^\alpha \dot{x}'^\beta \varepsilon(z^0)\,\delta(z^2). \tag{65}$$

The integrand can be expanded in τ' near $\tau'=\tau$ and represented in the form

$$\varepsilon_{\alpha\beta}\dot{x}^\alpha \dot{x}'^\beta \varepsilon(z^0)\,\delta(z^2) = -\varepsilon_{\alpha\beta}\dot{x}^\alpha \ddot{x}^\beta \delta(\tau-\tau'). \tag{66}$$

Here the equality $|x|\,\delta(x^2)=\delta(x)$ was used (see, for example, Ref. 14).

Then, integrating over τ' and expressing the self-acceleration

$$a(\tau) = \varepsilon_{\alpha\beta}\dot{x}^\alpha \ddot{x}^\beta \frac{f''}{2(f')^{3/2}} = \frac{d\ln f'(u)}{2d\tau} = \frac{d\tanh^{-1}\beta(\tau)}{d\tau} \tag{67}$$

in the form of the derivative of the rapidity with respect to the proper time, we obtain

$$\mathrm{Re}\,\Delta W_{\pm}^{B}=\pm\frac{1}{8\pi}\int_{-\infty}^{\infty}d\tau\,\varepsilon_{\alpha\beta}\dot{x}^{\alpha}\ddot{x}^{\beta}$$

$$=\pm\frac{1}{8\pi}\tanh^{-1}\beta_{21}=\pm\frac{\theta}{8\pi}. \tag{68}$$

Clearly,

$$\mathrm{Re}\,\Delta W_{\pm}^{F}=\pm\frac{1}{2}\int\int d\tau\,d\tau'\,\mathrm{Re}\,\Delta^{\pm}(z,\mu)=0 \tag{69}$$

because of the oddness of $\mathrm{Re}\,\Delta^{\pm}$ with respect to z.

The expressions obtained for $\mathrm{Re}\,\Delta W_{\mp}$ coincide to within the factor $(8\pi)^{-1}$ with the coefficients that are odd in θ for the terms proportional to u^{-1} and v^{-1} in the asymptotic expansions of $K(u)$ and $K(v)$, respectively [see (13) and (14) and the comment following Eq. (16)]. At the same time, $\mathrm{Re}\,\Delta W_f$ coincides to within the factor $(8\pi)^{-1}$ with the coefficient that is even in θ for the term proportional to u^{-1} or v^{-1} in the asymptotic expansion of $K(u)$ or $K(v)$. We note that all these coefficients, as well as the functions $K(u)$ and $K(v)$ themselves, are formed without any involvement of the parameter L, which eliminates the infrared divergence of the space-time integrals (9) for the mean number of particles emitted.

Thus, information on the interaction contained in $K(u)$ and $K(v)$, which determine $\mathrm{Im}\,\Delta W$, is conveyed to $\mathrm{Re}\,\Delta W$ owing to causality and the boundary conditions. In addition, $\mathrm{Re}\,\Delta W_f$ contains information on the interaction that propagates within the light cone, and $\mathrm{Re}\,\Delta W_{\pm}$ contains information on the interaction that propagates along the light cone and is therefore local owing to the timelike character of the trajectory.

Half the sum of the retarded and advanced fields is the self-field of the source,[4] and half their difference is the radiation field escaping to infinity. Since

$$\mathrm{Re}\,\Delta_f=\frac{1}{2}(\Delta^{\mathrm{ret}}+\Delta^{\mathrm{adv}}),$$

and

$$\mathrm{Re}\,\Delta^{+}=\frac{1}{2}(\Delta^{\mathrm{ret}}-\Delta^{\mathrm{adv}}),$$

$\mathrm{Re}\,\Delta W_f$ describes the self-energy shift of the source, and $\mathrm{Re}\,\Delta W_{+}$ describes the interaction with the radiation field, i.e., with real quanta. The boundary condition which eliminates the interaction with virtual quanta or pairs seems unnatural.

6. DISCUSSION AND PHYSICAL INTERPRETATION OF RESULTS

The proper-time representations for the changes in the self-interaction of a mirror upon acceleration in a two-dimensional vacuum of scalar and spinor fields can be considered the most significant results of this work. These representations coincided with the representations of the changes in the self-interaction of electric and scalar charges accelerated in four-dimensional space-time. In other words, the two were found to be identical functionals of the source trajectory.

This coincidence, first, confirms the interpretation given in Ref. 3 of the Bogolyubov coefficient $\beta^{*}_{\omega'\omega}$ as the source amplitude of a virtual pair of particles potentially emitted to the right and to the left with the frequencies ω and ω', with the timelike 2-momentum of the pair (5), the mass $m=2\sqrt{\omega\omega'}$, and a spin equal to 1 for a boson pair and 0 for a fermion pair.

Second, it means that the self-interaction of the mirror involves the creation and absorption of virtual pairs, rather than individual particles, and is transmitted from one point of the trajectory to another by the causal Green's function of the wave equation for four-dimensional, rather than two-dimensional space-time.

The action integral is formed by virtual pairs with mass $m=2\sqrt{\omega\omega'}$, which takes any positive values. Therefore, it is natural to expect that the effective propagation function of such pairs will be the integral of the propagation function of a massive particle in two-dimensional space-time over the mass m.

At the same time, it can be shown that the causal Green's functions for spaces of dimensionalities d and $d+2$, being functions of the invariant interval $s=\sqrt{-z^2}$ between two points and the mass μ, are related to one another by

$$\Delta_f^{(d+2)}(z,\mu)=\frac{1}{\pi}\frac{\partial}{\partial s^2}\Delta_f^{(d)}(z,\mu)$$

$$=\frac{1}{4\pi}\int_{\mu^2}^{\infty}dm^2\Delta_f^{(d)}(z,m) \tag{70}$$

and can be expressed in terms of the modified Bessel function of the second kind with index specified by the dimensionality of the space-time:

$$\Delta_f^{(d)}(z,\mu)=\frac{i\mu^{2\nu}}{(2\pi)^{\nu+1}(i\mu s)^{\nu}}K_{\nu}(i\mu s),\qquad \nu=\frac{d-2}{2}. \tag{71}$$

The second equality in (70) for $d=2$ confirms the emergence of a causal function characteristic of four-dimensional space-time as an effective propagation function of virtual pairs with different masses m in two-dimensional space-time. Now the small mass parameter μ, which was introduced in Sec. 2 to eliminate the infrared divergence, can be interpreted as the lower bound of the masses of the virtual pairs that mediate the self-interaction of a mirror.

A virtual pair can escape to infinity, since one of its particles is necessarily reflected from the mirror, after which the pair becomes real and massless. The emission of such pairs shapes $\mathrm{Im}\,\Delta W$. Owing to their masslessness, the emission of an arbitrarily large number of soft quanta approaching infinity becomes possible on trajectories with $\beta_{21}\neq 0$, i.e., infrared divergence of $\mathrm{Im}\,\Delta W_f$ appears. By choosing a nonvanishing but sufficiently small value of μ, we eliminate the infrared divergence in $\mathrm{Im}\,\Delta W_f$ and assure ourselves that

$\operatorname{Re}\Delta W_f$ does not depend on μ at $\mu \ll |\kappa|$. This means that the main contribution to $\operatorname{Re}\Delta W_f$ comes from virtual pairs with a mass of order $|\kappa|$.

In the general case, where the mean number of pairs created is not small compared to 1, the quantity $2\,\operatorname{Im}\Delta W$ is no longer equal to the mean number of pairs $\operatorname{Tr}(\beta^+\beta)$. Due to the interference of two or more pairs during their creation, it equals

$$2\,\operatorname{Im}\Delta W = \pm\operatorname{Tr}\ln(1\pm\beta^+\beta)\big|_0^F = \pm\operatorname{Tr}\ln(\alpha^+\alpha)\big|_0^F . \quad (72)$$

The latter formula prompted De Witt[15] to consider

$$W = \pm i\,\operatorname{Tr}\ln\alpha \quad (73)$$

to be a natural expression for W. The matrix formulation of the Bogolyubov coefficients α and β has been adopted in these equations. In addition, Tr must be replaced by (1/2) Tr when the particle and antiparticle are identical.[3]

We are not aware of any specific results for $\operatorname{Re}\Delta W$ having been derived from (73).

The symmetry discussed here would be total if it were the case in Heaviside units that $e^2 = \hbar c$.

This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grant Nos. 96-15-96463 and 96-02-17314a).

*) E-mail: ritus@lpi.ac.ru

[1)] The natural system of units, the Heaviside units of charge, and the metrics of four-dimensional and two-dimensional space with traces 2 and 0 are used, so that $\hbar = c = 1$, $e^2/4\pi = 1/137$, $k_\alpha x^\alpha = \mathbf{k}\cdot\mathbf{x} - k^0x^0$, and $j_\pm = j^0 \pm j^1$. For the remaining notation see Refs. 1–3.

---

[1] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **108**, 1125 (1995) [JETP **81**, 615 (1995)].

[2] V. I. Ritus, Zh. Éksp. Teor. Fiz. **110**, 526 (1996) [JETP **83**, 282 (1996)].

[3] V. I. Ritus, Zh. Éksp. Teor. Fiz. **114**, 46 (1998) [JETP **87**, 25 (1998)]; Erratum, Zh. Éksp. Teor. Fiz. **115**, 384 (1999) [JETP **88**, 207 (1999)].

[4] W. E. Thirring, *Principles of Quantum Electrodynamics*, Academic Press, New York (1958).

[5] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. 2*, New York, Interscience (1962).

[6] A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, Wiley, New York (1965).

[7] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **56**, 2035 (1969) [Sov. Phys. JETP **29**, 1093 (1969)].

[8] V. I. Ritus, Zh. Éksp. Teor. Fiz. **80**, 1288 (1981) [Sov. Phys. JETP **53**, 659 (1981)].

[9] V. I. Ritus, Zh. Éksp. Teor. Fiz. **82**, 1375 (1982) [Sov. Phys. JETP **55**, 799 (1982)].

[10] *Bateman Manuscript Project. Tables of Integral Transforms, Vol. 1*, A. Erdélyi (ed.), McGraw-Hill, New York (1954).

[11] L. Lewin, *Polylogarithms and Associated Functions*, North Holland, New York (1981).

[12] V. I. Ritus, Zh. Éksp. Teor. Fiz. **75**, 1560 (1978) [Sov. Phys. JETP **48**, 788 (1978)].

[13] A. P. Prudnikov, Yu. A. Brychkov, O. I. Marichev, *Integrals and Series, Vol. 2, Special Functions*, Gordon and Breach, New York (1986) [Russ. original, Nauka, Moscow (1983)].

[14] D. Ivanenko and A. Sokolov, *Classical Field Theory* [in Russian], Gostekhizdat, Moscow (1949).

[15] B. S. De Witt, Phys. Rep. C **19**, 295 (1975).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Investigation of cross sections for the formation and destruction of negative boron ions

I. S. Dmitriev, Ya. A. Teplova, and Yu. A. Faĭnberg

D. V. Skobel'tsyn Research Institute for Nuclear Physics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 5 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1539–1550 (November 1999)

The electron loss and electron capture cross sections $\sigma_{i,i+m}$ and $\sigma_{i,i-m}$ for boron ions and atoms traveling at the velocities $V=1.19$ and 1.83 a.u. in H_2, He, N_2, Ne, Ar, and Xe are measured. The known experimental data on these cross sections at velocities near the cross-section maximum are analyzed. It is found that the electron loss cross sections can be described by a formula which was previously derived in the free-collision approximation and takes into account features of both the ions and the ambient atoms. As the nuclear charge Z_t of the ambient atoms increases, the cross sections vary nonmonotonically, increasing on average as $Z_t^{1/2}$. A formula based on the model of independent electrons is proposed for electron capture by ions with small values of the charge i. It describes the dependence of the electron capture cross section $\sigma_{i,i-1}$ on the mean binding energy of an electron in an ion with the charge $i-1$. The total electron capture cross section $\sigma_{i,i-1}$ is proportional to the number of vacancies in the unfilled electron shell nearest the nucleus. The cross sections $\sigma_{i,i-1}$ exhibit substantially nonmonotonic variation with Z_t, increasing on average as $Z_t^{1/3}$. *© 1999 American Institute of Physics.* [S1063-7761(99)00311-X]

1. INTRODUCTION

The objective of the present investigation is to obtain new experimental data on the electron loss and electron capture cross sections for fast atomic particles with outer electrons having very small binding energies, including negative ions, in various media for the purpose of establishing laws governing the dependence of these cross sections on the nuclear charge of the ambient atoms Z_t and the binding energy I_{nl} of the electrons removed or captured. Since fast negative ions readily transform into neutral atoms, the study of their properties is of considerable interest for several areas of physics. These data are needed for implementing the injection of heavy ions in accelerators, as well as for investigating various processes in high-temperature plasmas.[1]

The cross sections for the loss and capture of one or several electrons by fast atoms, as well as negative and positive ions, of boron with velocities $V=1.19V_0$ and $1.83V_0$ (where $V_0=2.19\times10^8$ cm/s) as they pass through gaseous H_2, He, N_2, Ne, Ar, and Xe are measured in the present work. The analogous cross sections known from the literature for several other ions with outer electrons having minimal binding energies are analyzed. The data obtained enable us to accurately estimate the corresponding cross sections for other ions and media, for which there have not been any direct measurements. The ion velocities V are henceforth presented in atomic units of V_0.

2. EXPERIMENTAL METHOD

The electron loss and electron capture cross sections for boron ions and atoms were determined using the experimental equipment described in Refs. 2 and 3. The volatile boron compound BBr_3 was introduced into the cyclotron ion source with a vapor pressure at room temperature of about 8 kPa, which is sufficient for operation of the ion source. Singly charged $^{11}B^+$ ions which were accelerated in a 72-cm cyclotron and had energies $E=35$ and 83 keV/nucleon, i.e., velocities $V=1.19$ and 1.83, were directed into a charge-exchange device, in which the primary B^+ ions were transformed into beams of B^i ions with charge $i=-1$, 0, 1, and 2. Boron atoms and negative boron ions were obtained as a result of the charge exchange of B^+ ions in a thin helium target. Since the probability of formation of these particles decreased rapidly with increasing velocity V, their flux intensity was corrected by altering the gas pressure in the helium target. To obtain the maximum number of negative B^- ions, the gas layer in the target was increased from 1.5×10^{15} to 3×10^{15} atoms/cm^2 when the velocity was increased by a factor of 1.5. Thus, the number of B^- ions in the vicinity of the detector varied from 10–50 particles per second at $V=1.19$ to 2–5 particles per second at $V=1.83$. The flux of B^0 atoms was set to $\sim10^3$ particles per second. In the work with beams of B^0 atoms, the charged components B^+ and B^- were deflected by an electrostatic analyzer.

Beams of boron ions with various values of i formed as a result of charge exchange were alternately directed into the collision chamber, which consisted of a cylinder 24 cm long with entrance and exit channels 0.5 cm high, 0.1–0.2 cm wide, and 2.6 cm long. The gas pressure of H_2, He, N_2, Ne, Ar, and Xe injected into the center of the collision chamber was determined to within 10% by ionization gauges calibrated for the various gases using a McLeod gauge. The charge distribution of the boron ions after passage through the collision chamber was measured by a detector consisting of a magnetic analyzer and counters at two or three values of the gas pressure in the chamber. The field intensity of the

TABLE I. Electron loss cross sections $\sigma_{i,i+m}$ (in units of 10^{-16} cm^2/atom)

$i, i+m$	H_2	He	N_2	Ne	Ar	Kr	Xe
			$V=1.19$				
−1, 0	10.0	6.3	10.0	17.0	30.0	-	70.0
−1, 1	2.8	2.5	5.0	10.0	14.0	-	23.0
−1, 2	0.2	0.3	1.3	0.75	2.2	-	4.4
0, 1	2.3	2.0	5.0	2.2	10.0	14.0	10.0
0, 2	0.2	0.3	1.0	0.3	1.3	-	0.8
0, 3	-	0.05	0.04	-	0.01	-	-
1, 2	0.8	1.0	2.0	0.7	3.0	3.2	2.3
1, 3	0.015	0.02	0.05	0.02	0.10	0.05	0.03
2, 3	0.12	0.18	0.15	0.01	0.3	0.28	0.2
			$V=1.83$				
−1, 0	4.0	4.0	6.6	11.2	14.0	-	23.0
−1, 1	1.6	1.7	3.8	4.0	9.0	-	10.0
−1, 2	0.30	0.15	0.64	0.35	2.6	-	1.15
0, 1	2.8	3.0	5.6	6.3	13.0	19.0	14.0
0, 2	0.20	0.30	1.4	1.1	3.2	-	3.3
0, 3	0.002	0.002	0.1	0.035	0.16	-	0.11
1, 2	0.7	1.4	2.5	2.3	6.0	5.0	6.3
1, 3	0.014	0.022	0.17	0.12	0.43	0.18	0.25
2, 3	0.18	0.35	0.54	0.4	1.2	1.1	1.0
2, 4	-	0.001	0.003	0.01	-	-	-

analyzing magnet was such that on average, particles of just one charge struck the middle part of each counter. The cross sections $\sigma_{i,i+m}$ and $\sigma_{i,i-m}$ for the loss and capture of one or several electrons were found for the resulting charge distributions by the method described in Ref. 2. The errors in the cross sections accrued mainly from errors in the thickness of the gas layer in the collision chamber (~10%) and the statistical spread of the results of several series of measurements, and amounted to 10–15% on average for the cross sections for the loss and capture of one electron, 20–30% for the cross sections for the loss and capture of two electrons, and about 50% for the cross sections for the loss and capture of three electrons. At the velocity $V=1.83$ the error in $\sigma_{1,-1}$ reached 50–100%, i.e., only an upper limit for the cross sections was obtained in certain cases.

The electron loss and electron capture cross sections for positively charged boron ions traveling at these velocities were previously measured in atomic (H) and molecular (H_2) hydrogen[4] and in He, N_2, and Ar, as well as in Kr, in Refs. 2 and 5–7. All the cross sections $\sigma_{i,i+m}$ for the negative boron ions and atoms, as well as the cross sections for positive boron ions in Ne and in Xe, were determined for the first time. The resulting values of $\sigma_{i,i\pm m}$ are listed in Tables I and II. Since the cross sections measured in the present work and in Refs. 1, 5 and 6 coincided to within 20–30%, the corresponding mean cross sections are given in the tables.

TABLE II. Electron capture cross sections $\sigma_{i,i-m}$ (in units of 10^{-16} cm^2/atom)

$i, i-m$	H_2	He	N_2	Ne	Ar	Kr	Xe
			$V=1.19$				
0, −1	0.08	0.05	0.05	0.045	0.09	-	0.1
1, 0	1.6	1.0	1.0	0.65	2.5	2.76	2.8
1, −1	0.005	0.003	0.005	0.0025	0.01	-	0.025
2, 1	4.3	4.0	6.2	2.0	10.0	15.0	13.0
2, 0	0.35	0.4	0.3	0.1	0.6	0.83	2.5
3, 2	5.5	-	-	-	-	-	-
3, 1	1.3	-	-	-	-	-	-
4, 3	0.06	-	-	-	-	-	-
			$V=1.83$				
1, −1	0.025	0.035	0.025	0.04	0.045	-	0.1
1, 0	0.5	0.5	0.50	0.4	0.95	1.6	1.6
1, −1	0.001	0.0016	0.001	0.004	0.01	-	-
2, 1	1.5	2.2	2.4	1.5	5.0	6.6	6.3
2, 0	0.04	0.03	0.045	0.04	0.14	0.28	0.3
3, 2	2.7	3.2	5.0	3.5	10.0	11.0	10.0
3, 1	0.25	0.25	0.55	0.4	1.3	2.4	2.0
3, 0	-	-	0.03	-	-	0.063	0.06
4, 3	5.0	5.0	6.3	6.5	14.0	-	-
4, 2	0.2	0.3	0.9	0.06	2.0	-	-

3. DISCUSSION OF RESULTS

3.1. Electron loss cross sections

The loss of one or several electrons by a many-electron particle is a result of the loss of individual nl electrons. The total cross section for the removal of a fixed electron from the subshell with quantum numbers n and l, with the remaining electrons in arbitrary states, is called the single-electron loss cross section σ_{nl}. When the probability of an Auger transition with filling of an inner-shell vacancy can be neglected, σ_{nl} can be represented in the form[8]

$$\sigma_{nl}=q_{nl}^{-1}\left[S_i-\sum q_{nl}^{\text{in}}\sigma_{nl}^{\text{in}}\right], \tag{1}$$

where q_{nl} and q_{nl}^{in} are the numbers of electrons in the outer and inner subshells of the ion, and $S_i=\Sigma m\sigma_{i,i+m}$. The sum S_i characterizes the increase in the charge of the ion (atom) as a result of collisions with ambient atoms of the medium. In studies of the ionization of ambient atoms by various particles, the quantity analogous to S_i is called the total ionization cross section, and it coincides with the cross section for the production of free electrons.[9]

An analysis of the experimental electron loss cross sections $\sigma_{i,i+m}$ for ions of light elements with a nuclear charge $Z\leqslant 18$ and with a number of electrons $N=Z-i$ from 1 to 17 shows[10] that for a given set of quantum numbers nl, electron binding energy I_{nl}, and relative velocity of the ions V, the cross section σ_{nl} for ions with various numbers of electrons N scarcely depends on q_{nl}. This enables us to obtain the values of σ_{nl} for some ions from the experimental values of $\sigma_{i,i+m}$ for other ions. The lack of a dependence of σ_{nl} on q_{nl} for given nl, I_{nl}, and V corresponds to the results of calculations based on the independent electron loss model, for which the mean probability and cross section for the removal of each of the electrons do not depend on the presence of the other electrons in the ion. In that case the cross sections σ_{nl} for the loss of a single electron and the cross section $\sigma_{i,i+m}$ for the loss of m electrons from the outer subshell with q_{nl} electrons have the form[7,10,11]

$$\sigma_{nl}=\int W_{nl}(p)\,dp,$$

$$\sigma_{i,i+m}=C_q^m\int W_{nl}^m(p)[1-W_{nl}(p)]^{q-m}dp, \tag{2}$$

where p is the impact parameter, $W_{nl}(p)$ is the probability of removal of a single electron, and $C_q^m=q!/m!(q-m)!$. Expressions similar to (2) were obtained for the electron loss cross sections $\sigma_{i,i+m}$ of the K and L shells, and have been used in semiclassical calculations of inner-shell ionization.[12]

One special feature of the electron loss cross sections for the negative boron ions is the substantial contribution of the multiple-ionization cross sections $\sigma_{-1,-1+m}$ $(m>1)$ to the total ionization cross section S_{-1}. The fraction of this contribution is specified by $R_i=1-[\sigma_{i,i-1}/S_i]$, which has the value $R_{-1}=0.50\pm 0.05$ for B^- ions. At the same time, for B^0 atoms and B^+ ions the contributions of multiple ionization to the total ionization cross section are considerably smaller and amount, on average, to $R_0=0.25\pm 0.10$ for B^0 and $R_1=0.05\pm 0.01$ for B^+.

The cross sections σ_{nl} for the loss of individual nl electrons from outer $1s$, $2s$, and $2p$ subshells were found using formula (1) from the experimental cross sections $\sigma_{i,i+m}$ with allowance for the possibility of electron loss from the inner shells. The cross sections for the loss of inner $1s$ electrons were equated to the electron loss cross sections for hydrogen-like particles with the same electron binding energy as in the ion (atom) under consideration. The values of σ_{2s} were assumed to be equal to the cross sections for the loss of $2s$ electrons by lithium-like and beryllium-like ions with the same electron binding energies. The values of I_{nl} were taken from Ref. 13.

The values calculated in the Born approximation for the cross sections σ_{1s} and σ_{2s} for the loss of K electrons by hydrogen-like and helium-like ions upon the passage of fast particles though hydrogen, helium, and nitrogen[14–16] and the experimental electron loss cross sections σ_{nl} for negative H^- ions and H^0 atoms,[17] negative Li^- ions,[18] and positive ions of light elements in heavier media[10] enabled us to establish the dependence of σ_{nl} on I_{nl} in the range $I_{nl}=(0.1-10)I_0$ $(I_0=13.6\,\text{eV})$.

In the range $I_{nl}\leqslant I_0$, where the Born approximation coincides with the free-collision approximation,[19] the cross sections vary slowly with I_{nl}, the ratio σ_{1s}/σ_{2s} differing from unity by no more than 20%.[16] In the range $I_{nl}>I_0$ the cross sections σ_{nl} decrease rapidly with increasing I_{nl}. The dependence of σ_{nl} on I_{nl} obtained was used to estimate the contribution Δ of the cross sections for electron loss from the inner subshells, $\Sigma q_{nl}^{\text{in}}\sigma_{nl}^{\text{in}}$, to the total ionization cross section for the boron ions $S_i=m\sigma_{i,i+m}$ in different media. On average, the values of Δ were roughly 20%, 50%, 3–5%, and $\leqslant 1\%$ for B^-, B^0, B^+, and B^{2+}, respectively.

The resulting dependence of the cross sections σ_{nl} for the loss of the outer $2s$ (for B^- ions and B^0 atoms) and $2p$ (for B^+ and B^{2+}) electrons during the passage of fast boron atoms and ions through gaseous media with the velocities $V=1.19$ and 1.83 on the nuclear charge of the medium atoms Z_t is presented in Fig. 1. The cross sections σ_i are proportional to Z_t^k. When $V=1.19$, k is greatest for negative B^- ions $(k\approx 0.5)$ and decreases to $k\approx 0.4$ for B^0 atoms and to $k\approx 0.2-0.15$ for B^+ and B^{2+} ions. When $V=1.83$, $k\approx 0.5$ for all ions. As m, the number of electrons removed, is increased and V is diminished, the deviation of $\sigma_i(Z_t)$ from a monotonic dependence increases significantly. In particular, when $V=1.83$, the cross section σ_2 for B^{2+} ions in neon is approximately one-tenth the corresponding cross sections in nitrogen and argon.

As shown in Ref. 6, the values of the velocity $V_{\max}$, at which electron loss cross sections reach their maximum values, can be found from the relation $V_{\max}\sim\gamma U_i$, where U_i is the mean orbital velocity of the electron being removed. As Z_t increases, the coefficient γ increases from $\gamma\approx 1.3$ in helium to $\gamma\approx 2$ in krypton. Accordingly, the maximum electron loss cross sections for B^- ions should be observed at $V_{\max}\approx 0.3$. The greatest electron loss cross sections for B^0 atoms are achieved at $V_{\max}\approx 1.5$, and the greatest values for

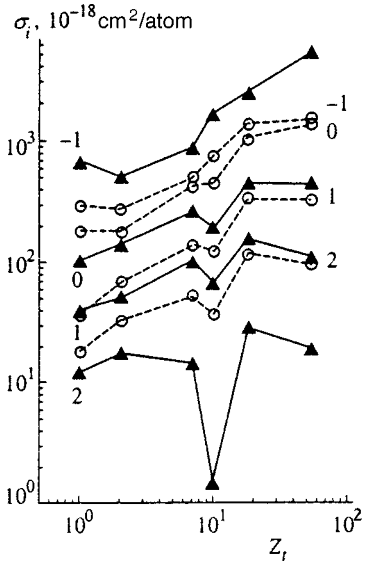

FIG. 1. Dependence of the single-electron loss cross sections σ_i for boron ions and atoms on Z_t at $V=1.19$ (solid lines and filled points) and at $V=1.83$ (dashed lines and unfilled points). The ion charge i is indicated near the curves.

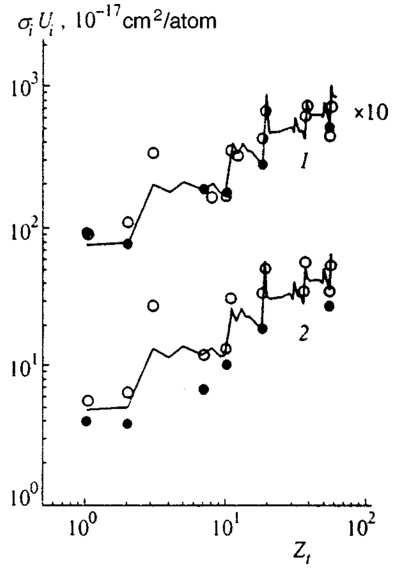

FIG. 2. Dependence of $\sigma_i U_i$ on Z_t for negative B^- ions (filled points) and H^- ions (unfilled points). Solid lines—calculation based on formula (3): *1*—$V=1.19$; *2*—$V=1.83$.

positive B^+ and B^{2+} ions are achieved at $V_{\max}\approx 2-3$. The values of σ_i obtained in the present work for boron atoms and ions with the charges $i=0$, 1, and 2 increase somewhat with increasing V, while the values for the negative ions decrease.

It was previously[20] shown in an investigation of the electron loss cross sections for negative hydrogen ions and hydrogen atoms that the dependence of σ_i on Z_t is has a steplike character. The sharp increase in the electron loss cross sections σ_i for hydrogen particles in alkali-metal vapors in comparison to the corresponding cross sections in inert gases results from the weaker screening of the nuclear Coulomb field in the alkali-metal atoms by the outer electrons. The following semiempirical formula, which takes into account the features of both the ions and the ambient atoms, was proposed in Ref. 20 to describe the dependence of σ_i on Z_t:

$$\sigma_i=\pi a_0^2 Z_t^{\alpha}/VU_iU(Z_t), \tag{3}$$

where $U_i=\sqrt{I_i/I_0}$ and $U(Z_t)=\sqrt{I(Z_t)/I_0}$ are the mean orbital velocities of the electrons in the ion and an ambient atom, and $I(Z_t)$ is the ionization potential of the ambient atoms. At $V\approx 1-2$ the exponent α is set to 0.5.

Formula (3), which was obtained in the free-collision approximation, is a modified version of the familiar Bohr formula for the loss of a weakly bound electron in media with $Z_t\geqslant V/2$ and is applicable in the range of ion velocities $V\geqslant(1-2)U_i$, where the cross sections fall off comparatively slowly, as V^{-1}.

In accordance with (3), the product σ_iU_i for identical values of $U(Z_t)$ and V should be the same for different ions. In fact, when $V=1.19$, the measured and calculated values of σ_iU_i for B^- and H^- ions coincide in all media to within 20–30% (with the exception of the H^- ions in lithium, where the experimental values of σ_iU_i are 1.5–2 times the calculated values). When $V=1.83$, the measured values of σ_iU_i for B^- ions are, on average, 30–40% less than the corresponding calculated and experimental values for H^- ions (Fig. 2).

The electron loss cross sections for the negative ions of light elements in the range of ion velocities $V\approx 1-2$ are known not only for the hydrogen and boron ions considered, but also for metastable helium ions in the $He^{m-}(1s2s2p)$ state,[21] $Li^-(1s^22s^2)$ ions,[18,22] $C^-(1s^22s^22p^3)$ ions, and $O^-(1s^22s^22p^2)$ ions.[23] The orbital velocities U_i of the outer electrons in these particles vary from $U_i\approx 0.077$ in He^- to 0.35 in O^- (Ref. 24). The contribution of the cross sections for the loss of two electrons to the total ionization cross section S_{-1} for H^-, He^-, and Li^- ions amounts to $R_{-1}=0.20\pm 0.05$ on average. At the same time, the cross sections for the loss of two and three electrons for C^- and O^- ions, just as for B^- ions, amount to ≈ 0.5 and ≈ 0.3 of the cross sections for the loss of one electron, and therefore $R_{-1}=0.60\pm 0.05$. The value of $V_{\max}$, where the electron loss cross sections peak, should vary from $V_{\max}\approx 0.2$ for helium ions to $V_{\max}\approx 0.7$ for carbon and oxygen ions.

Apart from the data just enumerated, for negative ions of light elements with outer electrons having a very small binding energy, the electron loss cross sections for metastable $He^{m0}(1s2s)$ (Ref. 22), $Li^0(1s^22s)$ (Ref. 25), and $H^0(1s)$ atoms, as well as the cross sections obtained in the present work for $B^0(1s^22s^22p)$ atoms, are now known. The values of U_i for these particles vary from $U_i=0.57$ for helium atoms to $U_i=1.0$ for hydrogen atoms, and the maxima of the electron loss cross sections are located in the range $V_{\max}\approx 1-2$.

An analysis of the electron loss cross sections σ_i for the negative ions and atoms of light elements showed that the cross sections σ_i decrease with increasing orbital velocities of the outer electrons U_i in proportion to U_i^{-1} and that the deviation of the measured cross sections from the values calculated from formula (3) is at most a factor of 1.5 over the

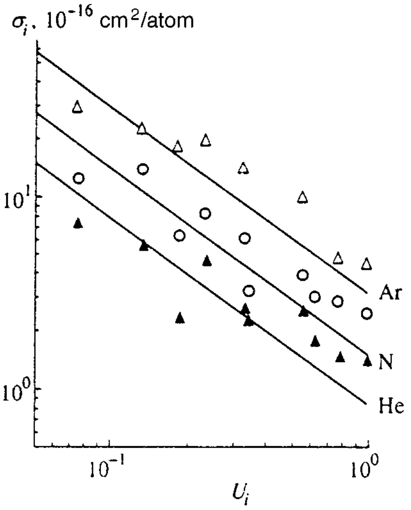

FIG. 3. Dependence of the electron loss cross sections σ_i for ions of light elements at $V=1.19$ on U_i in helium (filled triangles), in nitrogen (unfilled circles), and in argon (unfilled triangles). Straight lines—calculation based on formula (3).

entire range of U_i, which amounts to more than an order of magnitude (from 0.08 to 1.0). Figure 3 presents the data for helium, nitrogen, and argon (the most measurements were performed in these media). Clearly, the greatest deviation between the calculated and measured values occurs at $U_i \geqslant 0.5$. Somewhat better agreement with experiment is achieved over the full range of U_i, if a weaker dependence of σ_i on U_i, specifically $\sigma_i \propto U_i^{-2/3}$, is adopted in Eq. (3).

3.2. Electron capture cross sections

Scrutiny of the experimental data on the charge-exchange cross sections for atoms and singly charge ions of light elements[26,27] revealed that the cross sections for non-resonant charge exchange peak at the velocity $V_{\max}$, which satisfies Massey's adiabatic criterion[28]

$$V_{\max} \approx a|\Delta E|/h, \tag{4}$$

where ΔE is the change in the internal energy of the system, $a \approx 3 \times 10^{-8}$ for single-electron capture,[29] and $a \approx 1.5 \times 10^{-8}$ for double-electron capture.[26] In accordance with this criterion, the maximum electron capture cross sections for most of the ion–ambient-atom pairs studied fall in the range $V_{\max} \approx 0.5-2$. According to the experimental data for B^i ions in hydrogen,[4] when V systematically increases, the cross sections $\sigma_{i,i-1}$ for the ions with charges $i=1-5$ are proportional to V^{-3} over the range $V=1.5-2$. As the velocity increases further, the dependence of these cross sections on V strengthens, and we have $\sigma_{i,i-1} \propto V^{-5}$ by the time $V > 2-3$.

The investigation of the electron capture cross sections $\sigma_{i,i-1}$ for fast positive nitrogen and neon ions in Ref. 30 revealed that for ions with small values of i, which satisfy the condition

$$I_{i-1}(n) \leqslant I = I_v/3 + 3I_t, \tag{5}$$

the total electron capture cross section $\sigma_{i,i-1}$ is proportional to the number $p(n)$ of vacant states in the ion. In formula (5) $I_{i-1}(n)$ is the mean binding energy of an electron in the unfilled electron shell with principal quantum number n that is closest to the nucleus in an ion with charge $i-1$, I_t is the binding energy of the outer electrons in the ambient atoms, and $I_v = \mu V^2/2$ is the transport energy of an electron whose velocity V equals the velocity of a fast particle.

The proportionality between the electron capture cross section and the number of vacant states corresponds to the model of independent electrons,[31] according to which

$$\sigma_{i,i-m} = \pi a_0^2 C_p^m W^m, \tag{6}$$

where W is the mean probability of electron capture and $C_p^m = p!/m!(p-m)!$.

To estimate the charge-exchange cross sections at ion velocities close to the maximum in heavy media $[(I_{i-1}/I_0) < Z_t]$, where the effective number of electrons captured from ambient atoms $N \propto Z_t^{1/3}$ (Ref. 32), the value of W can be represented in the form

$$W \sim (I_{i-1}/I_0) Z_t^{1/3} V^{-3}. \tag{7}$$

On the basis of (6) and (7) the total single-electron capture cross section $\sigma_{i,i-1}$ at ion velocities near the maximum is given by

$$\sigma_{i,i-1} = p_{i-1}(n) \pi a_0^2 (I_{i-1}/I_0) Z_t^{1/3} V^{-3}. \tag{8}$$

When $V=1.19$, the electron capture cross sections $\sigma_{0,-1}$ for B^0 atoms depend weakly on Z_t, varying over the range from 0.5 to 1×10^{-17} cm^2/atom as Z_t increases from 1 to 54. The corresponding cross sections for positive B^+ and B^{2+} ions are 10–30 times the electron capture cross sections for boron atoms. The electron capture cross sections $\sigma_{i,i-1}$ for boron atoms and ions determined from formula (8) agree with the measured values to within a factor of 2–3. The greatest difference, a factor of 5, was obtained for $V=1.19$ in neon. In neon, however, the experimental values of $\sigma_{i,i-1}$ exhibit substantially nonmonotonic variation, increasing as $Z_t^{1/3}$ on average (Fig. 4).

According to formula (8), the single-electron capture cross sections $\sigma_{i,i-1}$ divided by the number of vacancies $p_{i-1}(n)$ in the unfilled shell nearest the nucleus, i.e., the values of $\sigma_c(n) = \sigma_{i,i-1}/p_{i-1}(n)$, should lie on straight lines corresponding to the linear dependence on $I_{i-1}(n)$. Figure 5 presents the dependence of $\sigma_c(n)$ on $I_{i-1}(n)$ for electron capture by atoms and positive ions of light elements with $Z \leqslant 10$ in a nitrogen target at $V=1.19$. The experimental cross sections were taken from Refs. 2, 17, 21, 22, 25, 30, and 33–35. The values of $\sigma_c(1)$ and $\sigma_c(2)$ for electron capture by H^0 atoms with a vacancy in the K shell and for electron capture by B^0, C^0, and O^0 atoms with vacancies in the L shell are located in the range of small values $I_{-1}(n) \approx 0.3-2$ eV. The values of $\sigma_c(1)$ for electron capture to vacant states in the K shell by H^+ ions (protons), $He^{1,2+}$ ions, and $Li^{2,3+}$ ions, as well as the values of $\sigma_c(2)$ for electron capture by Li^+, $B^{1,2+}$, $N^{1,2,3,4+}$, and $Ne^{1,2,3,4,5+}$ ions, lie in the range of $I_{i-1}(n)$ from 5 to 100 eV. As $I_{i-1}(n)$ increases, the experimental reduced cross sections $\sigma_c(1)$ and $\sigma_c(2)$ increase in proportion to $I_{i-1}(n)$ as $I_{i-1}(n_0)$ varies over a range of almost three orders of magnitude (from 0.3 to 100 eV). The σ_c determined from formula (8) are approxi-

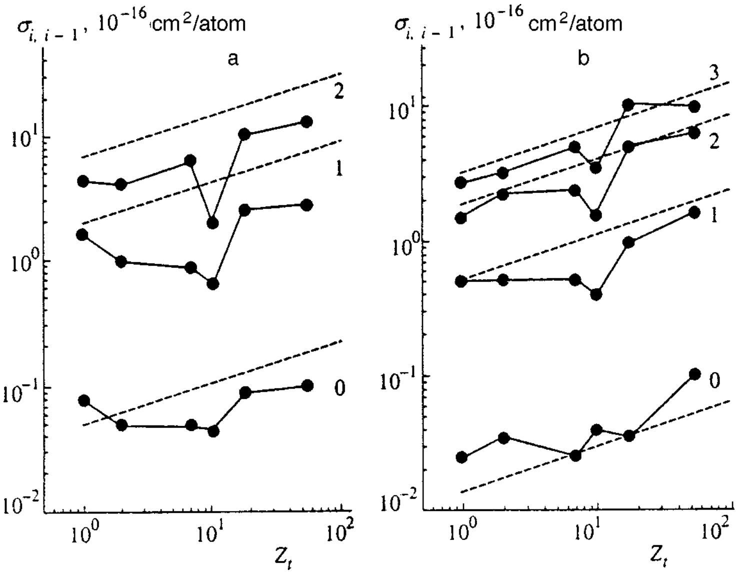

FIG. 4. Dependence of the electron capture cross sections $\sigma_{i,i-1}$ on Z_t: a—$V=1.19$, b—$V=1.83$. Dashed lines—calculation based on formula (8). The ion charge i is indicated near the curves.

mately triple the value of $\sigma_c(2)$, and are essentially identical to $\sigma_c(1)$. For ions with high values of i, for which $I_{i-1}(n)>I\approx 50$ eV, the dependence of the reduced cross sections $\sigma_c(1)$ and $\sigma_c(2)$ on $I_{i-1}(n)$ weakens. As shown in Ref. 30, the capture of an electron by these ions occurs predominantly to excited states of the ion; therefore, the total cross section $\sigma_{i,i-1}$ depends weakly on the number of electrons in the ion and is determined by the charge i and by the velocity V.

4. CONCLUSION

The results of the present work have enabled us to establish the basic behavior of electron loss and electron capture cross sections for ions of light elements in various media in the range of ion velocities V where the cross sections reach their maximum values.

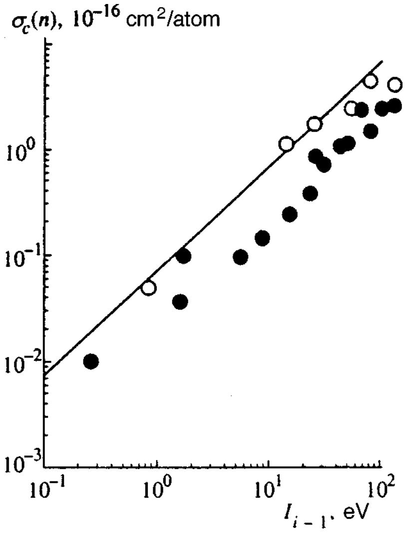

FIG. 5. Dependence of the values of $\sigma_c(n)=\sigma_{i,i-1}/p_{i-1}(n)$ on $I_{i-1}(n)$. The unfilled points show the values of $\sigma_c(1)$ for ions with a K vacancy, and the filled points show the values of $\sigma_c(2)$ for ions with an L vacancy. Solid line—calculation based on formula (8).

It has been shown that the experimental electron loss cross sections for boron ions and atoms, as well as the analogous published data for other ions of light elements, are satisfactorily described by the formula previously obtained in the free-collision approximation, which takes into account the features of both the ions and the ambient atoms.[20] The single-electron loss cross sections σ_i vary nonmonotonically, increasing as $Z_t^{1/2}$ on average. At a given ion velocity V the values of σ_i decrease with increasing orbital velocity U_i of the electron being removed as U_i^{-1}. When we go over from inert gases to alkali-metal vapors, the values of σ_i increase by a factor of 1.5.

A formula based on the model of independent electrons, which satisfactorily (to within a factor of 2–3) describes the dependence of the electron capture cross sections $\sigma_{i,i-1}$ on the magnitude I_{i-1} of the mean binding energy of an electron in an ion with charge $i-1$, has been proposed for ions with small values of i. As I_{i-1} increases, the cross sections increase proportionally to I_{i-1}. The total electron capture cross section $\sigma_{i,i-1}$ is proportional to the number of vacancies in the unfilled electron shell closest to the nucleus. Also, the values of $\sigma_{i,i-1}$ exhibit substantially nonmonotonic variation with Z_t, increasing as $Z_t^{1/3}$ on the average. As shown in Refs. 36 and 37 by investigating the charge-transfer cross sections for positive hydrogen and helium ions in various media, the nonmonotonic dependence of the cross sections on Z_t, which leads to the systematic alteration of maxima and minima, results from structural features of the ambient atoms and the resonant nature of the charge-exchange cross sections, and can be described qualitatively by the quantum-mechanical Oppenheimer–Brinkman–Kramers (OBK) approximation.

A preliminary analysis of data on the cross sections $\sigma_{i,i+m}$ and $\sigma_{i,i-m}$ for the loss and capture of two or more electrons by boron ions and atoms points to a significantly stronger dependence on Z_t.

We express our sincerest thanks to the service technicians for the 72-cm cyclotron—Yu. P. Druzhinin, V. A.

Kalita, S. A. Murav'ev, A. A. Obukhov, and V. A. Pronin—for their assistance in conducting the experiment.

This work was carried out with the financial support of the ''Controlled Thermonuclear Fusion and Plasma Processes'' program.

[1] *Present Status on Atomic and Modelling Data Relevant to Fusion Plasma Diagnostics and Modelling, NIES-Data 39*, H. Tawara (Ed.), Nagoya (1997).

[2] V. S. Nikolaev, I. S. Dmitriev, L. N. Fateeva, and Ya. A. Teplova, Zh. Éksp. Teor. Fiz. **40**, 989 (1961) [Sov. Phys. JETP **13**, 695 (1961)].

[3] Ya. A. Teplova and I. S. Dmitriev, in *Development of Scientific Research on Nuclear and Atomic Physics in the D. V. Skobel'tsyn Research Institute for Nuclear Physics at the M. V. Lomonosov Moscow State University* [in Russian], Izd. MGU, Moscow (1994), p. 132.

[4] T. V. Goffe, M. V. Shah, and H. B. Gilbody, J. Phys. B **12**, 3763 (1979).

[5] V. S. Nikolaev, L. N. Fateeva, I. S. Dmitriev, and Ya. A. Teplova, Zh. Éksp. Teor. Fiz. **41**, 89 (1961) [Sov. Phys. JETP **14**, 67 (1962)].

[6] I. S. Dmitriev, V. S. Nikolaev, L. N. Fateeva, and Ya. A. Teplova, Zh. Éksp. Teor. Fiz. **42**, 16 (1962) [Sov. Phys. JETP **15**, 11 (1962)].

[7] I. S. Dmitriev, V. S. Nikolaev, L. N. Fateeva, and Ya. A. Teplova, Zh. Éksp. Teor. Fiz. **43**, 361 (1962) [Sov. Phys. JETP **16**, 259 (1963)].

[8] Yu. A. Tashaev, I. S. Dmitriev, V. S. Nikolaev, and Ya. A. Teplova, J. Phys. B **11**, L223 (1978).

[9] E. W. McDaniel, *Collision Phenomena in Ionized Gases*, Wiley, New York (1964).

[10] V. S. Nikolaev, Usp. Fiz. Nauk **85**, 679 (1965) [Sov. Phys. Usp. **8**, 269 (1965)].

[11] J. H. McGuire and L. Weaver, Phys. Rev. A **16**, 1 (1977).

[12] D. Madison and E. Merzbacher, *Atomic Inner-Shell Processes*, Vol. 1, Academic Press, New York (1975).

[13] W. Lotz, J. Opt. Soc. Am. **58**, 915 (1968).

[14] I. S. Dmitriev, Ya. M. Zhileĭkin, and V. S. Nikolaev, Zh. Éksp. Teor. Fiz. **49**, 500 (1965) [Sov. Phys. JETP **22**, 352 (1966)].

[15] V. S. Senashenko, V. S. Nikolaev, V. Yu. Shafer, and I. S. Dmitriev, Vestn. Mosk. Univ., Fiz., Astron. **2**, 136 (1970).

[16] V. S. Nikolaev, V. S. Senashenko, and V. Yu. Shafer, Vestn. Mosk. Univ., Fiz., Astron. **3**, 357 (1973).

[17] H. Tawara and A. Russek, Rev. Mod. Phys. **45**, 178 (1973).

[18] R. W. McCullough, M. B. Shah, M. Lannon, and H. B. Gilbody, J. Phys. B **15**, 791 (1982).

[19] I. S. Dmitriev and V. S. Nikolaev, Zh. Éksp. Teor. Fiz. **44**, 660 (1963) [Sov. Phys. JETP **17**, 447 (1963)].

[20] I. S. Dmitriev, Ya. A. Teplova, and Yu. A. Faĭnberg, Zh. Éksp. Teor. Fiz. **107**, 55 (1995) [JETP **80**, 28 (1995)].

[21] J. Heinemeier, P. Hvelplund, and F. Simpson, J. Phys. B **9**, 2669 (1976).

[22] E. H. Pedersen and P. Hvelplund, J. Phys. B **6**, 1237 (1973).

[23] I. S. Dmitriev, V. S. Nikolaev, Ya. A. Teplova *et al.*, Zh. Éksp. Teor. Fiz. **50**, 1252 (1966) [Sov. Phys. JETP **23**, 832 (1966)].

[24] H. S. W. Massey, *Negative Ions*, Cambridge University Press (1976).

[25] E. Horsdal-Pedersen, J. Heinemeier, L. Larsen, and J. V. Mikkelsen, J. Phys. B **13**, 1167 (1980).

[26] Ya. M. Fogel', R. V. Mitin, and A. G. Koval', Zh. Éksp. Teor. Fiz. **31**, 397 (1956) [Sov. Phys. JETP **4**, 359 (1957)].

[27] Ya. M. Fogel', R. V. Mitin, V. F. Kozlov, and N. D. Romashko, Zh. Éksp. Teor. Fiz. **35**, 565 (1958) [Sov. Phys. JETP **8**, 390 (1959)].

[28] H. S. W. Massey, Rep. Prog. Phys. **12**, 248 (1948).

[29] Ya. M. Fogel', V. A. Ankudinov, and D. V. Pilipenko, Zh. Éksp. Teor. Fiz. **35**, 868 (1958) [Sov. Phys. JETP **8**, 601 (1959)].

[30] I. S. Dmitriev, Yu. A. Tashaev, V. S. Nikolaev, Ya. A. Teplova *et al.*, Zh. Éksp. Teor. Fiz. **73**, 1684 (1977) [Sov. Phys. JETP **46**, 884 (1977)].

[31] R. Hippler, D. Datz, P. R. Miller *et al.*, Phys. Rev. A **35**, 585 (1987).

[32] N. Bohr and J. Lindhard, Dan. Mat. Fys. Medd. **28(7)**, (1957).

[33] S. K. Allison, Phys. Rev. **109**, 76 (1958).

[34] S. K. Allison, J. Guevas, and M. Garcia-Munoz, Phys. Rev. **120**, 1266 (1960).

[35] L. I. Pivovar, Yu. Z. Levchenko, and G. A. Krivonosov, Zh. Éksp. Teor. Fiz. **36**, 491 (1966) [sic].

[36] I. S. Dmitriev, N. F. Vorob'ev, Zh. M. Konovalova *et al.*, Zh. Éksp. Teor. Fiz. **84**, 1987 (1983) [Sov. Phys. JETP **57**, 1157 (1983)].

[37] V. S. Nikolaev, I. S. Dmitriev, Ya. A. Teplova, and Yu. A. Faĭnberg, Vestn. Mosk. Univ., Fiz., Astron. **35(3)**, 84 (1994).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Rotational quasienergy states and alignment of molecules in a strong laser field

A. I. Andryushin and M. V. Fedorov*)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

(Submitted 5 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1551–1564 (November 1999)

The interaction of a molecule with a strong laser field is investigated. Raman-type transitions between rotational levels of a fixed vibrational state of the ground-state term are taken into account in the ''quantum rotator'' approximation. An initial problem of the evolution of the state of a molecule interacting with a pulsed field is solved. The dynamics of the degree of alignment of a molecule with respect to the direction of polarization of the field during the pulse is investigated. It is shown that for sufficiently long pulses with a smooth envelope the axis of the molecule adiabatically follows the time-varying field amplitude, and alignment is maximum when the field intensity is maximum. It is shown that alignment of molecules can be substantial only if the second-order composite matrix elements, determining the probability amplitudes of transitions between rotational levels of a molecule, are much greater than the dissociation broadening of the levels. The angular distribution of the dissociation products of a molecule in a strong laser field is investigated. *© 1999 American Institute of Physics.* [S1063-7761(99)00411-4]

1. INTRODUCTION

It has been found in a number of experimental works investigating the interaction of molecules with a strong laser field that under certain conditions the angular distribution of the photodissociation fragments of a molecule is extended in the direction of polarization of the field.[1,2] In such experiments, ordinarily, the object of investigation is a gas of diatomic molecules with an isotropic distribution of the directions of the axes. Motion of the dissociation products predominantly in the direction of polarization of the field could be a consequence of field-induced alignment of the axes of the molecules in this direction, which in turn could be due to restructuring of the rotational motion of the molecules in a strong light field.

The restructuring of the rotational motion of a molecule in a strong light field can occur for various reasons, such as resonant excitation and mixing of the vibrational states of the molecule,[3,4] photodissociation or photoionization broadening of rotational levels,[5–7] two-photon Raman transitions between the rotational levels of a molecule,[8,9] and so on. In certain special cases the photodissociation of the molecular ion H_2^+ by a laser pulse has been investigated numerically in a two-term model.[6,10,11] However, the reasons and the conditions for field-induced restructuring of the rotational motion of a molecule as well as the effect of the restructuring on the angular distribution of the dissociation products were not determined.

In Refs. 12 and 13 a molecule was treated as a classical rigid rotator, and the rotation and orientation of molecules in a strong light field were investigated in such a model. The applicability of the classical approach is briefly discussed in Sec. 4.1.

In the present paper resonant excitation of the vibrations and ionization of molecules is studied. Photodissociation broadening and two-photon Raman transitions between rotational levels of a molecule are taken into account in the quantum-mechanical model formulated below. The dynamics of reoccupation of the rotational levels of the ground term of a molecule, the time-dependent degree of alignment of an initially isotropic ensemble of molecules, and the angular distribution of the decay fragments of the dissociated molecule are investigated.

2. FORMULATION OF THE PROBLEM

We consider the interaction of the hydrogen molecular ion H_2^+ with a radiation pulse whose electric field we give in the form $\varepsilon(t)=\varepsilon_0 f(t)\cos(\omega t)$, where $f(t)$ is the envelope normalized to unity at the maximum, $f_{\max}=f(0)=1$. The time-dependent wave function $\Psi(t)$ of the molecule in the field satisfies the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi=[H_0-\mathbf{d}\cdot\boldsymbol{\varepsilon}(t)]\Psi,$$

where H_0 is the Hamiltonian of the free molecule in the Born–Oppenheimer approximation, $\mathbf{d}$ is its dipole moment operator, and $\hbar=|e|=1$. We seek a solution of this equation in the form of an expansion in terms of the complete basis of wave functions of the free molecule:[14]

$$\Psi=\sum_{n,\Lambda,\nu,N} C_{n,\nu,N,\Lambda,M}(t)|n\nu N\Lambda M\rangle\exp(-iE_{n\nu N}t)$$
$$+\sum_{n',\Lambda,N}\int dE\, C_{n',E,N,\Lambda,M}(t)|n'EN\Lambda M\rangle$$
$$\times\exp(-iEt), \tag{1}$$

where ν is the vibrational quantum number, Λ is the projection of the electron angular momentum on the axis of the molecule, N is the total angular momentum quantum number, M is the projection of the total angular momentum in the direction of polarization of the field, and n are other term quantum numbers of the molecule. The first term on the right-hand side of the Eq. (1) takes account of the expansion in terms of all discrete levels of the coupled terms of the molecule, and the second is an expansion in continuum states of all terms, E is the energy of the continuum states, $E_{vN}^{(n)} = E_v^{(n)} + B_e^{(n)} N(N+1)$ is the energy of the vibrational–rotational states in the nth term, $E_v^{(n)}$ are the energies of the corresponding vibrational levels, and $B_e^{(n)}$ is the rotational constant of the nth term. We assume the initial state of the molecule to be one of the vibrational–rotational states of the electronic ground term $1s\sigma_g(^2\Sigma_g^+)$: $|gvNM\rangle = |g\rangle|v\rangle|NM\rangle$, where $|g\rangle$ is the electron wave function and $|v\rangle$ and $|NM\rangle$ are, respectively, the vibrational and angular functions of the nuclear motion. In the present paper we assume that in the initial state $v=1$, i.e., the molecule is in the lowest vibrational level of the ground term. For the ortho- modification of the molecular H_2^+ ion, only odd rotational quantum numbers, $N=1,3,5,\ldots$ are possible in the ground term, and for the para- modification only even numbers are possible, $N=0,2,4,\ldots$.[14,15] For a linearly polarized field the quantum number M is an integral of the motion. Therefore the problem can be solved separately for each number M, whereupon the results can be averaged with respect to the distribution function of the unperturbed molecules over the states $|NM\rangle$.

We examine single- and multiphoton dissociation of a molecule. In accordance with the Franck–Condon principle, for a k-photon transition to the repulsive term $2p\sigma_u(^2\Sigma_u^+)$ from the lowest vibrational level ($v=1$) of the ground-state term $1s\sigma_g(^2\Sigma_g^+)$ the frequency of the field ω must satisfy the condition $\omega \geqslant \omega_{\text{res}}^{(k)} = [U(R_e) - E_N]/k$, where $U(R_e)$ is the potential energy of the nuclei in the repulsive term, R_e is the equilibrium internuclear distance of the ground term, and $E_N \equiv E_{1N}^{(0)}$.

Substituting the expansion of the wave function (1) into the Schrödinger equation, we obtain a system of differential equations for the probability amplitudes $C_{n,v,N,\Lambda,M}(t)$ and $C_{n,E,N,\Lambda,M}(t)$. Using the procedure for adiabatic exclusion of the continuum,[16] we express $C_{n,E,N,\Lambda,M}(t)$ in terms of $C_{n,v,N,\Lambda,M}(t)$ and obtain the following system of equations for the probability amplitudes of finding a molecule in the vibrational–rotational states of the ground term $C_N(t) \equiv C_N^{(M)}(t)$ (where the index M is dropped to simplify the equations)

$$i\dot{C}_N(t) = \sum_{N'} C_{N'}(t) \exp\{i(E_N - E_{N'})t\} \times \left[-\frac{1}{4}\alpha_{NN'} f^2(t)\varepsilon_0^2 - \frac{i}{2} f^{2k}(t) \sum_k \Gamma_{NN'} \right], \quad (2)$$

where $\Gamma_{NN'}$ and $\alpha_{NN'}$ are, respectively, the tensor of dissociation widths and the polarizability (scattering) tensor, respectively. The ionization-width tensor can be determined in explicit form as the product of k-photon composite matrix elements of the dipole moment operator

$$\Gamma_{NN'} = 2\pi(-\mathbf{d}\cdot\boldsymbol{\varepsilon}_0/2)_{NE}^{(k)}(-\mathbf{d}\cdot\boldsymbol{\varepsilon}_0/2)_{EN'}^{(k)}\Big|_{E=E_N+k\omega}. \quad (3)$$

For single-photon dissociation, $k=1$, the components of the dissociation-width tensor $\Gamma_{NN'}$ and the polarizability tensor $\alpha_{NN'}$ can be expressed in terms of the real and imaginary parts of the composite second-order matrix elements $M_{N'N}^{(2)}$:

$$-\frac{1}{4}\alpha_{NN'}\boldsymbol{\varepsilon}_0^2 - \frac{i}{2}\Gamma_{NN'} = M_{NN'}^{(2)}$$

$$\equiv \left(\int dE + \sum_i \right) \frac{(-\mathbf{d}\cdot\boldsymbol{\varepsilon}_0)_{N,E(i)}(-\mathbf{d}\cdot\boldsymbol{\varepsilon}_0)_{E(i),N'}}{E(E_i) - E_N - \omega + i\delta}\Bigg|_{\delta\to 0}, \quad (4)$$

where the sum over i is a sum over all intermediate states of the molecule from which the integral over the energy E of the molecule in the repulsive term, the transition to which corresponds to dissociation, is singled out. According to Eq. (4), for $k=1$ $\Gamma_{NN'}$ and $-\alpha_{NN'}\varepsilon_0^2/4$ are of the same (second) order in the intensity of the field ε_0 and, generally speaking, are comparable in magnitude, and the actual ratio between the dissociation-width tensor and the amplitudes of the Raman transitions depends on the frequency of the field. In the present case at single-photon transitions from the lowest vibrational level of the molecule, in accordance with the Franck–Condon principle the dissociation width $\Gamma_{NN'}(\omega)$ of the level E_N has a maximum near exact resonance $\omega \sim \omega_{\text{res}} \equiv \omega_{\text{res}}^{(1)} = U(R_e) - E_N$. Under the same conditions the dispersion dependence of the Stark shift of the level E_N, $-\alpha_{NN}(\omega)\boldsymbol{\varepsilon}_0^2/4$, has a form similar to the derivative of $\Gamma_{NN'}(\omega)$ and, therefore, $|\alpha_{NN}(\omega)|\boldsymbol{\varepsilon}_0^2/4 \ll \Gamma_{NN'}(\omega)$ near ω_{res}.[17,18] Conversely, far from ω_{res} there exists a region where the shift of a level is relatively large, and $|\alpha_{NN}(\omega)|\boldsymbol{\varepsilon}_0^2/4 \gg \Gamma_{NN'}(\omega)$.[12,13] For multiphoton dissociation of a molecule ($k \geqslant 2$) the probability amplitudes of Raman transitions once again are of second order in the field intensity, while the components of the dissociation-width tensor are proportional to the field intensity to the power $2k$ and therefore, as a rule, $|\alpha_{NN}(\omega)|\boldsymbol{\varepsilon}_0^2/4 \gg \Gamma_{NN'}(\omega)$ in this case.

Note that there is a difference between situations involving short and long pulses. If the pulse duration is short compared with the reciprocal of the spacing between the nearest rotational levels, $\tau < 1/B_e$, where $B_e \equiv B_e^{(0)}$, then the radiation spectrum is broad, $\Delta\omega \sim 1/\tau > B_e$, and Raman transitions between different rotational levels can occur in strong and weak fields. In contrast, for pulses of long duration, $\tau > 1/B_e$, the spectral width is small, $\Delta\omega \sim 1/\tau < B_e$, and Raman transitions in a weak field are forbidden by energy conservation considerations. However, in a strong field this stricture is lifted and transitions can occur.

The criterion for a strong field is

$$\frac{1}{4}|\alpha_{N'N}(\omega)|\boldsymbol{\varepsilon}_0^2 > B_e. \quad (5)$$

This criterion is a common one, and does not depend on the pulse duration. Although for a short pulse ($\tau < 1/B_e$) Raman transitions can also occur in weaker fields, they are inefficient in that case. Raman transitions become efficient for large and small τ only if (5) is satisfied. Numerically, for the ground state of the molecular hydrogen ion H_2^+ (5) corresponds to field intensities $I \geqslant 10^{10}\,\mathrm{W \cdot cm^{-2}}$ or field intensities $\varepsilon_0 \geqslant 3 \times 10^6\,\mathrm{V \cdot cm^{-1}} \equiv \varepsilon_B$, where ε_B is the threshold intensity for the appearance of the strong-field effects considered below. This estimate of ε_B can be easily shown to be correct either by approximating the polarizability tensor $\alpha_{NN'}(\omega)$ by the static limit of its diagonal part $|\alpha_{NN}(\omega=0)| \approx 5$ (in atomic units) or by estimating $\alpha_{NN'}(\omega)$ by its maximum frequency dependence $|\alpha_{NN}(\omega)|_{\max} \approx 15^{12,13}$ for $B_e = 30\,\mathrm{cm}^{-1}$.[19]

Considering weak and strong fields [according to the criterion (5)], we assume that the field intensity is still bounded from above by the conditions

$$|\alpha_{NN'}|\boldsymbol{\varepsilon}_0^2 \ll \omega_e, \quad \Gamma_{NN'} \ll \omega_e, \tag{6}$$

where ω_e is the frequency of a vibrational quantum in the ground term of the molecule, $\omega_e = 2297\,\mathrm{cm}^{-1}$.[19] Since $|\alpha_{NN}(\omega)|_{\max} \approx 15^{12,13}$ and $[\Gamma_{NN'}(\omega)]_{\max} \approx 10$,[17,19] we find that (6) holds for $I \leqslant 10^{12}\,\mathrm{W \cdot cm^{-2}}$, which is obviously consistent with the condition for the strong-field regime (5). The condition (6) justifies the possibility of taking account of only transitions between rotational sublevels of the lowest vibrational level of the ground term of the molecule without excitation of the higher vibrational levels. This approximation is reflected explicitly in Eqs. (2), whose right-hand sides do not contain a summation over the vibrational states of the molecule. The region of even stronger fields, where (6) is not satisfied, is certainly of interest, but such an extension of the problem falls outside the scope of the present work.

Equations (2)–(4) can be further simplified by singling out in the matrix elements of the dipole moment the parts calculated in terms of the rotational functions of a free molecule, and assuming that the remaining factors are either weakly dependent on the rotational quantum number in the initial and final states and can be approximated by constants. In this approximation, all components of the ionization-width tensor and the polarizability tensor are characterized only by the constants β_0 and α_0, respectively,

$$\Gamma_{NN'}^{(M)} = \beta_0 D_{NN'}^{(M)} \varepsilon_0^2, \quad \alpha_{NN'}^{(M)} = \alpha_0 D_{NN'}^{(M)}. \tag{7}$$

For single-photon dissocation the coefficients $D_{NN'}^{(M)}$ are

$$D_{NN}^{(M)} = \frac{N^2 - M^2}{4N^2 - 1} + \frac{(N+1)^2 - M^2}{4(N+1)^2 - 1},$$

$$D_{NN'}^{(M)} = \sqrt{\frac{(N_{\max}^2 - M^2)((N_{\max}-1)^2 - M^2)}{(4N_{\max}^2 - 1)(4(N_{\max}-1)^2 - 1)}}\;\delta_{NN' \pm 2}, \tag{8}$$

where $N_{\max} = \max\{N, N'\}$. For three-photon dissociation of a molecule, naturally the expression (7) for the polarizability tensor $\alpha_{NN'}^{(M)}$ remains the same, and the formula for the dissociation-width tensor becomes

$$\Gamma_{NN'}^{(M)} = \tilde{\beta}_0 \tilde{D}_{NN'}^{(M)} \varepsilon_0^6, \tag{9}$$

where the coefficients $\tilde{D}_{NN'}^{(M)}$ are given in Table I. The constants α_0, β_0, and $\tilde{\beta}_0$ are not calculated in the present paper, but instead are given phenomenologically.

TABLE I.

D_{NN+3M}	$\sqrt{\frac{(N+1)^2-M^2}{4(N+1)^2-1}\,\frac{(N+2)^2-M^2}{4(N+2)^2-1}\,\frac{(N+3)^2-M^2}{4(N+3)^2-1}}$
D_{NN+1M}	$\sqrt{\frac{(N+1)^2-M^2}{4(N+1)^2-1}}\left(\frac{(N+2)^2-M^2}{4(N+2)^2-1}+\frac{(N+1)^2-M^2}{4(N+1)^2-1}+\frac{N^2-M^2}{4N^2-1}\right)$
D_{NN-1M}	$\sqrt{\frac{N^2-M^2}{4N^2-1}}\left(\frac{(N+1)^2-M^2}{4(N+1)^2-1}+\frac{N^2-M^2}{4N^2-1}+\frac{(N-1)^2-M^2}{4(N-1)^2-1}\right)$
D_{NN-3M}	$\sqrt{\frac{(N-1)^2-M^2}{4(N-1)^2-1}\,\frac{(N-2)^2-M^2}{4(N-2)^2-1}\,\frac{N^2-M^2}{4N^2-1}}$

The solutions $C_N^{(M)}(t)$ of the system (2) can be used to find as a function of time t the probability density of the orientation of the molecular axis in any given direction characterized by the angles θ and φ of a spherical coordinate system with z axis oriented parallel to the polarization vector of the field ε_0:

$$P_N^{(M)}(t,\theta,\varphi) = \left|\sum_{N'} C_{N'}^{(M)}(t) Y_{N'M}^*(\theta,\varphi)\right|^2, \tag{10}$$

where as before the quantum numbers N and M characterize the initial state of the molecule. If the system under study is initially an ensemble of molecules with distribution function $\rho_N^{(M)}$ over the quantum numbers N and M, then the time-dependent distribution function over the directions of the axes of the molecules in the ensemble is

$$\begin{aligned} P(t,\theta,\varphi) &= \sum_{NM} \rho_N^{(M)} P_N^{(M)}(t,\theta,\varphi) \\ &= \sum_{NM} \rho_N^{(M)} \left|\sum_{N'} C_{N'}^{(M)}(t) Y_{N'M}^*(\theta,\varphi)\right|^2. \end{aligned} \tag{11}$$

Using the solutions of the system of equations (2), the angular distribution function of the dissociation products of the ensemble of molecules can also be found similarly to Eq. (11):

$$P_{fr}(\theta,\varphi) = \sum_{NM} \rho_N^{(M)} \int dE \left|\sum_{N'} C_{EN'}^{(M)}(T) Y_{N'M}^*(\theta,\varphi)\right|^2. \tag{12}$$

Here $C_{EN}^{(M)}(t)$ are the probability amplitudes of finding a molecule at time t in continnum states of the repulsive term with energy E and quantum numbers of the angular momentum and its projection N and M:

$$\begin{aligned} C_{EN}^{(M)}(t) = -i \int_0^t dt' \sum_{N'} \exp\{i(E - E_{N'} - \omega)t'\} \\ \times \left(-\frac{\mathbf{d} \cdot \boldsymbol{\varepsilon}_0}{2}\right)_{ENM;N'M}^{(k)} C_{N'M}(t'), \end{aligned} \tag{13}$$

where $(-\mathbf{d} \cdot \boldsymbol{\varepsilon}_0/2)_{ENM;N'M}^{(k)}$ is the k-photon matrix element of a dipole transition from the state of the ground term $|g, v$

$=1,N',M\rangle$ to the continuum state $|u,E,N,M\rangle$; the time T in the argument of $C_{EN}^{(M)}$ on the right-hand side of Eq. (12) is the moment at which the pulse ceases.

3. ROTATIONAL QUASIENERGY STATES OF A MOLECULE AND ALIGNMENT EFFECTS WITH THE INTERACTION TURNED ON ADIABATICALLY

For a constant, time-independent envelope of the field $f(t)\equiv 1$, the system of equations (2) can be transformed to a system of equations with constant coefficients for the functions $A_N(t)=\exp(-iE_Nt)C_N(t)$

$$i\dot{A}_N(t)-E_NA_N(t)=\sum_{N'} A_{N'}(t) \times\left(-\frac{1}{4}\alpha_{NN'}\varepsilon_0^2-\frac{i}{2}\sum_k \Gamma_{NN'}\right), \tag{14}$$

which obviously has a solution of the form $A_N(t)=\exp(-i\gamma t)a_N$, where a_N is a constant and γ are complex quasienergies of the system. The equations for the coefficients a_N have the form

$$\left(\gamma-E_N+\frac{\alpha_0+2i\beta_0}{4}\varepsilon_0^2D_{N,N}\right)a_N =-\frac{\alpha_0+2i\beta_0}{4}\varepsilon_0^2(D_{N,N+2}a_{N+2}+D_{N,N-2}a_{N-2}), \tag{15}$$

where the approximate formulas (7) were used for the polarizability and ionization-width tensors. The single-photon dissociation channel is assumed to be open, and in this case the constants $D_{N,N'}$ are determined by Eqs. (8).

As noted in the Introduction, for single-photon dissociation of a molecule the ratio of the constants α_0 and β_0 depends on the frequency of the external field. For $\alpha_0\ll\beta_0$ the perturbation of the levels of the molecule in a field is determined mainly by the dissociation broadening, and this case has been investigated in detail in our previous work.[7] In the present paper this case is not studied, since appreciable alignment of molecules in a strong field does not occur with strong dissociation broadening of the levels. From this standpoint, the opposite case, where the perturbation of the rotational levels of a molecule as a result of direct two-photon Raman transitions is much greater than their dissociation broadening, $|\alpha_{NN'}^{(M)}|\gg|\Gamma_{NN'}^{(M)}|$ or $\alpha_0\gg\beta_0(\tilde{\beta}_0)$, is much more interesting. These equations (15) are equivalent to the equations for the expansion coefficients of the prolate spheroidal function of the first kind in spherical harmonics.[8,9,20]

The solutions of the system of equations (15) determine the expansion coefficients a_N of the quasienergy wave functions Ψ_γ in terms of the rotational states of a free molecule and the field-dependent quasienergies γ of the system. The real and imaginary parts of the quasienergies determine the positions and widths of the quasienergy levels. Although the quasienergy states are a superposition of the rotational states of a free molecule with different N, it is nonetheless convenient to characterize each quasienergy state by a rotational quantum number N_0 of the level of the free molecule (E_{N_0}) into which a given quasienergy $\gamma_{N_0}(\varepsilon_0)$ passes in the formal limit of zero field intensity, $\varepsilon_0\to 0$. The value and desirability of solving for the quasienergies and quasienergy wave functions of a molecule are due to the fact that when the field is turned on sufficiently slowly (adiabatically), i.e., for sufficiently long pulse durations (see below for estimates), the system always remains in a quasienergy state, satisfying a given initial condition, and its characteristics at any time are determined by the parameters of this quasienergy state, in which it is only necessary to replace ε_0 by a weakly time-dependent envelope $\varepsilon_0 f(t)$ of the light pulse. In other words, in the adiabatic approximation a replacement of the parametric dependence of γ and a_N on ε_0 by a dependence on $\varepsilon_0 f(t)$ makes it possible to solve the initial problem, i.e., to find the solution $\Psi(t)=\Psi_\gamma[\varepsilon_0 f(t)]$ of the Schrödinger equation that satisfies the initial conditions $\gamma[\varepsilon_0 f(t)]\to E_{N_0}$ and $\Psi_\gamma \to|N_0,M_{N_0}\rangle$ as $t\to-\infty$.

If the system is in a quasienergy state Ψ_γ, then in particular the probability density $P^{(\gamma)}$ of the orientation of the axis of the molecule in this state can also be determined using Eq. (10):

$$P^{(\gamma,M)}(t,\theta,\varphi)=\left|\sum_N a_N^{(\gamma,M)}[\varepsilon_0 f(t)] \times\exp\{-i(E_N+\gamma)t\}Y^*_{NM}(\theta,\varphi)\right|^2. \tag{16}$$

For an interaction turned on adiabatically, the quasienergy γ and the corresponding coefficients $a_N^{(\gamma,M)}$ in Eq. (16) are chosen from the solutions of Eqs. (15) and the initial conditions $\gamma[\varepsilon_0 f(t)]\to E_{N_0}$ and $\Psi_\gamma\to|N_0,M_{N_0}\rangle$ as $t\to-\infty$. Assuming the adiabaticity conditions to be satisfied and, in addition, $|\alpha_{NN'}^{(M)}|\gg|\Gamma_{NN'}^{(M)}|$ ($\alpha_0\gg\beta_0$), and $\Gamma_{NN'}^{(M)}T\ll 1$ we solve Eq. (15) and investigate the dynamics of the alignment of an ensemble of molecules excited by a field into quasienergy states Ψ_γ.

We assume as usual that the object of investigation is an equilibrium molecular gas at room temperature. Under these conditions only rotational states with small N of the molecules are populated, and for the ortho- configuration of the molecular ion H_2^+ only the states with $N=1$ are populated, all three sublevels with $M=1$, 0, and -1 being equally populated, $\rho_1^{(M)}=1/3$. In accordance with Eq. (11) the distribution function over directions of the axes of such an ensemble of molecules has the form

$$P(t,\theta,\varphi)=\frac{1}{3}[P_1^{(1)}(t,\theta,\varphi)+P_1^{(0)}(t,\theta,\varphi) +P_1^{(-1)}(t,\theta,\varphi)], \tag{17}$$

where the functions $P_1^{(M)}(t,\theta,\varphi)$ must be found from Eqs. (16). In the limit $t\to-\infty$, $P(t,\theta,\varphi)\Big|_{t\to-\infty}=\text{const}=1/4\pi$ follows from Eq. (17). This explicitly confirms that the distribution of the molecular axes in the unperturbed state of the ensemble is isotropic.

Figure 1 shows the directional distribution function $P(\theta)$ (17) of the axes of the ensemble of molecules under

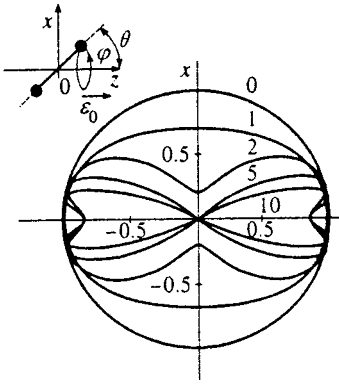

FIG. 1. Angular distribution of the axes of an ensemble of molecules in a radiation field. The distribution results from adiabatic switching of the interaction. The numbers show the field intensity ε_0 in units of ε_B.

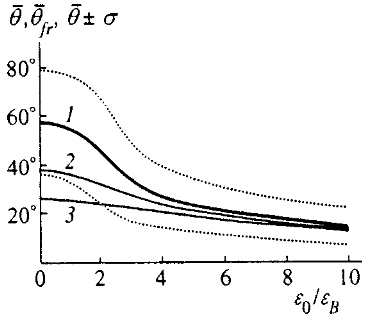

FIG. 2. Average angle $\overline{\theta}(\varepsilon_0)$ between the axis of a molecule and the field intensity with the field turned on adiabatically (curve *1*) and the average emergence angle $\overline{\theta}_{fr}(\varepsilon_0)$ of the fragments with single-photon (curve *2*) and three-photon (curve *3*) dissociation; dotted curves—$\overline{\theta}(\varepsilon_0) \pm \sigma(\varepsilon_0)$.

study after the interaction is turned on adiabatically (i.e., at $t=0$, when the field in the pulse has reached its maximum value), calculated for several peak values of the field intensity, $\varepsilon_0/\varepsilon_B = 0$, 1, 2, 5, and 10. We note that because of axial symmetry the distribution function P does not depend on the azimuthal angle φ (in a plane perpendicular to $\boldsymbol{\varepsilon}_0$). According to the figure, as the field increases the distribution function $P(\theta)$ becomes narrower and is localized at small values of θ and $\pi-\theta$, i.e., close to the direction of field polarization. In addition, in a strong field a dip also appears in the distribution function precisely at $\theta=0$ and $\theta=\pi$. This dip becomes increasingly noticeable as the field intensity increases.

The distribution function $P(\theta)$ found above can be used to calculate average values, for example, the average angle $\overline{\theta}$ between the axis of a molecule and the direction of polarization of the field and the quantum variance $\sigma=[\overline{\theta^2}-(\overline{\theta})^2]^{1/2}$, where

$$\overline{\theta^n} = \int_0^{\pi/2} \theta^n P(\theta) \sin\theta \, d\theta \Bigg/ \int_0^{\pi/2} P(\theta) \sin\theta \, d\theta. \tag{18}$$

We note that the integration in Eq. (18) extends from 0 to $\pi/2$ and not from 0 to π, since the positions of the molecule with the axis making angles θ and $\pi-\theta$ with $\boldsymbol{\varepsilon}_0$ are assumed to be equivalent. It is also interesting to note that in the absence of a field (for $P=\text{const}$) $\overline{\theta}=1$ rad and $\sigma \approx 0.4$ rad. In this case these values of $\overline{\theta}$ and σ are independent of the choice of direction for the quantization axis.

Figure 2 shows the average angle $\overline{\theta}$ and the quantities $\overline{\theta} \pm \sigma$ as a function of the peak amplitude ε_0 of the field in a pulse with the interaction turned on adiabatically. The average angle $\overline{\theta}(\varepsilon_0)$ decreases with increasing field, and in the strong-field limit it reaches a constant value of only several degrees. The relative standard deviation of the angle θ from its average value remains more or less constant, $\sigma(\varepsilon_0)/\overline{\theta}(\varepsilon_0) \approx 40\%$. Therefore, in a strong field the average angle between the axes of the molecules and the direction of ε_0 as well as the standard deviation σ of the average angle are substantially less than their values for a weak field. This indicates unequivocally that the axes of the molecules are aligned in the direction of polarization of a strong light field.

The alignment of the molecules in the direction of polarization of a strong field can largely determine the angular directivity of the dissociation fragments of the molecule. Photodissociation itself can serve as a competing mechanism for the narrowing of the directional pattern of the molecular fragments. As is well known,[19,5,21] even in a weak field and with an isotropic distribution of the directions of the molecular axes, the directional pattern of the fragments in single-photon dissociation as calculated in first-order perturbation theory is governed by the factor $\cos^2\theta_{fr}$, where θ_{fr} is the angle between the direction of emergence of the fragments and the field intensity vector $\boldsymbol{\varepsilon}_0$. For three-photon dissociation the corresponding factor is $\cos^6\theta_{fr}$, an extremely narrowly directed function.

In the theoretical model under study the distribution function $P(\theta_{fr})$ of the fragments of the molecule over the directions of emergence and the average emergence angle $\overline{\theta}_{fr}$, determined similarly to Eq. (18), were calculated using Eqs. (13) and (14) for single- and three-photon dissociation. The calculation was performed in the adiabatic approximation used in the present section. The computed dependence of the average emergence angle on the peak field intensity ε_0 in the pulse is presented in Fig. 2. As one can see from the figure, the average emergence angle $\overline{\theta}_{fr}(\varepsilon_0)$ of the fragments of a molecule is always less than the average angle $\overline{\theta}(\varepsilon_0)$ between the axes of the undissociated molecules and $\boldsymbol{\varepsilon}_0$. For one- and three-photon ionization, in the weak field limit $\overline{\theta}_{fr} \approx 2\overline{\theta}/3$ and $\overline{\theta}_{fr}=0.45\overline{\theta}$, respectively. In a strong field $\overline{\theta}_{fr} \approx \overline{\theta}$ in both cases. This result essentially suggests a change in the mechanism whereby the directional pattern of the dissociation products becomes narrower as the field increases. In a weak field the distribution of the axes of the molecules over the angles θ is almost isotropic, and the narrowing of the directional pattern of the dissociation products results solely from the dissociation process itself, i.e., the transition to the continuum. In a strong field the dominant mechanism becomes alignment of the undissociated molecules along the field. The degree of directivity of the dissociation products is essentially the same as the degree of orientation of the axes of the molecules.

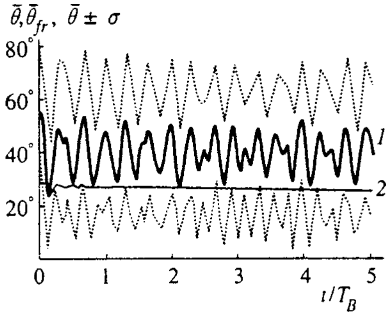

FIG. 3. Time-dependence of the average angle $\overline{\theta}(t)$ between the axis of the molecules and the field intensity (curve *1*) and the average emergence angle $\overline{\theta}_{fr}(t)$ of the fragments with single-photon dissociation (curve *2*) with the interaction turned on instantaneously with $\varepsilon_0 = 10\varepsilon_B$; dotted curves—$\overline{\theta}(t) \pm \sigma(t)$.

4. SOLUTION OF THE INITIAL PROBLEM AND ALIGNMENT EFFECTS INDUCED BY VARIOUSLY-SHAPED PULSES

4.1. Square pulse

We consider first the often used model of a square pulse, i.e., the interaction is turned on and off instantaneously. Solving Eqs. (2) and (4) and assuming once again that initially the ensemble of molecules is characterized by an isotropic directional distribution function of the axes of the molecules, we find using Eqs. (11) and (18) the time-dependent average angle $\overline{\theta}(t)$ between the axis of a molecule and the polarization direction of the field and the quantum variance $\sigma(t)$ fo this angle. Figure 3 shows the functions $\overline{\theta}(t)$ and $\overline{\theta}(t) \pm \sigma(t)$ calculated for a square pulse of duration $T = 5T_B$ (where $T_B = 1/B_e$) at peak field strength seven times the threshold value for the appearance of strong-field effects, $\varepsilon_0 = 7\varepsilon_B$ (see Eq. (5) for the definition of ε_B). As is evident from the figure, the average angle $\overline{\theta}(t)$ undergoes complicated nonharmonic but periodic oscillations between the initial value $\overline{\theta}(0) = 1$ rad $\approx 57°$ and the minimum value $\approx 27°$ with frequencies equal to the differences of the occupied quasienergy states. Expanding the function $\overline{\theta}(t)$ in a Fourier (integral) series, we find that oscillations with frequencies equal to the differences of the quasienergies of the quasienergy states of the molecule that are populated initially with the interaction turned on instantaneously are represented in the spectrum of the periodic function $\overline{\theta}(t)$. However, the large quantum variance of the angle θ, much greater than the amplitude of the oscillations of the average angle, can make it very difficult to observe these oscillations. We note that the observed large quantum variance of the angle θ gives a far from obvious possibility of describing the rotations of a molecule in a field using the classical theory,[12,13] where the dependence $\overline{\theta}(t)$ is a strictly deterministic function.

Since the dissociation of a molecule (by assumption, slow, $\Gamma_{M,M}(T) \ll 1$) occurs during the entire time of the interaction with the field, the oscillations in the time dependence $\overline{\theta}_{fr}(t)$ of the average emergence angle of the decay products of the molecule are smoothed out as t increases, as is clearly seen in Fig. 3.

4.2. Gaussian pulse

The model with the interaction turned on instantaneously is attractive solely because of its simplicity. It does not correspond to the envelope of real laser pulses, which is best approximated by a Gaussian function $f(t) = \exp[-(t/T)^2]$. For sufficiently long pulse duration T, the solution of the problem should correspond to the adiabatically switched model studied in the preceding section. However, for values of T comparable to $T_B = 1/B_e$ the deviations from adiabaticity can be appreciable. In the language of quasienergy functions a deviation from adiabaticity means that on the leading edge of the pulse several quasienergy states of the molecule in the field, and not only one state as in the adiabatic case, become occupied as the field is turned on. Solving the equation for the time-dependent coefficients in the expansion of the wave function of the molecule in the field of a Gaussian pulse, it can be shown that the adiabaticity parameter is

$$n_0 = \frac{(6+4N_0)B_e T}{\pi}\left[\ln\left(\frac{|\alpha_0 + 2i\beta_0|\varepsilon_0^2}{(6+4N_0)B_e}\right)\right]^{-1/4}, \tag{19}$$

which can be interpreted as the number of oscillations of functions of the type $\theta(t)$ on the leading edge of a Gaussian pulse. For $n_0 < 1$ there is not enough time for even one oscillation of $\overline{\theta}(t)$ to occur as the leading edge is turned on. In this case the interaction is turned on abruptly and the process resembles more an instantaneously switched process. Many quasienergy states of the molecule become occupied as the interaction is turned on. In contrast, for $n_0 \gg 1$ the switching time is long, and many oscillations of the function $\overline{\theta}(t)$ occur during this time, but the amplitude of these oscillations is extremely small, and the switching process is nearly adiabatic. These behaviors are characterized by the computational results presented in Fig. 4, which shows the functions $\overline{\theta}(t)$ and $\overline{\theta}(t) \pm \sigma(t)$ calculated in the same model as in all preceding calculations but for a Gaussian pulse with duration $T = 0.1T_B$, T_B, and $5T_B$ and peak field strength $\varepsilon_0 = 10\varepsilon_B$. For $T = 5T_B$, according to Eq. (19), $n_0 \approx 8$, i.e., the pulse is turned on gradually, essentially adiabatically. The average angle $\overline{\theta}(t)$ decreases with time t to the minimum value, which is reached at $t = 0$, i.e., at the maximum of the time-dependent amplitude $\varepsilon_0 f(t)$ of the field. For small values of n_0 the deviations from adiabaticity are large. In short and long pulses the quantum variance σ is greater than the amplitude of the oscillations of the average angle $\overline{\theta}(t)$.

In contrast to a pulse with a rectangular envelope, in a Gaussian pulse the dissociation of molecules occurs mainly near the field maximum at $t = 0$. Figure 5 shows as a function of the field strength the pulse-averaged angle of the axes of an ensemble of molecules at the center of the pulse and the average emergence angles of the photodissociation products for single- and three-photon decay of molecules in an ensemble for several pulse durations. For very short pulses, $n_0 \ll 1$, the difference from perturbation theory is small. For very long pulses, $n_0 \gg 1$, the difference from the adiabatic switching model is small. The angular distribution of the axis of a molecule at the center of a pulse and the angular distributions of the fragments of a molecule for single- and three-

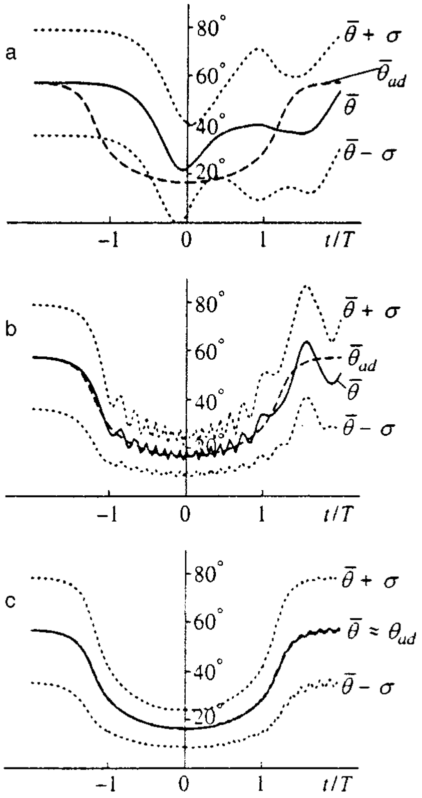

FIG. 4. Time-dependence of the average angle $\overline{\theta}(t)$ between the axis of the molecules and the intensity of the field for a Gaussian pulse with peak intensity $\varepsilon_0 = 10\varepsilon_B$ and pulse duration $T = 0.1T_B$ (a), $T = T_B$ (b), and $T = 5T_B$ (c); dotted curves—$\overline{\theta}(t) \pm \sigma(t)$; dashed curves—$\overline{\theta}(t)$ in the model with the interaction turned on adiabatically; $\overline{\theta}_{ad}(t)$—average angle between the field intensity and the axis of molecules in a quasienergy state with an adiabatically slowly varying field amplitude $\varepsilon_0(t)$.

photon decay are essentially the same as the corresponding distributions with the field turned on adiabatically. For sufficiently long pulses, $n_0 \gg 1$, two regions can be singled out in the functions $\overline{\theta}(\varepsilon_0)$, $\overline{\theta}_{fr}(\varepsilon_0)$, and $\theta_{fr}^{(3)}(\varepsilon_0)$: the region of relatively weak fields, where the properties of the angular distribution of the fragments are determined by the orienting properties of the phototransition to the continuum of the unstable term, and the region of strong fields, where the angular distribution of the fragments is the same as the angular distribution of the axes of the molecules in the ensemble. Here $\overline{\theta}_{fr}(F) \approx \overline{\theta}(F)$ (Fig. 5). We note that, other conditions being the same, as the fraction ΓT of molecules that decay during a pulse increases, the width of the angular distribution of the fragments over the emergence angles increases.

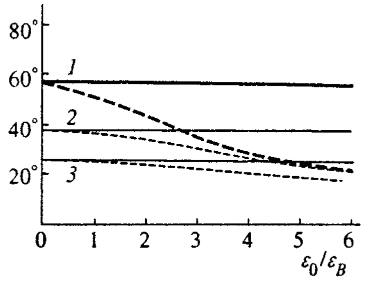

FIG. 5. Average angle $\overline{\theta}(t)$ between the axis of a molecule and the direction of the field intensity (*1*) and the average emergence angle of the fragments for single-photon (*2*) and three-photon (*3*) dissociation for short (solid lines) and long (dashed lines) pulses, for which, respectively, $n_0 \ll 1$ and $n_0 \gg 1$, calculated for a Gaussian pulse at $t = 0$ as a function of the peak field strength $\varepsilon_{\max} = \varepsilon_0(t = 0)$.

5. CONCLUSIONS

The main results of this work are the conclusion that substantial alignment of molecules in the direction of polarization of the field is possible and a formulation of the corresponding conditions. It was shown that alignment can be substantial only if the field is turned on adibatically, i.e., if the duration of a pulse with a smooth envelope is sufficiently long. Then the alignment is maximum at the maximum of the field. The second important condition for alignment is that the dissociation broadening of the rotational levels of a molecule must be relatively small compared with the perturbation of the levels as a result of direct two-photon Raman transitions.

The second important result of this work concerns the angular distribution of the products of dissociation of a molecule. It was shown that the narrowing of the directional pattern of the dissociation fragments and its orientation in the direction of polarization of the field can be determined by two factors: the orientation of the axes of the molecules in the direction of a strong field and by the transition itself from a bound state of the molecule to the continuum of an unstable term. If the conditions of alignment of molecules in a bound state as formulated above are not satisfied, then the directivity of the dissociation products of the molecules can be determined only by the properties of the transition to the continuum. Specifically, the directivity of the angular distribution of the fragments along the field increases with the number of photons involved in the transition to the repulsive term. However, if the strong field satisfies the conditions of substantial alignment of the axes of the molecules in their bound state, then the angular distribution of the dissociation products is essentially the same as the angular distribution of the axes of the molecules.

These results can be compared qualitatively with experimental data.[2] For example, under the experimental conditions of Ref. 2 the pulse duration was short, so that $n_0 \ll 1$ (Cl_2, $T \approx 100$ fs, $B \approx 1.7\,\mathrm{cm}^{-1}$, $\alpha \approx 30$ a.u.). Therefore, in our view, the directivity of the angular distribution of the dissociation products of the molecules along the field was determined only by the orienting properties of the multiphoton transition to the photoionization continuum with formation of a molecular ion and a subsequent transition to the photodissociation continuum with decay of the molecules into fragments. Alignment of the axis of a molecule and the axis of a molecular ion as a result of Raman-type transitions before transitions to a state in the continuum was small. This is indicated by the dependence of the alignment on the pulse duration, the field intensity, and the wavelength. According to the data of Ref. 2, variations in the pulse duration and field intensity had only a minor effect on the angular distribution of the dissociation fragments, and an increase in wavelength narrowed the angular distribution. The latter indicates directly that an increase in the number of photons involved in the transition has a controlling effect on the angular distribution of the fragments.

We reiterate that our analysis is limited to fields that are not too strong (4). To go beyond these restrictions it is necessary to take account of the possibility of excitation of molecular vibrations, which greatly complicates the problem.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 99-02-18034).

*)E-mail: fedorov@gon.ran.gpi.ru

---

[1] A. Zavriyev, P. H. Bucksbaum, H. G. Muller, and D. W. Schumacher, Phys. Rev. A **42**, 5500 (1990).

[2] D. Normand, S. Dobosz, M. Lezius, P. D'Oliveira, and M. Schmidt, in *Multiphoton Processes 1996*, P. Lambropoulos and H. Walther (eds.), IOP, Bristol (1997), p. 287.

[3] V. P. Makarov and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **70**, 1185 (1976) [Sov. Phys. JETP **43**, 615 (1976)].

[4] M. V Fedorov, Zh. Éksp. Teor. Fiz. **73**, 134 (1977) [Sov. Phys. JETP **46**, 69 (1977)].

[5] A. D. Bandrauk and G. Turcott, J. Phys. Chem. **87**, 5098 (1983).

[6] A. D. Bandrauk, J. M. Gauthier, and E. Aunabel, Phys. Rev. A **48**, 2145 (1993).

[7] A. I. Andriushin and M. V. Fedorov, Laser Phys. **7**, 280 (1996).

[8] B. A. Zon and B. G. Katsnel'son, Zh. Éksp. Teor. Fiz. **69**, 1166 (1975) [Sov. Phys. JETP **42**, 595 (1975)].

[9] B. Friedrich and D. Herschbach, Phys. Rev. Lett. **74**, 4623 (1995).

[10] E. Charon, A. Giusti-Suzor, and F. H. Mies, Phys. Rev. A **49**, R641 (1994).

[11] A. Giusti-Suzor, F. H. Mies, L. F. DiMauro, E. Charron, and B. Yang, J. Phys. B **28**, 309 (1995).

[12] M. E. Sukharev and V. E. Kraĭnov, Zh. Éksp. Teor. Fiz. **113**, 573 (1998) [JETP **86**, 318 (1998)].

[13] B. A. Zon, private communication (1999).

[14] G. H. Dunn, Phys. Rev. **172**, 1 (1968).

[15] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atomic and Molecular Physics* [in Russian], Atomizdat, Moscow (1980).

[16] M. V. Fedorov and A. E. Kazakov, Prog. Quantum Electron. **13**, 1 (1989).

[17] J. M. Schulman and R. Detrano, J. Chem. Phys. **62**, 2111 (1975).

[18] A. V. Gol'tsov and I. I. Tugov, Trudy FIAN **146**, 76 (1984).

[19] N. Zare, *Angular Momentum. Understanding Spatial Aspects in Chemistry and Physics*, Wiley, New York (1988).

[20] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions*, Dover, New York (1965).

[21] S. J. Singer, K. F. Freed, and Y. B. Band, J. Chem. Phys. **79**, 6060 (1983).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Subbarrier conversion in $^{125}\mathrm{Te}^{45+}$

F. F. Karpeshin

Institute of Physics, St. Petersburg State University, 198904 St. Petersburg, Russia

I. M. Band and M. B. Trzhaskovskaya*)

St. Petersburg Nuclear Physics Institute, Russian Academy of Sciences, 188350 Gatchina, Leningrad Region, Russia

(Submitted 12 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1565–1574 (November 1999)

We discuss the process of resonant subbarrier internal conversion of γ rays, where the converted electron is transferred to one of the atomic orbitals. For the first time we study how this process is affected by residual interactions: the splitting of the atomic terms in the total angular momentum of the atom, configuration mixing, and the magnetic interaction between the electrons of the atom. The calculations are done by the relativistic multiconfiguration Dirac–Fock method with allowance for the Breit interaction for the $M1$ transition with an energy of 35 492 eV in multiply charged ions of ^{125}Te. We show that allowing for the residual interaction is obligatory if we want to calculate the conversion rate in the vicinity of a resonance correctly. *© 1999 American Institute of Physics.* [S1063-7761(99)00511-9]

1. INTRODUCTION

The use of multiply charged ions in modern accelerators allows observing processes that are, at least in principle, possible and have been studied theoretically, but are extremely rare in neutral atoms. One of these exotic processes is the subbarrier (or discrete or resonant) internal conversion of γ rays, where the electron is (virtually) transferred to one of the discrete levels of the atom. Such a process was considered in Ref. 1 and 2. Since discrete conversion is definitely a resonant process, it was suggested in Ref. 2 that resonant conversion could be used to accelerate nuclear transitions by balancing the resonance defect by an external field of laser radiation of appropriate frequency.

A different experimental study of the $M1$ transition with an energy of $35\,491.9 \pm 0.5$ eV in multiply charged ions of ^{125}Te was carried out at the University of Bordeaux.[3,4] In the neutral atom this transition occurs primarily due to conversion on the K-shell, and the corresponding internal conversion coefficient α_K^{M1} is equal to 11.6. As the shell becomes ionized, the electron binding energy increases, and at ionization multiplicity $q=45$ ordinary conversion on the K-shell becomes impossible. However, the expected decrease in lifetime was not observed in experiments: it remained constant even at $q=46$; only with further ionization does the lifetime exhibit a smooth increase. This experiment stimulated a theoretical investigation[4] of the process of subbarrier resonant conversion that occurs at $q \geqslant 45$. The Feynman diagram of this process is depicted in Fig. 1. We see that the electron is (virtually) transferred to one of the states belonging to the discrete spectrum, which then decays via a radiative electron (or hole) transition.

In the present paper we refine the calculation done in Ref. 4 by allowing for the residual interactions. On particular, we allow for the splitting of atomic terms in the total angular momentum of the atom, take into account the magnetic Breit interaction between the atomic electrons, and allow for configuration mixing. As expected, allowing for these effects strongly influences the resonant conversion rate.

2. ALLOWANCE FOR THE MAGNETIC INTERACTION OF ELECTRONS AND THE SPLITTING OF TERMS IN TOTAL ANGULAR MOMENTUM OF THE ATOM

The rate of resonant conversion depends very strongly on the energy of the intermediate electron state (see Fig. 1). Hence in calculations it is very important to allow for the splitting of atomic levels in the total angular momentum and for the magnetic Breit interaction (in addition to the Coulomb interaction). The effect of the Breit interaction on the energy of $1s \to ns$ transitions in the $^{125}\mathrm{Te}^{45+}$ ion is illustrated by Table I. The first column lists the values of the energies of $1s \to ns$ transitions, $n=9-18$, calculated by the Dirac–Fock method for the median (''average'') level ($\mathrm{DF_{av}}$) without allowance for the Breit interaction and the total angular mo-

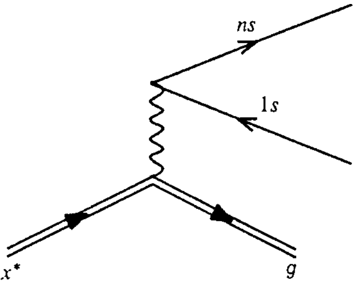

FIG. 1. Feynman diagram for discrete conversion. A hole forms in the $1s$ shell, and the corresponding electron goes to the ns state. The double line corresponds to a nuclear transition from the excited (ex^*) state to the ground (g) state. The arrow corresponding to the hole state is directed against the flow of time.

TABLE I. Energies of $1s \to ns$ transitions calculated by the $\mathrm{DF_{av}}$ method without allowance for the splitting in the total angular momentum J of the atom, by the $\mathrm{DF_{td}}$ with allowance for the splitting $(E(J_1,J_2))$, and by the $\mathrm{DF_{td}}$ method with allowance for the magnetic interaction $(MI(J_1,J_2))$. The energies are in electronvolts and are measured from 35000 eV.

ns	$\mathrm{DF_{av}}$	$E(3/2,5/2)$	$E(1/2,3/2)$	$MI(3/2,5/2)$	$MI(1/2,3/2)$
9*s*	331	320	349	252	283
10*s*	402	391	419	323	354
11*s*	454	443	471	375	406
12*s*	493	483	511	414	445
13*s*	524	513	541	445	476
14*s*	548	537	563	469	500
15*s*	568	557	585	488	519
16*s*	584	573	601	504	535
17*s*	597	586	614	517	548
18*s*	608	597	625	528	559

mentum of the atom. This approximation corresponds to the model used in Ref. 4. The total angular momentum J of the method can be taken into account by using the term-dependent Dirac–Fock method ($\mathrm{DF_{td}}$) or the multiconfiguration Dirac–Fock method (MCDF). Such calculations lift the degeneracy in the total angular momentum J of the atomic terms. As a result, the excited ns levels split into two doublets, each of which remains almost doubly degenerate. The levels of the lower doublet with the angular momenta $J = 3/2$ and 5/2 differ by roughly 0.1 eV in energy. The levels of the upper doublet with angular momenta $J = 1/2$ and 3/2 are separated by a narrower gap, ~ 0.02 eV.

The energies of the $1s \to ns$ transitions calculated by the different methods are compared in Table I. In addition to the energies obtained by the $\mathrm{DF_{av}}$ method, the third and fourth columns list the energies of the doublets calculated by the $\mathrm{DF_{td}}$ method. Since inside each doublet the transition energies do differ by an amount smaller than the accuracy of the data in Table I, these components are paired. The corresponding values of J are listed in the first row of the table. The energies of the transitions calculated by the $\mathrm{DF_{td}}$ method with allowance for the Breit interaction are listed in the fifth and sixth columns.

The results presented in Table I show that allowance for the residual interaction radically changes the arrangement of the electron levels. Splitting of the levels is almost constant and amounts to 27–31 eV. This value exceeds the separation of median levels, so that the levels belonging to neighboring doublets overlap. The magnetic interaction increases the total energy E_{tot} of the ground state of the $^{125}\mathrm{Te}^{45+}$ ion (with the electron configuration $1s^2 2s^2 2p_{1/2}^2 2p_{3/2}$) by 90.7 eV. For highly excited states with the electron configuration $1s2s^2 2p^2 2p_{3/2} ns$, the energy E_{tot} increases by roughly 22 eV. Thus, the entire energy of transitions for ns states with large values of n decreases by roughly 70 eV, which is much larger than the level separation.

3. INTERACTION OF CONFIGURATIONS

What we said in Secs. 1 and 2 readily implies that the rate of discrete conversion strongly depends on the level density, which usually increases significantly when the mixing of configurations is taken into account. As noted earlier, in the case at hand the allowance for the interaction of configurations causes splitting of each ns state into two levels in the final configurations $(1s2s^2 2p_{1/2}^2 2p_{3/2} ns)$, and each component of the split levels remains almost doubly degenerate. In the present work the configuration interaction is taken into account within the MCDF method.

Using our RAINE package of computer programs,[5] devised for relativistic calculations of atomic structures and the interaction of electromagnetic radiation and the nucleus with the atomic electrons, we developed a program that realized the MCDF method. The main concepts were taken from the theory of Grant *et al.*,[6] but in our calculations by the MCDF method we took for the base functions the Slater determinants corresponding to states with a well-defined projection M of the total angular momentum J. The energy levels of the atom are determined by diagonalizing the Hamiltonian, and each level corresponds to an eigenfunction of the operator of total angular momentum, i.e., to a definite value of the angular momentum J.

There are three types of configuration interaction.[6] Figure 2a depicts the Feynman diagram of the mixing amplitude, which occurs due to the one-particle operator of kinetic energy and the Coulomb interaction with the nucleus. Figure

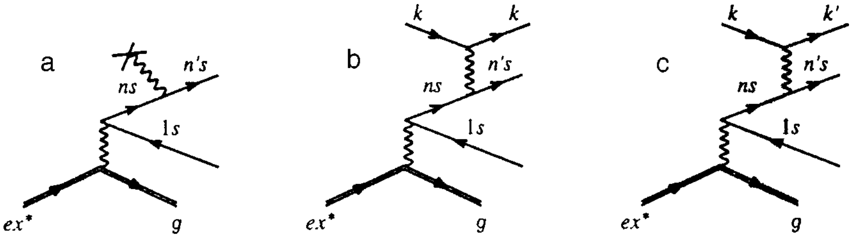

FIG. 2. Three types of configuration interaction: (a) mixing due to the one-particle operator of kinetic energy and the Coulomb interaction with the nucleus; (b) mixing due to the interaction with the core; and (c) interaction with excitation of the core.

TABLE II. Matrix elements of the two-particle interaction of configurations, $\langle k\, ns|H'|8s\, k\rangle$ (Fig. 2b), rows 1–4; M_{tot} is the total matrix element of two-particle interaction; M_1 is the one-particle matrix element $\langle ns|H_{\mathrm{one}}|8s\rangle$ (Fig. 2a); and M_{res} is the resulting matrix element. All values of matrix elements are in electronvolts.

$k \setminus n$	6	7	9
$1s_{1/2}$	8.518	9.403	7.355
$2s_{1/2}$	13.405	15.9473	12.770
$2p_{1/2}$	14.186	16.559	13.183
$2p_{3/2}$	7.058	8.252	6.573
M_{tot}	43.167	50.161	39.881
M_1	-42.887	-49.944	-39.702
M_{res}	0.280	0.217	0.179

2b depicts the diagram representing the two-particle interaction of the electron and the field generated by the other electrons, whose states do not change, so that they can be regarded as comprising the core. This interaction also leads to mixing of the final ns states in the principal quantum number *n*. The state of the core does not change if one allows for this type of interaction. Finally, Fig. 2c illustrates two-particle interaction with more complicated many-particle configurations of the core-excitation type.

The typical matrix elements of the type of interaction of Fig. 2a or b for Te^{45+} are listed in Table II. The first four rows in the table illustrate the fragmentation of the $8s$ states in the $6s$, $7s$, and $9s$ shells, fragmentation that occurs due to the interaction with the other electrons of the *K*- and *L*-shells. The results show that the interaction matrix elements are almost the same for all electrons of the *L*-shell and are somewhat larger from the *K*-shell. The magnitude of each matrix element is about 10 eV. They add up coherently, yielding, for instance, $\Sigma_k\langle k\, 9s|H'|8s\, k\rangle = 39.881$ eV. However, this value is balanced almost perfectly by the contributions of the interaction of the type depicted in Fig. 2a, which is approximately equal in value but has the opposite sign, as the sixth and seventh rows in Table II show. This perfect balance has a simple physical meaning. The electron in the highly excited state $9s$ ''detects'' basically the nearly Coulomb potential with a charge $Z-q+1$, which depends very weakly on the specific configuration of the core. Therefore allowance for the given type of interaction gives rise only to a slight mixing of configurations with an amplitude of less than $\sim 1\%$.

However, as noted earlier, calculations by the MCDF method yield splitting of levels in total angular momentum, with the value of this splitting being approximately 30 eV. The level density changes when the interaction represented by Fig. 2c is taken into account. The typical matrix elements of this type are listed in Table III. Since they are off-diagonal in all four states, they prove to be smaller by factor of 10 to 100. Note that the matrix elements with numbers from 3 to 6 with excitation of the hole states $2p_{1/2}$ in the core add up coherently. The total value of such matrix elements amounts to roughly 6 eV, which is larger by a factor of ten than the matrix element M_{res} of one-particle mixing listed in Table II. And since the difference of the corresponding energy is roughly 50 eV, allowance for this interaction can lead to mixing of these states with an amplitude close to 10% and to a corresponding change in the transition energy and the conversion matrix elements. Thus, although the density of the atomic levels is not really as high as in the case of the multiply charged Au^{24+} ion studied by Gribakin *et al.*,[7] allowance for configuration mixing, as we will show shortly, proves to be very important if we want to calculate the rate of discrete conversion correctly.

4. CONVERSION TRANSITIONS TO *nd*-SHELLS

The selection rules for the conversion $M1$ transition also allow for transitions in which the electron orbital angular momentum l changes by 2, i.e., $\Delta l = 2$. In the case at hand this means that there can be transitions of the $1s$ electron not only to the ns states but also to the discrete states $nd_{3/2}$ of shells, transitions with $\Delta J = 0, \pm 1$ which are close in energy to the energy of the nuclear transition. However, we expect the probability of these transitions to be low due to the smallness of the wave functions of the d electrons in the region of the nucleus. Since the electron wave functions behave at the origin as $(pr)^l$, it is indeed true that in the case of an $M1$ transition the internal conversion coefficients are at their

TABLE III. Matrix elements of the interaction of configurations, $\langle jn|H'|n'j'\rangle$ (Fig. 2c). The one-particle states are specified by the values of the principal quantum number, the orbital and total angular momenta, and the projection of the total angular momentum.

	Configuration				Matrix
Number	j	n	j'	n'	element, eV
1	$2p_{3/2}\frac{3}{2}$	$6s_{1/2}-\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2p_{1/2}\frac{1}{2}$	-0.184
2	$2p_{3/2}\frac{1}{2}$	$6s_{1/2}\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2p_{1/2}\frac{1}{2}$	-0.106
3	$2p_{3/2}\frac{1}{2}$	$6s_{1/2}-\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2p_{1/2}-\frac{1}{2}$	-0.144
4	$2p_{3/2}-\frac{1}{2}$	$6s_{1/2}\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2p_{1/2}-\frac{1}{2}$	-0.106
5	$2p_{3/2}\frac{1}{2}$	$7s_{1/2}\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2p_{1/2}\frac{1}{2}$	-0.083
6, 7	$2p_{3/2}\frac{1}{2}$	$6p_{1/2}\pm\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2s_{1/2}\mp\frac{1}{2}$	-0.401
8	$2p_{3/2}-\frac{1}{2}$	$6p_{1/2}\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2s_{1/2}-\frac{1}{2}$	0.401
9	$2p_{3/2}\frac{3}{2}$	$6p_{1/2}-\frac{1}{2}$	$8s_{1/2}\frac{1}{2}$	$2s_{1/2}\frac{1}{2}$	0.695
10	$2p_{3/2}-\frac{3}{2}$	$7p_{1/2}\frac{3}{2}$	$8s_{1/2}-\frac{1}{2}$	$2s_{1/2}\frac{1}{2}$	-0.126

TABLE IV. Energies $\omega_{n\kappa}$ and the discrete conversion coefficients α_d^{M1} for the $M1$ transitions $1s\rightarrow nd_{3/2}$ calculated by the $\mathrm{DF_{av}}$ method.

$nd_{3/2}$	$\omega_{n\kappa}$, eV	α_d^{M1}, eV
$11d_{3/2}$	35 458	0.481
$12d_{3/2}$	35 496	0.368
$13d_{3/2}$	35 526	0.286
$14d_{3/2}$	35 550	0.223
$15d_{3/2}$	35 569	0.175
$16d_{3/2}$	35 585	0.137

maximum in the $1s\rightarrow ns$ transition, for which the wave function of the ns electron is finite at the origin (see, e.g., Ref. 8). On the other hand, the wave functions of the d states have a second-order zero near the nucleus and are localized essentially outside the K-shell. In Table IV we list the internal conversion coefficients for transitions to the discrete states of the $nd_{3/2}$-shell. We see that the values of these coefficients prove to be smaller by a factor of 1000 than the values of the internal conversion coefficients for transitions to states of the ns-shells given in Ref. 4 (see also Table V).

TABLE V. Results of calculations by the MCDF method: the energies $\omega_{n\kappa}$ of the $1s\rightarrow ns$ transitions for ns levels with different angular momenta J, the weights S of the "pure" configurations $1s2s^22p_{1/2}^22p_{3/2}ns$, the discrete internal conversion coefficients $\alpha_d^{M1}\Delta_J$, and the conversion factors $R^{n\kappa}S$.

ns	J	$\omega_{n\kappa}$, eV	S	$\alpha_d^{M1}\Delta_J$, eV	$R^{n\kappa}S$
12s	1/2	35 443.8	0.9967	67.31	0.0230
	3/2	35 443.9	0.9954	83.74	0.0287
	3/2	35 410.6	0.9878	51.03	0.0060
	5/2	35 410.4	0.9887	200.90	0.0237
13s	1/2	35 474.4	0.9899	51.90	0.1309
	3/2	35 474.4	0.9934	64.57	0.1633
	3/2	35 441.1	0.9864	39.34	0.0119
	5/2	35 440.9	0.9823	154.90	0.0465
14s	1/2	35 498.5	0.9893	40.41	0.6384
	3/2	35 498.5	0.9881	50.27	0.7932
	3/2	35 465.2	0.9817	30.63	0.0333
	5/2	35 465.0	0.9800	120.61	0.1289
15s	1/2	35 517.8	0.9744	31.56	0.0362
	3/2	35 517.9	0.9737	39.27	0.0446
	3/2	35 484.5	0.9645	23.93	0.3011
	5/2	35 484.5	0.9672	94.21	1.1605
16s	1/2	35 533.6	0.9370	24.57	0.0105
	3/2	35 533.7	0.9362	30.56	0.0130
	3/2	35 500.3	0.9255	18.62	0.1785
	5/2	35 500.2	0.9305	73.33	0.7223
17s	1/2	35 547.9	0.9472	18.92	0.0045
	3/2	35 547.9	0.9465	23.54	0.0057
	3/2	35 514.6	0.9368	14.34	0.0205
	5/2	35 514.4	0.9404	56.46	0.0825
R_{tot}					4.6077

Note that a similar situation occurs in ^{229}Th when one deals with the de-excitation of the nucleus through the electron bridge $7s\rightarrow 8s\rightarrow 7p$ (see Ref. 2). A situation that is the opposite of the one discussed above was studied by Strizhov and Tkalya,[9] who took into account only the contribution to the electron bridge that starts either at the $6d_{3/2}$ state or at the $7s$ state but passes through the excited $6d_{5/2}$ level. As a result, the effect was underestimated by a factor of 1000. Unfortunately, this serious error of past years found its way into Ref. 10, with the result that the interpretation of the experimental data was based on an erroneous theoretical calculation. Indeed, the analysis of subsequent experimental data[11,12] suggested that the UV photon emission observed by Irwin and Kim[10] and by Richardson *et al.*[13] is a possible consequence of the fluorescence of nitrogen molecules in the air surrounding the radioactive sources rather than the decay of the 3.5-eV level in ^{229}Th.

5. RESULTS OF CALCULATIONS OF THE INTERNAL CONVERSION COEFFICIENTS

As noted earlier, as a result of calculations by the MCDF method, for each electron configuration of the final state, $1s2s^22p_{1/2}^22p_{3/2}ns$, four levels form with total angular momenta $J=5/2$, 3/2, 1/2, and 3/2. The angular momenta of the levels are ordered according to increasing energy. Using the data of our calculations by the MCDF method, we estimated the fractions of the internal conversion coefficients, Δ_J, related to each of these levels. The results of the calculations yield the following distribution of internal conversion coefficients among the four levels:

$$\Delta_{J=5/2}=0.5,\quad \Delta_{J=3/2}=0.127,$$
$$\Delta_{J=1/2}=0.168,\quad \Delta_{J=3/2}=0.209, \tag{1}$$

We see that the sum over all fractions is roughly unity, $\Sigma_J\Delta_J\approx 1$. The internal conversion coefficients for the $1s\rightarrow ns$ transitions, $n=12-17$, are listed in Table V. Note that the coefficients have been multiplied by the respective fraction Δ_J.

In the case of ordinary conversion, the coefficient $\alpha^{n\kappa}$ is defined as the ratio of the rates of conversion and radiative transitions of the nucleus. When we are dealing with discrete conversion, the coefficient $\alpha_d^{n\kappa}$ has the dimensionality of energy on account of the different normalization of the wave function of the conversion electron.[4] However, we can determine the conversion factor $R^{n\kappa}$, which is still the ratio of the rates of conversion and radiative nuclear transitions:[4]

$$R^{n\kappa}=\alpha_d^{n\kappa}\frac{1}{2\pi}\frac{\Gamma}{(\omega_\gamma-\omega_{n\kappa})^2+(\Gamma/2)^2}, \tag{2}$$

where Γ is the total energy of the hole state, ω_γ is the energy of the nuclear transition, and $\omega_{n\kappa}$ is the energy of the electron transition. If, following Ref. 4, we assume that the width Γ of the $1s$ vacancy is 5 eV and allow for the weights S of the "pure configurations" $1s2s^22p_{1/2}^22p_{3/2}ns$, which are determined in the calculation by the MCDF method, we obtain the values of the conversion factor $R^{n\kappa}S$ listed in Table V.

The total resonant conversion factor R_{tot} is obtained by summing over all the final states:

$$R_{\mathrm{tot}}=\sum_n R^{n\kappa}S. \tag{3}$$

Table V shows that the maximum contribution to the total factor of resonant conversion, $R_{\mathrm{tot}}^{\mathrm{MCDF}}$, is provided by the levels with angular momenta $J=3/2$ and 5/2 with energies 35 484.5 eV and 35 484.4 eV, respectively, which arise in the splitting of the final state 15s, and also by the levels with angular momenta $J=1/2$ and 3/2 with an energy 35 498.5 eV, which arise in the splitting of the final state 14s. This result should be compared with the results of Ref. 4, where the calculations done by the $\mathrm{DF}_{\mathrm{av}}$ method without allowance for the residual interactions suggested that the main contribution to $R_{\mathrm{tot}}^{\mathrm{DF}}$ is provided by the final state 18s. The total conversion factor $R_{\mathrm{tot}}^{\mathrm{MCDF}}$ is equal to 4.36, which differs from $R_{\mathrm{tot}}^{\mathrm{DF}}=5.9$ obtained in Ref. 4 by roughly 30%. The results listed in Table V show that the total weight of the impurity configurations may be substantial and reaches 7% for the given shell. The corresponding shift in the transition energies due to configuration mixing can be estimated by comparing the data of Tables V and I. This shift reaches 3–4 eV, in accordance with the quantitative reasoning done in Sec. 3. Note that core excitation (Fig. 2c) has a strong effect on the value of R_{tot} and on the corresponding transition energies. Allowing for this mechanism in a proper manner gives rise to an energy shift of the levels with $J=3/2$, 5/2 (15s) from the value 35 488 eV provided by the $\mathrm{DF}_{\mathrm{td}}$ method to the value 35 484.5 eV. Allowing (or not allowing) for this mechanism and the corresponding shifts of the transition energy by 3 eV changes the value of the partial factor of the conversion transition to the given level by a factor of 2.5 and that of the total conversion factor by a factor of 1.8.

This work was made possible by grants from the Russian Fund for Fundamental Research (Grants Nos. 96-02-18039 and 99-02-17550), the contract DSWA (USA) No. DSWA 01-98-C-0040, and a grant of the Portuguese Committee for Science in the PRAXIS XXI Program.

*⁾E-mail: trzhask@thd.pnpi.spb.ru

---

[1] D. F. Zaretskiĭ and F. F. Karpeshin, Yad. Fiz. **29**, 306 (1979) [Sov. J. Nucl. Phys. **29**, 151 (1979)].

[2] F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, and M. A. Listengarten, Phys. Lett. B **372**, 1 (1996).

[3] F. Attallah, M. Aiche, J. F. Chemin, J. N. Scheurer, W. E. Meyerhof, J. P. Grandin, P. Aguer, G. Bogaert, J. Kiener, A. Lefebvre, J. P. Thibaud, and C. Grunberg, Phys. Rev. Lett. **75**, 1715 (1995).

[4] F. F. Karpeshin, M. R. Harston, F. Attallah, J. F. Chemin, J. N. Scheurer, I. M. Band, M. B. Trzhaskovskaya, Phys. Rev. C **53**, 1640 (1996).

[5] I. M. Band, M. A. Listengarten, M. B. Trzhaskovskaya, and V. I. Fomichev, Preprints of Leningrad Nuclear Physics Institute Nos. 289 (1976), 298–300(1977), 498 (1979), and 1479 (1989).

[6] I. P. Grant, B. J. McKenzie, P. H. Norrington, D. F. Mayers, and N. C. Pyper, Comput. Phys. Commun. **21**, 207 (1980).

[7] G. F. Gribakin, A. A. Gribakina, and V. V. Flambaum, Aust. J. Phys. **52**, 443 (1999).

[8] I. M. Band, M. A. Listengarten, and A. P. Feresin, *Anomalies in the Coefficients of Internal Conversion of Gamma Rays* [in Russian], Nauka, Moscow (1976).

[9] V. F. Strizhov and E. V. Tkalya, Zh. Éksp. Teor. Fiz. **99**, 697 (1991) [Sov. Phys. JETP **72**, 387 (1991)].

[10] G. M. Irwin and K. H. Kim, Phys. Rev. Lett. **79**, 900 (1997).

[11] S. B. Utter, P. Beiersdorfer, A. Barnes, R. W. Lougheed, J. R. Crespo Löpez-Urrutia, J. A. Becker, and M. S. Weiss, Phys. Rev. Lett. **82**, 505 (1999).

[12] R. W. Shaw, J. P. Young, S. P. Cooper, and O. F. Webb, Phys. Rev. Lett. **82**, 1109 (1999).

[13] D. S. Richardson, D. M. Benton, D. E. Evans, J. A. R. Griffith, and G. Tungate, Phys. Rev. Lett. **80**, 3206 (1998).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Polarization of the vacuum in a relativistic hydrogenlike atom: the Lamb shift

S. G. Karshenboĭm*)

State Scientific Center ''D. I. Mendeleyev National Metrological Research Institute,''
198005 St. Petersburg, Russia

(Submitted 26 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1575–1586 (November 1999)

This paper examines the shift of energy levels in a hydrogenlike atom induced by vacuum polarization effects. The contribution of free polarization is found for the ground state and several excited states in a closed analytical form. For the first time an expression is derived for the radiative correction to the energy in the form of an explicit function of the parameter $Z\alpha$. The results are valid for states nl_j with the largest values of orbital and total angular momenta ($l=n-1$ and $j=l+1/2$). The final expression, found in terms of generalized hypergeometric functions, is a function of three variables, $Z\alpha$, n, and the ratio of the particle masses on the orbit and in the vacuum loop, i.e., the result is valid for ordinary atoms and for muonic atoms. Several useful asymptotic expressions are also derived. © *1999 American Institute of Physics.* [S1063-7761(99)00611-3]

1. INTRODUCTION

The diagram that allows for vacuum polarization effects (Fig. 1) is the simplest one, but there is no closed analytical expression even for its contribution to the energy levels. Some coefficients in the expansion in the strength of the Coulomb interaction $Z\alpha$ are known for various energy levels in an (electronic) hydrogenlike atom with a point nucleus.[1–4] An analytical expression for muonic atoms has been found only in the leading nonrelativistic (Schrödinger) approximation.[5] There are also some numerical results (see, e.g., Ref. 6).

The goal of the present paper is to derive closed analytical expressions for the corrections to the energy levels in electronic and muonic atoms. We will discuss the case of a hydrogenlike atom with an infinitely heavy point nucleus and a bound particle of mass m. In real applications the particle may be either an electron or a muon. We will study the contribution of the free polarization of vacuum (the Uehling potential); the particle in the loop in cases of interest is an electron ($m_l=m_e$).

Note that there are several reasons why this problem is so important. First, in some cases the expansion in powers of $Z\alpha$ is not very efficient because terms with high powers of logarithms (e.g., $\ln^2(1/Z\alpha)\approx 24$ and $\ln^3(1/Z\alpha)\approx 120$) arise[7] and so one is forced to do exact calculations in $Z\alpha$ or at least to better understand the structure of the series in $Z\alpha$. The calculations done in the present paper make it possible to study the structure of the expansion in powers of this parameter. Second, a number of useful asymptotic expressions for both electronic and muonic atoms are derived. For instance, for not-too-light muonic atoms ($l=n-1$ and $j=l+1/2$) we have

$$E_U(nl_j)=-\frac{\alpha}{\pi}\,\frac{(Z\alpha)^2m/n^2}{\sqrt{1-(Z\alpha)^2/n^2}}\left\{\left[\frac{2}{3}\ln(2\kappa_n)+\frac{2}{3}(\psi(1)-\psi(2-3\epsilon_n))-\frac{5}{9}\right]+\frac{\pi}{2}\,\frac{1-\epsilon_n}{\kappa_n}+\frac{2-3n(1-\epsilon_n)-10n^2(1-\epsilon_n^2)}{6}\,\frac{1}{\kappa_n^2}+O\left(\frac{1}{\kappa_n^3}\right)\right\}, \tag{1}$$

where the relativistic effects are described by the parameter

$$\epsilon_n=1-\sqrt{1-\frac{(Z\alpha)^2}{n^2}} \tag{2}$$

and where we have introduced the ratio of the atomic momentum (γ) to the mass of the particle in the loop (m_l):

$$\kappa_n=\frac{\gamma}{m_l n}=\frac{Z\alpha m}{m_l n}. \tag{3}$$

In particular, for muonic atoms we have $\kappa_n\simeq 1.5Z/n$. Here we used a system of units in which $\hbar=c=1$.

Third, the correction obtained in this paper is unquestionably the simplest quantum-electrodynamics correction, but the methods developed here can be used when more complicated diagrams are involved.[8]

We begin by discussing the general expression for the ground state, then we turn to the asymptotic expressions, and finally determine the Uehling relativistic correction for the levels nl_j with the largest values of angular momenta ($l=n-1$ and $j=l+1/2$). We will conclude the paper with a discussion of the results and the possible applications of the methods developed here.

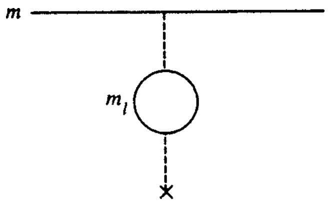

FIG. 1. Uehling potential for an atom with a particle of mass m. The mass of the particle in the loop is m_l.

2. DERIVATION OF THE GENERAL EXPRESSION

We begin with the derivation of a convenient representation for the Uehling potential. Allowing for the polarization in the momentum representation amounts to the well-known substitution

$$\frac{1}{\mathbf{q}^2}\rightarrow\frac{\alpha}{\pi}\int_0^1 dv\,\frac{v^2(1-v^2/3)}{1-v^2}\,\frac{1}{\mathbf{q}^2+\lambda^2},\tag{4}$$

where

$$\lambda(v)=\frac{2m_l}{\sqrt{1-v^2}}\tag{5}$$

(the reader will recall that the mass of the particle in the loop is m_l). In the coordinate representation the substitution has the form

$$\frac{1}{r}\rightarrow\frac{\alpha}{\pi}\int_0^1 dv\,\frac{v^2(1-v^2/3)}{1-v^2}\,\frac{\exp\{-\lambda r\}}{r}.\tag{6}$$

The correction to the energy can be written in the form of a product,

$$E_U=\frac{\alpha}{\pi}E_C R_C(\kappa),\tag{7}$$

where E_C is the average potential energy of the Dirac electron and

$$R_C(\kappa)=\int_0^1 dv\,\frac{v^2(1-v^2/3)}{1-v^2}\,\mathscr{R}_C(\lambda(v)).\tag{8}$$

The potential energy can easily be expressed in terms of the total energy E_D of the bound Dirac particle (see, e.g., Ref. 9),

$$E_C(nl_j)=\left\langle nl_j\left|-\frac{Z\alpha}{r}\right|nl_j\right\rangle=Z\alpha\frac{\partial}{\partial(Z\alpha)}E_D(nl_j),\tag{9}$$

and $\mathscr{R}_C$ in (8) can be written in the form of a ratio,

$$\mathscr{R}_C(\lambda)=\frac{\displaystyle\int_0^\infty dr\,r^2[f^2(r)+g^2(r)]\,\frac{\exp\{-\lambda r\}}{r}}{\displaystyle\int_0^\infty dr\,r^2[f^2(r)+g^2(r)]\,\frac{1}{r}}.\tag{10}$$

Here we have employed the notation adopted in Ref. 10 for the radial part of the large $[f(r)]$ and small $[g(r)]$ components of the Dirac wave function. The above expressions are valid for all states, and in the case of the $1s$ level we have

$$E_C(1s)=-\frac{(Z\alpha)^2m}{\sqrt{1-(Z\alpha)^2}}\tag{11}$$

and

$$\mathscr{R}_C(\lambda)=\frac{\displaystyle\int_0^\infty dr\,r\exp\{-(2\gamma+\lambda)r\}\,r^{-2\epsilon}}{\displaystyle\int_0^\infty dr\,r\exp\{-2\gamma r\}\,r^{-2\epsilon}}.\tag{12}$$

The relativistic effects for the ground state are determined by the value of the parameter

$$\epsilon=1-\sqrt{1-(Z\alpha)^2},$$

which vanishes in the Schrödinger approximation, and they are also taken into account in the relativistic potential energy (11). To simplify the expressions, we discard the label 1 for the ground state and use the notation ϵ and κ.

The radial integration proves to be trivial and yields a simple result:

$$\mathscr{R}_C(\lambda)=\left(\frac{2\gamma}{2\gamma+\lambda}\right)^{2(1-\epsilon)}.\tag{13}$$

What is not trivial is that the resulting single integral

$$\begin{aligned}R_C(\kappa)=\int_0^1 dv\,v^2\left(1-\frac{v^2}{3}\right)&\left(\frac{\kappa}{1+\kappa\sqrt{1-v^2}}\right)^2\\ &\times\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{-2\epsilon}\end{aligned}\tag{14}$$

can be evaluated in closed form,[11] (more precisely, in terms of the generalized hypergeometric function ${}_3F_2$). It is convenient to express the results of integration in terms of the base integrals

$$E_U(1s)=-\frac{\alpha}{\pi}\,\frac{(Z\alpha)^2m}{\sqrt{1-(Z\alpha)^2}}\left\{I_{122}-\frac{1}{3}I_{222}\right\},\tag{15}$$

where

$$\begin{aligned}I_{abc}=&\frac{\kappa^{c-2\epsilon}}{2}B\left(a+\frac{1}{2},1-\frac{b}{2}+\frac{c}{2}-\epsilon\right){}_3F_2\left(\frac{c}{2}-\epsilon,\frac{c}{2}+\frac{1}{2}\right.\\ &\left.-\epsilon,1-\frac{b}{2}+\frac{c}{2}-\epsilon;\frac{1}{2},a+\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon;\kappa^2\right)\\ &-\frac{(c-2\epsilon)}{2}\,\kappa^{c+1-2\epsilon}B\left(a+\frac{1}{2},\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon\right)\\ &\times{}_3F_2\left(\frac{c}{2}+1-\epsilon,\frac{c}{2}+\frac{1}{2}-\epsilon,\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon;\frac{3}{2},a+2\right.\\ &\left.-\frac{b}{2}+\frac{c}{2}-\epsilon;\kappa^2\right).\end{aligned}\tag{16}$$

The evaluation of these integrals is given in the Appendix. Clearly, the expression is finite if

$$Z\alpha<1.\tag{17}$$

Below we will discuss some asymptotic expressions for (15).

3. UEHLING CORRECTION FOR $Z\alpha \ll 1$

We wish to discuss the nonrelativistic limit without making any assumptions concerning the size of the parameter κ. Two special cases are important here: an electronic atom ($\kappa = Z\alpha$) and a not-too-light muonic atom ($\kappa \simeq 1.5Z \gg 1$).

When Z is small, the expression for the correction to the energy can be written

$$R_C(\kappa) = r_C(\kappa) - 2\epsilon p_C(\kappa) + O(\epsilon^2), \tag{18}$$

where

$$2\epsilon \simeq (Z\alpha)^2 \tag{18}$$

in the nonrelativistic approximation. We will first calculate r_C and p_C and then discuss the corrections to the expressions.

In the Schrödinger approximation, the result incorporates only the first term of the expansion (18), which is well known:[5]

$$r_C(\kappa) = \frac{1}{3}\left[-\frac{4+\kappa^2-2\kappa^4}{\kappa^3}\mathscr{A}(\kappa) + \frac{4+3\kappa^2}{\kappa^3}\frac{\pi}{2} - \frac{12+11\kappa^2}{3\kappa^2} \right], \tag{20}$$

where

$$\mathscr{A}(\kappa) = \frac{\arccos\kappa}{\sqrt{1-\kappa^2}} = \frac{\ln(\kappa + \sqrt{\kappa^2-1}\,)}{\sqrt{\kappa^2-1}}, \tag{21}$$

with $\mathscr{A}(0) = \pi/2$.

When dealing with the electronic atom ($\kappa = Z\alpha$), we note that for small values of Z the power expansion up to terms of order $\alpha(Z\alpha)^7 m$ is of practical interest. With appropriate accuracy, the normalization factor is

$$E_C = -(Z\alpha)^2 m\left[1 + \tfrac{1}{2}(Z\alpha)^2 + \tfrac{3}{8}(Z\alpha)^4\right]. \tag{22}$$

When κ is small, we have the following expressions for the terms in (18):

$$r_C(\kappa) = \frac{4}{15}\kappa^2 - \frac{5\pi}{48}\kappa^3 + \frac{12}{35}\kappa^4 - \frac{7\pi}{64}\kappa^5, \tag{23}$$

$$p_C(\kappa) = \left[\frac{4}{15}\ln(2\kappa) - \frac{77}{225}\right]\kappa^2 - \left(\frac{5\pi}{48}\ln\frac{\kappa}{2} + \frac{47\pi}{576}\right)\kappa^3. \tag{24}$$

Finally,

$$R_C = \frac{4}{15}\kappa^2 - \frac{5\pi}{48}\kappa^3 + \frac{12}{35}\kappa^4 - \frac{7\pi}{64}\kappa^5 - (Z\alpha)^2 \times\left[\left(\frac{4}{15}\ln(2\kappa) - \frac{77}{225}\right)\kappa^2 - \left(\frac{5\pi}{48}\ln\frac{\kappa}{2} + \frac{47\pi}{576}\right)\kappa^3\right]. \tag{25}$$

As a result we arrive at an expansion (for $Z\alpha$ small) for the Uehling correction to the energy level of the ground state in an (electronic) hydrogenlike atom:

$$E_U = \frac{\alpha}{\pi}(Z\alpha)^4 m\left\{ -\frac{4}{15} + \frac{5\pi}{48}Z\alpha + \left[\frac{4}{15}\ln(2Z\alpha) - \frac{1289}{1575}\right] \times (Z\alpha)^2 + \left[\frac{5\pi}{48}\ln\frac{2}{Z\alpha} + \frac{23\pi}{288}\right](Z\alpha)^3 \right\}. \tag{26}$$

This expansion reproduces all the coefficients known so far (see Refs. 1 and 2 and the papers cited therein) and one unknown coefficient (we denote it by A_{70}).

4. UEHLING CORRECTION FOR MUONIC ATOMS

Now we examine the case of large values of κ. First we note that in muonic atoms this limit is reached for fairly small values of $Z\alpha$. In the case of very small $Z\alpha$ ($Z = 1-3$, $\kappa(\mathrm{H}) \simeq 1.5$, $\kappa(\mathrm{He}^+) \simeq 3$, and $\kappa(\mathrm{Li}^{++}) \simeq 4.5$) the results obtained in Ref. 5 are sufficient [this result is shown in (20)]. For larger values of Z we can expand in powers of $1/\kappa$. Let us go back to (18) and discuss the terms in it. The analytical result of Ref. 5 represented in (20) can easily be expanded:

$$r_C(\kappa) = \frac{2}{3}\ln(2\kappa) - \frac{11}{9} + \frac{\pi}{2\kappa} + O\left(\frac{1}{\kappa^2}\right). \tag{27}$$

The expansion of the second term in (18) we find directly:

$$p_C(\kappa) = \frac{2}{3} - \frac{\pi^2}{9} + O\left(\frac{1}{\kappa}\right) \tag{28}$$

To obtain the asymptotic expressions, we only need to bear in mind that the contribution stems from a small domain of integration with respect to v near the upper limit $v \simeq 1$. Introducing the substitutions

$$v^2\left(1 - \frac{v^2}{3}\right) \to \frac{2v}{3} \tag{29}$$

in (14), we can easily evaluate the integrals with respect to v. The result of all this is

$$R_C(\kappa) = \left[\frac{2}{3}\ln(2\kappa) - \frac{11}{9} + \frac{\pi}{2\kappa}\right] - \left(\frac{2}{3} - \frac{\pi^2}{9}\right)(Z\alpha)^2 + O((Z\alpha)^4) + O\left(\frac{(Z\alpha)^2}{\kappa}\right) + O\left(\frac{1}{\kappa^2}\right). \tag{30}$$

The substitution (29) is valid if Z is small and κ is large. However, (29) can easily be generalized. In particular, when $\kappa \gg 1$, we can use the transformation

$$\frac{v^2(1-v^2/3)}{1-v^2} = \frac{2}{3}\frac{v}{1-v^2} - \frac{v(1-v)(2+v)}{3(1+v)} \tag{31}$$

to evaluate the integral

$$I_{bc} = \int_0^1 dv\, \frac{v^2(1-v^2/3)}{(1-v^2)^{b/2}}\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{c-2\epsilon}. \tag{32}$$

Only one of these integrals is needed in calculating the contribution of the Uehling potential,

$$R_C(\kappa)=I_{22}. \tag{33}$$

In accordance with (31), we split this integrals into two parts, $I^{(1,2)}$. The first term reduces to the incomplete beta function:

$$\begin{aligned} I^{(1)}_{bc}&=\frac{2}{3}\int_0^1 dv\,\frac{v}{(1-v^2)^{b/2}}\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{c-2\epsilon}\\ &=\frac{2}{3}\kappa^{b-2}\int_\delta^1\frac{dz}{z}(1-z)^{1-b+c-2\epsilon}\\ &=\frac{2}{3}\kappa^{b-2}B_{1-\delta}(2-b+c-2\epsilon,b-2), \end{aligned} \tag{34}$$

where

$$\frac{1}{z}=1+\kappa\sqrt{1-v^2}, \tag{35}$$

and

$$\delta=\frac{1}{1+\kappa}\ll 1.$$

Now it is fairly simple to find the necessary asymptotic expression for the incomplete beta function:

$$\begin{aligned} B_{1-\delta}(c-2\epsilon,0)&=\ln(1+\kappa)+[\psi(1)\\ &\quad-\psi(c-2\epsilon)]+\frac{c-2\epsilon-1}{\kappa}\\ &\quad-\frac{(c-2\epsilon-1)(c+2-2\epsilon)}{2\kappa^2}+O\left(\frac{1}{\kappa^3}\right). \end{aligned}$$

The second term for $b=2$ can be expanded,

$$\begin{aligned} I^{(2)}_{2c}&=\int_0^1 dv\,\frac{v(1-v)(2+v)}{3(1+v)}\left(1-\frac{c-2\epsilon}{1+\kappa\sqrt{1-v^2}}\right.\\ &\quad\left.+\frac{(c-2\epsilon)(c-1-2\epsilon4)}{2\kappa^2(1-v^2)}\right)+O\left(\frac{1}{\kappa^3}\right), \end{aligned}$$

and integrated:

$$\begin{aligned} I^{(2)}_{2c}&=-\left(\frac{5}{9}-\frac{2}{3}\ln 2\right)+\left(\frac{\pi}{4}-\frac{2}{3}\right)\frac{c-2\epsilon}{\kappa}\\ &\quad-\frac{(c-2\epsilon)(c-1-2\epsilon)}{12\kappa^2}+O\left(\frac{1}{\kappa^3}\right). \end{aligned} \tag{36}$$

The result is an expression for the base integral

$$\begin{aligned} I_{2c}&=\left[\frac{2}{3}\ln(2\kappa)+\frac{2}{3}(\psi(1)-\psi(c-2\epsilon))-\frac{5}{9}\right]+\frac{\pi}{4}\,\frac{c-\epsilon}{\kappa}\\ &\quad+\left[\frac{1}{3}-\frac{c-2\epsilon}{4}-\frac{5}{12}(c-2\epsilon)^2\right]\frac{1}{\kappa^2}+O\left(\frac{1}{\kappa^3}\right), \end{aligned} \tag{37}$$

which for $c=2$ yields

$$\begin{aligned} R_C&=\left[\frac{2}{3}\ln(2\kappa)+\frac{2}{3}(\psi(1)-\psi(2-2\epsilon))-\frac{5}{9}\right]\\ &\quad+\frac{\pi}{2}\,\frac{1-\epsilon}{\kappa}+\left[\frac{1}{3}-\frac{1-\epsilon}{2}-\frac{5}{3}(1-\epsilon)^2\right]\frac{1}{\kappa^2}+O\left(\frac{1}{\kappa^3}\right). \end{aligned} \tag{38}$$

Formula (30) is reproduced when Z is small (i.e., ϵ is small), which is a check on the validity of (38).

The convergence of the expression is determined by the value of the argument of the ψ-function (the logarithmic derivative of the Γ-function), which must not vanish. This condition again leads to the finiteness of the correction for the charge Z, which is given by the inequality (17).

5. UEHLING CORRECTION FOR SOME EXCITED STATES

Now we examine the excited states nl_j with the largest values of orbital and total angular momenta ($l=n-1$ and $j=l+1/2$). The wave functions and energy levels of the Dirac atom prove to be the simplest in this case. In particular, the normalization factor in (7) has the form

$$E_C(n)=-\frac{(Z\alpha)^2m/n^2}{\sqrt{1-(Z\alpha)^2/n^2}}, \tag{39}$$

and in the Schrödinger approximation

$$E_C(n)\approx-(Z\alpha)^2m/n^2.$$

The quantity $R(n,n\epsilon_n,\kappa_n)$ assumes the form [cf. (14)]

$$\begin{aligned} R(n,n\epsilon_n,\kappa_n)&=\int_0^1\frac{dv\,v^2}{1-v^2}\left(1-\frac{v^2}{3}\right)\\ &\quad\times\left(\frac{\kappa_n\sqrt{1-v^2}}{1+\kappa_n\sqrt{1-v^2}}\right)^{2n(1-\epsilon_n)}. \end{aligned} \tag{40}$$

Above we examined the case $n=1$ in detail using the simplified notation $\epsilon_1=\epsilon$, $\kappa_1=\kappa$, and $R(1,\epsilon,\kappa)=R_C(\kappa)$. One of the ways to pass to arbitrary values of n is to use the recurrence relation

$$\begin{aligned} &\left(\frac{\sqrt{1-v^2}}{1+\kappa_n\sqrt{1-v^2}}\right)^{2n(1-\epsilon_n)}\\ &\quad=\frac{1}{(2n-1-2n\epsilon_n)(2n-2-2n\epsilon_n)}\,\frac{\partial^2}{\partial\kappa_n^2}\\ &\quad\times\left(\frac{\sqrt{1-v^2}}{1+\kappa_n\sqrt{1-v^2}}\right)^{2(n-1-n\epsilon_n)}, \end{aligned} \tag{41}$$

which gives rise to the expression

$$\begin{aligned} &\left(\frac{\sqrt{1-v^2}}{1+\kappa_n\sqrt{1-v^2}}\right)^{-2n(1-\epsilon_n)}\\ &\quad=\frac{\Gamma(2-2n\epsilon_n)}{\Gamma(2n-2n\epsilon_n)}\,\frac{\partial^{(2n-2)}}{\partial\kappa_n^{(2n-2)}}\left(\frac{\sqrt{1-v^2}}{1+\kappa_n\sqrt{1-v^2}}\right)^2. \end{aligned} \tag{42}$$

This readily yields

$$R(n,n\epsilon_n,\kappa_n)=\frac{\kappa_n^{2n}}{(2n-1-2n\epsilon_n)(2n-2-2n\epsilon_n)}\times\frac{\partial^2}{\partial\kappa_n^2}\,\frac{R(n-1,n\epsilon_n,\kappa_n)}{\kappa_n^{2(n-1)}}\tag{43}$$

and

$$R(n,n\epsilon_n,\kappa_n)=\kappa_n^{2n}\,\frac{\Gamma(2-2n\epsilon_n)}{\Gamma(2n-2n\epsilon_n)}\times\frac{\partial^{(2n-2)}}{\partial\kappa_n^{(2n-2)}}\,\frac{R(1,n\epsilon_n,\kappa_n)}{\kappa_n^2}.\tag{44}$$

In the Schrödinger approximation, the results for the $1s$ level [see (20)] and for the $2p$ and $3d$ levels were found in Ref. 5. In the general case for levels with quantum numbers $l=n-1$ and $j=l+1/2$ the result assumes the form[11] [cf. (15)]

$$R(n,n\epsilon_n,\kappa_n)=I_{122n}(n\epsilon_n,\kappa_n)-\tfrac{1}{3}\,I_{222n}(n\epsilon_n,\kappa_n).\tag{45}$$

The calculation of the base integrals I_{abc} [see (16)] is discussed in the Appendix.

For excited states in muonic atoms obeying the condition $\kappa\simeq1.5Z/n\gg1$ we have the expression

$$\begin{aligned}R(n,n\epsilon_n,\kappa_n)=&\left[\frac{2}{3}\ln(2\kappa_n)+\frac{2}{3}(\psi(1)-\psi(2n(1-\epsilon_n)))\right.\\&\left.-\frac{5}{9}\right]+\frac{\pi}{2}\,\frac{n(1-\epsilon_n)}{\kappa_n}+\left[\frac{1}{3}-\frac{n(1-\epsilon_n)}{2}\right.\\&\left.-\frac{5n^2}{3}(1-\epsilon_n)^2\right]\frac{1}{\kappa_n^2}+O\left(\frac{1}{\kappa_n^3}\right).\end{aligned}\tag{46}$$

This expression was derived by a method similar to the used in deriving the expression (38) for the ground state.

We now examine the nonrelativistic expansion for a hydrogenlike atom with a bound electron ($\kappa_n=Z\alpha/n$). We can easily write the first two terms:

$$\begin{aligned}E_U=&-\frac{\alpha}{\pi}\left(\frac{Z\alpha}{n}\right)^{2(1+n)}\frac{m}{3}\left\{(n+1)B\left(\frac{5}{2},n\right)\right.\\&\left.-Z\alpha(2n+3)B\left(\frac{5}{2},n+\frac{1}{2}\right)+\cdots\right\}.\end{aligned}$$

The first term in the expansion reproduces the well-known result of Ref. 4.

6. DISCUSSION

There are several points that must be mentioned in discussing the results. First, we note that for large values of κ the logarithmic terms can easily be found without doing explicit calculations. Simple considerations associated with a running coupling constant,

$$Z\alpha\to Z\alpha(\kappa_n)=Z\alpha\left(1+\frac{2}{3}\,\frac{\alpha}{\pi}\ln\kappa_n\right),$$

immediately lead to

$$E_U=E_C\,\frac{2}{3}\,\frac{\alpha}{\pi}\ln\kappa_n\,,$$

and this is true for all values of $Z\alpha$. The fact this result coincides with the result of explicit calculations serves as additional proof of their validity. Another proof of their validity is that the results agree with those of numerical calculations done in Ref. 6 for the ground state. Other confirmation can be found in the present paper: for instance, we reproduced all known analytical results for the contribution of the free polarization of vacuum to the energy levels of electronic and muonic atoms.

Here is a brief discussion of the aspects related to the accuracy of the calculations. First, we note that the accuracy can easily be increased. To do this, in the expressions corresponding to the Schrödinger approximation we need only replace the mass of the Dirac particle with the reduced mass and retain the particle mass only in the relativistic corrections. Another problem is the finite size of the nucleus in muonic atoms. In this connection it must be noted that the results obtained in this paper refer to states with large angular momenta, and the structure of the nucleus has a small effect on such states. States with large orbital momenta are interesting objects of investigation in muonic atoms. Note, for example, the high-precision measurements of the $3d_{5/2}$–$2p_{3/2}$ transitions in muonic^{24}Mg and ^{28}Si, which made it possible to determine the exact value of the mass of the negative muon.[12] A review of the data on $5g$–$4f$, $4f$–$3d$, and $3d$–$2p$ transitions can be found in Ref. 13. The investigation of such states is of interest in connection with the problem of antiproton helium, where only states with large values of l have a fairly long lifetime (see Ref. 14).

The corrections discussed in this paper are those to the Lamb shift. However, the methods developed here can be used for hyperfine splitting,[8] and we intend to continue our studies in this direction.

The author is grateful to U. Jentschura, V. G. Ivanov, H. Pilkhun, and V. G. Shabaev for useful discussions. The work was partially supported by the Fundamental Metrology Program.

APPENDIX DERIVATION OF THE GENERAL EXPRESSION

To integrate in (14), it is convenient to introduce some base integrals:

$$R_C(\kappa)=I_{122}-\tfrac{1}{3}\,I_{222}\,,$$

where[11]

$$I_{abc}=\int_0^1 dv\;\frac{v^{2a}}{(1-v^2)^{b/2}}\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{c-2\epsilon}.$$

Integration can be carried out for the case $\kappa<1$:

$$I_{abc}=\int_0^1 dv\ \frac{v^{2a}}{(1-v^2)^{b/2}}\left(\frac{\kappa\sqrt{1-v^2}}{1+\kappa\sqrt{1-v^2}}\right)^{c-2\epsilon}$$

$$=\frac{1}{2}\kappa^{c-2\epsilon}\int_0^1 dt\, t^{a-1/2}(1-t)^{-b/2+c/2-\epsilon}$$

$$\times\sum_{n=0}^{\infty}(-1)^n\kappa^n\frac{(c-2\epsilon)_n}{n!}(1-t)^{n/2}$$

$$=\frac{1}{2}\kappa^{c-2\epsilon}\sum_{n=0}^{\infty}(-1)^n\kappa^n\frac{(c-2\epsilon)_n}{n!}B$$

$$\times\left(a+\frac{1}{2},1-\frac{b}{2}+\frac{c}{2}-\epsilon+n/2\right)$$

$$=\frac{1}{2}\kappa^{c-2\epsilon}\Gamma\left(a+\frac{1}{2}\right)\sum_{n=0}^{\infty}(-1)^n\kappa^n\frac{(c-2\epsilon)_n}{n!}$$

$$\times\frac{\Gamma(1-b/2+c/2-\epsilon+n/2)}{\Gamma(a+3/2-b/2+c/2-\epsilon+n/2)},$$

where $(a)_n=a(a+1)\cdots(a+n-1)$ is the Pochhammer symbol.

For further calculations it is convenient to sum the even and odd terms in the series separately:

$$\mathscr{E}=\frac{1}{2}\kappa^{c-2\epsilon}\Gamma\left(a+\frac{1}{2}\right)$$

$$\times\sum_{k=0}^{\infty}\kappa^{2k}\frac{(c-2\epsilon)_{2k}}{(2k)!}\,\frac{\Gamma(1-b/2+c/2-\epsilon+k)}{\Gamma(a+3/2-b/2+c/2-\epsilon+k)},$$

$$\mathscr{O}=\frac{1}{2}\kappa^{c+1-2\epsilon}\Gamma\left(a+\frac{1}{2}\right)$$

$$\times\sum_{k=0}^{\infty}\kappa^{2k}\frac{(c-2\epsilon)_{2k+1}}{(2k+1)!}\,\frac{\Gamma(3/2-b/2+c/2-\epsilon+k)}{\Gamma(a+2-b/2+c/2-\epsilon+k)},$$

Using the identities

$$\Gamma(z+k)=(z)_k\Gamma(z),$$

$$(2k)!=2^{2k}(1/2)_k k!,$$

$$(2k+1)!=2^{2k}(3/2)_k k!,$$

$$(z)_{2k}=2^{2k}(z/2)_k(z/2+1/2)_k,$$

$$(z)_{2k+1}=2^{2k}z(z/2+1)_k(z/2+1/2)_k,$$

we obtain

$$\mathscr{E}=\frac{1}{2}\kappa^{c-2\epsilon}\Gamma\left(a+\frac{1}{2}\right)\frac{\Gamma(1-b/2+c/2-\epsilon)}{\Gamma(a+3/2-b/2+c/2-\epsilon)}$$

$$\times\sum_{k=0}^{\infty}\kappa^{2k}\frac{(c/2-\epsilon)_k}{(1/2)_2}\,\frac{(c/2+1/2-\epsilon)_k}{k!}$$

$$\times\frac{(1-b/2+c/2-\epsilon)_k}{(a+3/2-b/2+c/2-\epsilon)_k}$$

$$=\frac{1}{2}\kappa^{c-2\epsilon}B\left(a+\frac{1}{2},1-\frac{b}{2}+\frac{c}{2}-\epsilon\right)$$

$$\times{}_3F_2\left(\frac{c}{2}-\epsilon,\frac{c}{2}+\frac{1}{2}-\epsilon,1-\frac{b}{2}+\frac{c}{2}\right.$$

$$\left.-\epsilon;\frac{1}{2},a+\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon;\kappa^2\right),$$

$$\mathscr{O}=-\frac{c-2\epsilon}{2}\kappa^{c+1-2\epsilon}\Gamma\left(a+\frac{1}{2}\right)\frac{\Gamma(3/2-b/2+c/2-\epsilon)}{\Gamma(a+2-b/2+c/2-\epsilon)}$$

$$\times\sum_{k=0}^{\infty}\kappa^{2k}\frac{(c/2+1-\epsilon)_k}{(3/2)_k}\,\frac{(c/2+1/2-\epsilon)_k}{k!}$$

$$\times\frac{(3/2-b/2+c/2-\epsilon)_k}{(a+2-b/2+c/2-\epsilon)_k}$$

$$=-\frac{c-2\epsilon}{2}\kappa^{c+1-2\epsilon}B\left(a+\frac{1}{2},\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon\right)$$

$$\times{}_3F_2\left(\frac{c}{2}+1-\epsilon,\frac{c}{2}+\frac{1}{2}-\epsilon,\frac{3}{2}-\frac{b}{2}+\frac{c}{2}-\epsilon;\frac{3}{2},a+2\right.$$

$$\left.-\frac{b}{2}+\frac{c}{2}-\epsilon;\kappa^2\right).$$

---

*)E-mail: ksg@hm.csa.ru

[1] G. W. Erickson and D. R. Yennie, Ann. Phys. **35**, 271 (1965); **35**, 447 (1965).

[2] P. J. Mohr, Phys. Rev. Lett. **34**, 1050 (1975).

[3] V. G. Ivanov and S. G. Karshenboĭm, Yad. Fiz. **60**, 333 (1997) [Phys. At. Nucl. **60**, 270 (1997)].

[4] N. L. Manakov, A. A. Nekipelov, and A. G. Faĭnshteĭn, Zh. Éksp. Teor. Fiz. **95**, 1167 (1989) [Sov. Phys. JETP **68**, 673 (1989)].

[5] G. E. Pustovalov, Zh. Éksp. Teor. Fiz. **32**, 1519 (1957) [Sov. Phys. JETP **5**, 1234 (1957)]; D. D. Ivanenko and G. E. Pustovalov, Usp. Fiz. Nauk **61**, 27 (1957).

[6] W. R. Johnson and G. Soff, At. Data Nucl. Data Tables **33**, 405 (1985).

[7] S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **103**, 1105 (1993) [JETP **76**, 541 (1993)].

[8] S. G. Karshenboim, V. G. Ivanov, and V. M. Shabaev, Can. J. Phys. **76**, 503 (1998).

[9] J. Epstein and S. Epstein, Am. J. Phys. **30**, 266 (1962).

[10] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).

[11] S. G. Karshenboim, Can. J. Phys. **76**, 168 (1998).

[12] I. Beltrami, B. Aas, W. Beer *et al.*, Nucl. Phys. A **451**, 679 (1986).

[13] E. Borie and G. A. Rinker, Rev. Mod. Phys. **54**, 67 (1982).

[14] T. Yamazaki, E. Widmann, J. Eades, M. Kumakura, N. Morita, H. A. Torii, M. Hori, T. Ishikawa, F. E. Maas, H. Tamura, R. S. Hayano, L. Suhai, Y. Fijita, B. Ketzer, H. Daniel, F. J. Hartmann, M. Hasinoff, R. Pohl, R. Schmidt, T. von Egidy, and D. Horváth, Phys. Rev. A **55**, R3295 (1997).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Anomalous light-induced drift of lithium atoms in a mixture of noble gases

A. I. Parkhomenko*)

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 4 June 1999)
Zh. Éksp. Teor. Fiz. **116**, 1587–1600 (November 1999)

This paper is a theoretical study of the spectral features of the velocity of light-induced drift (LID) of lithium atoms (^{7}Li and ^{6}Li) in a binary mixture of noble gases: Ne+Ar, Ne+Kr, and Ne+Xe. The spectral shape of the LID signal is predicted to depend strongly on the fraction ξ of neon in the buffer mixture in the range $\xi \approx 0.8$–0.9 ($\xi = N_{\mathrm{Ne}}/N_b$, where N_{Ne} is the neon concentration, and N_b is the total concentration of the buffer particles). When the velocity of anomalous LID is treated as a function of the radiation frequency, it is found to have one, three, five, or seven zeros and to differ substantially from the dispersion-curve-like behavior with one zero predicted by the standard LID theory with velocity-independent transport collision rates. The reason for these additional zeros of the drift velocity is the alternating-sign dependence on the lithium-atom velocity of the relative difference of transport rates of collisions between buffer particles and excited and unexcited atoms. What is also established is that the anomalous LID of lithium atoms can be observed at almost all temperatures, depending on the value of ξ. At a fixed temperature, anomalous LID can be observed only in a narrow range of values of the fraction of neon in the buffer mixture, $\Delta\xi \approx 0.02$. The results make possible highly precise testing in the LID experiments of the interatomic potentials used in calculations of the velocity spectrum of anomalous LID. © *1999 American Institute of Physics.* [S1063-7761(99)00711-8]

1. INTRODUCTION

Light-induced drift (LID) in gases[1,2] is one of the strongest effects of the action of radiation on the translational motion of the particles of the gas. Theoretically, under laser excitation the drift velocities can reach the value of the thermal velocity.[3] Experiments have shown that as a result of LID the atoms can move with a velocity of several tens of meters per second and gather in a layer with a thickness less than 1 mm (see Refs. 4 and 5). It has now been experimentally established that almost two dozen objects, atoms and molecules, exhibit the LID effect (see, e.g., the reviews cited in Refs. 6–9).

The size of the LID effect is proportional to the relative difference of the transport rates of collisions of resonant particles in the ground and excited states with buffer particles. Until fairly recently, all data on LID experiments could be described satisfactorily by a LID theory in which the transport collision rates were velocity independent.[6–13] This theory yielded a dispersion-curve-like frequency dependence of the drift velocity $u_L(\Omega)$ with one zero where the detuning Ω of the radiation frequency vanishes (see Fig. 4 below). In 1992, while studying LID in C_2H_4 molecules with Kr acting as a buffer gas, van der Meer *et al.*[14] detected an unexpectedly large departure of the frequency dependence of the drift velocity $u_L(\Omega)$ from a dispersionlike curve: an anomalous spectral profile of LID velocity with three zeros instead of the one zero predicted by the then existing theory. The deviation from the predictions of the theory was so large that effect became known as anomalous LID. Today we have a vast body of data, both experimental[14–20] and theoretical,[16,20–28] gathered in anomalous-LID studies. It has been found that anomalous LID is entirely due to the dependence of transport collision rates on the velocity v of the resonant particles, and the anomaly can arise only if the difference $\Delta\nu(v)$ of the transport collision rates on the combining (i.e., affected by radiation) levels changes its sign as a function of v.

When molecules are involved, this behavior of $\Delta\nu(v)$ is due to inelastic collisional transitions between rotational levels, with the result that a regular pattern in the manifestation of anomalous LID of molecules may be present. In Ref. 28 it was shown that for all linear molecules with moderate values of the rotational constant, anomalous LID can always by observed under the proper experimental conditions.

The situation with atoms is quite different. The alternating-sign behavior of $\Delta\nu(v)$, which is necessary if we want to observe anomalous LID, is due only to certain features in the behavior of the potentials of the interaction of atoms in the combining states and the buffer particles. Hence to calculate anomalous LID of atoms we must know the interaction potentials for each specific system of colliding particles (a resonant particle and a buffer particle).

The anomalous LID of atoms was theoretically predicted in Refs. 25 and 27 for the Li–Ne and Rb–Kr systems with the Li and Rb atoms resonantly excited. So far no consistent research into anomalous LID in such systems has been done, since high temperatures ($T \sim 1000$ K) are needed if we want to observe anomalous LID in them.

Among atoms potassium atoms are the only objects for

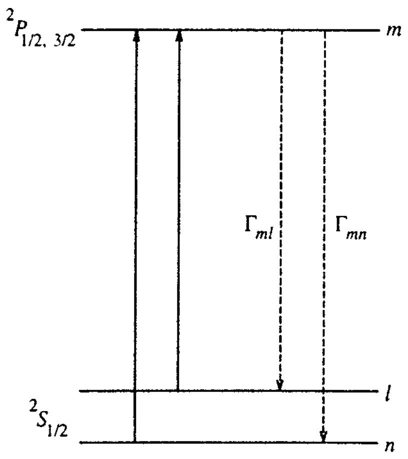

FIG. 1. Energy level diagram. Solid arrows stand for transitions initiated by radiation, and dashed arrows designate spontaneous radiative transitions.

which anomalous LID has been observed. In 1996, Yahyaei-Moayyed and Streater[19] detected in their experiment the anomalous LID of potassium atoms in a buffer medium consisting of a mixture of neon and another noble gas. However, no theoretical studies of the anomalous dependence of the drift velocity u_L on the detuning Ω of the radiation frequency for such systems have been done, with the result is still impossible to directly compare the results of that experiment with the anomalous-LID theory.

In the present paper we find a system that can be treated theoretically. In this system anomalous LID of atoms is possible at almost any temperature (including room temperature). It is found that for ^{7}Li or ^{6}Li atoms that are in a buffer mixture of neon and a different noble gas of the Ar, Kr, or Xe type there is always anomalous LID if the fraction of Ne atoms in the buffer mixture is selected appropriately (the selection is temperature-dependent). The anomalous spectral profile of the LID velocity $u_L(\Omega)$ can have three, five, or even seven zeros instead of the one zero predicted by the standard LID theory with velocity-independent transport collision rates

2. GENERAL RELATIONSHIPS

We will discuss the interaction between of a traveling monochromatic wave and three-level absorbing particles placed in a gas mixture with buffer particles. The level diagram for the absorbing particles is depicted in Fig. 1. Here the levels n and l are the components of the hyperfine structure of the ground state, and the level m corresponds to the excited electronic state. A particle radiatively relaxes from the level m to the levels n and l with the rate constants Γ_{mn} and Γ_{ml}, respectively. We will ignore the collisions between the absorbing particles, assuming that the buffer gas concentration N_b is much higher than the concentration N of the absorbing gas.

This level diagram is a good representation of the real structure of the ground and first excited states of lithium atoms (^{7}Li or ^{6}Li). Indeed, the ground level $^2S_{1/2}$ of these atoms is split into two hyperfine components. For ^{7}Li atoms the hyperfine splitting of the ground state, $\omega_{ln}=5.049\times 10^9\,\mathrm{s}^{-1}$ (Ref. 29) is comparable to the Doppler width of the resonant line and so the ground state can be modeled by two levels, n and l. For ^{6}Li atoms the hyperfine splitting of the ground state, $\omega_{ln}=1.434\times 10^9\,\mathrm{s}^{-1}$ (Ref. 29) is several times smaller than the Doppler width of the resonant line, but still we will model the ground state by two levels. When ^{7}Li atoms are involved (the nuclear spin is 3/2), the statistical weight of the level n is $g_n=3$ (the total angular momentum of the atom $F=1$), and the statistical weight of the level l is $g_l=5$ $(F=2)$. When ^{6}Li atoms (nuclear spin 1) are involved, we have $g_n=2$ $(F=1/2)$ and $g_l=4$ $(F=3/2)$.

The level m (with the statistical weight g_m) models a group of levels that are components of the hyperfine structure of the excited states, $^2P_{1/2}$ and $^2P_{3/2}$. Such modeling of a group of levels by a single level is possible since for the ^{7}Li and ^{6}Li atoms the hyperfine splitting in these excited states is very small in comparison to the Doppler width of the resonant line. The radiation involves only one of the fine components of the excited state, $^2P_{1/2}$ or $^2P_{3/2}$.

The lithium atoms comply with the limiting case of strong collisional coupling between the fine components $^2P_{1/2}$ and $^2P_{3/2}$ (the Massey parameter is much smaller than unity). In the excitation of lithium atoms to state $^2P_{1/2}$ and to state $^2P_{3/2}$ the scattering cross sections are almost the same.[30] Therefore, in terms of collisions the pair of the fine components can be interpreted as a single level, and the LID velocity considered a function of the detuning of the radiation frequency is the same in the excitations of the D_1 and D_2 lines of the lithium atoms. This result, i.e., the existence of universal transport characteristics for the lithium atom in the ^2P state, has also been corroborated by LID experiments involving sodium atoms,[5] for which, as for lithium atoms, the limiting case of strong collisional coupling between the fine components $^2P_{1/2}$ and $^2P_{3/2}$ holds.

The interaction of particles and radiation in steady-state, spatially homogeneous conditions are describes by the following equations for the density matrix:[10,31]

$$\begin{gathered}\Gamma_m\rho_m(\mathbf{v})=S_m(\mathbf{v})+N[P_n(\mathbf{v})+P_l(\mathbf{v})],\\ \Gamma_{mi}\rho_m(\mathbf{v})+S_i(\mathbf{v})=NP_i(\mathbf{v}),\\ \left[\frac{\Gamma_m}{2}-i(\Omega_{0i}-\mathbf{k}\cdot\mathbf{v})\right]\rho_{mi}(\mathbf{v})\\ =S_{mi}(\mathbf{v})+iG\left[\rho_i(\mathbf{v})-\frac{g_i}{g_m}\rho_m(\mathbf{v})\right],\end{gathered}\tag{1}$$

where

$$\begin{gathered}NP_i(\mathbf{v})=-2\,\mathrm{Re}[iG^*\rho_{mi}(\mathbf{v})],\\ |G|^2=\frac{BI}{2\pi},\quad B=\frac{\lambda^2\Gamma_m}{4\hbar\omega}\,\frac{g_m}{g_n+g_l},\\ \Gamma_m=\Gamma_{mn}+\Gamma_{ml},\quad \Omega_{0i}=\omega-\omega_{mi},\quad i=n,l.\end{gathered}\tag{2}$$

Here $\rho_i(\mathbf{v})$ is the velocity distribution of the particles at the level i; $N=N_n+N_l+N_m$ is the concentration of the absorbing particles ($N_i=\int\rho_i(\mathbf{v})\,d\mathbf{v}$); $S_m(\mathbf{v})$, $S_i(\mathbf{v})$, and $S_{mi}(\mathbf{v})$ are the collision integrals; ω, λ, and $\mathbf{k}$ are the frequency, wavelength, and wave vector of the radiation; Γ_{mi} is the rate of spontaneous relaxation of the excited level m through the $m \to i$ channel; ω_{mi} is the frequency of the $m-i$ transition; and I is the radiation intensity. In the formula for the second Einstein coefficient B in (2) we have notet that in the adopted three-level model the ratio of the radiative transition rates from the level m to the hyperfine components n and l is determined by the ratio of the statistical weights:[32] $\Gamma_{mn}/\Gamma_{ml} = g_n/g_l$. The probability of radiation absorption per unit time, $P_i(\mathbf{v})$ (or absorption rate), in the $m-i$ transition by a particle with a fixed velocity $\mathbf{v}$ is determined by the off-diagonal element of the density matrix, $\rho_{mi}(\mathbf{v})$ (or coherence). Note that the last equation in (1) for the off-diagonal element $\rho_{mi}(\mathbf{v})$ is valid only if we ignore the coherence $\rho_{ln}(\mathbf{v})$ between the hyperfine components n and l. For lithium atoms this approximation holds only if the radiation intensity is not too high, or $I \ll 10\,\mathrm{W/cm^2}$ (Ref. 33).

For atoms of alkali metals in an atmosphere of noble buffer gases, the cross sections of the collisional transitions $n \to l$ and $l \to n$ between the component of the hyperfine structure of the ground state are extremely small—six to ten orders of magnitude smaller than the gaskinetic cross sections.[34] Bearing this in mind, below we will examine the case where there is no collisional exchange between the hyperfine components n and l, i.e., we will assume that the diagonal collision integrals $S_i(\mathbf{v})$ $(i=n,l,m)$ in (1) are due only to elastic scattering.

We integrate with respect to the velocity $\mathbf{v}$ the second equation in (1) and note that in elastic collisions $\int S_i(\mathbf{v})\,d\mathbf{v} = 0$. Combining the resulting equation, we get

$$\frac{P_n}{P_l} = \frac{\Gamma_{mn}}{\Gamma_{ml}} = \frac{g_n}{g_l}. \tag{3}$$

where $P_i = \int P_i(\mathbf{v})\,d\mathbf{v}$. This equation shows that the ratio of the integral rates of absorption in the $m-n$ and $m-l$ transitions depends neither on the intensity of the radiation nor the radiation frequency. The ratio characterizes the process of optical pumping of the hyperfine components of the ground state and follows from the absence of collisional exchange between the hyperfine components n and l.

In the absence of phase memory in collisions involving optical transitions (which is a natural assumption in nonlinear atomic spectroscopy), the off-diagonal collision integral has the form

$$S_{mi}(\mathbf{v}) = -[\gamma_{mi}(v) + i\Delta_{mi}(v)]\rho_{mi}(\mathbf{v}), \quad i=n,l, \tag{4}$$

where $\gamma_{mi}(v)$ and $\Delta_{mi}(v)$ are the collision broadening and shift of the levels, respectively. Combining (4) with the absorption rate $P_i(\mathbf{v})$ specified in (2), we get

$$NP_i(\mathbf{v}) = \frac{BI}{\pi} Y_i(\mathbf{v}) \left[\rho_i(\mathbf{v}) - \frac{g_i}{g_m}\rho_m(\mathbf{v}) \right], \tag{5}$$

where

$$Y_i(\mathbf{v}) = \frac{\Gamma_j(v)}{\Gamma_j^2(v) + [\Omega_i(v) - \mathbf{k}\cdot\mathbf{v}]^2},$$
$$\Gamma_j(v) = \frac{\Gamma_m}{2} + \gamma_{mi}(v), \tag{6}$$
$$\Omega_i(v) = \Omega_{0i} - \Delta_{mi}(v), \quad i=n, \ j=1; \quad i=l, \ j=2.$$

Here $\Gamma_1(v)$ and $\Gamma_2(v)$ are the homogeneous halfwidths of the absorption lines in the $m-n$ and $m-l$ transitions, respectively. In alkali-metal atoms the homogeneous halfwidths $\Gamma_1(v)$ and $\Gamma_2(v)$ are almost the same, so that we will assume $\Gamma_1(v) = \Gamma_2(v) \equiv \Gamma(v)$.

For the diagonal collision integral we use the model of particle ''arrival'' that is isotropic in velocities:

$$S_i(\mathbf{v}) = -\nu_i(v)\rho_i(\mathbf{v}) + S_i^{(2)}(v), \quad i=n,l,m, \tag{7}$$

where the arrival term $S_i^{(2)}(v)$ is a function of the absolute value of velocity, $v = |\mathbf{v}|$, and $\nu_i(v)$ in (7) is the transport collision rate.[28,35] The collisional model (7) allows for a velocity dependence of the collision rate and at the same time makes it possible to solve the problem analytically.

The relationship between the transport collision rate $\nu_i(v)$ in (7) and the characteristics of an elementary scattering act are given by the formula[10]

$$\nu_i(v) = \frac{q}{v^3}\int_0^\infty u^2 \exp\left\{ -\frac{u^2+v^2}{\bar{v}_b^2} \right\} F(uv)\sigma_i(u)\,du, \tag{8}$$

where

$$F(uv) = \frac{2uv}{\bar{v}_b^2}\cosh\frac{2uv}{\bar{v}_b^2} - \sinh\frac{2uv}{\bar{v}_b^2},$$
$$q = \frac{\mu}{M}\frac{N_b \bar{v}_b}{\sqrt{\pi}}, \quad \mu = \frac{MM_b}{M+M_b}, \quad \bar{v}_b = \sqrt{\frac{2k_BT}{M_b}}, \tag{9}$$

with N_b and M_b the concentration and mass of the buffer particles, M the mass of the particles that absorb radiation, k_B is the Boltzmann constant, T is the temperature, u is the relative velocity of the colliding particles before collision, and $\sigma_i(u)$ is the transport cross section of scattering of an absorbing particle in state i by a buffer particle. We can assume, to a high accuracy, that the transport rates of collisions involving the hyperfine components n and l are the same:[27]

$$\nu_l(v) = \nu_n(v). \tag{10}$$

The velocity of LID of the absorbing particle is defined by the relationship

$$\mathbf{u}_L \equiv \frac{\mathbf{j}_n + \mathbf{j}_l + \mathbf{j}_m}{N}, \quad \mathbf{j}_i = \int \mathbf{v}\rho_i(\mathbf{v})\,d\mathbf{v}, \tag{11}$$

where $\mathbf{j}_i$ is the partial flux of particles in state i. By combining (1), (7), and (11) with allowance for (10), we arrive at a formula for the LID velocity:

$$\mathbf{u}_L = \int \tau(v)\mathbf{v}[P_n(\mathbf{v}) + P_l(\mathbf{v})]\,d\mathbf{v},$$
$$\tau(v) = \frac{\nu_n(v) - \nu_m(v)}{\nu_n(v)[\Gamma_m + \nu_m(v)]}. \tag{12}$$

As noted in the Introduction, anomalous LID can arise only if the difference of the transport collision rates, $\Delta\nu(v) \equiv \nu_m(v) - \nu_n(v)$ [or, which is the same, the factor $\tau(v)$] changes its sign as a function of v. Here the absorbing particles with both positive values of $\tau(v)$ and negative values of $\tau(v)$ contribute to the drift velocity $\mathbf{u}_L$. This can cause a large deviation of the frequency dependence of $\mathbf{u}_L$ from the dispersion-curve-like dependence (as was the case when $\tau(v)=\text{const}$), including the occurrence of several additional zeros instead of one zero.

3. LOW RADIATION INTENSITIES

To calculate the drift velocity, we will limit ourselves to the case of low radiation intensities:

$$I \ll \Gamma(\Gamma_m+\nu_m)/B\,, \tag{13}$$

where ν_m and Γ are the characteristic (average) values of $\nu_m(v)$ and $\Gamma(v)$. If condition (13) is met, then in (5) we can ignore the population of the excited level ($\rho_m(\mathbf{v})=0$), and the population velocity distributions in the hyperfine components $i=n,l$, can be considered essentially Maxwellian ($\rho_i(\mathbf{v})=N_iW(\mathbf{v})$, where $W(\mathbf{v})$ is the Maxwellian distribution). Here from (5) we find that

$$P_i(\mathbf{v})=\frac{BI}{\pi}\frac{N_i}{N}Y_i(\mathbf{v})W(\mathbf{v}). \tag{14}$$

Combining (3) and (14) and using the normalization condition $N_n+N_l\simeq N$ (here we have noted that $N_m\ll N$ holds if the conditions (13) are met), we find that

$$\begin{gathered}\frac{N_n}{N}=\frac{w_n\langle Y_l\rangle}{w_n\langle Y_l\rangle+w_l\langle Y_n\rangle},\quad \frac{N_l}{N}=1-\frac{N_n}{N},\\ \langle Y_i\rangle=\int Y_i(\mathbf{v})W(\mathbf{v})\,d\mathbf{v},\quad w_i=\frac{g_i}{g_n+g_l},\quad i=n,l.\end{gathered} \tag{15}$$

Plugging (14) and (15) into (12) and integrating over the directions of $\mathbf{v}$, we arrive at the final expression for the LID velocity $\mathbf{u}_L$, which we write as

$$\mathbf{u}_L\equiv\frac{\mathbf{k}}{k}u_L\,,\quad u_L=u_0u(x), \tag{16}$$

where we have introduced the parameter u_0 with dimensions of velocity,

$$u_0=\frac{2BI}{\pi^{3/2}k\Gamma_m}, \tag{17}$$

and the dimensionless velocity u as a function of the dimensionless detuning x of the radiation frequency,

$$\begin{aligned}u(x)=&\frac{1}{w_n\langle Y_l\rangle+w_l\langle Y_n\rangle}\int_0^\infty t\tau_0(t)\exp(-t^2)\\ &\times[w_n\langle Y_l\rangle f_n(t)+w_l\langle Y_n\rangle f_l(t)]\,dt.\end{aligned} \tag{18}$$

Here we have introduced the following functions of the dimensionless velocity $t=v/\bar{v}$:

$$\begin{gathered}f_i(t)=x_i(t)\psi_i(t)+\frac{y(t)}{2}\ln\frac{y^2(t)+[t-x_i(t)]^2}{y^2(t)+[t+x_i(t)]^2},\\ \psi_i(t)=\arctan\frac{t+x_i(t)}{y(t)}+\arctan\frac{t-x_i(t)}{y(t)},\\ \tau_0(t)=\frac{\nu_n(t)-\nu_m(t)}{\nu_n(t)[1+\nu_m(t)/\Gamma_m]},\quad y(t)=\frac{\Gamma(t)}{k\bar{v}},\\ x_i(t)=\frac{\Omega_i(t)}{k\bar{v}},\quad \bar{v}=\sqrt{\frac{2k_BT}{M}},\quad i=n,l.\end{gathered} \tag{19}$$

In Eq. (18), the quantities $\langle Y_i\rangle$, which are defined in (15), assume the form

$$\langle Y_i\rangle=\frac{2}{\sqrt{\pi}\,k\bar{v}}\int_0^\infty t\exp(-t^2)\,\psi_i(t)\,dt,\quad i=n,l. \tag{20}$$

For the dimensionless detuning x of the radiation frequency it is convenient to use in (18) the quantity

$$x=\frac{\Omega}{k\bar{v}}, \tag{21}$$

where

$$\Omega=\omega-\omega_0\,,\quad \omega_0=w_n\omega_{mn}+w_l\omega_{ml}\,. \tag{22}$$

The frequency ω_0 corresponds to the ''center of gravity'' of the transition frequencies ω_{mn} and ω_{ml} with allowance for the statistical weights of the levels *n* and *l*. In the region of ''normal'' LID [where we can put $\tau_0(t)=\text{const}$ in (18)], the drift velocity $\mathbf{u}_L$ vanishes only at one point, $x=0$ (see Ref. 36 and Fig. 4 below).

When calculating the LID velocity of atoms placed in a mixture of two different buffer gases, in the formula (19) for $\tau_0(t)$ we must put

$$\nu_i(t)=\nu_{1i}(t)+\nu_{2i}(t),\quad i=n,m, \tag{23}$$

where the subscripts 1 and 2 denote the type of buffer particles.

4. ANOMALOUS LID OF LITHIUM ATOMS

The LID of lithium atoms in noble buffer gases was studied numerically via formulas (8) and (18)–(23). The transport collision rates $\nu_i(t)\equiv\nu_i(t\bar{v})\equiv\nu_i(v)$ for Li–X systems, where X is an atom of a noble gas, were calculated numerically using (8) with the transport cross sections $\sigma_i(u)$ taken from Ref. 25, where they were calculated using the semiempirical interaction potential of Pascale and Vandeplanque.[37]

As noted earlier, the anomalous behavior of LID is completely determined by the alternating-sign dependence of the factor τ_0 on the velocity $t=v/\bar{v}$. Allowance in (18)–(20) for the dependence of the homogeneous halfwidth Γ and the frequency detuning Ω_i on the velocity t can only refine the spectrum $u(x)$ of the velocity of anomalous LID, without changing it qualitatively. But when there is Doppler broadening of the absorption line (in the limit $\Gamma\ll k\bar{v}$), the effect of $\Gamma(t)$ and $\Omega_i(t)$ on the spectrum $u(x)$ of the LID velocity is usually small and can be ignored.

Indeed, when there is Doppler broadening of the absorption line, the parameter $y(t)$ in (19) is small: $y(t)\ll 1$. This makes it possible to ignore the fact that the parameter y in (19) is t-dependent, since in the limit $y(t)\to 0$ the functions $f_i(t)$ and $\psi_i(t)$ in (19) is in no way dependent on the param-

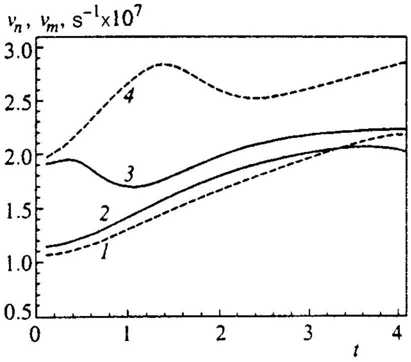

FIG. 2. Transport collision rate as a function of the dimensionless velocity $t=v/\overline{v}$ of lithium atoms at $T=300$ K and a buffer gas pressure $P=1$ torr: curve *1*, $\nu_m(t)$ for the ^{7}Li–Ne system; curve *2*, $\nu_n(t)$ for the ^{7}Li–Ne system; curve *3*, $\nu_n(t)$ for the ^{7}Li–Kr system; and curve *4*, $\nu_m(t)$ for the ^{7}Li–Kr system.

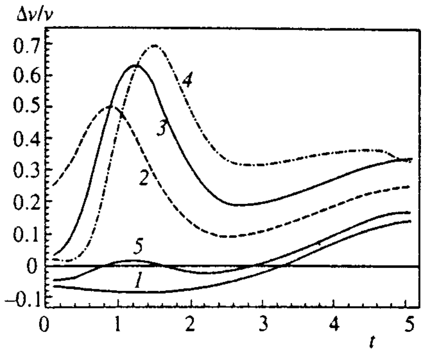

FIG. 3. Relative difference of the transport collision rates, $\Delta\nu/\nu$, as a function of the dimensionless velocity $t=v/\overline{v}$ at $T=300$ K for different systems: curve *1*, the ^{7}Li–Ne system; curve *2*, the ^{7}Li–Ar system; curve *3*, the ^{7}Li–Kr system; curve *4*, the ^{7}Li–Xe system; and curve *5*, the ^{7}Li–(Ne + Kr) system with the neon fraction $\xi=0.875$.

eter $y(t)$ [in this limit the function $\psi_i(t)$] has the shape of a step beginning at $t=|x_i(t)|$, i.e., $\psi_i(t)=\pi$ for $t\geqslant|x_i(t)|$ and $\psi_i(t)=0$ for $0\leqslant t<|x_i(t)|$). In view of what we have just said, in calculating the LID velocity in Doppler broadening conditions we ignored the *t*-dependence of the homogeneous halfwidth of the absorption line, Γ, and put $\Gamma(t)=\Gamma=\text{const}$ in (19). The specific values of $\Gamma=\Gamma_m/2+\gamma$ for different Li–X systems were found from the data of Allard and Kielkopf[38] on collision broadening γ of the absorption line ($\gamma=3.86$ MHz/torr for Li–Ne, $\gamma=5.31$ MHz/torr for Li–Ar, $\gamma=7.00$ MHz/torr for Li–Kr, and $\gamma=7.96$ MHz/torr for Li–Xe.

The *t*-dependence of the dimensionless frequency detuning x_i can affect the functions $f_i(t)$ and $\psi_i(t)$ in the integral (18) for the LID velocity $u(x)$ only if $|\Omega_{0i}|\lesssim\Delta_{mi}(t)$. Since the collision shift $\Delta_{mi}(t)$ of the levels is usually several times smaller than the collision broadening γ (Refs. 38 and 39) and so $|\Delta_{mi}(t)|<\gamma<\Gamma$ holds, in the event of Doppler broadening of the absorption line (for $y=\Gamma/k\overline{v}\ll 1$) the effect of $x_i(t)$ on the functions $f_i(t)$ and $\psi_i(t)$ manifests itself only if $|x_i(t)|\lesssim y\ll 1$, i.e., in a narrow range of detuning near the zero values of $x_i(t)$. In view of this, in calculating the LID velocity we also ignored the *t*-dependence of the frequency detunings $\Omega_i(t)$ in (19).

At $x_i(t)=x_i=\text{const}$ the dimensionless frequency detuning x given by (21) is related to the detunings x_i by the following relationships:

$$x_n=x-\delta w_l\,,\quad x_l=x+\delta w_n\,,\quad \delta=\frac{\omega_{ln}}{k\overline{v}},\tag{24}$$

where δ is the dimensionless distance between the components of the hyperfine structure of the ground state.

Figures 2 and 3 depict the transport collision rates ν_i and the relative difference of transport collision rates

$$\frac{\Delta\nu}{\nu}\equiv\frac{\nu_m(t)-\nu_n(t)}{\nu_n(t)},\tag{25}$$

calculated according to Eq. (8) for ^{7}Li in different buffer gases as functions of the dimensionless velocity $t=v/\overline{v}$. A characteristic feature of the behavior of the collision rates $\nu_i(t)$ for lithium atoms in Ar, Kr, or Xe as a buffer gas is the nonmonotonic dependence on the velocity *t*. A characteristic feature of the ^{7}Li–Ne system is the decrease in the transport collision rate within a certain velocity interval in the event of lithium-atom excitation (the curves *1* and *2* in Fig. 2 and the curve *1* in Fig. 3). It is this feature that ensures a controllable (as required by the experimenter) dependence of the factor $\Delta\nu/\nu$ [and hence the factor $\tau_0(t)$ in (18)] needed for observing anomalous LID in the buffer mixture of neon with another noble gas (Ar, Kr, or Xe) by varying the fraction ξ of neon in the buffer mixture:

$$\xi=\frac{N_{\text{Ne}}}{N_b},\quad N_b=N_{\text{Ne}}+N_{\text{X}},\tag{26}$$

where N_{Ne} is the neon concentration, N_{X} is the concentration of the other noble buffer gas (Ar, Kr, or Xe), and N_b is the total concentration of the buffer gas. The alternating-sign factor $\Delta\nu/\nu$ may vanish three times at velocities depending on the fraction ξ of neon in the buffer mixture (see curve *5* in Fig. 3 for the buffer mixture Ne+Kr; for the buffer mixtures Ne+Ar and Ne+Xe the corresponding curves are similar to

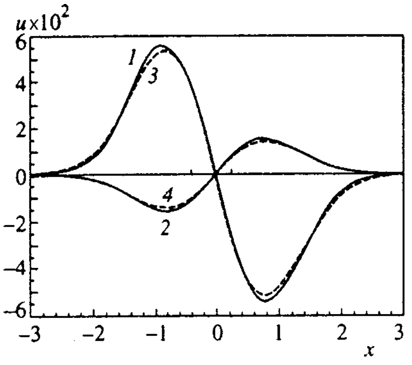

FIG. 4. Dimensionless drift velocity $u(x)$ as a function of the dimensionless frequency detuning $x=\Omega/k\overline{v}$ of the radiation for ^{7}Li atoms in Kr or Ne as a buffer gas in the case of normal LID at $T=300$ K and a buffer-gas pressure $P=5$ torr. The short vertical lines in the middle of the diagram indicate the frequencies in resonance with the m–n and m–l transitions. The dashed curves *3* and *4* represent the results of calculations by formulas (18)–(23) where $\nu_i(t)$ is replaced by $\langle\nu_i\rangle$. Curves *1* and *3* correspond to the ^{7}Li–Kr system and curves *2* and *4*, to the ^{7}Li–Ne system.

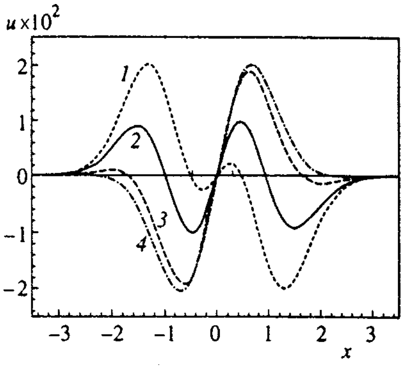

FIG. 5. Drift velocity u as a function of the frequency detuning x for ^{7}Li atoms in Ne as a buffer gas at a pressure $P=5$ torr for different temperatures: curve *1* corresponds to $T=1600$ K, curve *2* to 1250 K, curve *3* to 800 K, and curve *4* to 600 K.

curve *5*). This behavior of the factor $\Delta\nu/\nu$ means that the anomalous LID of lithium atoms in a buffer mixture can be observed at almost all temperatures by specifying the value of the parameter ξ, which can easily be controlled in the experiment.

The results of numerical calculations of the LID velocity $u(x)$ by formulas (18)–(23) are illustrated in Figs. 4–8.

Figure 4 depicts the dependence of the LID velocity of ^{7}Li atoms on the frequency detuning in a one-component buffer gas Ne or Kr at $T=300$ K. In this case the drift velocity spectrum $u(x)$ corresponds to normal LID and is described satisfactorily by a LID theory in which the transport collision rates are independent of the velocity t, i.e., $\nu_i(t)$ in (19) is replaced by the average transport rate

$$\langle \nu_i\rangle = \frac{2}{\overline{v^2}}\int (\mathbf{n}\cdot\mathbf{v})^2 W(\mathbf{v})\,\nu_i(v)\,d\mathbf{v} = \frac{8}{3\sqrt{\pi}}\,\frac{\mu}{M}\,\frac{N_b}{\bar{u}^5}\int_0^\infty u^5 \exp\left(-\frac{u^2}{\bar{u}^2}\right)\sigma_i(u)\,du, \qquad (27)$$

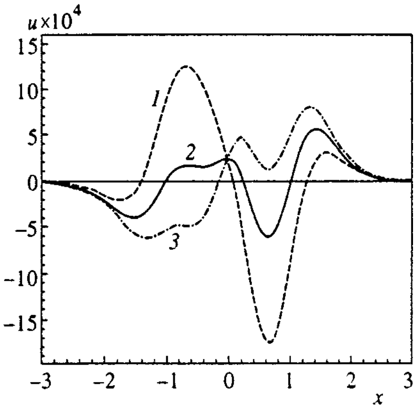

FIG. 6. Drift velocity u as a function of the frequency detuning x for ^{7}Li atoms in the buffer mixture Ne+Kr at $T=300$ K, pressure $P=5$ torr, and different neon fractions ξ: curve *1* corresponds to $\xi=0.865$, curve *2* to $\xi=0.875$, and curve *3* to $\xi=0.81$. The vertical lines indicate the frequencies in resonance with the $m-n$ and $m-l$ transitions.

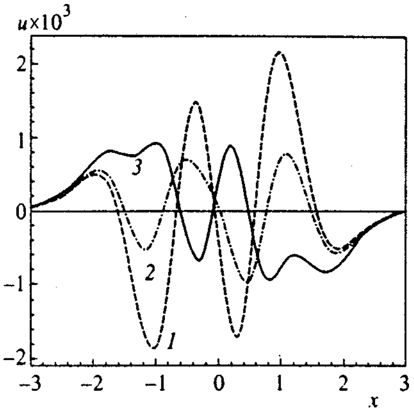

FIG. 7. Drift velocity u as a function of the frequency detuning x for ^{7}Li atoms in mixtures of two buffer gases at $T=600$ K and pressure $P=5$ torr: curve *1* corresponds to ^{7}Li–(Ne+Ar) and $\xi=0.765$, curve *2* to ^{7}Li–(Ne+Kr) and $\xi=0.865$, and curve *3* to ^{7}Li–(Ne+Xe) and $\xi=0.9$. The vertical lines indicate the frequencies in resonance with the $m-n$ and $m-l$ transitions.

where $\bar{u}=(2k_BT/\mu)^{1/2}$, and $\mathbf{n}$ is a unit vector pointing in an arbitrarily chosen direction. The average transport rate $\langle\nu_i\rangle$ is related to the diffusion coefficient D_i of particles in state i by the simple relationship $D_i=\overline{v^2}/2\langle\nu_i\rangle$.

Formula (18) clearly shows that, due to the factor $t\exp(-t^2)$, the principal contribution to the integral for the drift velocity $u(x)$ is provided by particles with velocities $t\sim1$. In the vicinity of $t\sim1$ at $T=300$ K, the sign of the factor $\Delta\nu/\nu$ for the systems ^{7}Li–Ne and ^{7}Li–Kr does not change (see Fig. 3), which is the reason why normal LID is observed.

As the temperature grows, the thermal velocity $\bar{v}$ and the region near $t=v/\bar{v}\sim1$ in Figs. 2 and 3 move to the right. Now for the ^{7}Li–Ne system the relative difference of the transport collision rates, $\Delta\nu/\nu$, in the region where $t\sim1$ holds becomes an alternating-sign function (curve *1* in Fig. 3) and anomalous LID arises, which is illustrated by Fig. 5. The curves *1*, *2*, and *3* in Fig. 5 correspond to anomalous LID, with the curves *1* and *4* outlining the limits of anomalous LID diagrams and the curve *2* ($T=1250$ K) corresponding to the maximum anomalous LID effect (in the ranges where the x are either positive or negative the amplitudes of

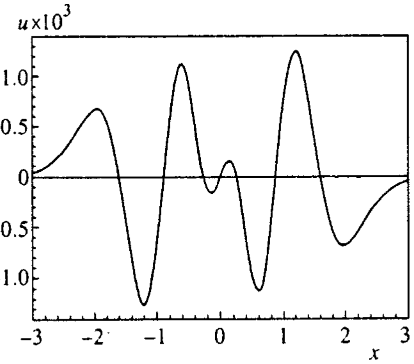

FIG. 8. Drift velocity u as a function of the frequency detuning x for ^{6}Li atoms in the buffer mixture Ne+Kr at $T=600$ K, pressure $P=5$ torr, and a neon fraction $\xi=0.865$.

the maximum and minimum drift velocities $u(x)$ are approximately the same). Note that in the case of anomalous LID the drift velocity decreases insignificantly in comparison to the case of normal LID (for the curves *2* in Fig. 4 and 5 the maxima of $u(x)$ differ only by a factor of 1.5). This feature of anomalous LID in the case at hand is due to the specific behavior of the relative difference of the transport collision rates, $\Delta\nu/\nu$ (curve *1* in Fig. 3): as the function $\Delta\nu/\nu$ passes through zero, it rather soon (in the interval $\Delta t \sim 1$) reaches a plateaulike curve, where it is weakly dependent on the velocity t ($\Delta\nu/\nu \approx \mathrm{const}$). Hence in the case of anomalous LID the region $\Delta\nu/\nu \approx \mathrm{const}$, which ensures normal LID, contributes significantly to the integral (18) for LID velocity, and this results in a small decrease in anomalous LID.

Figure 6 depicts the spectrum $u(x)$ of the drift velocity of ^{7}Li atoms at different neon fractions ξ in the buffer mixture Ne+Kr. We see that at a given temperature, anomalous LID can be observed only within a narrow range of values of the neon fraction in the buffer mixture, $\Delta\xi \sim 0.02$. Numerical analysis demonstrates the same strong sensitivity of anomalous LID to changes in the neon concentration in other buffer mixtures, Ne+Ar and Ne+Xe. A characteristic feature of the anomalous spectrum $u(x)$ of the drift velocity in a buffer mixture of two gases is the total lack of antisymmetry in the spectrum $u(x)$, in contrast to the cases of normal and anomalous LID in a one-component buffer gas (see Figs. 4 and 5). In a binary buffer mixture, the shape of the velocity spectrum $u(x)$ may be quite complicated (see, e.g., the curve *3* in Fig. 6).

Examples of the spectra of anomalous LID in ^{7}Li in different binary buffer mixtures are depicted in Fig. 7.

If we compare Fig. 4 (curve *2*) and Figs. 6 and 7, we see that in the case of anomalous LID of ^{7}Li atoms in buffer mixtures of two gases the drift velocity decreases in comparison to normal LID by a factor of 10 to 20.

The anomalous LID spectrum for ^{7}Li atoms can have one (curve *3* in Fig. 6), three, or five zeros (see Figs. 5–7). For ^{6}Li atoms the anomalous LID spectrum can even have seven zeros (Fig. 8).

5. CONCLUSION

We have studied the anomalous LID of lithium atoms, which in the case of normal LID in noble buffer gases have the simplest dispersion-curve-like antisymmetric spectrum $u(x)$ of the drift velocity, a spectrum which coincides with that for two-level particles. Although lithium atoms are described by a three-level model, the velocity spectrum of normal LID has a simple ‘‘two-level’’ shape due to optical pumping of the hyperfine components of the ground state, since there is no collision exchange between the hyperfine components in the case of noble buffer gases. The extreme simplicity of the velocity spectrum of normal LID of lithium atoms makes it possible to easily notice in experiments any small deviations from the normal drift-velocity spectrum.

The anomalous LID velocity spectrum is extremely sensitive to alternating-sign dependence of the difference $\nu_m - \nu_n$ in the transport rates of collisions of buffer particles and the excited (m) and unexcited (n) atoms on the velocity v of the resonant atoms. Since the transport collision rates $\nu_i(v)$ are completely determined by the potentials of the interaction between resonant and buffer particles, the velocity spectrum of anomalous LID is highly sensitive to the difference in the potentials of the interaction of buffer particles and resonant atoms in the m and n states. This makes possible highly precise testing in the LID experiments of the interatomic potentials used in calculations of the velocity spectrum of anomalous LID and hence relatively simple experimental testing of the accuracy of various theoretical methods used in calculating the interaction potentials.

A good criterion for determining the possibility of anomalous LID arising in a one-component buffer gas is the alternating-sign temperature dependence of the difference of averaged transport rates (27) of collisions, $\langle\nu_m\rangle - \langle\nu_n\rangle$ (or, what the same, of the difference of the diffusion coefficients, $D_m - D_n$, in the excited and ground states) of the resonant atoms and buffer particles. In the case of a binary buffer mixture, anomalous LID can occur if the differences of the averaged transport collision rates, $\langle\nu_m\rangle - \langle\nu_n\rangle$ for the resonant atoms in the two buffer gases have opposite signs. Anomalous LID is expected to exist at temperatures in a one-component buffer gas and at relative fractions of the gases in a binary buffer mixture at which the difference $\langle\nu_m\rangle - \langle\nu_n\rangle$ (or $DR_m - D_n$) vanishes. These criteria were used in Refs. 19, 25, and 27 and in the present paper to determine possible objects to manifest anomalous LID.

Using the above criteria to analyze the results of calculations of transport characteristics of alkali-metal atoms via the Pascale–Vandeplanque interaction potentials[37] (the results of these calculations are partially discussed in Ref. 25), we can conclude that anomalous LID should be expected to manifest itself in the following cases:

1. In the excitation of the D_1 line of Rb atoms in the mixtures Ne+Kr, Ne+Xe, He+Kr, and He+Xe at almost all temperatures (depending on the choice of the buffer-gas fractions in the binary buffer mixture);

2. In the excitation of the D_1 line of Cs atoms in the mixtures Ne+Xe, He+Xe, and Kr+Xe at almost all temperatures; and

3. In the excitation of the D_2 line of Cs atoms in the mixtures He+Ar, He+Kr, He+Xe, and He+Ne at temperatures $T \gtrsim 300\,K$, and in the following one-component buffer gases: He at $T \sim 300$ K, Ar at $T \sim 1000$ K, and Kr at $T \sim 1600$ K.

In contrast to lithium atoms, Rb and Cs atoms satisfy the limiting case of weak collisional coupling between the fine components $^2P_{1/2}$ and $^2P_{3/2}$. Hence the transport characteristics of Rb and Cs atoms in the excited states $^2P_{1/2}$ and $^2P_{3/2}$ are different (see, e.g., Ref. 40), and so are the LID velocity spectrum in the cases of excitation of D_1 and D_2 lines.

The author would like to thank Prof. J. Pascale for sending detailed tables listing the potentials of the interaction of alkali-metal atoms and atoms of noble gases. The investigation whose results are presented in this paper was sponsored by the Russian Fund for Fundamental Research (Grant No. 98-02-17924).

*)E-mail: shalagin@iae.nsk.su

---

[1] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).

[2] A. M. Shalagin, ''Light-induced drift,'' in *Physics Encyclopedia* [in Russian], Vol. 4, A. M. Prokhorov (Ed.), Bol'shaya Rossiĭskaya Éntsiklopediya, Moscow (1994), p. 468.

[3] A. K. Popov, A. M. Shalagin, V. M. Shalaev, and V. Z. Yakhnin, Zh. Éksp. Teor. Fiz. **80**, 2175 (1981) [Sov. Phys. JETP **53**, 1134 (1981)].

[4] S. N. Atutov, X. St. Lesjak, S. P. Pod'yachev, and A. M. Shalagin, Opt. Commun. **60**, 41 (1986).

[5] S. N. Atutov, I. M. Ermolaev, and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **92**, 1215 (1987) [Sov. Phys. JETP **65**, 679 (1987)].

[6] G. Nienhuis, Phys. Rep. **138**, 151 (1986).

[7] H. G. C. Werij and J. P. Woerdman, Phys. Rep. **169**, 145 (1988).

[8] E. R. Eliel, Adv. At., Mol., Opt. Phys. **30**, 199 (1992).

[9] L. J. F. Hermans, Int. Rev. Phys. Chem. **11**, 289 (1992).

[10] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Non-linear Spectroscopy*, North-Holland, Amsterdam (1991).

[11] V. P. Mironenko and A. M. Shalagin, Izv. Akad. Nauk SSSR, Ser. Fiz. **45**, 995 (1981).

[12] P. L. Chapovskiĭ, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 1069 (1989).

[13] A. E. Bakarev, A. L. Makas', and P. L. Chapovskiĭ, Kvant. Elektron. (Moscow) **13**, 30 (1986) [Sov. J. Quantum Electron. **16**, 16 (1986)].

[14] G. J. van der Meer, J. Smeets, S. P. Pod'yachev, and L. J. F. Hermans, Phys. Rev. A **45**, R1303 (1992).

[15] P. L. Chapovsky, G. J. van der Meer, J. Smeets, and L. J. F. Hermans, Phys. Rev. A **45**, 8011 (1992).

[16] G. J. van der Meer, J. Smeets, E. R. Eliel, P. L. Chapovsky, and L. J. F. Hermans, Phys. Rev. A **47**, 529 (1992).

[17] E. J. van Duijn, H. I. Bloemink, E. R. Eliel, and L. J. F. Hermans, Phys. Lett. A **184**, 93 (1993).

[18] E. J. van Duijn, R. Nokhai, and L. J. F. Hermans, J. Chem. Phys. **105**, 6375 (1996).

[19] F. Yahyaei-Moayyed and A. D. Streater, Phys. Rev. A **53**, 4331 (1996).

[20] B. Nagels, P. L. Chapovsky, L. J. F. Hermans, G. J. van der Meer, and A. M. Shalagin, Phys. Rev. A **53**, 4305 (1996).

[21] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Phys. Lett. A **162**, 45 (1992).

[22] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Zh. Éksp. Teor. Fiz. **102**, 424 (1992) [Sov. Phys. JETP **75**, 225 (1992)].

[23] F. Kh. Gel'mukhanov, G. V. Kharlamov, and S. G. Rautian, Opt. Commun. **94**, 521 (1992).

[24] I. Kuščer, L. J. F. Hermans, P. L. Chapovsky, J. J. M. Beenakker, and G. J. van der Meer, J. Phys. B **26**, 2837 (1993).

[25] F. Kh. Gel'mukhanov and A. I. Parkhomenko, J. Phys. B **28**, 33 (1995).

[26] A. I. Parkhomenko, Opt. Spektrosk. **80**, 604 (1996) [Opt. Spectrosc. **80**, 541 (1996)].

[27] F. Kh. Gel'mukhanov, A. I. Parkhomenko, T. I. Privalov, and A. M. Shalagin, J. Phys. B **30**, 1819 (1997).

[28] A. I. Parkhomenko and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **115**, 1664 (1999) [JETP **88**, 913 (1999)].

[29] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions*, Springer-Verlag, Berlin (1985).

[30] R. H. G. Reid, J. Phys. B **8**, L493 (1975).

[31] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).

[32] I. I. Sobelman, *Atomic Spectra and Radiative Transitions*, Springer-Verlag, Berlin (1979).

[33] A. D. Streater and J. P. Woerdman, J. Phys. B **22**, 677 (1989).

[34] W. Happer, Rev. Mod. Phys. **44**, 169 (1972).

[35] L. V. Il'ichev and A. I. Parkhomenko, Zh. Éksp. Teor. Fiz. **112**, 856 (1997) [JETP **85**, 462 (1997)].

[36] S. N. Atutov, A. I. Parkhomenko, S. P. Pod''yachev, and A. M. Shalagin, J. Phys. B **25**, 2943 (1992).

[37] J. Pascale and J. Vandeplanque, J. Chem. Phys. **60**, 2278 (1974).

[38] N. Allard and J. Kielkopf, Rev. Mod. Phys. **54**, 1103 (1982).

[39] I. I. Sobelman, L. A. Vainshtein, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines*, Springer-Verlag, Berlin (1981).

[40] W. A. Hamel, J. E. M. Haverkort, H. G. C. Werij, and J. P. Woerdman, J. Phys. B **19**, 4127 (1986).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Effect of thermophoretic forces on the formation of ordered structures of macroparticles in a thermal plasma

V. E. Fortov, A. P. Nefedov, O. F. Petrov,* A. A. Samaryan, Ya. K. Khodataev,
and A. V. Chernyshev

*Research Center for the Thermal Physics of Pulsed Interactions, Russian Academy of Sciences,
127412 Moscow, Russia*

(Submitted 2 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1601–1615 (November 1999)

The thermophoretic interaction of macroparticles and its effect on the formation of ordered structures of macroparticles in plasma was studied. It was shown that coexistence of regions with a chaotic arrangement of particles and regions of ordered structures is typical of a thermal plasma with strong interaction of the macroparticles. Computer simulation of a system of strongly interacting macroparticles, taking account of the thermophoretic interaction of the particles, was performed. The results showed that the thermophoretic attractive forces explain the form of the spatial nonuniformity associated with the grouping of particles in small domains. The experimentally obtained correlation function was very close to the correlation function obtained in the computer simulation. © *1999 American Institute of Physics.* [S1063-7761(99)00811-2]

1. INTRODUCTION

A thermal plasma with macroscopic particles is a low-temperature plasma containing small liquid or solid particles of matter. The macroparticles effectively interact with the charged components of the plasma and therefore they strongly influence the properties of the plasma. Thus, electron-emitting particles can increase the electron density and electric conductivity of the gas phase. However, if the particles trap electrons, then the opposite effect arises. In the limit of an un-ionized gas the presence of macroparticles completely determines the electrophysical properties of the plasma. Effects associated with the presence of particles were observed in the very first experiments[1,2] studying the plasma of a hydrocarbon flame.

An interesting feature of such a plasma is that because of their relatively large sizes (from hundredths of a micron to some tens of microns), the particles can carry extremely large charges (of the order of 10^2-10^5 electron charges). As a result, under certain conditions strong interparticle correlation leads to gas—liquid—solid phase transitions and the appearance of spatially ordered structures in the arrangement of the macroscopic particles, similar to the structures in a liquid or solid.[3–7] Such structures have also been observed for CeO_2 particles in a thermal plasma at atmospheric pressure.[8–10] The plasma formation was characterized by large sizes (the plasma volume was about $30\,cm^3$, which corresponded to $\sim 10^8$ particles with density $10^7\,cm^{-3}$), uniformity, and a lack of external electric and magnetic fields.

These investigations were primarily of a phenomenological character, having set aside a number of questions concerning the mechanism of particle interaction and the anomalous value of the kinetic energy of the particles.[7,11] The interaction between the macroparticles plays a large role in the dynamics of the formation of ordered structures in a dusty system, for example, accompanying the formation of structures with pronounced short-range order over the comparatively short lifetimes (of the order of several milliseconds) of a thermal plasma in a laboratory setup.[9]

Our objective in the present work was to study the nonelectric (thermophoretic) interaction of macroparticles and its effect on the formation of ordered structures in a thermal plasma. This particle interaction mechanism results from the decoupling of the particle temperature from the temperature of the ambient gas. In the present work experimental measurements of the temperature of the gas and macroparticles in a thermal plasma and numerical simulation of the interaction of macroparticles in a thermal plasma, taking account of the thermophoretic attractive forces, were performed. In what follows the results of the analytic calculations of the interaction between macroparticles and the data from the numerical simulation of the formation of dust structures with various types of interactions of the macroparticles are discussed. The correlation functions obtained in the laboratory and in the numerical simulation are compared.

2. INTERACTION OF MACROPARTICLES IN PLASMA

A Debye pairwise interaction of macroparticles is ordinarily assumed in plasma:[12–14]

$$\Phi(r)=\Phi_D(r)=-Z_d e\,\frac{\partial \phi_D}{\partial r}, \tag{1}$$

where

$$\phi_D=\frac{Z_d e}{r}\exp\left(-\frac{r}{r_D}\right). \tag{2}$$

Here r_D is the Debye radius, ϕ_D is the Debye potential, e is the elementary charge, and Z_d is the charge of a macroparticle in units of the elementary charge. This expression for ϕ_D pertains to the case of interest, with $R_d \ll r_D$ (R_d is the radius of a macroparticle).

At short distances the electric interaction force is the Coulomb force, and at large distances (compared with r_D) it decreases exponentially because of plasma screening and is therefore a short-range force.

As established recently, besides the electric interaction of macroparticles there also exists a nonelectric interaction resulting from flows of microparticles.[15] Ordinarily, this type of interaction results from the breakdown of thermodynamic equilibrium on the surface of the macroparticles. Specifically, for the case most often encountered in practice—negatively charged particles—each macroparticle gives rise to a plasma flow (of positive ions) onto itself. Neighboring particles interact by frictional forces resulting from these flows, which results in an effective attraction of macroparticles. It has been shown that such a nonelectric attraction can be strong enough to compete with the electrostatic repulsion and give rise to the formation of molecules of macroparticles as well as crystals of macroparticles with a free boundary, which exist independently of the external forces.

In our case this mechanism, due to bombardment of the surface of the macroparticles, is ineffective, since the particles are positively charged and therefore the flux of positive ions onto their surface is relatively small. However, in this case, if the temperature of gas is different from that of the surface of the particles, a different mechanism can arise, due to flows of neutral gas molecules. A temperature difference can, for example, result from radiative cooling of the macroparticle material. Being colder than the gas, a particle gives rise to a heat flow onto its surface and a temperature gradient in the ambient gas. A neighboring particle will move under the thermophoretic force into a region of lower temperature, i.e., it will be attracted. In this manner, a special mechanism of attraction arises. The corresponding interaction force will be calculated below.

We consider two macroparticles of radius R_d and surface temperature T_{ds} in a neutral gas. The temperature and density of the gas far from the particles are T_g and n_g, respectively. Let

$$R_d \ll \lambda \ll l, \tag{3}$$

where λ is the mean-free path of the neutral molecules in the gas and l is the distance between particles. To calculate the heat flux onto a particle, the effect of the other particle can be neglected. Then, assuming perfect accommodation, it is easy to find the distribution of gas molecules on the surface of a particle:

$$f_s=\begin{cases} \dfrac{n_g m^{3/2}}{(2\pi T_g)^{3/2}}\exp\left(-\dfrac{mV^2}{2T_g}\right), & \mathbf{V}\cdot\mathbf{e}_n<0,\\[2ex] \dfrac{n_{ds} m^{3/2}}{(2\pi T_{ds})^{3/2}}\exp\left(-\dfrac{mV^2}{2T_{ds}}\right), & \mathbf{V}\cdot\mathbf{e}_n>0,\end{cases} \tag{4}$$

where $\mathbf{e}_n$ is the normal to the surface of the macroparticle and m is the mass of a gas molecule. Since particles do not absorb gas molecules in our case, the parameter n_{ds} is determined by the requirement that the total flux of neutral molecules onto the particle surface be zero, which gives

$$n_{ds}=n_g\sqrt{\frac{T_g}{T_{ds}}}. \tag{5}$$

The heat flux, corresponding to the distribution (4), onto the surface of a macroparticle is given by the formula

$$\mathbf{q}_s=\int f_s\frac{mV^2}{2}(\mathbf{V}\cdot\mathbf{e}_n)\mathbf{e}_n d^3V$$

$$=\mathbf{e}_n\frac{n_g}{\sqrt{2\pi}}\sqrt{\frac{T_g}{\sqrt{2\pi}}}(T_{ds}-T_g). \tag{6}$$

Since the heat conduction equation is linear, the heat flux in the gas will be a simple sum of heat fluxes produced by each particle. The magnitude of the temperature gradient due to one particle can be written in the stationary case as

$$q=q_s\frac{R_d^2}{r^2}=\frac{R_d^2}{r^2}\frac{n_g}{\sqrt{2\pi}}\sqrt{\frac{T_g}{m}}(T_{ds}-T_g). \tag{7}$$

The corresponding temperature gradient is

$$\frac{\partial T}{\partial r}=\frac{q}{\kappa}, \tag{8}$$

where κ is the thermal conductivity. A neighboring particle interacts with the thermophoretic force, which for $R_d \ll \lambda$ is

$$\Phi=-\frac{32}{15}\sqrt{\frac{\pi m}{8T_g}}R_d^2\kappa\frac{\partial T}{\partial r}. \tag{9}$$

Substituting the expressions (7) and (8) into Eq. (9), we obtain the pairwise interaction force of the macroparticles

$$\Phi_T=\frac{16}{15}n_g\frac{R_d^4}{l^2}(T_{ds}-T_g). \tag{10}$$

It can plausibly be assumed that this force is indeed a pairwise force: because the heat conduction equation is linear, the heat fluxes resulting from different particles and the corresponding thermophoretic forces can be calculated independently. The magnitude of this force is inversely proportional to the squared distance, i.e., it is analogous to the force of gravity, and it is directly proportional to the temperature difference ΔT. Therefore if $T_{ds}<T_g$, then macroparticles attract one another, and in the opposite case they repel one another.

The relative role of the electric and thermophoretic mechanisms can be characterized by the parameter ξ:

$$\xi\equiv\left|\frac{\Phi_T(l)}{\Phi_C(l)}\right|=\frac{16}{15}n_g\frac{R_d^4}{Z_d^2}(T_{ds}-T_g). \tag{11}$$

Ordered structures of macroparticles in a thermal plasma are investigated in Refs. 8–10. The thermophoretic interaction was not studied in the analysis of ongoing processes, since the temperature difference between the gas and dispersed phase was small and could be explained completely by the error in the temperature measurements ($\sim 2\%$). However, if the resulting temperature difference (about 50 K) is taken into account, then under the experimental conditions $\xi=0.9$.

In summary, since it is long-range, the thermophoretic force obviously strongly affects the formation of ordered structures. To check this, the temperature difference between the gas and the macroparticles must be measured to high accuracy.

3. MEASUREMENTS OF PARTICLE AND GAS TEMPERATURES

The experimental test stand for studying the thermal plasma included a plasma generator and diagnostic equipment for determining particle and gas parameters. The plasma source consisted of a burner, a gas-supply system, and a system for introducing particles. The particles were fed to the internal flame of the burner by a special-purpose fluidized bed cyclone device. This burner design made it possible to produce a laminar plasma jet with uniformly distributed parameters—the temperature and the electron and ion densities. Polydisperse CeO_2 particles were used in our experiments, with a distribution half-width, as our measurements showed, of at most 30%. The average Sauter diameter of the particles was about 0.8 μm. In the working regime, the velocity V_g of the plasma jet varied over the range 4–6 m/s, the electron and ion densities varied over the range $10^9-10^{12}\,\mathrm{cm}^{-3}$, and the gas was kept at atmospheric pressure. The electron and ion temperatures were the same, and varied over the range $T_i=T_e=T_g=1700-2200\,\mathrm{K}$. The test stand is described in greater detail in Refs. 10 and 17. Here we discuss in detail only the modified methods for measuring the gas and particle temperatures, and the diagnostic instrumentation based on them, which made it possible to determine the temperature difference to within 10%.

The conventional method for determining the temperature in a flame is a generalized reversal method, but the presence of dispersed-phase particles in gas flows can strongly influence the optical and radiative characteristics of the flows and, in consequence, substantially distort the value of the temperature when using conventional optical methods in a two-phase medium. In Ref. 17 it is shown that to high accuracy, the temperature of the gas phase in a dusty thermal plasma can be obtained by measuring three signals at two wavelengths: λ_0 (within the line profile) and λ (next to the spectral line)—the signal S_P from the plasma with macroparticles, the signal S_L from a standard lamp with temperature T_L, and the signal S_{LP} due to the lamp radiation emitted at temperature T and transmitted through the plasma:

$$T_g=\left(\frac{1}{T_L}-\frac{\lambda}{c_2}\ln\left(\frac{k_0(\tau_0F_0-\tau F)}{\tau_0-\tau}\right)\right)^{-1},$$

$$F=\frac{S_P}{S_L+S_P-S_{LP}}, \tag{12}$$

where τ is the optical density of a plasma with particles, and the subscript 0 corresponds to parameters determined within the spectral line profile (sodium atoms).

Color pyrometry is ordinarily used to determine the particle temperature.[18] Measurements of the spectral intensity ratios of the radiation of a dispersed medium at two or more wavelengths make it possible to determine the color temperature of the particles:

$$\frac{S_P(\lambda_1)}{S_P(\lambda_j)}=\frac{\varepsilon(\lambda_1)I^B(T_P,\lambda_1)}{\varepsilon(\lambda_j)I^B(T_P\lambda_j)}, \tag{13}$$

where $I^B(T,\lambda)$ is the Planck function and $\varepsilon(\lambda_j)$ is the emissivity of the dispersed phase. In the simplest case of ''gray'' particles, when the emissivity of the dispersed phase does not depend on wavelength ($\varepsilon(\lambda_j)=\varepsilon$), these temperatures are the same. Otherwise, to determine the true particle temperature it is necessary to know the form of the spectral dependence $\varepsilon(\lambda)$. In many cases of practical importance, when multiple scattering processes can be neglected, the directional emissivity of a dispersed medium is given by[19]

$$\varepsilon(\lambda)=(1-\omega(\lambda))(1-\exp\{-\tau(\lambda)\}), \tag{14}$$

where $\omega(\lambda)$ is the single scattering albedo. In this case, since the optical density can be calculated from the formula

$$\tau(\lambda)=-\ln(\{S_{LP}(\lambda)-S_P(\lambda)\}/S_L(\lambda)), \tag{15}$$

determination of the particle temperature can be reduced to choosing an accurate approximation for the spectral dependence of the single scattering albedo. In our experiments a function of the form

$$\omega(\lambda)=1-\mathrm{const}/\tau\lambda^{0.2}, \tag{16}$$

proposed in Ref. 20, was used to determine the particle temperature.

The optical layout of the diagnostic setup to determine the gas and particle temperatures is presented in Fig. 1. A standard VL (SI-10-300) tungsten lamp TL, whose radiation is interrupted by a chopper wheel CH and redirected by the beamsplitter BS, is used as the radiation source.

A characteristic feature of the apparatus for determining the gas temperature is the use of a spectral instrument SI of high resolving power (a DFS-452 spectrograph) and the use of a linear CCD array, each pixel of which corresponds to a spectral interval of 0.06 Å, as a detector. This makes it possible to record the line profile, rather than the integrated radiation of the gas in the atomic lines. The small spectral range covered by the L2 linear array does not permit investigating the particle characteristics (but it is possible to take account of the influence of the particles on the spectral lines). The advantage lies in the possibility of obtaining good resolution of lines (in our case, the sodium doublet) and therefore determining the temperature of the gas phase more accurately.

The optical layout of the method for determining the particle temperature of the dispersed phase (see Fig. 1) consists of a tungsten lamp TL, a chopper wheel CH, a beamsplitter BS, a mirror M, objectives O1 and O2, and a lightguide LG1. This scheme is similar to one used for diagnostics of the gas phase and differs from it by the angle of convergence of the probe beam and the use of a novel spectrophotometer, consisting of a slit S, a system of objectives O3 and O4, and a 300 line/mm diffraction grating DG, as the spectral instrument. A linear CCD array L1 was also used in this scheme. Such a system made it possible to per-

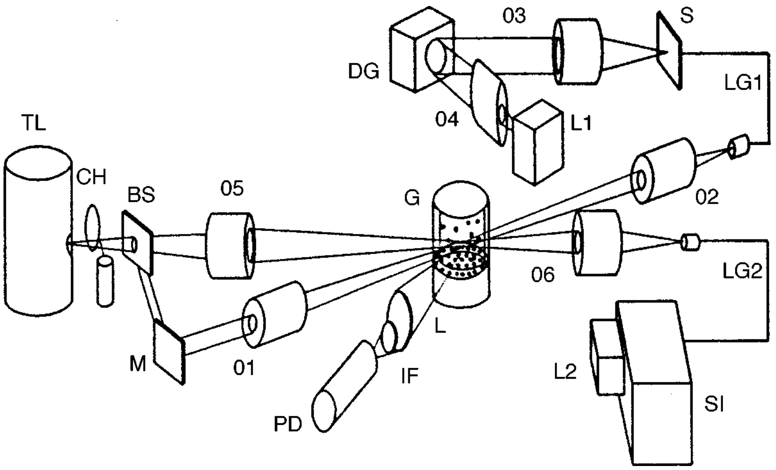

FIG. 1. Optical layout of the diagnostic system: TL—tungsten lamp, CH—chopper wheel, O1, O2, O3, O4, O5, O6—objectives, PD—photodetector, L—lens, L1, L2—linear CCD arrays, LG1, LG2—lightguides, DG—diffraction grating, S—slit, G—plasma generator, M—mirror, BS—beamsplitter, SI—spectral instrument, IF—interference filter.

form spectral analysis of the collected radiation in the wavelength range 450–900 nm, which is sufficient to determine parameters of macroparticles such as their size, absorption coefficient, temperature, and density.

The main channels are used for spectral measurements of the intensities of the radiation from the plasma layer S_P, the reference lamp S_L, and the lamp signal S_{LP} transmitted through the medium. A chopper wheel CH is introduced into the scheme to measure the signals S_P and S_{LP}. The measurements were performed in an automatic mode, the signals from the linear arrays were processed with a computer, and the signals from the photodetector PD were fed into an L-1250 ADC board and then processed with a computer.

A reference channel is introduced into the scheme to correct errors introduced into measurements of the optical density and intensity of the characteristic radiation of the layer of macroparticles as a result of fluctuations of the particle density. The reference channel consists of a lens L, an interference filter IF, and a photodetector PD (FÉU-84). Radiation from the measurement volume is directed onto the photodetector PD (signal S_{LP}). Measurement of the signals S_{LP}^0 and S_{LP} simultaneously with the signals S_P and S_L makes it possible to monitor the particle density during the experiment and either to reject unsuccessful measurements or to correct the signal from the plasma:

$$S_P^* = S_P S_{LP} / S_{LP}^0 .$$

Systematic errors in determining the emissivity from the relative signal measurements S_P/S_L were corrected by preliminary calibration of the instrument.

The observation of ordered structures with simultaneous diagnostics of the plasma was conducted in the zone of stabilization of the temperature of the plasma jet at height $h = 25 - 40\,\mathrm{mm}$ above the cover of the burner for various plasma temperatures and particle densities. The plasma temperature was varied by varying the propane/air ratio over the range 0.95–1.47. In measurements with CeO_2 particles the particle density n_p was varied over the range $(0.2 - 5.0) \times 10^7\,\mathrm{cm}^{-3}$, the plasma temperature T_g was varied over the range 1700–2200 K, the temperature difference ΔT between the gas and macroparticles was varied from 20 to 70 K, the ion density n_i was varied from $0.42 \times 10^{10}\,\mathrm{cm}^{-3}$ to $4.0 \times 10^{10}\,\mathrm{cm}^3$, and the electron density n_e was varied over the range $(2.5 - 7.2) \times 10^{10}\,\mathrm{cm}^{-3}$. On the basis of these data it was found from the quasineutrality relation $Z_p n_p + n_i = n_e$ that the CeO_2 particles are positively charged to about $10^3 e$, to within a factor of 2. The ordered structures were analyzed using a two-point correlation function $R(r)$. The resulting plasma parameters were used for numerical modeling.

4. NUMERICAL MODELING IN A THERMAL PLASMA

Computer simulation of the behavior of an ensemble of macroparticles in a thermal plasma was performed in order to investigate the influence of the thermophoretic interaction of the macroparticles on the formation of ordered structures in a thermal plasma. The simulation parameters corresponded to the experimental conditions of the plasma with CeO_2 particles, for which a liquid-type structure was observed. Numerical modeling was carried out using the Mathcad system by the molecular dynamics method in $2D$ geometry. It consisted of solving the equation of motion for each macroparticle assuming a pairwise interaction law. Frictional forces and Brownian motion were also taken into account:

$$m_d \frac{d^2 \mathbf{r}_k}{dt^2} = \sum_j \Phi(r)\big|_{r = |\mathbf{r}_k - \mathbf{r}_j|} \frac{\mathbf{r}_k - \mathbf{r}_j}{|\mathbf{r}_k - \mathbf{r}_j|} - m_d \nu_{fr} \frac{d\mathbf{r}_k}{dt} + \mathbf{F}_{br} , \tag{17}$$

where m_d is the particle mass, ν_{fr} is the friction decrement, and F_{br} is a random force giving rise to Brownian motion.

The number of particles was $N = 200$. A square with periodic boundary conditions was used as the simulation region. The significance of the periodic boundary conditions is that in the first place, when any macroparticle leaves the simulation region an identical particle enters the opposite side of the square and, in the second place, interactions with periodic images of the particles are taken into consideration. These images are constructed in a strip of width $L_0/2$ around the square, where L_0 is the edge length of the square. Such periodic boundary conditions make is possible to avoid boundary effects and set the average macroparticle density.

Initially, the particles occupy random locations inside the simulation region, after which formation of an ordered structure commences. Here a problem arises that is charac-

TABLE I. Parameters used in the simulation.

$T_g = 1700$ K	$\Delta T = 50$ K	$n_d = 5\times 10^7$ cm^{-3}
$P_g = 1$ bar	$m_d = 1.16\times 10^{-12}$ g	$\Gamma_S = 30$
$n_e = 7\times 10^{10}$ cm^{-3}	$\tilde{m}_d = 3.2\times 10^{-10}$ g	$\gamma_p = 150$
$n_i = 4\times 10^9$ cm^{-3}	$l = 17$ μm	$N = 200$
$r_D = 11$ μm	$Q = 500e$	$\tau = 0.3$ ms
$R_d = 0.4$ μm	$\nu_{fr} = 1.14\times 10^5$ s^{-1}	$\tilde{\nu}_{fr} = 570$ s^{-1}

teristic of $2D$ simulation of the behavior of a real system of macroparticles. The average interparticle distance l for a $2D$ system of macroparticles is equal to the value calculated from the density in a real experiment:

$$l = (4\pi n_d/3)^{-1/3}.$$

In this case it is to be expected that the simulation of the processes due to short-range electric forces will be correct. However, the effect of the long-range nonelectric interaction cannot be completely accurately taken into account in a $2D$ geometry. To simulate a structure of macroparticles with a long-range interaction correctly, simulation must be performed. Nonetheless, $2D$ simulation for long-range interaction is helpful to demonstrate the effect due to attraction between particles.

Two systems were simulated:

1. A system of macroparticles taking account of only the electric interaction.

2. A system that incorporated electrostatic repulsion and thermophoretic attraction. Its purpose was to demonstrate the influence of long-range attractive forces.

In practice, the actual interaction law must be cut off at very short and very long distances. It is obvious that initially some macroparticles can be very close to one another, and the initial interaction force will be very large, which will require a very short time step. To avoid this problem the interaction force was truncated at distances shorter than $l_0 = 0.4l$. This does not affect the process leading to ordering of the macroparticles, since neighboring particles, separated by a distance of less than l_0, disperse quite rapidly, after which there are no longer likely to be any particles separated by such a short distance.

On the other hand, the long-range interaction of particles can give rise to unrealistic effects induced by the periodic boundary conditions. For this reason, the interaction of macroparticles at distances greater than the width of the strip of the periodic samples (i.e., $L_0/2$), is cut off gradually. This does not affect the local processes of interest, which ensures the validity of the simulation. Thus, the following interaction law was used for numerical modeling instead of the actual one:

$$\tilde{\Phi}(l) = \begin{cases} \Phi(l_0), & l < l_0, \\ \Phi(l), & l_0 < l < 0.5L_0, \\ \Phi(l)\dfrac{L_0 - l}{0.1L_0}, & 0.5L_0 < l < L_0, \\ 0, & l > L_0. \end{cases} \tag{18}$$

The simulation parameters are given in Table I. Some remarks concerning this table are made below. For several parameters in the table two values are presented. The values marked with a tilde were used in the simulation.

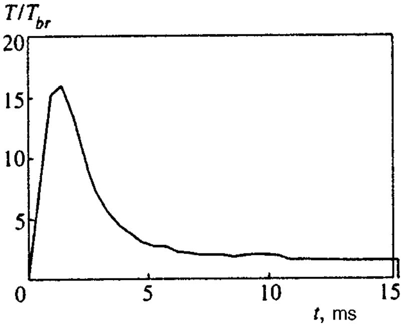

FIG. 2. Time dependence of the temperature of a system of macroparticles in a simulation with a purely electrostatic interaction.

The mass of a macroparticle is $m_d = (4\pi/3)\rho_d R_d^3$, where ρ_d [g/cm^3] is the density of CeO$_2$. We used the approximation

$$\nu_{fr} = \frac{6\pi\eta R_d}{m_d\{1 - (\lambda/R_d)(1.257 + 0.400\exp(-1.10R_d/\lambda))\}}, \tag{19}$$

where η is the viscosity of the gas. This complicated expression is necessary because the experimental conditions correspond to a transitional regime, where neither the Stokes formula nor the expressions for the collisionless regime are accurate.

The mean-free path and the viscosity were taken to be $\lambda = 1.56$ μm and $\eta = 1.73\times 10^{-3}$ g/cm·s. Substituting these values into the expression (19) yields $\nu_{fr} = 1.4\times 10^5$ s^{-1}. This is a very large value, considering that the characteristic formation time t_f of an ordered structure is tens of seconds. This means that the inertia of the particles has virtually no effect on their regular motion determined by the interaction forces. Unfortunately, it is impossible to eliminate the inertial term from Eq. (17), because Brownian motion is impossible without it.

The time step τ must be less than $1/\nu_{fr}$. To avoid numerical instability $\tau \approx 0.15/\nu_{fr}$ was used in practice. For this reason, simulation with the real value of m_d required too many time steps. However, a much larger particle mass can be used. Specifically, the mass $\tilde{m}_d = 200 m_d$ was used. The regular motion of macroparticles remains inertialess if

$$\tilde{\nu}_{fr} t_f \ll 1. \tag{20}$$

Here $\tilde{\nu}_{fr}$ was calculated from Eq. (19), substituting $\tilde{m}_d$ for m_d. Therefore the regular motion of macroparticles is insensitive to such an increase in their mass. However, this results in an initial increase in the temperature of the macroparticles (Fig. 2). This effect, which does not occur in reality, is due to the rapid liberation of heat at the initial stage of evolution of the structure from disorder to order. In reality, T_d remains constant, since friction is strong enough to dissipate the energy released and to prevent a temperature increase. Nevertheless, note that the growth time T_d is short compared with t_f, so this effect can be neglected.

As one can see from Table I, the ion density is low compared with the electron density. Note that the question of

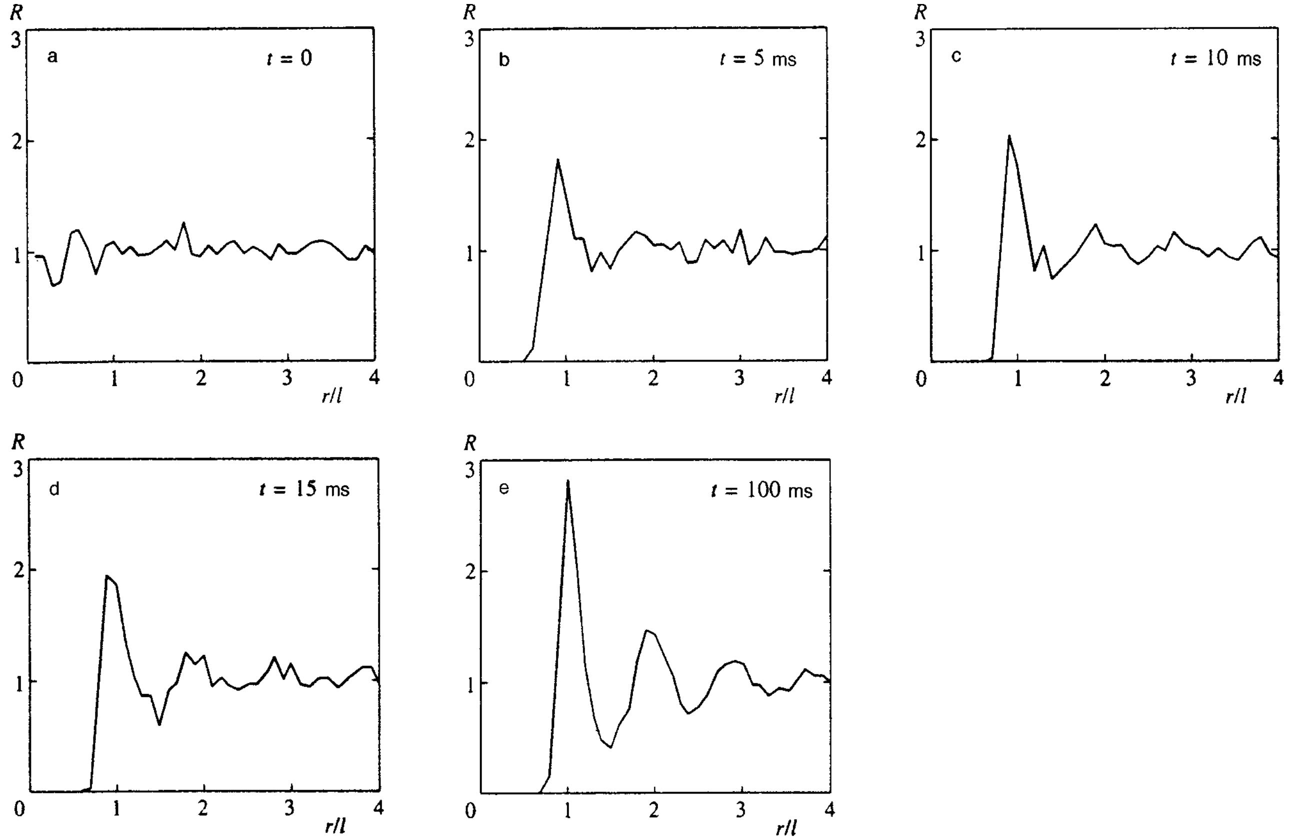

FIG. 3. Evolution of the correlation function in a simulation with a purely electrostatic interaction.

determining the Debye length in a plasma with such a high macroparticle density is unclear. This problem requires special consideration, which is beyond the scope of the present work. In the present work r_D was calculated neglecting the ion screening:

$$r_D=\sqrt{T_e/4\pi e^2 n_e}.$$

The first simulation with a purely electrostatic interaction of the macroparticles showed that after a relaxation process the system arrives in a final, statistically stationary state corresponding to a liquid structure. This agrees with the values of the nonideality parameters Γ_s and γ_p.

Figure 3 shows the dynamics of the correlation function $R(r)$, calculated according to the instantaneous positions of the macroparticles. The arrangement of the macroparticles in the simulation region at the end of the experiment is presented in Fig. 4. It is evident from Fig. 3 that the particles contributing to the correlation function at small r vanish. This process occurs very rapidly because of the strong electric repulsion forces at short distances. Then, the first pronounced peak appears at $t=10$ ms and grows. Subsequently, this peak continues to grow and higher-order peaks appear at the same time. The final correlation function is characterized by a large number of sharp oscillations. The formation time t_f of the structure is an ill-defined quantity, since it depends on the range of distances where the evolution of the correlation function is studied. The greater this distance, the greater the time required to establish the correlation function at this distance.

In practice, a range of distances where oscillations of the final correlation have an appreciable magnitude can be considered. We consider the first five peaks. In this case $t_f\approx 50$ ms. It is also convenient to introduce the time t_1 at which the first peak emerges. This is actually the time required for order of any type to appear in the system. The current simulation yields $t_1\approx 10$ ms, consistent with the analytic estimate

$$t_1\approx\frac{l m_d \nu_{fr}}{\Phi(l)}, \tag{21}$$

which yields $t_1=28$ ms. As noted above, in this simulation the transit time of macroparticles is $t_{\mathrm{exp}}=10$ ms. This means

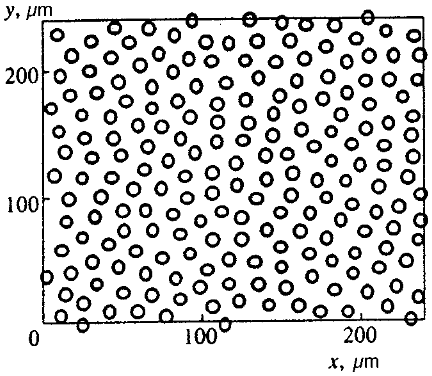

FIG. 4. Final form obtained for a system of macroparticles in a simulation with a purely electrostatic interaction, $t=100$ ms.

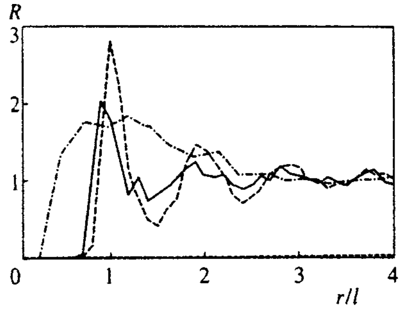

FIG. 5. Comparison of the correlation functions obtained in numerical simulation with purely electric interaction of macroparticles and measured experimentally: solid line—simulation, $t=10$ ms (corresponds to the transit time of the plasma through the experimental setup), dashed curve—simulation, final stationary state; dot-dashed curve—experiment.

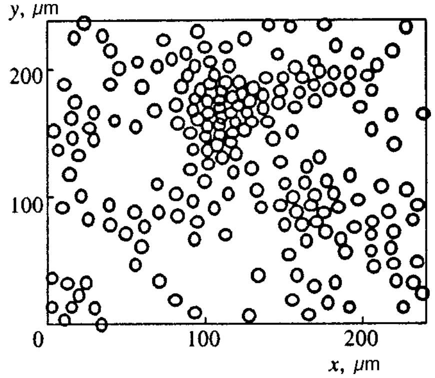

FIG. 7. System of macroparticles, numerical simulation taking account of the thermophoretic attraction, $t=21$ ms.

that the experimentally measured correlation function corresponds to the moment at which the structure is formed. The real transit time is sufficient for short-range order to appear, but insufficient for many oscillations to appear in the correlation function.

Figure 5 shows the experimental correlation function compared with two correlation functions obtained by numerical simulation, one of which corresponds to the real transit time, while the other corresponds to the final stage. Two characteristic features distinguish the experimental correlation function. First, it has but a single peak, and second, that peak is very wide. The lack of long-range order peaks is not surprising, since the foregoing tells us that the actual transit time is too brief for them to develop. The broadening of the main peak can probably be explained in terms of local irregularities of the macroparticle structure. One possible mechanism leading to these nonuniformities is thermophoretic attraction of particles.

This mechanism is studied in the second simulation, where long-range thermophoretic attraction of macroparticles is included in the interaction law. It turns out that the attraction of macroparticles leads to a nonperiodic Jeans-type instability. Figure 6 shows the evolution of the correlation function in the second numerical simulation. The final arrangement of macroparticles in the simulation square is shown in Fig. 7.

It can be seen from the last figure that the macroparticles tend to gather into a cloud. Thus, the attraction of macroparticles might account for the strong local irregularities observed in the laboratory experiment. Figure 8 shows the ex-

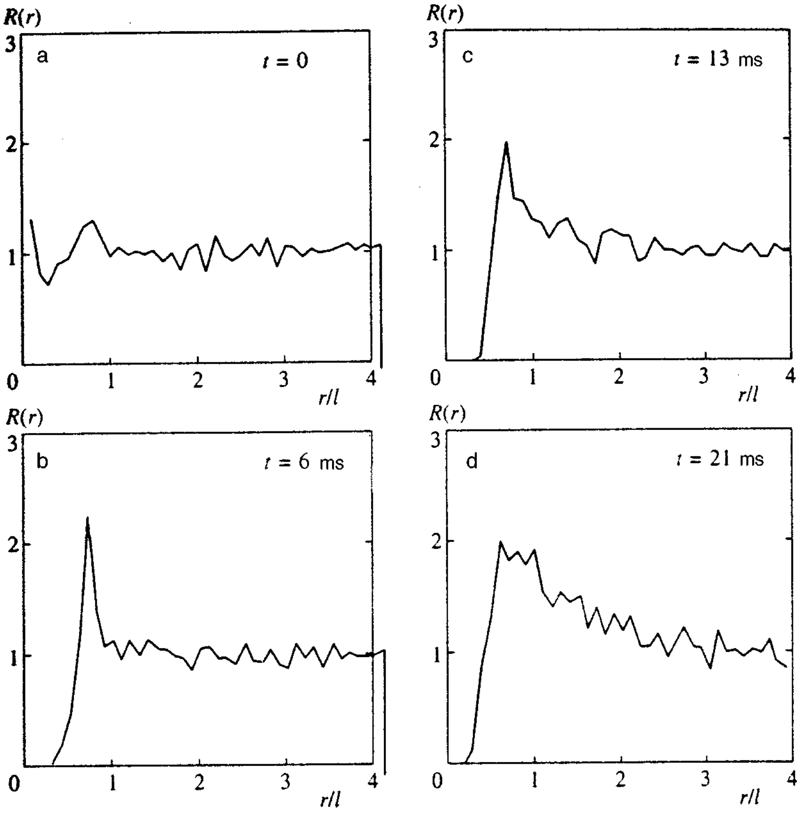

FIG. 6. Temporal evolution of the correlation function obtained by simulation, taking account of the thermophoretic attraction.

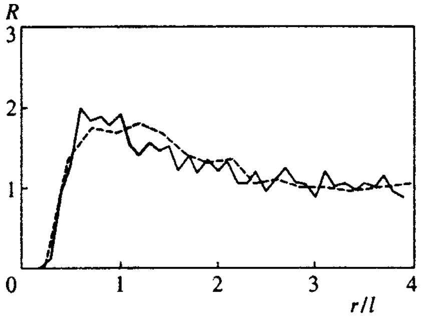

FIG. 8. Comparison of the correlation function obtained via numerical simulation, taking account of the thermophoretic interaction of macroparticles, and the experimentally measured correlation function: solid line—simulation, $t=16$ ms; dashed line—experiment.

perimental correlation function and the correlation function obtained by simulating the nonlinear stage of instability. These are clearly in good agreement.

5. CONCLUSIONS

In summary, in the present work the thermophoretic force and its effect on the formation of ordered structures were determined for various values of the plasma parameters. It was shown that the coexistence of regions with a chaotic arrangement of particles and regions of ordered structures (domains) is typical of a thermal plasma with strongly interacting macroparticles. Particles in the domains can be separated by less than the mean distance. A system of strongly interacting macroparticles was numerically simulated, taking account of thermophoretic interactions among the particles. The results showed that the thermophoretic attractive forces explain the form of the spatial nonuniformity associated with particle clustering into small domains (whose size is of the order of the mean separation). The experimentally measured correlation function was found to be very close to the correlation function obtained by numerical simulation.

[*)] E-mail: ipdustpl@redline.ru

---

[1] T. M. Sugden and B. A. Thrush, Nature (London) **168**, 703 (1951).
[2] K. E. Shuler and J. Weber, J. Chem. Phys. **22**, 491 (1954).
[3] V. E. Fortov and I. T. Yakubov, *Nonideal Plasma* [in Russian], Énergoatomizdat, Moscow (1994).
[4] S. Ichimaru, Rev. Mod. Phys. **54**, 1017 (1982).
[5] H. Ikezi, Phys. Fluids **29**, 1764 (1986).
[6] J. H. Chu and Lin I, Phys. Rev. Lett. **72**, 4009 (1994).
[7] H. Thomas, G. E. Morfill, V. Demmel *et al.*, Phys. Rev. Lett. **73**, 652 (1994).
[8] V. E. Fortov, A. P. Nefedov, O. F. Petrov *et al.*, JETP Lett. **63**, 181 (1996).
[9] V. E. Fortov, A. P. Nefedov, O. F. Petrov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 467 (1997) [JETP **84**, 256 (1997)].
[10] V. E. Fortov, A. P. Nefedov, O. F. Petrov *et al.*, Phys. Lett. A **219**, 89 (1996).
[11] A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).
[12] M. O. Robbins, K. Kremer, and G. S. Grest, J. Chem. Phys. **88**, 3286 (1988).
[13] M. J. Stevens and M. O. Robbins, J. Chem. Phys. **98**, 2319 (1992).
[14] R. T. Farouki and S. Hamaguchi, Appl. Phys. Lett. **61**, 2973 (1992).
[15] V. Tsytovich, Ya. Khodataev, and R. Bingham, Comments Plasma Phys. Control. Fusion **17**, 249 (1996).
[16] A. M. Ignatov, Fiz. Plazmy **22**, 648 (1996) [Plasma Phys. Rep. **22**, 585 (1996)].
[17] O. S. Vaulina, A. P. Nefedov, O. F. Petrov, A. A. Samaryan, and A. V. Chernyshev, Fiz. Plazmy **24** (1998) [Plasma Phys. Rep. **24** (1998)].
[18] D. Ya. Svet, *Optical Measurement of True Temperatures* [in Russian], Nauka, Moscow (1982).
[19] A. P. Nefedov, O. F. Petrov, and O. S. Vaulina, J. Quant. Spectrosc. Radiat. Transf. **54**, 453 (1995).
[20] O. S. Vaulina, A. P. Nefedov, O. F. Petrov, A. A. Samaryan, and A. V. Chernyshev, Teplofiz. Vys. Temp. **35** (6) (1997).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Convective heat conduction and diffusion in one-dimensional hydrodynamics

V. A. Gorodtsov*

Institute of Mechanics, Russian Academy of Sciences, 117526 Moscow, Russia

(Submitted 25 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1616–1629 (November 1999)

The impurity concentration in localized structures is described on the basis of analytic solutions of model equations for convective diffusion in the one-dimensional hydrodynamic approximation without pressure. The simplicity of the derivation of the analytic results depends on the ratio of the kinetic coefficients of the liquid (the Prandtl numbers). For the same kinetic coefficients, any time-dependent problem can be reduced to problems for the conventional heat conduction equation. For integer Prandtl numbers the problem of time-dependent convective diffusion in the flow field of a uniformly moving shock wave likewise reduces to problems for the heat conduction equation. Relations are established between problems whose Prandtl numbers differ by an integer. Various representations of the Green's functions for the equations of convective diffusion are analyzed. For integer Prandtl numbers they can be expressed in terms of error functions. The asymptotic character of the solutions depends strongly on the satisfaction of global conservation laws. For global conservation of the impurity mass, coalescence of shock waves corresponds to merging of impurity solitons, i.e., clustering. © *1999 American Institute of Physics.* [S1063-7761(99)00911-7]

1. INTRODUCTION

The dynamical and stochastic description of the evolution and interaction of many degrees of freedom, leading to the formation of structures on account of the characteristic features of the collective behavior, has been substantially developed on the basis of the theory of continuous media. The basic characteristic features, such as nonlinearity, viscous dissipation, and dispersion, on whose balance nonequilibrium structure formation is based, can be effectively taken into account in the hydrodynamic approximation. Even simplified one-dimensional hydrodynamic models reveal important features of phenomena and, above all else, the decisive role of localized formations in the form of shock waves and solitons. The one-dimensional model of a viscous liquid without pressure, known as the Burgers model,[1] has long attracted a great deal of attention for describing deterministic and stochastic flows in aerodynamics and plasma physics.[1–3] For all of the model simplification, it retains the inertial nonlinearity and high dissipation, which play a leading role in the formation of turbulent flow. The quasilinear Burgers differential equation is especially attractive because it can be reduced to a linear diffusion (heat conduction) equation by means of a nonlinear Cole–Hopf change of variables.[4,5] The new burst of interest in this equation and its multidimensional modifications is due to the further elaboration of methods for solving problems of turbulence theory[6–8] and to the substantial expansion of its physical applications in theories of transport flows,[3] structure formation in the evolution of the Universe,[9] growth of interfaces,[10,11] and convective diffusion.[12]

The one-dimensional hydrodynamic model without pressure remains simple even when it is extended to the problem of convective diffusion (or/and heat conduction), where molecular and convective transport in the carrying liquid are combined. The case of equal kinetic coefficients (unit Prandtl number) is especially simple. In this case there exists an extension of the Cole–Hopf transformation whereby the nonlinear equations of motion and convective diffusion can be reduced to a linear heat conduction equation.[13] The situation becomes somewhat more complicated when the kinetic coefficients are different, and a reasonably complete analytic investigation is possible only for particular types of flows. For flow in the form of a uniformly propagating shock wave the equations of time-dependent convective diffusion can once again be reduced to a linear heat conduction equation with constant coefficients for any integer Prandtl numbers (in the general case, the relation between the solutions of problems whose Prandtl numbers differ by an integer will be established). The asymptotic character of the impurity distribution in a shock wave depends on whether or not a global conservation law for the impurity is satisfied for the equations of the model. The model without complete conservation of the impurity will be studied first in Sec. 2. Then, based on this model, a model with a global conservation law will be studied completely analogously in Sec. 4. In the latter case, an ''impurity soliton'' will ultimately (in the limit of long times) accompany the shock wave. The well-known effect of perfectly inelastic collisions of one-dimensional shock waves for Burgers' equation[2,3] will thereby correspond to merging of impurity solitons and an increasing local impurity concentration. This completes the one-dimensional description of impurity clustering, which has attracted a great deal of attention in the literature in the more complicated multidimensional and stochastic situations.[9,14–16] In Sec. 3 the Green's function of the convective heat conduction equation is analyzed on the basis of the solution of the well-known spectral problem of quantum mechanics for the

Pöschl–Teller potential.[17,18] It is demonstrated that the simplication of the analysis for integer Prandtl numbers is due to the nonreflective character of the potential in this case.

2. DIFFUSION EQUATIONS FOR A PASSIVE IMPURITY

The equation for convective heat conduction (diffusion) through a velocity field is coupled with the equation of motion, which in one-dimensional hydrodynamics without pressure is Burgers' equation. In the passive-impurity approximation (in this approximation there is no sense of distinguishing between heat conduction and diffusion, so that either term will be used), which does not exert a back effect on the fluid flow, one variant of such coupling (another is studied below in Sec. 4) is represented by the differential equations

$$\partial_t u + u\partial_x u = \nu\partial_x^2 u, \quad \partial_t\theta + u\partial_x\theta = \chi\partial_x^2\theta, \tag{2.1}$$

for which the solution of the initial problem presupposes the initial conditions

$$u(x,t)|_{t=0} = \partial_x\phi_0(x), \quad \theta(x,t)|_{t=0} = \theta_0(x) \tag{2.2}$$

for the velocity and density fields of the passive impurity.

In Ref. 13 it was shown that this system of equations with identical kinetic coefficients (unit Prandtl number $P\equiv\nu/\chi=1$) is just as simple as a separate Burgers equation. Using the generalized Cole–Hopf change of variables,

$$u(x,t) = -2\nu\partial_x \ln\varepsilon(x,t), \quad \theta(x,t) = \psi(x,t)/\varepsilon(x,t), \tag{2.3}$$

it reduces to two ordinary linear heat conduction equations

$$\partial_t\varepsilon = \nu\partial_x^2\varepsilon, \quad \partial_t\psi = \chi\partial_x^2\psi \quad (\nu=\chi), \tag{2.4}$$

$$\varepsilon(x,t)|_{t=0} = \exp\left(-\frac{\phi_0(x)}{2\nu}\right),$$

$$\psi(x,t)|_{t=0} = \theta_0(x)\exp\left(-\frac{\phi_0(x)}{2\nu}\right).$$

We now analyze the equations of the one-dimensional model (2.1) for arbitrary (always positive in the sense of thermodynamic constraints) Prandtl numbers. Without distinguishing heat conduction and diffusion, we use a single term for the dimensionless ratio of the coefficient of kinematic viscosity to the other kinetic coefficients (in the literature, a special term—the Schmidt number—is used for the ratio of the coefficients of viscosity and diffusion). It is convenient to use dimensionless variables constructed using the characteristic velocity u_0 and the coefficient of viscosity. Then the system of equations assumes the form

$$\partial_\tau v + 2Pv\partial_\xi v = P\partial_\xi^2 v, \quad \partial_\tau\theta + 2Pv\partial_\xi\theta = \partial_\xi^2\theta, \tag{2.5}$$

$$\xi\equiv\frac{u_0 x}{2\nu}, \quad \tau P\equiv\frac{u_0^2 t}{4\nu}, \quad v\equiv\frac{u}{u_0} = v(\xi,\tau).$$

Since the second equation is linear in the impurity concentration, there is no special need to reduce the latter to dimensionless form.

In the general case this system of equations cannot be transformed to simple separate equations of the type (2.4), as is obvious from the fact that the Painlevé test for Eqs. (2.1) is satisfied only for unit Prandtl number.[13] Nonetheless, this is possible in special cases.

The generalized Cole–Hopf transformation

$$v = -\partial_\xi \ln\varepsilon, \quad \theta = \psi/\varepsilon^P \tag{2.6}$$

once again reduces the nonlinear equation of motion to a linear heat conduction equation with constant coefficients

$$\partial_\tau\varepsilon = P\partial_\xi^2\varepsilon,$$

and the equation for the impurity concentration reduces to a linear equation with a variable coefficient and without the first spatial derivative of the unknown ψ:

$$\hat{L}_{P-1}\psi = 0, \quad \hat{L}_P \equiv \partial_\tau - \partial_\xi^2 + P(P+1)v_\xi. \tag{2.7}$$

The differential operator arising here satisfies the important commutation relation

$$\hat{L}_{P-m}\hat{M}_{P-m} - \hat{M}_{P-m}\hat{L}_{P-1-m} = \frac{m(P-m)}{P}v_\tau,$$

$$\hat{M}_P \equiv \partial_\xi + Pv. \tag{2.8}$$

Besides the overt P-dependence, here there is also an implicit dependence (through the form of the dimensional velocity). Stationary flow is an exception, as is obvious from the equation of motion in Eq. (2.5). For stationary flow the commutation relation becomes homogeneous and makes it possible to use the lowering operator $\hat{M}_P$ to reduce the solution of the problem under study for some Prandtl number to the solution of a similar problem with a Prandtl number smaller by 1. Thus, in the modern terminology, this operator is a Darboux operator for the differential operator of the problem. Because of the symmetry property of the latter, $\hat{L}_{-P} = \hat{L}_{P-1}$, the raising operator will be $\hat{M}_{-P}$. Any change in Prandtl number by an integer can be achieved by repeated application of the lowering or raising operators. Specifically, for integer Prandtl numbers the problem is reducible to the simple problem mentioned above for the heat conduction equation (the problem with $P=1$). For stationary flow and integer P, a relation between the differential operators of these problems can be easily established using (2.8):

$$\hat{L}_{P-1}\hat{M}_{P-1}\hat{M}_{P-2}\dots\hat{M}_1 = \hat{M}_{P-1}\hat{M}_{P-2}\dots\hat{M}_1\hat{L}_0,$$

$$\hat{L}_0 = \partial_\tau - \partial_\xi^2. \tag{2.9}$$

Starting with the heat conduction equation, left-multiplying it by the operator $\hat{M}_1$ and using the commutation relation (2.8) with $v_\tau = 0$, we obtain

$$\hat{L}_1\psi^{(2)} = 0, \quad \hat{L}_0\varphi = 0; \quad \psi^{(2)} = \hat{M}_1\varphi. \tag{2.10}$$

Continuing in this manner we prove that the solution of an equation with an integer Prandtl number can be represented in terms of the solution of the heat conduction equation as

$$\hat{L}_{P-1}\psi^{(P)} = 0, \quad \hat{L}_0\varphi = 0; \quad \psi^{(P)} = \hat{M}_{P-1}\hat{M}_{P-2}\dots\hat{M}_1\varphi. \tag{2.11}$$

The explicit form of the required one-dimensional stationary flow can be found quite easily. The stationary solution, bounded at infinity, of the equation of motion (2.1) has the form

$$u(x)=-u_0\tanh\frac{u_0x}{2\nu};\qquad v=-\tanh\xi. \tag{2.12}$$

Since the equations exhibit Galilean invariance, such a flow corresponds to a uniformly moving shock wave with a viscous transitional layer in a coordinate system moving together with the wave (in what follows, all calculations are performed in such a coordinate system). We note that the potential ε corresponding to the indicated stationary velocity distribution in the Cole–Hopf transformation is not stationary:

$$\varepsilon(\xi,\tau)=C_0e^{P\tau}\cosh\xi, \tag{2.13}$$

since it must satisfy the time-dependent heat conduction equation (2.4). However, under the Cole–Hopf transformation, separating the time and space variables causes the time (together with the arbitrary constant C_0) to drop out of the result for the velocity (2.12).

Here it is still necessary to complete the proof that any solution of the equation under study for ψ with integer Prandtl numbers can be represented in the form (2.11). For this we consider a solution in the form $\psi^{(p)}=\hat{M}_{p-1}\chi$ and use for the transformations the commutation relation (2.8) with $v_\tau=0$:

$$\hat{L}_{p-1}\psi^{(p)}=0;\qquad \hat{L}_{p-1}\hat{M}_{p-1}\chi=\hat{M}_{p-1}\hat{L}_{p-2}\chi=0;$$

$$(\partial_\xi-(P-1)\tanh\xi)\hat{L}_{p-2}\chi=0.$$

Integrating the latter equation once we arrive at an inhomogeneous equation whose solution can be conveniently represented as a sum of the particular solution and the general solution of the homogeneous equation:

$$\hat{L}_{p-2}\chi=C\cosh^{P-1}\xi;$$

$$\chi=-\frac{C}{(P-1)^2}\cosh^{P-1}\xi+\psi^{(p-1)}.$$

Applying the operator $\hat{M}_{p-1}$ to this sum, the contribution from the first term vanishes, and we obtain finally a formula expressing the relation between the solutions of two problems with Prandtl numbers differing by 1:

$$\psi^{(p)}=\hat{M}_{p-1}\psi^{(p-1)}. \tag{2.14}$$

Applying this formula repeatedly we once again arrive at the result (2.11), where $\varphi\equiv\psi^{(1)}$.

Next, to study time-dependent convective diffusion of a passive impurity in the field of a uniformly moving shock wave, we focus our attention on the initial problem with the impurity concentrated at a point (the solution of other initial problems can be found by superposing such singular solutions):

$$(\partial_\tau-2P\tanh\xi\,\partial_\xi-\partial_\xi^2)\theta=0,\qquad \theta(\xi,\tau)|_{\tau=0}=\delta(\xi-\xi_0). \tag{2.15}$$

The impurity distribution at subsequent times with integer Prandtl numbers will be expressed, in accordance with Eqs. (2.6), (2.11), and (2.13), as follows (in the product, the operators are assumed to be ordered so that the number increases from right to left)

$$\begin{aligned}\theta(\xi,\tau)&=(C_0e^{P\tau}\cosh\xi)^{-P}\prod_{m=1}^{P-1}(\partial_\xi-m\tanh\xi)\varphi\\&=(C_0e^{P\tau})^{-P}\left(\frac{\partial}{\partial\sinh\xi}\right)^{P-1}\frac{\varphi}{\cosh\xi},\end{aligned} \tag{2.16}$$

in terms of the solution of the initial problem for the heat conduction equation of the form (the Heaviside step function enters into the initial condition)

$$\begin{aligned}&(\partial_\tau-\partial_\xi^2)\varphi=0,\\&\varphi|_{\tau=0}=C_0^P\cosh\xi_0\\&\qquad\times\cosh\xi\frac{(\sinh\xi-\sinh\xi_0)^{P-2}}{(P-2)!}H(\xi-\xi_0).\end{aligned} \tag{2.17}$$

A representation of the impurity distribution in terms of a convolution integral between the Green's function and the initial distribution then follows for the heat conduction equation:

$$\begin{aligned}&\theta(\xi,\tau)=e^{-P^2\tau}\cosh\xi_0\,\mathrm{sech}^P\,\xi\prod_{m=1}^{P-1}(\partial_\xi-m\tanh\xi)I_{p-2},\\&I_n\equiv\int_{\xi_0}^{\infty}d\eta\cosh\eta\frac{(\sinh\eta-\sinh\xi_0)^n}{n!}D_0(\xi-\eta,\tau),\\&D_0(\xi,\tau)\equiv\frac{1}{2\sqrt{\pi\tau}}\exp\left(-\frac{\xi^2}{4\tau}\right).\end{aligned} \tag{2.18}$$

These integrals can be expressed in terms of the error function $\operatorname{erf}z=1-\operatorname{erfc}z$. For example,

$$\begin{aligned}&I_0=\frac{e^\tau}{4}(2\cosh\xi+S_1),\\&S_1\equiv\sum_\pm(\pm e^{\pm\xi})\operatorname{erf}\left(\sqrt{\tau}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right),\\&I_1+I_0\sinh\xi_0=\frac{e^{4\tau}}{8}(2\sinh2\xi+S_2),\\&S_2=\sum_\pm e^{\pm2\xi}\operatorname{erf}\left(2\sqrt{\tau}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right),\end{aligned} \tag{2.19}$$

and we have for the concentration distributions of a passive impurity in the cases $P=2$ and $P=3$, respectively,

$$\begin{aligned}&\theta(\xi,\tau)=e^{-3\tau}\frac{\cosh\xi_0}{4}\mathrm{sech}^2\,\xi(\partial_\xi-\tanh\xi)S_1,\\&\theta(\xi,\tau)=e^{-5\tau}\frac{\cosh\xi_0}{8}\mathrm{sech}^3\,\xi(\partial_\xi-2\tanh\xi)(\partial_\xi-\tanh\xi)\\&\qquad\times(S_2-2e^{-3\tau}S_1\sinh\xi_0).\end{aligned} \tag{2.20}$$

Since the expressions for S_n are asymptotically constant for fixed ξ and ξ_0 at long times $(\tau\rightarrow\infty)$, concentration perturbations will vanish in this limit. Therefore, in the model at hand the convective transport of an impurity as a result of fluid flow in a shock wave cannot compensate for diffusion spreading.

For what follows, it is also convenient to rewrite the last result (we confine ourselves to the case $P=2$) in the form

$$\theta(\xi,\tau)=e^{-4\tau}\left(\frac{\cosh\xi_0}{\cosh\xi}\right)^2 D(\xi,\xi_0,\tau;P=2),$$

$$\hat{L}_1 D=0,\quad D|_{\tau=0}=\delta(\xi-\xi_0),$$

$$D(\xi,\xi_0,\tau;P=2)=\frac{e^{\tau}}{4}\operatorname{sech}\xi_0(\partial_\xi-\tanh\xi)S_1. \tag{2.21}$$

3. GREEN'S FUNCTION OF THE EQUATION OF CONVECTIVE HEAT CONDUCTION (SPECTRAL EXPANSION)

The equation of convective heat conduction in the presence of a flow produced by a uniformly moving shock wave can be solved for any Prandtl number using expansions in terms of the eigenfunctions of the stationary part of its operator. The latter problem has been studied in detail in quantum mechanics and is known as the spectral problem for the modified Pöschl–Teller potential. The simplicity of the case of integer Prandtl numbers corresponds to the case of nonreflective potentials and is due to the additional internal symmetry of the problem.

Seeking the solution of Eq. (2.7) (in a coordinate system comoving with the wave), rewritten for the specific case of a shock wave, in the form

$$\hat{L}_{p-1}\psi(\xi,\tau)\equiv(\partial_\tau-\partial_\xi^2-P(P-1)\operatorname{sech}^2\xi)\psi=0, \tag{3.1}$$

using the method of separation of variables (Fourier's method)

$$\psi(\xi,\tau)=a(\tau)\Psi(\xi);\quad a(\tau)=e^{-\lambda\tau},$$

$$(\partial_\xi^2+P(P-1)\operatorname{sech}^2\xi)\Psi=-\lambda\Psi, \tag{3.2}$$

depending on the sign of the separation constant a discrete (for $\lambda=-\mu^2<0$) and a continuous (for $\lambda=k^2>0$) spectrum arise. The complete solution can be represented as a linear combination of eigenfunctions of the discrete spectrum and an integral of the eigenfunctions of the continuous spectrum,

$$\psi(\xi,\tau)=\sum_{n=1}^{N}a_n e^{-\lambda_n\tau}\Psi_n(\xi)+\int_{-\infty}^{\infty}\frac{dk}{2\pi}b(k)\Psi_+(\xi,k)e^{-k^2\tau}, \tag{3.3}$$

with the exception of the situation with $0<P<1$, where a contribution only of the continuous spectrum occurs (in quantum mechanics this corresponds to the passage of a particle above the potential barrier $U(\xi)=P(1-P)\operatorname{sech}^2\xi>0$).

For the Green's function of the operator under study (we call the D function solving the initial singular problem a Green's function also)

$$\hat{L}_{p-1}G=\delta(\xi-\xi_0)\delta(\tau);$$

$$G(\xi,\xi_0,\tau)=H(\tau)D(\xi,\xi_0,\tau),$$

$$\hat{L}_{p-1}D=0,\quad D|_{\tau=0}=\delta(\xi-\xi_0) \tag{3.4}$$

such an expansion has the form

$$D(\xi,\xi_0,\tau)=\sum_{n=1}^{N}e^{-\lambda_n\tau}\Psi_n(\xi)\Psi_n(\xi_0)+\int_{-\infty}^{\infty}\frac{dk}{2\pi}\Psi_+(\xi,k)\Psi_+^*(\xi_0,k)e^{-k^2\tau}. \tag{3.5}$$

For known eigenfunctions and eigenvalues this yields an alternative method for solving the problem of the evolution of an impurity distribution, initially concentrated at a point, in the fluid flow created by a uniformly moving shock wave. The spectral data for the Schrödinger equation with a modified Pöschl–Teller well are well known.[17,18] The discrete spectrum for such a potential is finite:

$$\lambda_n=-(P-n)^2<0,\quad n=1,2,\ldots,N<P, \tag{3.6}$$

and the eigenfunctions of the discrete part of the spectrum have a polynomial form:

$$\begin{aligned}\Psi_n(\xi)/A_n&=(1-\sigma^2)^{(P-n)/2}F(-n+1,-n+2P;1+P-n;(1-\sigma)/2)\\&=(1-\sigma^2)^{(P-n)/2}\frac{\Gamma(n)\Gamma(1+P-n)}{\Gamma(P)}P_{n-1}^{(P-n,P-n)}(\sigma)\\&=(1-\sigma^2)^{(P-n)/2}\frac{\Gamma(n)\Gamma(1+2P-2n)}{\Gamma(2P-n)}C_{n-1}^{P-n+1/2}(\sigma)\\&=2^{P-n}\Gamma(1+P-n)P_{P-1}^{-(P-n)}(\sigma).\end{aligned} \tag{3.7}$$

Here equivalent expressions are presented for the eigenfunctions in terms of finite hypergeometric series, Jacobi polynomials, ultraspherical (Gegenbauer) polynomials, and associated Legendre functions.

The orthonormality of the eigenfunctions of the discrete spectrum of the Schrödinger equation,

$$\int_{-\infty}^{\infty}d\xi\Psi_n(\xi)\Psi_m(\xi)=\delta_{nm},$$

makes it possible to normalize the amplitude:

$$A_n=2^{n-P}\sqrt{\frac{(P-n)\Gamma(2P-n)}{\Gamma(n)\Gamma^2(1+P-n)}}. \tag{3.8}$$

Likewise, for the eigenfunctions of the continuous spectrum the Jost functions singled out in accordance with the asymptotic behavior $\Psi_+(\xi,k)\rightarrow\exp(ik\xi)$ as $\xi\rightarrow\infty$, we have a representation in terms of the hypergeometric function

$$\Psi_+(\xi,k)=e^{ik\xi}F\left(-P+1,P;1-ik;\frac{1-\sigma}{2}\right),$$

where the latter reduces to a Jacobi polynomial for integer Prandtl numbers (the hypergeometric series cuts off for nonpositive integer values of the first argument):

$$\Psi_+(\xi,k)=e^{ik\xi}\frac{\Gamma(P)\Gamma(1-ik)}{\Gamma(P-ik)}P_{P-1}^{(-ik,ik)}(\sigma).$$

Using the well-known formulas for transforming hypergeometric functions, the Jost function under study can be rewritten in a form convenient for estimating its asymptotic behavior at the other boundary $\xi\to-\infty$:

$$\Psi_+(\xi,k)=c_{21}(k)e^{ik\xi}F\left(-P+1,P;1+ik;\frac{1+\sigma}{2}\right)$$
$$+c_{22}(k)e^{-ik\xi}F\left(-P+1,P;1-ik;\frac{1+\sigma}{2}\right)$$
$$\xrightarrow[\xi\to-\infty]{}c_{21}(k)e^{ik\xi}+c_{22}(k)e^{-ik\xi},$$

$$c_{21}(k)=\frac{\Gamma(1-ik)\Gamma(-ik)}{\Gamma(P-ik)\Gamma(1-P-ik)},$$

$$c_{22}(k)=i\frac{\sin\pi(P-1)}{\sinh\pi k}.$$

Hence, for integer Prandtl numbers, the reflected wave $\sim\exp(-ik\xi)$ clearly vanishes:

$$c_{22}(k)=0,\quad c_{21}(k)=\prod_{n=1}^{P-1}\frac{ik+n}{ik-n},$$

i.e., the potential $U(\xi)=-P(P-1)\operatorname{sech}^2\xi<0$ in this situation is ''nonreflective'' in the language of quantum mechanical scattering theory.

For the smallest Prandtl numbers, $0<P<1$, the spectral problem, which has only a continuous spectrum, reduces to the problem of scattering above a positive barrier and the expansion of the Green's function with respect to this spectrum assumes the form

$$D=\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-k^2\tau}e^{ik(\xi-\xi_0)}F\left(1-P,P;1\right.$$
$$\left.-ik;\frac{1-\sigma}{2}\right)F^*\left(1-P,P;1-ik;\frac{1-\sigma_0}{2}\right),$$

$$\sigma\equiv\tanh\xi,\quad\sigma_0\equiv\tanh\xi_0.$$

In the limiting cases $P=0$ and $P=1$ the hypergeometric functions here reduce to 1, and the integral reduces to the well-known expression for the Green's function of the heat conduction equation

$$D_0(\xi-\xi_0,\tau)=\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-k^2\tau}e^{ik(\xi-\xi_0)}$$
$$=\frac{1}{2\sqrt{\pi\tau}}\exp\left(-\frac{(\xi-\xi_0)^2}{4\tau}\right).$$

In the next range, $2\geqslant P>1$, a single discrete eigenvalue with $n=1$ appears and the corresponding eigenfunction assumes the form (especially simple for $P=2$)

$$n=1;\quad\lambda_1=-(P-1)^2,$$

$$\Psi_1(\xi)=\frac{\sqrt{2\Gamma(2P-2)}}{2^{P-1}\Gamma(P-1)}\operatorname{sech}^{P-1}\xi\xrightarrow[P\to2]{}\frac{1}{\sqrt{2}}\operatorname{sech}\xi,\tag{3.9}$$

and the eigenfunction of the continuous spectrum (the Jost function with the asymptotic form $\exp(ik\xi)$ as $\xi\to\infty$) will be

$$\Psi_+(\xi,k)=e^{ik\xi}F\left(-P+1,P;1\right.$$
$$\left.-ik;\frac{1-\sigma}{2}\right)\xrightarrow[P\to2]{}e^{ik\xi}\left(1+\frac{1-\sigma}{ik-1}\right).\tag{3.10}$$

The eigenfunction expansion of the Green's function will be (for what follows, it is useful to indicate the parametric Prandtl number dependence explicitly in the D function)

$$D(\xi,\xi_0,\tau;P)=e^{(P-1)^2\tau}\frac{\Gamma(P-1/2)}{\sqrt{\pi}\Gamma(P-1)}$$
$$\times\operatorname{sech}^{P-1}\xi\operatorname{sech}^{P-1}\xi_0$$
$$+\int_{-\infty}^{\infty}\frac{dk}{2\pi}\Psi_+(\xi,k)\Psi_+^*(\xi_0,k)e^{-k^2\tau}.$$

In the limit $P\to2$ the integral of the eigenfunctions of the continuous spectrum can be expressed in terms of error functions. Combining them with the function of the discrete spectrum we obtain

$$D(\xi,\xi_0,\tau;P=2)=D_0(\xi-\xi_0,\tau)+\frac{e^\tau}{4}\operatorname{sech}\xi\operatorname{sech}\xi_0$$
$$\times\sum_{\pm}\operatorname{erf}\left(\sqrt{\pi}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right).$$

The latter expression, pertaining to the case of an integer Prandtl number, corresponds to the expression obtained in the preceding section by the ''algebraic'' method. This can be shown by performing the differentiation and simple transformations in Eq. (2.21).

For Prandtl numbers in the range $3\geqslant P>2$ there are two discrete levels

$$n=1:\quad\lambda_1=-(P-1)^2,$$

$$\Psi_1(\xi)=\sqrt{\frac{\Gamma(P-1/2)}{\sqrt{\pi}\Gamma(P-1)}}\operatorname{sech}^{P-1}\xi,$$

$$n=2:\quad\lambda_1=-(P-2)^2,$$

$$\Psi_2(\xi)=\sqrt{\frac{2\Gamma(P-1/2)}{\sqrt{\pi}\Gamma(P-2)}}\tanh\xi\operatorname{sech}^{P-2}\xi,$$

and in the limit $P\to3$ the normalized eigenfunctions of the discrete spectrum assume the simple form

$$\Psi_1(\xi)=\frac{\sqrt{3}}{2}\operatorname{sech}^2\xi,\quad\Psi_2(\xi)=\sqrt{\frac{3}{2}}\tanh\xi\operatorname{sech}\xi.$$

For a function of the discrete spectrum the representation

$$\Psi_{+}(\xi,k)=e^{ik\xi}F\left(-P+1,P;1-ik;\frac{1-\sigma}{2}\right)$$

$$\xrightarrow[P\to 3]{} e^{ik\xi}\left(1+\frac{\sigma(1-\sigma)}{ik-1}+2\,\frac{(1-\sigma)^2}{ik-2}\right).$$

is valid in this range of Prandtl numbers. Now the eigenfunction expansion of the Green's function is

$$D(\xi,\xi_0,\tau;P)=e^{(P-1)^2\tau}\frac{\Gamma(P-1/2)}{\sqrt{\pi}\Gamma(P-1)}\operatorname{sech}^{P-1}\xi\operatorname{sech}^{P-1}\xi_0$$

$$+e^{(P-2)^2\tau}\frac{2\Gamma(P-1/2)}{\sqrt{\pi}\Gamma(P-2)}\,\frac{\tanh\xi\tanh\xi_0}{(\cosh\xi\cosh\xi_0)^{P-2}}$$

$$+\int_{-\infty}^{\infty}\frac{dk}{2\pi}\Psi_{+}(\xi,k)\Psi_{+}^{*}(\xi_0,k)e^{-k^2\tau},$$

and in the limit $P\to 3$ it can be expressed in terms of error functions

$$D(\xi,\xi_0,\tau;P=3)=D_0(\xi-\xi_0,\tau)+\frac{3e^{4\tau}}{8}\operatorname{sech}^2\xi\operatorname{sech}^2\xi_0$$

$$\times\sum_{\pm}\operatorname{erf}\left(2\sqrt{\tau}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right)$$

$$+\frac{3e^{\tau}}{4}\operatorname{sech}\xi\operatorname{sech}\xi_0\tanh\xi\tanh\xi_0$$

$$\times\sum_{\pm}\operatorname{erf}\left(\sqrt{\tau}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right).$$

A generalization of the three expressions presented for the Green's function for $P=1$, 2, and 3 will be a formula that is valid for arbitrary integer Prandtl numbers (the complete proof is given in Ref. 19 for imaginary time, i.e., for the propagator of the time-dependent Schrödinger equation)

$$D(\xi,\xi_0,\tau;P)=D_0(\xi-\xi_0,\tau)+\sum_{n=1}^{P-1}e^{(P-n)^2\tau}\frac{\Psi_n(\xi)\Psi_n(\xi_0)}{2}$$

$$\times\sum_{\pm}\operatorname{erf}\left((P-n)\sqrt{\tau}\pm\frac{\xi-\xi_0}{2\sqrt{\tau}}\right).$$

Since for integral Prandtl numbers the eigenfunctions of the continuous spectrum are proportional to the sum of simpler pole contributions

$$\Psi_{+}(\xi,k)=e^{ik\xi}\left(1+\sum_{n=1}^{P-1}\frac{a_n^{(p)}(\sigma)}{ik-n}\right),$$

the integrals of their products in Eq. (3.5) reduce to sums of residues proportional to error functions.

The generalized Cole–Hopf change of variables (2.6) establishes between the Green's function of the operator $\hat{L}_{p-1}$ and the Green's function of the operator of the initial convective heat conduction equation (2.5) in the presence of a flow produced by a shock wave

$$\hat{K}_p\equiv\partial_\tau-2P\tanh\xi\partial_\xi-\partial_\xi^2,\quad \hat{K}_pG^{(\theta)}=\delta(\xi-\xi_0)\,\delta(\tau),$$

the simple relation

$$G^{(\theta)}=e^{-P^2\tau}\left(\frac{\cosh\xi_0}{\cosh\xi}\right)^{P}D(\xi,\xi_0,\tau;P)H(\tau).$$

Using the expression (2.13), the Green's function of the singular initial heat conduction problem can be written in the form

$$\theta(\xi,\tau)=e^{-P^2\tau}\left(\frac{\cosh\xi_0}{\cosh\xi}\right)^{P}\left\{D_0(\xi-\xi_0,\tau)+\frac{1}{2}\right.$$

$$\times\sum_{m=1}^{P-1}e^{m^2\tau}\Psi_{P-m}(\xi)\Psi_{P-m}(\xi_0)$$

$$\times\left[\operatorname{erf}\left(m\sqrt{\tau}+\frac{\xi-\xi_0}{2\sqrt{\tau}}\right)\right.$$

$$\left.\left.+\operatorname{erf}\left(m\sqrt{\tau}-\frac{\xi-\xi_0}{2\sqrt{\tau}}\right)\right]\right\}. \tag{3.11}$$

It follows from this result that for long times ($\tau\to\infty$) a temperature perturbation produced at any given location ($\xi=\mathrm{const}$) by an initial event at a point (at the point ξ_0) rapidly vanishes. Indeed, the first exponential factor decreases more rapidly than the secondary exponential factors, of which the most rapidly growing one refers to the lowest discrete level. Therefore diffusion spreading predominates here.

4. CONVECTIVE DIFFUSION WITH GLOBAL CONSERVATION OF AN IMPURITY

We now consider a model system of equations of the type (2.1) for the convection of a passive impurity, modified so that the equation for the impurity concentration admits a conservation law for the total amount of impurity:

$$\partial_t c+\partial_x(uc)=\chi\partial_x^2c;\quad \partial_t\int c\,dx=0. \tag{4.1}$$

For a fluid flow produced by a uniformly moving shock wave with a viscous transitional layer, we obtain for the impurity distribution in a coordinate system comoving with the shock wave, using dimensionless variables,

$$(\partial_\tau-\partial_\tau^2-2P\tanh\xi\partial_\xi-2P\operatorname{sech}^2\xi)c=0, \tag{4.2}$$

which differs from the equation of the model analyzed above only by the addition of the last term, which is proportional to the concentration. It is remarkable that after the previous generalized Cole–Hopf change of variables (2.6) it reduces to an equation for a new unknown function ψ of the same type as that obtained previously:

$$\psi=\varepsilon^{P}c;\quad \hat{L}_p\psi\equiv(\partial_\tau-\partial_\xi^2-P(P+1)\operatorname{sech}^2\xi)\psi=0. \tag{4.3}$$

However, here an important difference arises in the asymptotic behavior of the solutions for long times, even for identical initial conditions, in connection with the increase in the number of the operator here. For example, perturbations of the concentration which are caused by an initial action of the type (2.15) at a point (the Green's function problem)

$$c(\xi,\tau)|_{\tau=0}=\delta(\xi-\xi_0), \tag{4.4}$$

can be represented similarly to the previously examined model

$$c(\xi,\tau)=e^{-P^2\tau}\left(\frac{\cosh\xi_0}{\cosh\xi}\right)^P D(\xi,\xi_0,\tau;P+1). \tag{4.5}$$

Here, at first glance, the small but decisive difference of the Prandtl numbers by 1 occurs in the arguments of the Green's function. Since the asymptotic behavior of the Green's function for different Prandtl numbers is different, this results in different long-time asymptotic behavior of the solutions of the two types of model equations being analyzed.

According to the general representation of the D function from the preceding section, its asymptotic behavior for fixed coordinates ξ, ξ_0 and long times $\tau\to\infty$ is determined by the eigenfunction with the lowest discrete eigenvalue (the ground state)

$$D(\xi,\xi_0,\tau;P+1)\sim e^{P^2\tau}\Psi_1^{(P+1)}(\xi)\Psi_1^{(P+1)}(\xi_0), \tag{4.6}$$

and in accordance with Eq. (4.5) the final impurity distribution, in contrast to the situation studied earlier, does not vanish at long times and possesses a localized soliton-like form (the asymptotic behavior for fixed coordinate ξ, i.e., for the distribution accompanying a shock wave):

$$c(\xi,\tau)\xrightarrow[\tau\to\infty]{}\frac{2\Gamma(2P)}{\Gamma^2(P)2^{2P}}\operatorname{sech}^{2P}\xi. \tag{4.7}$$

It is easily verified that such a limiting distribution is a stationary solution of the convective diffusion equation under study. The archetypical (sech^2) form is obtained for unit Prandtl number. For large Prandtl numbers, the soliton becomes more peaked.

In the situation of unit Prandtl number, the general time-dependent case admits a simple analysis with reduction to the heat conduction equation.[13] The interaction (collision) of shock waves in the approximation characterized by Burgers' model is of an perfectly inelastic character.[2,3] If the mean free path is sufficiently long, to each shock wave there will correspond an asymptotically simplified impurity distribution—an ''impurity soliton.'' We emphasize that despite the classical form, impurity solitons differ radically from classical solitons (for example, the solitons of the Korteweg–de Vries equation) in terms of the character of their interactions. Coalescence of colliding shock waves corresponds to perfectly inelastic merging of impurity solitons. For large Prandtl numbers, individual impurity solitons are more strongly peaked, and a similar inelastic merging can be expected to be the asymptotically simplified result of collisions of shock wave—impurity soliton pairs. The result of such nonlinear interactions will be a peaked impurity concentration field, a strong localization of an impurity with general conservation of its total amount.

5. CONCLUSIONS (IMPURITY CLUSTERING EFFECT)

The analytic investigation of convective diffusion in a one-dimensional hydrodynamics without pressure performed above makes it possible to determine the overall physics of the phenomenon, depending strongly on the Prandtl number. For a medium with the same kinetic coefficients, just as for the model Burgers equation, time-dependent convective-diffusion problems turn out to be solvable. They were reduced to problems for a linear heat conduction equation. For integer Prandtl numbers, the equations of convective diffusion in the field of a uniformly moving shock wave likewise reduce to a heat conduction equation with constant coefficients, and the Green's function for time-dependent diffusion in the presence of stationary flow can be expressed in terms of error functions. The solutions of these deterministic initial problems can provide the basis for understanding impurity accumulation and localization.

Above all else, a passive impurity performs the role of a marker for determining the localized dynamical structures of the velocity field of a fluid flow. It successfully reflects the characteristic features of the formation, interaction, and restructuring of localized structures. However, restructuring of the markers themselves occurs as a result of the competition between diffusion and convection. The formation of stable, asymptotically simplified, impurity structures against a background of dynamical structures turns out to be, as is clear from the preceding exposition, very sensitive to the detailed features of the coupling of the impurity and velocity fields and to the satisfaction of the global conservation laws. When a global conservation law exists, impurity clusterization occurs.

Hydrodynamic turbulence in the one-dimensional Burgers model (''Burgerlence'') possesses the structure of a gas of inelastically colliding shock waves which is quite rarefied at the long-time stage of evolution.[20,21] For this reason, analysis of the elementary process of the collision of shock waves, the dynamical foundation of a stochastic wave ensemble, becomes decisive.

The dynamical picture of clustering in the one-dimensional approximation turns out to be especially simple for unit Prandtl number. In this case the time-dependent problem of the collision of shock waves with the formation of a larger wave as a result of the coalescence of the initial waves can be solved exactly. Asymptotically, each shock wave is accompanied by an impurity soliton, and merging of the colliding impurity solitons will correspond to perfectly inelastic collision of shock waves. As the shock waves merge, the impurity increasingly concentrates in individual locations. The sharpness of the impurity localization (local heating) depends on the ratio of the kinetic coefficients of the medium; it increases with the Prandtl number.

Note that from the standpoint of the evolution of a stochastic wave ensemble, such enlargement of structures is typical of an inverse cascade process, whose possibility is closely related to the low dimensionality of the problem.

This work was supported by the Russian Fund for Fundamental Research (Project 99-01-00435).

*)E-mail: teodor@ipmnet.ru

---

[1] J. M. Burgers, *The Nonlinear Diffusion Equation*, Reidel, Boston (1974).
[2] M. J. Lighthill, in *Surveys in Mechanics*, G. K. Batchelor and R. M. Davies (eds.), University Press, Cambridge (1956).
[3] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York (1974).
[4] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).

[5] J. D. Cole, Quart. Appl. Math. **9**, 225 (1951).
[6] A. M. Polyakov, Phys. Rev. E **52**, 6183 (1995).
[7] V. Gurarie and A. Migdal, Phys. Rev. E **54**, 4908 (1996).
[8] W. A. Woyczynsky, *Burgers-KPZ Turbulence, Göttingen Lectures*, Springer-Verlag, Berlin (1998).
[9] S. F. Shandarin and Ya. B. Zeldovich, Rev. Mod. Phys. **61**, 185 (1989).
[10] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[11] É. V. Teodorovich, Zh. Éksp. Teor. Fiz. **109**, 506 (1996) [JETP **82**, 268 (1996)].
[12] A. I. Saichev and W. A. Woyczynsky, Physica D **100**, 119 (1997).
[13] V. A. Gorodtsov, PMM **62**, 1021 (1998).
[14] S. N. Gurbatov, A. N. Malakhov, and A. I. Saichev, *Nonlinear Random Waves in Nondissipative Media* [in Russian], Nauka, Moscow (1990).
[15] V. I. Klyatskin and A. I. Saichev, Zh. Éksp. Teor. Fiz. **111**, 1297 (1997) [JETP **84**, 716 (1997)].
[16] S. N. Gurbatov and G. V. Pasmanik, Zh. Éksp. Teor. Fiz. **115**, 564 (1999) [JETP **88**, 309 (1999)].
[17] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Nonrelativistic Theory)*, 3rd ed., Pergamon Press, New York (1977).
[18] S. Flügge, *Practical Quantum Mechanics*, Springer-Verlag, Berlin (1971).
[19] R. E. Crandall, J. Phys. A **16**, 3005 (1983).
[20] J. M. Burgers, in *Statistical Models and Turbulence*, M. Rosenblatt and C. Van Atta (eds.), Springer-Verlag, New York (1972).
[21] T. Tatsumi and S. Kida, J. Fluid Mech. **55**, 659 (1972).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Structure of turbulent flows of incompressible fluids and the parametrization of turbulence

S. N. Gordienko*) and S. S. Moiseev†)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Moscow; Space Research Institute, Russian Academy of Sciences, 117810 Moscow, Russia

(Submitted 19 October 1999)
Zh. Éksp. Teor. Fiz. **116**, 1630–1647 (November 1999)

The structure of stationary isotropic, homogeneous turbulence in an incompressible fluid with Re $\gg 1$ set into motion by a force with amplitude f_0 and spatial and temporal time scales of r_0 and τ_0, respectively, is examined. It is found that, depending on the magnitude of the force that sets the fluid into motion, three fundamentally different turbulent stationary states of the fluid can develop and the dimensionless parameters responsible for transitions from one state to another, $\gamma = f_0\tau_0^2/r_0$ and $\Gamma = \gamma^{4/3}\,\mathrm{Re}$, are determined. It is shown that for $\gamma \ll 1$ and $\Gamma \ll 1$ a Kolmogorov spectrum with $E(k) \propto 1/k^{5/3}$ develops in the inertial range. During the transition to turbulent flows driven by large amplitude forces f_0, i.e., during the transition to a regime with $\gamma \ll 1$ and $\Gamma \gg 1$, a segment of the spectrum with $E(k) \propto 1/k^2$ develops near the viscous range and ''detaches'' the Kolmogorov spectrum from the viscous range. Further increases in the amplitude f_0 of the force, i.e., approaching the parameter range with $\gamma \gg 1$ and $\Gamma \gg 1$, causes the entire inertial range to be ''occupied'' by a spectrum $E(k) \propto 1/k^2$, and outside the inertial range, large scale structures with a characteristic size extending to $\gamma^{2/5} r_0$ begin to be generated. In the regime with $\Gamma \ll 1$, the power dissipated per unit mass of fluid is independent of the viscosity, but on going to turbulent regimes with $\Gamma \gg 1$, the viscous losses begin to depend on the viscosity of the fluid. The ''turn-off'' of viscous dissipation for $\Gamma \gg 1$ shows that a drag crisis can occur simply as the source power is increased, without any further conditions. With this method for the excitation of turbulence, the Loitsyanskii integral diverges for arbitrary values of γ and Γ. A physical mechanism is proposed to explain the readjustment of the spectrum of the turbulent fluctuations at different γ and Γ. These results have all been obtained neglecting intermittency. © *1999 American Institute of Physics.* [S1063-7761(99)01011-2]

1. INTRODUCTION

The transition of liquids or plasmas to a turbulent state is of fundamental importance for understanding the physics of a number of processes (including transport phenomena) taking place in these media. An enormous number of theoretical and experimental papers have been published so far on various aspects of turbulence in widely varying situations. This has led to an understanding of weak turbulence, i.e., the case when the dynamics of the system can be described in the language of weakly interacting linear waves, but the theory of strong turbulence is still far from complete.

The best known example of strong turbulence is the turbulence in flows of incompressible fluids at high Reynolds numbers. Although the main result of hydrodynamic turbulence theory is the Kolmogorov spectrum, which was derived more than fifty years ago and has been repeatedly verified experimentally,[1] attempts to justify this spectrum theoretically, even relying on such usually efficient methods of theoretical physics as the renormalization group technique,[2–4] diagram techniques,[5,6] and the functional formulation of the problem proposed so long ago by Hopf,[7] have not yielded entirely satisfactory results (see the explanations given below). It can be stated that the method of obtaining the spectrum of turbulent fluctuations going back to Kolmogorov is still the clearest approach to the problem.

The basis of the ideology proposed by Kolmogorov is the hypothesis that the only dimensional constant in the theory of developed hydrodynamic turbulence in the inertial range is the rate of energy dissipation per unit mass, that is, a quantity with the dimensions of $\mathrm{cm}^2/\mathrm{s}^3$. It remains unclear why a quantity with this dimensionality should be better than one with dimensions of $\mathrm{cm}^n/\mathrm{s}^m$, where $3n \neq 2m$, i.e., what distinguishes the case $3n = 2m$, where m and n are real numbers.

The Kolmogorov spectrum and the concepts of universality, constant energy flux over the spectrum, and inertial range associated with it are the basis of the modern understanding of the nature of homogeneous turbulence in incompressible fluids. Meanwhile, the Kolmogorov spectrum is not the only spectrum of turbulent fluctuations that has been observed experimentally in turbulent incompressible fluids. Besides the Kolmogorov spectrum $E(k) \propto 1/k^{5/3}$, for example, the spectrum $E(k) \propto 1/k^{7/3}$ associated with the helicity flux is well known.[8] It is, therefore, necessary to determine the conditions under which the rate of energy dissipation is a dimensional constant that determines the turbulent fluctuation spectrum and to examine the question of how, and under what conditions, it is replaced by a dimensional parameter of different character. Unfortunately, this problem has not received the required elucidation in any of the above cited papers devoted to justifying the Kolmogorov spectrum. Thus, the question of the conditions for the formation of one or another spectrum arises, that is, of finding the parameters that determine the spectral composition of the turbulent fluctuations in one or another range of wave vectors. This paper is devoted to the problem of determining the parts of the spectrum with universal behavior and the parameters governing their onset. A number of related questions have been examined briefly before,[9] but part of the results of that paper needed to be made more precise and interpreted physically.

Before discussing the specific results, let us emphasize specially that the entire analysis presented here neglects intermittency. Recall that intermittency leads to the appearance of an additional factor of the form $(L/r)^{\delta}$, where L is a characteristic size and δ is the dimensional anomaly. Thus, intermittency effects can change the numerical value of the correlator substantially, even for small δ, if L is sufficiently large. In addition, if δ is small, then the additional factor associated with intermittency is a function that varies slowly compared to the Kolmogorov part of the correlator. In other words, neglecting intermittency is somewhat analogous to the WKB approximation: only the rapidly varying function is taken into account, although the slow part can change substantially in magnitude over long intervals.

The possibility of neglecting intermittency is the hypothesis under which the results obtained below are valid. Thus, here we do not touch upon the applicability of the Kolmogorov theory under intermittency conditions, but study the need to modify the Kolmogorov theory when the finite correlation time of the force that drives fluid motion is taken into account.

2. STATEMENT OF THE PROBLEM AND THE DESCRIPTION OF TURBULENCE IN TERMS OF THE EULER EQUATION

As we are interested only in fluid flows with large Reynolds numbers, as a first step we consider the Euler equation (Re$=+\infty$)

$$\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}=-\nabla p+\mathbf{f}(\mathbf{r},t),\qquad \operatorname{div}\mathbf{v}=0, \tag{1}$$

with an external force $\mathbf{f}(\mathbf{r},t)$ that obeys

$$\langle \mathbf{f}(\mathbf{r},t)\rangle=0,$$

$$\langle \mathbf{f}(\mathbf{r}_1,t_1)\cdot\mathbf{f}(\mathbf{r}_2,t_2)\rangle=f_0^2K\left(\frac{t_1-t_2}{\tau_0},\frac{\mathbf{r}_1-\mathbf{r}_2}{r_0}\right), \tag{2}$$

where K describes the correlation properties of the force $\mathbf{f}$ that sets the fluid into motion and, by the definition of f_0^2,

$$\int K\left(\frac{t}{\tau_0},\frac{\mathbf{r}}{r_0}\right)d\mathbf{r}\,dt=r_0^3\tau_0.$$

For the following discussion, it is important that the last integral converge and be nonzero; that is, the limits imposed on the force that brings the liquid into motion reduce to this. In order to avoid misunderstandings, we emphasize specially that the equality $\langle \mathbf{f}(\mathbf{r},t)\rangle=0$ does not imply that this integral of the function K goes to zero. Without loss of generality, we can assume that the force $\mathbf{f}$ is purely solenoidal and include the potential part in the pressure gradient.

First, we rewrite Eq. (1) in Lagrangian coordinates.[12] To do this, it is necessary to begin by expressing the pressure in terms of the distribution of the velocities in the liquid. Calculating the divergence of Eq. (1) subject to the incompressibility condition $\operatorname{div}\mathbf{v}=0$, we have

$$\triangle p=-\frac{\partial^2}{\partial x_i\partial x_j}(v_iv_j), \tag{3}$$

here and everywhere, if the contrary is not stated, repeated indices denote summation.

From Eq. (3) we obtain

$$p(\mathbf{r},t)=-\frac{1}{4\pi}\int\frac{1}{|\mathbf{r}-\mathbf{r}'|}\frac{\partial^2}{\partial x_i'\partial x_j'}(v_i'v_j')d\mathbf{r}', \tag{4}$$

where $\mathbf{v}_i'=\mathbf{v}(\mathbf{r}',t)$.

Using Eqs. (3) and (4), we find

$$\frac{\partial p}{\partial x_k}=\frac{1}{4\pi}\int(v_i'-v_i)(v_j'-v_j)\frac{\partial^3}{\partial x_i'\partial x_j'\partial x_k'}\frac{1}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \tag{5}$$

We now transform to Lagrangian variables.[12] Let a particle of liquid be at the point $\mathbf{r}_0$ at time $t_0=0$. (Note that the value of $\mathbf{r}_0$ can be regarded simply as a label for the particle.) Then the velocity of the particle with number $\mathbf{r}_0$ at time t, which we shall denote by $\mathbf{w}(\mathbf{r}_0,t)$ below, is simply equal to the velocity at time t at the point $\mathbf{r}(\mathbf{r}_0,t)$, i.e.,

$$\mathbf{w}(\mathbf{r}_0,t)=\mathbf{v}(\mathbf{r}(\mathbf{r}_0,t),t). \tag{6}$$

Noting that, according to the incompressibility condition

$$\det\left|\frac{\partial\mathbf{r}(\mathbf{r}_0,t)}{\partial\mathbf{r}_0}\right|=1, \tag{7}$$

i.e., the motion does not change the total number of particles, we rewrite Eq. (1) in the form

$$\frac{d}{dt}w_k=-\int G_{ijk}(\mathbf{r}(\mathbf{r}_0,t)-\mathbf{r}(\mathbf{r}_0',t))\times(w_i'-w_i)(w_j'-w_j)d\mathbf{r}_0'+\mathbf{f}(\mathbf{r},t), \tag{8}$$

with

$$\frac{d}{dt}r_k(\mathbf{r}_0,t)=w_k(\mathbf{r}_0,t). \tag{9}$$

Here the incompressibility condition is simply written as a limitation on the allowable initial conditions:

$$\operatorname{div}_{\mathbf{r}_0}\mathbf{w}(\mathbf{r}_0,t)=0. \tag{10}$$

Equations (8) and (9) are the equations of classical continuum mechanics for classical particles, although with a rather exotic interaction. This naturally brings up a few fundamental questions. First, note that, in the absence of an external force, the Euler equation (1), whose solutions are assumed to have a sufficient number of derivatives (i.e., to belong to the corresponding class of smoothness), conserves the total kinetic energy of the fluid, which coincides with the kinetic energy of the dynamic system (8) and (9). This last point is extremely nontrivial: the kinetic energy is conserved in a system of interacting particles. Here the interaction is such that it cannot be stated that the force acting on a particle at any given time is perpendicular to the velocity of that particle. The question therefore arises of what property of the dynamic system (8) and (9) distinguishes it from Hamiltonian systems, in which only the total energy (and not just the kinetic energy, specifically) energy is conserved, or from systems with dissipation (frictional forces), in which no energy of any kind is conserved. In order to understand this, it is necessary to consider the initial assumptions which led to the Euler equation. First of all, we note that, if Newton's third law is valid in the system of particles but only particles with equal velocity vectors interact, then it is the total kinetic energy which is conserved in such a system. In the derivation of the Euler equation it was assumed that only adjacent Lagrangian particles interact, while Newton's third law (action equals reaction) is valid and the velocities of the Lagrangian particles depend smoothly on position. Thus, if we consider two Lagrangian particles a and b interacting with one another, then the force of particle a acting on particle b, $\mathbf{F}_{ab}$, is related to the force $\mathbf{F}_{ba}$ by $\mathbf{F}_{ab}=-\mathbf{F}_{ab}$ (Newton's third law), and their velocities coincide, i.e., $\mathbf{v}_a=\mathbf{v}_b$, to within infinitesimally small quantities which vanish for zero distance between particles a and b (there is only a contact interaction); that is, the work done by particle a on particle b for an infinitesimal time interval dt equals $\mathbf{v}_b\cdot\mathbf{F}_{ab}dt$, while the work done by particle b on particle a over the same time interval, $\mathbf{v}_a\cdot\mathbf{F}_{ba}dt$, is of the same magnitude, but opposite sign. It is clear from these remarks why the kinetic energy is conserved for a fluid whose motion is described by smooth solutions of the Euler equation. If the necessary smoothness is lacking, then the infinitesimal differences in the velocities of the interacting particles cannot be neglected in the above discussion. Besides, this can be confirmed in a purely formal way by not turning to intuitive considerations associated with the sources of the Euler equation: note that, if the integrand in

$$\int (\mathbf{v}(\mathbf{r},t)\cdot\nabla p(\mathbf{r},t))d\mathbf{r}$$

is a smooth function of position, then the integral reduces to a surface integral in the case of an incompressible fluid and the Euler equation conserves the kinetic energy of the fluid. Therefore, conservation of the kinetic energy in the dynamic system (8) and (9) is a property of a particular class of its solutions that corresponds to the solutions of the Euler equation with a sufficiently high class of smoothness. This latter discussion provide an interpretation for the concept of a nonviscous limit with finite dissipation introduced in Ref. 8. Note that solutions corresponding to a nonviscous limit with finite dissipation should arise naturally in the dynamic system of Eqs. (8) and (9) if we do not require the existence of higher order derivatives with respect to the label of the Lagrangian particles, the parameter $\mathbf{r}_0$. Thus, obtaining a closed, irreversible kinetic equation for the Lagrangian particles of the fluid makes it possible to formalize, in a natural way, the intuitive concept of turbulent viscosity, which is often used in the theory of turbulence.

Of course, a chain of coupled equations for many particle distribution functions can be written down for this system of Eqs. (8) and (9). Unfortunately, because of the strong and unusual interaction between the particles, it is impossible to truncate the resulting infinite chain of equations correctly. In the meantime, there is a way of proceeding to an understanding of the dynamics of this system. It is, indeed, understandable that finding simultaneous distribution functions for the velocity and pressure gradient would be of great interest for hydrodynamics. However, in the language of the dynamic system of Eqs. (8) and (9), the pressure gradient at any point is proportional to the acceleration of the Lagrangian fluid particle located at that point. Given this, we can use a method similar to that in Ref. 14: any description of the system dynamics ultimately implies a description of typical single particle trajectories comprising the dynamic system of particles. Here the use of distribution functions raises the question of what variables these functions should depend on in order for a closed description of the dynamics to be attainable in their language. In the case of gases, it is well known that we can limit ourselves to considering distribution functions that depend only on positions and momenta (the Boltzmann equation), while generalized distribution functions that also depend on higher derivatives must be considered in the case of media with more complicated dynamic properties.[14]

3. KINETIC EQUATIONS FOR DESCRIBING TURBULENCE

In order to study the spectral properties of the problem stated by Eqs. (1) and (2), we shall use a kinetic equation for the Lagrangian particles of an incompressible fluid[13] in the limit Re→+∞ (see explanations below):

$$\left(\frac{\partial}{\partial t}+\hat{L}_1+\hat{L}_2\right)F_2(t,1,2)=P(t,1,2)+P(t,2,1). \tag{11}$$

Here $F_2(t,1,2)=F_2(t,\mathbf{r}_1,\mathbf{v}_1,\mathbf{a}_1,\mathbf{r}_2,\mathbf{v}_2,\mathbf{a}_2)$ is the generalized (i.e., dependent on the accelerations resulting from the interaction of the Lagrangian particles, as well as on their velocities) two-particle distribution function,

$$\hat{L}_i=\mathbf{v}_i\frac{\partial}{\partial\mathbf{r}_i}+(\mathbf{a}_i+\mathbf{f}(\mathbf{r}_i,t))\frac{\partial}{\partial\mathbf{v}_i},\quad i=1,2,$$

$$P(t,1,2)=\int_0^{+\infty}\hat{R}(t,t-\tau)\left(\int\int\hat{Q}(1,3)\hat{Q}(2,4)F_2(t-\tau,1,3)F_2(t,2,4)d3d4\right)d\tau, \tag{12}$$

$$\begin{aligned}\hat{Q}(1,2)=&\,2(a_{2,l}-a_{1,l})(v_{2,m}-v_{1,m})\\&\times\frac{\partial^3}{\partial x_{1,l},\partial x_{1,m},\partial x_{1,p}}\,\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\,\frac{\partial}{\partial a_{1,p}}\\&-(v_{2,m}-v_{1,m})(v_{2,l}-v_{1,l})(v_{2,n}-v_{1,n})\\&\times\frac{\partial^4}{\partial x_{1,l},\partial x_{1,m},\partial x_{1,n},\partial x_{1,p}}\,\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\,\frac{\partial}{\partial a_{1,p}},\end{aligned}$$

where the repeated subscripts in the last formula imply summation and $v_{2,l}$, $a_{2,l}$, and $v_{1,l}$, $a_{1,l}$ are the lth components of the vectors $\mathbf{v}_2$, $\mathbf{a}_2$, and $\mathbf{v}_1$, $\mathbf{a}_1$, respectively. The operator $\hat{R}(t,t-\tau)$ is defined as

$$\hat{R}(t,t-\tau)=\hat{U}^{-1}(t)\hat{U}(t-\tau), \tag{13}$$

where the operator $\hat{U}(t)$ obeys the differential equation

$$\frac{d\hat{U}(t)}{dt}=\hat{U}(t)\hat{L}, \quad \hat{U}(0)=1, \tag{14}$$

and the operator $\hat{U}^{-1}(t)$ obeys the following equation which follows from Eq. (14),

$$\frac{d\hat{U}^{-1}(t)}{dt}=-\hat{L}\hat{U}^{-1}(t), \quad \hat{U}^{-1}(0)=1.$$

The operators $\hat{U}(t)$ and $\hat{U}^{-1}(t)$ can be represented as time-ordered exponentials, which, for a time-independent force $\mathbf{f}$, reduce to the usual exponentials, in which case the operator $\hat{R}$ is easily calculated. In the general case, the operator $\hat{R}$ can be calculated only for special cases, but this is of no significance for the following discussion, since the scaling properties of $\hat{R}$, which are of importance there, follow from Eqs. (13) and (14).

Let us clarify the significance of introducing distribution functions that contain the accelerations as additional arguments. Usually, in constructing kinetic equations, one begins with the distribution functions and tries to describe processes associated with the interparticle interaction. A proposal[14] has been made that kinetic equations be obtained by solving the inverse problem: assuming that the distribution of forces in space is given (in the case of a liquid, the pressure gradient), calculate many-particle distribution functions that are consistent with the given distribution of force vectors. In this way distribution functions appear which depend on the accelerations. Since no restrictions have been imposed on the initial distribution of forces in space, except those which follow from Newton's laws, this approach permits a correct description of fluctuations in the framework of the kinetic equation.[15] Furthermore, any generalized distribution function also contains information of the sort which is contained only in the entire infinite set of standard distribution functions (for example, any generalized distribution function can be used to calculate $\langle \mathbf{a}^{2n}\rangle$ for arbitrary n). In connection with this last point, it is clear why, in a number of problems where the standard BBGKY chain cannot be truncated, it is possible to obtain a closed equation for the generalized two-particle distribution function.[13] A kinetic equation of this form has been used successfully for solving a number of problems in plasma physics.[16]

If the function $F_2(t,1,2)$ is a solution of Eq. (11), then the function

$$\begin{aligned}F_2^{(\Delta,\lambda)}(t,1,2)=&\,\lambda^{-18\Delta-6}F_2(\lambda^{1+\Delta}t,\lambda\mathbf{r}_1,\lambda^{-\Delta}\mathbf{v}_1,\lambda^{-1-2\Delta}\mathbf{a}_1,\\&\times\lambda\mathbf{r}_2,\lambda^{-\Delta}\mathbf{v}_2,\lambda^{-1-2\Delta}\mathbf{a}_2),\end{aligned} \tag{15}$$

is also a solution of Eq. (11), but with another external force

$$\mathbf{f}^{(\Delta,\lambda)}(\mathbf{r},t)=\lambda^{2\Delta+1}\mathbf{f}(\lambda\mathbf{r},\lambda^{1+\Delta}t), \tag{16}$$

where $\lambda>0$ and Δ are arbitrary numbers. This is easily confirmed by substituting Eqs. (15) and (16) in Eq. (11).

A number of expressions from this section were given in Ref. 9 with some errors that did not affect the results of that paper.

Here we note a fundamental feature of the kinetic equation formalism in the theory of turbulence. The kinetic equation formalism makes it possible to describe turbulence in terms of an evolution problem, i.e., proceeding from arbitrarily specified initial conditions it is possible to follow the time evolution of the solution. If the initial conditions are forgotten in the course of events, then the system approaches a stationary solution. Here the remarkable thing is the possibility of finding a stationary solution by proceeding from the evolution problem. There is no need to invert a linear or nonlinear operator, i.e., the possibility of studying turbulence in the framework of an evolution problem allows us to avoid, in a natural way, the problem of zero modes which always shows up during the inversion of operators. The complexity of the zero mode problem can be traced, for example, in Kraichnan's model for passive scalar diffusion.[10] Although a closed linear equation for the four point correlator can be obtained in this model, the nonevolutionary character of the resulting boundary value problem means that the operator must be inverted, i.e., it leads to a zero mode problem, which is one of the central problems[10] in the diffusion of passive scalars. The zero mode problem in descriptions of turbulence in incompressible fluids has been examined elsewhere.[11] Note that the zero mode problem is considerably more complicated for passive scalars than for hydrodynamic turbulence. This follows, as well, from the fact that passive scalars are described by linear inhomogeneous partial differential equation, while hydrodynamic turbulence is associated with an intrinsically nonlinear problem. This difference is fundamental: in fact, the solution of an inhomogeneous linear equation transforms to a solution of the same equation if an arbitrary solution of the corresponding homogeneous equation is added to it, but in nonlinear problems there is no analog of this property of linear equations. Another fundamental feature of the passive scalar problem, which makes it more complicated, in a certain sense, than the hydrodynamic turbulence problem is that, in order to study steady-state diffusion (i.e., to describe diffusion by time independent correlation functions) it is necessary to introduce two types of random variables–the source of the passive scalar and the velocity field.[10]

4. LARGE-SCALE STRUCTURE OF A TURBULENT FLOW AND THE LOITSYANSKII INTEGRAL

In order to study the large-scale structure of a flow, we choose the number Δ so that in the limit $\lambda\rightarrow+\infty$ the function $F_2^{(\Delta,\lambda)}(t,1,2)$ approaches a finite limit. This requires that the correlator of the force (16) have a finite limit as $\lambda\rightarrow+\infty$, i.e., in connection with the equality

$$\lim_{\lambda\rightarrow+\infty}\langle\mathbf{f}^{(\Delta,\lambda)}\cdot(\mathbf{r}_1,t_1)\cdot\mathbf{f}^{(\Delta,\lambda)}(\mathbf{r}_2,t_2)\rangle$$

$$=f_0^2r_0^3\tau_0\delta(\mathbf{r}_1-\mathbf{r}_2)\delta(t_1-t_2)\lim_{\lambda\rightarrow+\infty}\lambda^{3\Delta-2},\tag{17}$$

we must set $\Delta=\Delta_\infty=2/3$. We note that, precisely because of our examination of the limit $\lambda\rightarrow+\infty$, in the right hand side of Eq. (17) the correlator K for the external force can be replaced by a product of δ functions. Here it is significant that the resulting $\Delta_\infty=2/3>-1$.

We therefore conclude that the function $\lim_{\lambda\rightarrow+\infty}F_2^{(2/3,\lambda)}$ is homogeneous (to prove this it is sufficient to note that, according to the definition of the limit, $\lim_{\lambda\rightarrow+\infty}F_2^{(2/3,\lambda)}=\lim_{\lambda\rightarrow+\infty}F_2^{(2/3,\lambda\lambda')}$, where λ' is an arbitrary positive number) and that the relation

$$\langle\mathbf{v}(\mathbf{r},t)\cdot\mathbf{v}(0,t)\rangle=C_1\left(\frac{f_0^2r_0^3\tau_0}{r^2}\right)^{2/3}$$

$$\text{for } r\gg\max\left(r_0,\left(\frac{f_0\tau_0^2}{r_0}\right)^{2/5}r_0\right)\tag{18}$$

holds, where C_1 is a universal constant. In this stage of the analysis, we cannot explain the domain of applicability for this correlator given by the inequality in Eq. (18). Later we shall examine the domain of applicability of Eq. (18), but for now we can confirm that the correlator (18) actually falls off as $1/r^{4/3}$ at large distances. In fact, writing the left hand side of Eq. (18) in the form of an integral of the function $\lim_{\lambda\rightarrow+\infty}F_2^{(2/3,\lambda)}$ and using the homogeneity of this function, it is possible to establish that the correlator has a power law dependence on r and to determine the power. Note that a power law behavior on the right hand side of Eq. (18) follows from the homogeneity of the function $\lim_{\lambda\rightarrow+\infty}F_2^{(2/3,\lambda)}$ and the power to which r is raised on the right of Eq. (18) is uniquely determined by the number $\Delta_\infty=2/3$. Here, of course, the functional dependence in Eq. (18) is consistent with the dimensionality of the parameter in Eq. (17). We emphasize specially that we have not postulated, but proven, the power law character of the right hand side of Eq. (18). Equation (18) describes the creation of large vortices (inverse cascade) and of the long wave length correlations associated with them. The correlator (18) leads to a spectrum $E(k)\propto k^{1/3}$ in the long wavelength region, as confirmed experimentally.[17]

The correlator (18) shows up in the so-called Loitsyanskii integral[12]

$$\Lambda=-\frac{1}{4\pi}\int r^2\langle\mathbf{v}(\mathbf{r},t)\cdot\mathbf{v}(0,t)\rangle d\mathbf{r}.$$

According to Eq. (18), the Loitsyanskii integral diverges, i.e., $\Lambda=\infty$. Thus, a correct description of the inverse cascade for the method of exciting turbulence being studied here leads to a divergent Loitsyanskii integral. It is curious that prior to this, the convergence of the Loitsyanskii integral had been studied only for isotropic turbulent motion with an exponentially rapid decrease in the correlator (18) at large distances at the initial time, i.e., assuming that it is possible to create isotropic turbulence with negligible long wavelength correlations. Under this assumption it is possible to show that even for subsequent free evolution, the Loitsyanskii integral remains finite.[12] Equation (18) shows that a similar assumption that the role of large scale vortices (the inverse cascade) is negligible is not always satisfied for isotropic turbulence.

Recall that the Loitsyanskii integral is related to the square of the angular momentum of a turbulent fluid, i.e., in a certain sense it characterizes the appearance of spontaneous rotation of the entire fluid as a whole. It is easy to verify that the square of the total angular momentum $\mathbf{M}$ of the fluid contained within a large volume V (isolated within an unbounded fluid) is $M^2=4\pi\rho^2\Lambda V$, where ρ is the fluid density. The fact that $M^2\propto V^{20/9}$ confirms that long-wavelength correlations are important in the dynamics of a turbulent fluid.

5. UNIVERSAL PROPERTIES OF SMALL-SCALE FLOW STRUCTURE

It turns out that the small-scale structure of the flow is also universal. In fact, an examination of the small-scale structure of the flow of Eqs. (1) and (2) reduces to choosing $\Delta=\Delta_0=-1/2$ in Eq. (16), which ensures the existence of a finite homogeneous function $\lim_{\lambda\rightarrow+0}F_2^{(-1/2,\lambda)}$. This implies that

$$\langle(\mathbf{v}(\mathbf{r},t)-\mathbf{v}(0,t))^2\rangle=C_2f_0r\quad\text{for } r\ll\min(r_0,f_0\tau_0^2),\tag{19}$$

where C_2 is a universal constant. In the following, we shall explain the physical meaning of the inequality that determines the domain of applicability of the correlator (19).

A numerical study has been made of turbulent flows driven by a constant force applied to the fluid.[18] It was found that in certain directions (in the simulations the external force was chosen to have a spatial period and not depend on time) a spectrum $E(k)\propto 1/k^2$ is formed, rather than the Kolmogorov spectrum $E(k)\propto 1/k^{5/3}$. This is in accord with Eq. (19) and, we believe, validates the ideas developed here.

To study the flow at scales from r_0 to $f_0\tau_0^2$, we shall discuss two limiting cases, $\gamma=f_0\tau_0^2/r_0\ll 1$ and $\gamma=f_0\tau_0^2/r_0\gg 1$.

5.1. The condition for formation of a Kolmogorov spectrum: $\gamma=f_0\tau_0^2/r_0\ll 1$

Let $\gamma\ll 1$. We shall examine F_2 for $|\mathbf{r}_1-\mathbf{r}_2|\ll r_0$. In this range of distances, the force correlator (2) can be set equal to

$$\langle\mathbf{f}(\mathbf{r}_1,t_1)\cdot\mathbf{f}(\mathbf{r}_2,t_2)\rangle=f_0^2K\left(\frac{t_1-t_2}{\tau_0},0\right).\tag{20}$$

We emphasize specially that equation (20) for the correlator does not allow us to write down the form of the spectrum from dimensionality arguments alone. In fact, the dimensional parameters f_0 and τ_0 can be used to construct a combination with the dimensions of length, $f_0\tau_0^2$, and, therefore, dimensional arguments only determine the spectrum to within a factor of $\phi(f_0\tau_0^2 k)$, where ϕ is an arbitrary function.

Let us apply the considerations that led to Eq. (18) to the correlator (20). Note that using Eq. (20) guarantees the discussion for distances $|\mathbf{r}_1-\mathbf{r}_2|\ll r_0$, while the limit $\lambda\to+\infty$ ensures it for $|\mathbf{r}_1-\mathbf{r}_2|\gg\gamma r_0$. Thus, we find that for $\gamma r_0 \ll|\mathbf{r}_1-\mathbf{r}_2|\ll r_0$, the flow can be described by a uniform function $\lim_{\lambda\to+\infty}F_2^{(-1/3,\lambda)}$ and the equation

$$\langle(\mathbf{v}(\mathbf{r},t)-\mathbf{v}(0,t))^2\rangle=C_3(f_0^2\tau_0 r)^{2/3} \text{ for } \gamma r_0\ll r\ll r_0 \tag{21}$$

holds, i.e., the Kolmogorov-Obukhov law is satisfied within this scaling interval. The question of the limit of zero viscosity, but finite dissipation, in the Euler equation and the development of a Kolmogorov spectrum in this case has been discussed elsewhere.[8]

Therefore, for the method of exciting turbulence discussed here, a Kolmogorov spectrum develops only when certain conditions are imposed on the external force that drives the fluid into motion, $\gamma\ll 1$. Therefore, only in this case does a constant with the dimensions of power dissipated per unit mass actually end up being the parameter which determines the behavior of the fluid within a certain spectral interval. Later we shall discuss the physical processes taking place at the boundaries of that interval in detail, but for now we proceed to examine the case $\gamma\gg 1$.

5.2. The structure of the intermediate scaling region for $\gamma\gg 1$

Let $\gamma\gg 1$. Note that the characteristic time corresponding to fluctuations with a characteristic scale of r_0 and the corresponding [to these fluctuations, according to Eq. (19)] characteristic velocity $(f_0 r_0)^{1/2}$, is estimated to be $\gamma^{-1/2}\tau_0$. Let $r(\tau_0)$ (we shall estimate it below) be the characteristic fluctuation scale length, to which the characteristic time τ_0 corresponds. We shall consider F_2 for $r_0\ll|\mathbf{r}_1-\mathbf{r}_2|\ll r(\tau_0)$. For this scaling region, the force correlator (2) can be rewritten as

$$\langle\mathbf{f}(\mathbf{r}_1,t_1)\cdot\mathbf{f}(\mathbf{r}_2,t_2)\rangle=f_0^2 K\left(0,\frac{\mathbf{r}_1-\mathbf{r}_2}{r_0}\right), \tag{22}$$

since the characteristic velocity of the Lagrangian fluid particles [Eq. (19)] is at least $(f_0 r_0)^{1/2}$ and the distance moved by a Lagrangian particle over a time on the order of τ_0 is of order $\gamma^{1/2} r_0$.

Arguments similar to those of the previous section show that for $r_0\ll|\mathbf{r}_1-\mathbf{r}_2|\ll r(\tau_0)$, a turbulent flow of the fluid can be described by the homogeneous function $\lim_{\lambda\to+\infty}F_2^{(1/4,\lambda)}$, with

$$\langle\mathbf{v}(\mathbf{r},t)\cdot\mathbf{v}(0,t)\rangle=C_4\left(\frac{f_0^2 r_0^3}{r}\right)^{1/2} \text{ for } r_0\ll r\ll r(\tau_0)=\gamma^{2/5} r_0. \tag{23}$$

The following considerations were used to estimate the term $r(\tau_0)$ in Eq. (23). Note that, according to Eq. (23), the characteristic velocity corresponding to fluctuations with a scale length of order $r(\tau_0)$ are roughly $[f_0^2 r_0^3/r(\tau_0)]^{1/4}$. Thus, the size $r(\tau_0)$ can be found from the equation

$$r(\tau_0)\sim\tau_0(f_0^2 r_0^3/r(\tau_0))^{1/4},$$

which yields $r(\tau_0)=\gamma^{2/5} r_0$.

Note that the inequality in Eq. (18) follows from the inequalities defining the region of applicability of the correlators (19) and (23) for $\gamma\ll 1$ and $\gamma\gg 1$.

In the case examined in this section, there is no segment of the spectrum with a Kolmogorov power law dependence of the form $E(k)\propto 1/k^{5/3}$. Note, also, the slow decrease of the velocity correlator, $\propto 1/\sqrt{r}$ for $r_0\ll r\ll\gamma^{2/5} r_0$, which is replaced by a rapid decrease of universal character, $\propto 1/r^{4/3}$ for $\gamma^{2/5} r_0\ll r$ [see Eq. (18)].

6. THE ROLE OF FINITE VISCOSITY

We now take account of a finite viscosity ν in the problem, i.e., the existence of yet another characteristic scale length, the viscous length. We shall show that the energy dissipated per unit time per unit mass (specific power dissipation ϵ), is determined by the new dimensionless parameter

$$\Gamma=\gamma^{4/3}\,\mathrm{Re}, \tag{24}$$

where $\mathrm{Re}=r_0(f_0^2 r_0\tau_0)^{1/3}/\nu$ for $\gamma\ll 1$ and $\mathrm{Re}=r_0(f_0 r_0)^{1/2}/\nu$ for $\gamma\gg 1$. We shall examine three cases that can occur in turbulent flows ($\mathrm{Re}\gg 1$) in detail.

It makes sense to examine the character of the dissipation for different values of Γ in some detail, by isolating different special cases, since an examination of the nature of the dissipation yields some important qualitative results concerning the structure of turbulent flows.

6.1. The physics of turbulent dissipation for $\gamma\ll 1$

For $\gamma\ll 1$ and $\Gamma\ll 1$, according to Eqs. (19) and (24) the viscous length exceeds $f_0\tau_0^2$, i.e., the Kolmogorov spectrum is immediately adjacent to the viscous range. The specific power dissipation ϵ in this case can be estimated as the viscous loss over scale lengths on the order of the viscous length,

$$\epsilon\sim f_0^2\tau_0, \tag{25}$$

that is, it is independent of the viscosity [cf. Eq. (21)]. This result is entirely consistent with the, by now, classical statement that viscous losses in a turbulent fluid are independent of the magnitude of the viscosity; however, as we shall now verify, this situation is only valid as long as the Kolmogorov spectrum is immediately adjacent to the viscous range.

Let $\gamma\ll 1$, as before, but $\Gamma\gg 1$. Then, Eqs. (19) and (24) imply that the size $f_0\tau_0^2$ substantially exceeds the viscous length, i.e., the Kolmogorov spectrum is separated from the viscous range by the spectrum (19). The specific dissipation power is then roughly

$$\epsilon\sim\nu^{1/3} f_0^{4/3}, \tag{26}$$

i.e., approximately $\mathrm{Re}_*^{1/3}$ times smaller than the energy flux along the Kolmogorov segment of the spectrum, where Re_*

is the Reynolds number for the short wavelength end of the Kolmogorov spectrum. Thus, the segment (19) of the spectrum shields the viscous range from the energy flux propagating along the Kolmogorov segment of the spectrum and turns on the viscous dissipation.

We now provide a physical interpretation of this last result. The intuitive qualitative discussions given below make it possible to understand its physical content. In order to understand the physical significance of Eq. (26), we note that the external force $\mathbf{f}(\mathbf{r},t)$ interacts most efficiently with fluctuations of the fluid which either have a resonant wave vector $k\sim 1/r_0$ or a resonant frequency $\omega\sim 1/\tau_0$ ($k\sim 1/f_0\tau_0^2$). According to Eq. (25), fluctuations in spatial resonance with the external force are efficiently excited (i.e., are in phase with the external force and the latter pumps them efficiently). Energy moves along the Kolmogorov segment of the spectrum to modes which are in temporal resonance with the external force and which vibrate in counterphase to the external force, i.e., the external force quenches oscillations with $k\sim 1/f_0\tau_0^2$ and gathers energy from these modes. The resulting collective energy sink is an extremely important physical phenomenon. In other words, we can say that the external force acts simultaneously as a source of energy and as a sink for the energy. Thus, when $\gamma\ll 1$ and $\Gamma\gg 1$ dissipation is shut off in the fluid and this should be of some significance in theories of flapping flight and the drag crisis.

From a practical standpoint, there is some importance in the fact that the reduction in drag (drag crisis) associated with the turnoff of viscous dissipation in accordance with Eq. (26) takes place even while the source power is increasing, i.e., without any additional requirements for a change in the structure of the boundary.

Given the significance of the scaling (19) in turning off dissipation in a turbulent fluid, we now clarify which integral of motion this scaling can be associated with, i.e., indicate the quantity which is conserved for an ideal fluid, whose density has a unit of measurement the same as that for f_0. This quantity is the helicity, $G=\int \mathbf{v}\cdot \mathrm{curl}\,\mathbf{v}\,d\mathbf{r}$. Thus, the formation of Eq. (19) and the turn off of dissipation can be associated with fluctuations in the helicity near the viscous range (the so-called I-invariant[19]). To avoid misunderstanding, we note specially that, besides the $E(k)\propto 1/k^2$ spectrum associated with the helicity near the viscous range examined here, a helical spectrum of the form $E(k)\propto 1/k^{7/3}$ associated with a constant helicity flux is often examined in the literature.

The Kolmogorov theorem in the case $\gamma\ll 1$ and $\Gamma\gg 1$ for a two point (velocity, velocity)–velocity correlator is discused in the Appendix.

6.2. Turbulent flows with $\gamma\gg 1$

When $\gamma\gg 1$, Γ always obeys the inequality $\Gamma\gg 1$. It is easy to verify that in this case the specific dissipation power is given by Eq. (26), i.e., the dissipative losses depend on the viscosity.

The physics of a turbulent flow changes fundamentally in this limiting case. Note that for $\gamma\gg 1$, all the inertial range is filled by a helical spectrum and outside the inertial range the velocity-velocity correlator begins to fall off slowly. It is well known, however, that helicity formation in a turbulent flow can lead to spontaneous generation of large scale structures.[8] Thus, it should be expected that for $\gamma\gg 1$ the spontaneous generation of structures begins in a turbulent flow. The slow decrease in the correlator (23) for $r_0\ll r\ll\gamma^{2/5}r_0$ is convincing confirmation of this fundamentally important statement. It is important to note that the generation of large scale structures in turbulent flows does not begin for $\gamma\ll 1$ and $\Gamma\gg 1$, when a helical segment shows up in the spectrum of the turbulent fluctuations, but when the helical spectrum $E(k)\propto 1/k^2$ fills the whole inertial range, i.e., has completely expelled the Kolmogorov spectrum.

This picture, of course, is consistent with existing ideas regarding the role of spatial and temporal resonances in energy transfer between turbulent fluctuations with different scales and an external source (force) that makes the fluid move. From this standpoint, the energy is delivered to fluctuations with a characteristic scale length of order r_0 corresponding to the maximum velocities $(f_0r_0)^{1/2}$ (spatial resonance). The energy delivered to these modes is transferred to long wavelength fluctuations, i.e., it goes into generating large scale structures (inverse cascade), thereby reaching turbulent fluctuations with a characteristic scale length of $\gamma^{2/5}r_0$, which are in a temporal resonance with the external force (have a characteristic time scale of order τ_0). By oscillating in counterphase, fluctuations with a size on the order of $\gamma^{2/5}r_0$ can efficiently deliver energy to the external source; that is, the force which brings the liquid into motion effectively damps their oscillations.

It is interesting to note that the spatial scale length of the resulting structures can substantially (by a factor $\gamma^{2/5}$) exceed the characteristic scale length for the force that brings the fluid into motion. This may have interesting geophysical applications, since it provides a natural mechanism for the formation of hydrodynamic structures with a horizontal scale length greatly exceeding the thickness of the atmosphere (depth of the oceans), while the spatial scale length of the force that brings the air into motion may not exceed the thickness of the atmosphere (depth of the oceans).

7. HIERARCHY OF STATES OF A TURBULENT FLOW

In order to exhibit the physical content of these results, let us consider a simple example. We shall consider a liquid set into motion by a force with amplitude f_0, a spatial scale length r_0, and a characteristic time scale τ_0. Let us see how the structure of the flow changes with varying f_0 when the spatial and temporal scales of the forces are fixed. If the amplitude f_0 of the force is very small, then the fluid flow is laminar and this case is not the subject of this paper. When the amplitude of the force is raised to some level that is low, given the low viscosity of the moving fluid, a transition to turbulence takes place. If the viscosity is low enough, then a transition takes place for $\gamma\ll 1$ and $\Gamma\ll 1$; that is, in this case the Kolmogorov spectrum is adjacent to the viscous range and the dissipated energy is independent of the viscosity and equals the energy flux over the Kolmogorov spectrum, $E(k)\propto 1/k^{5/3}$. As the force amplitude f_0 is increased further, the picture changes: a transition occurs in the region $\gamma \ll 1$ and $\Gamma \gg 1$, i.e., a segment of helical spectrum $E(k) \propto 1/k^2$ grows near the inertial range and this leads to a reduction in the dissipated energy by a factor of $\mathrm{Re}^{1/3}$ compared to the energy flux over the Kolmogorov segment of the spectrum. With further increases in f_0, the helical part of the spectrum expands owing to the narrowing of the Kolmogorov region within the inertial range. Further increases in f_0 lead to a regime with $\gamma \gg 1$, when the entire spectrum in the inertial range is helical, and in the region outside the inertial range large scale structures with scale lengths of $r_0 \ll r \ll \gamma^{2/5} r_0$ begin to be generated.

Therefore, there is an entire hierarchy of turbulent states of the fluid with qualitatively different properties and which succeed one another as the amplitude of the force driving the fluid motion is raised.

8. OTHER APPROACHES TO PARAMETRIZING TURBULENCE

Since the results obtained in this paper rely on the kinetic equation formalism of the theory of strong turbulence, which itself is quite complicated, a question arises as to the extent to which they can be obtained using the traditional formalism of strong turbulence theory. It is important to consider this question because it also allows us to look at the physics of the processes taking place in a turbulent medium from a new point of view.

First, we note that nontrivial physics of turbulent fluid behavior has arisen from an examination of a force with a finite correlation time. In other words, we have taken account of the fact that turbulent fluctuations can be in both spatial and temporal resonance with harmonics of the external force which drives the fluid motion. The resulting turbulence regime is determined by the characteristic scales of these two types of resonant fluctuations. Since the two types of resonant wave vectors divide length space into three regions, exactly three ranges of wave vectors develop, with different kinds of turbulent fluctuations. It turns out that each of these regions corresponds to a distinct dimensional constant which uniquely determines the turbulent fluctuation spectrum. The possibility of such a simple description of the spectra is utterly nontrivial, since the nonlinearity of the Euler equation means that fluctuations on different scales interact significantly with one another and the appearance of universal dimensional parameters that describe the physics on the different scales is quite unexpected.

Let us consider the possibility of obtaining the results of this paper using a chain of equations for the moments. If we start from the very beginning and specify the force correlator by one of the simplified forms, e.g., Eq. (17), (20), or (22), rather than the general expression (2), then the corresponding spectra will follow immediately from dimensional considerations or a more extended scaling formalism, such as that developed in Ref. 8. An approach of this sort, however, does not allow matching of the calculated spectrum with an interval of wave vectors when the external force has the correlator (2). An attempt to directly transfer the method used in this paper to a system of equations for the moments encounters the following difficulty: correlators of the force–force type, $\langle f_i(\mathbf{r}_1,t) f_j(\mathbf{r}_2,t)\rangle$, and velocity–force type, $\langle f_i(\mathbf{r}_1,t) v_j(\mathbf{r}_2,t)\rangle$, arise naturally in the system of equations, but it is not clear from general considerations whether it is required that this kind of correlator be finite when using the scaling technique of this paper. From this standpoint, the advantage of the kinetic equation method lies in the fact that the kinetic equation only contains one random quantity, the external force, while the velocity is simply an argument of the distribution function. In addition, if we assume that the main correlator to which all the others adjust themselves is the force–force correlator, then all the results of this paper can also be obtained by the method of moments. The other advantage of the kinetic equation method compared to the method of moments is the clear interpretation of the zero viscosity limit with finite dissipation permitted by the Euler equation.

9. CONCLUSION

One of the main results of this paper is a proof that the concepts of the universality of developed turbulence in incompressible fluids arising from the classical work of Kolmogorov do need substantial refinement. It turns out that a number of segments of universal character actually appear in the turbulent fluctuation spectrum, but the fact that these appear, and their relative position and extent in wave vector space are determined by the parameters of the source that delivers energy to the fluid and drives it into motion. Thus, the three fundamentally different classes of stationary homogeneous isotropic flows of incompressible fluids examined here will develop. The classification of turbulent flows established here for the first time may be of interest in a number of geophysical and technological problems.

We emphasize, specially, that here we have not considered the radical changes in the Kolmogorov picture of turbulence which might arise owing to intermittency. This analysis has had an entirely different goal: to study the changes in the energy structure of a turbulent flow (in the second velocity correlator) when the finite correlation time of the force driving the fluid is taken into account.

An important new result obtained by taking the finite correlation time of the driving source into account is some nontrivial physics of the energy flux over the spectrum. In particular, under certain conditions (found above), the energy flux flows only over a limited segment of the spectrum which lies entirely within the inertial range. In other words, a situation can arise in which the energy flux does not reach the viscous range, but is repelled to a segment of the spectrum associated with helicity; then, energy is gained by the external force at the boundary of the Kolmogorov and helicity spectra (the corresponding harmonics of the liquid oscillate in counterphase to the external force). The resulting structure of the correlators is consistent with the Kolmogorov theorem. (See the Appendix.)

(Note added in press.) In connection with the results involving the long-wavelength part of the spectrum it should be mentioned that the question of the rule of the Loĭtsyanskiĭ integral was treated previously in Ref. 21.

We thank S. I. Anisimov and É. I. Yurchenko for their interest in this work and for a useful discussion.

This work was supported by the Russian Fund for Fundamental Research (Grant Nos. 98-02-17229 and 98-02-17441) and the Council of the Program for Support of Leading Scientific Schools (Grant No. 96-15-96448).

APPENDIX A

Here we point out some features of the derivation of the Kolmogorov theorem for the case $\gamma \ll 1$. The expression $\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle$ arises in a proof of the Kolmogorov theorem for a two-point (velocity, velocity)-velocity correlator.[20] It transforms to

$$\begin{aligned}\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle &= \langle(\mathbf{f}(\mathbf{r}_2,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle+\langle(\mathbf{f}(\mathbf{r}_1,t)-\mathbf{f}(\mathbf{r}_2,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle\\ &= \langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_1,t))\rangle+\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t)-\mathbf{v}(\mathbf{r}_1,t))\rangle.\end{aligned} \quad (A1)$$

Since we are only interested in points lying within the inertial range, the average in the second term of Eq. (A1) contains an additional small term of the form given by the ratio of $|\mathbf{r}_1-\mathbf{r}_2|$ to the characteristic scale length r_0 for the force. Thus, the second term in Eq. (A1) can be neglected compared to the first, and this immediately leads to the Kolmogorov theorem. Note, however, that this procedure for estimating the various terms on the right hand side of Eq. (A1) is mathematically rigorous only when the averaged function is fixed in sign: when the averaged function changes sign, a small absolute magnitude of the correction to it still does not imply that the contribution of this correction to the average will be small.

We can now show that the standard argument leading to the Kolmogorov theorem requires considerable refinement if a finite source correlation time is taken into account. We emphasize specially that neglecting the second term compared to the first in Eq. (A1) owing to this smallness in the second term is justified only when the first term is not exactly equal to zero and does not contain an additional smallness compared to the second because of some other small parameter. In the limit $\gamma \ll 1$, the argument that usually leads to the Kolmogorov theorem has to be refined, precisely for the following reason: in this case, the smallest parameter of the problem is the correlation time τ_0, so it is necessary to study the dependence of $\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle$ on τ_0 in detail. For this purpose, we write

$$\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle=\lim_{T\to+\infty}\frac{1}{2T}\times\int_{-T}^{T}(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))dt. \quad (A2)$$

Let us write the velocity $\mathbf{v}(\mathbf{r}_2,t)$ in the sum of two terms,

$$\mathbf{v}(\mathbf{r}_2,t)=\mathbf{v}^{(1)}(\mathbf{r}_2,t)+\mathbf{v}^{(2)}(\mathbf{r}_2,t), \quad (A3)$$

where $\mathbf{v}^{(1)}(\mathbf{r}_2,t)$ is the contribution to the velocity owing to harmonics with wave vectors $k \ll k_*$, where $1/r_0 \ll k_* \ll 1/f_0\tau_0^2$, while $\mathbf{v}^{(2)}(\mathbf{r}_2,t)$ is the contribution from all the remaining harmonics.

Thus, we can rewrite Eq. (A3) in the form

$$\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle=\lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T}(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}^{(1)}\times(\mathbf{r}_2,t))dt+\lim_{T\to+\infty}\frac{1}{2T}\times\int_{-T}^{T}(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}^{(2)}(\mathbf{r}_2,t))dt. \quad (A4)$$

Note that $|\mathbf{v}^{(1)}| \gg |\mathbf{v}^{(2)}|$, but the characteristic times associated with the harmonics which appear in $\mathbf{v}^{(1)}$ are substantially longer than the characteristic time scale for the force, τ_0, while $\mathbf{v}^{(2)}$ contains harmonics with characteristic times on the order of τ_0. Thus, the inequality $|\mathbf{v}^{(1)}| \gg |\mathbf{v}^{(2)}|$ does not justify neglecting the second term Eq. (A4) compared to the first because of the different oscillatory properties of the integrands. Thus, the contribution of the harmonics with wave vectors $k \sim 1/f_0\tau_0^2$ to Eq. (A4) may be comparable to the contribution of the long wavelength part.

However, the role of harmonics with $k \sim 1/f_0\tau_0^2$ differs in the cases of $\Gamma \ll 1$ and $\Gamma \gg 1$. The case $\Gamma \gg 1$, where the harmonics with $k \sim 1/f_0\tau_0^2$ lie outside the viscous range, is of greatest interest. In this case it is to be expected that for $|\mathbf{r}_1-\mathbf{r}_2| \gg f_0\tau_0^2$, the major contribution to Eq. (A1) is from long-wavelength harmonics, since at such large distances the short-wavelength fluctuations are not correlated with the external force. For $|\mathbf{r}_1-\mathbf{r}_2| \ll f_0\tau_0^2$, i.e. at distances comparable to the amplitude of the oscillations of a Lagrangian particle acted on by the external force, however, we should expect short wave length fluctuations to make a significant contribution to Eqs. (A4) and (A1). Thus, the standard derivation of the Kolmogorov theorem is valid when a finite correlation time of the external force with $\gamma \ll 1$ and $\Gamma \gg 1$ is included, only if $|\mathbf{r}_1-\mathbf{r}_2| \gg f_0\tau_0^2$. Thus, the Kolmogorov theorem does not prohibit a substantial realignment of the (velocity, velocity)–velocity correlator for $|\mathbf{r}_1-\mathbf{r}_2| \ll f_0\tau_0^2$. According to Eqs. (19) and (20), the second term in Eq. (A4) for $|\mathbf{r}_1-\mathbf{r}_2| < f_0\tau_0^2$ (when it is to be expected that the correlations between the external force and the small scale component of the velocity are substantial) can be approximated by $f_0^2\tau_0$, since $v^{(2)} \sim f_0\tau_0$, which is comparable to the energy flux ϵ over the Kolmogorov segment of the spectrum according to Eq. (25).

In the case $\Gamma \ll 1$, the harmonics with $k \sim 1/f_0\tau_0^2$ fall within the viscous range, so that these fluctuations are suppressed by viscous effects. However, if $v^{(2)} \ll f_0\tau_0$, then the contribution of the short wavelength harmonics with $k \sim 1/f_0\tau_0^2$ to Eq. (A1) is clearly small. Thus, in this case the standard proof of the Kolmogorov theorem encounters no difficulty.

Note that $\langle(\mathbf{f}(\mathbf{r}_1,t)\cdot\mathbf{v}(\mathbf{r}_2,t))\rangle$ can also be represented in the form of a volume average. It can therefore be verified that the external force also interacts significantly with fluctuations that are in spatial resonance with it. Thus, resonant harmonics also have a distinct role in the proof of the Kolmogorov theorem.

*⁾E mail: gord@itp.ac.ru
†⁾E mail: moiseev@mx.iki.rssi.ru

---

[1] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics*, vols. 1 and 2, MIT Press, Cambridge, MA, 1971, 1975 [Russian orig., Vol. 2, Nauka, Moscow (1967)].
[2] D. Foster, D. R. Nelson, and M. J. Stephen, Phys. Rev. A **19**, 732 (1977).
[3] C. DeDominicis and D. C. Martin, Phys. Rev. A **19**, 419 (1979).
[4] K. G. Wilson, Phys. Rev. B **4**, 3175, 3184 (1971).
[5] V. E. Zakharov and V. S. L'vov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **28**, 1470 (1975).
[6] S. S. Moiseev, A. V. Tur, and V. V. Yanovskiĭ, Dokl. Akad. Nauk SSSR **279**, 96 (1984) [Sov. Phys. Dokl. **29**, 926 (1984)].
[7] E. Hopf, J. Rat. Mech. Anal. **1**, 87 (1952).
[8] S. Moiseev and O. Onishenko, Physica B **228**, 83 (1996).
[9] S. N. Gordienko and S. S. Moiseev, JETP Lett. **68**, 204 (1998).
[10] M. Chertkov, G. Falkovich, I. Kolokolov, and V. Lebedev, Phys. Rev. E **52**, 4924 (1995).
[11] R. Z. Sagdeev, ed., *Nonlinear Phenomena in Plasma Physics and Hydrodynamics*, Mir, Moscow (1986).
[12] L. D. Landau and E. M. Lifshitz, *Hydrodynamics* [in Russian], Nauka, Moscow (1986).
[13] S. N. Gordienko, Fiz. Plazmy **25**, 137 (1999) [Plasma Phys. Rep. **25**, 119 (1999)].
[14] S. N. Gordienko, Zh. Éksp. Teor. Fiz. **106**, 436 (1994) [JETP **79**, 241 (1994)].
[15] M. A. Leontovich, Zh. Éksp. Teor. Fiz. **5**, 211 (1935).
[16] S. N. Gordienko and É. I. Yurchenko, JETP Lett. **67**, 668 (1998).
[17] H. Branover, A. Eidelman, M. Nagorny, and M. Kireev, *Progress in Turbulence Research, AIAA* **162**, 64 (1994).
[18] V. Borue and S. Orszag, J. Fluid Mech. **306**, 293 (1995).
[19] A. V. Belyan, S. S. Moiseev, and O. G. Chkhetiani, Preprint No. 1845, IKI RAN, Moscow (1992).
[20] U. Frisch, *Turbulence: the Legacy of A. N. Kolmogorov*, Cambridge University Press, New York (1995).
[21] E. Kuznetsov, A. C. Newell, and V. E. Zakharov, Phys. Rev. Lett. **67**, 3243 (1991).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Surface magnetism of Sc-substituted Ba–M hexaferrites

A. S. Kamzin*)

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted 13 January 1999; resubmitted 21 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1648–1663 (November 1999)

A technique of simultaneous gamma-ray, x-ray, and electron Mössbauer spectroscopy is used to study the magnetic structure of the surface layer with direct comparison to the magnetic structure inside single crystal samples of hexagonal Ba–M ferrites, in which part of the iron ions have been replaced by diamagnetic Sc ions (chemical formula $BaFe_{12-\delta}Sc_{\delta}O_{19}$). It is found that when the diamagnetic Sc ions are introduced into the crystal lattice of $BaFe_{12-\delta}Sc_{\delta}O_{19}$ at concentrations ($x=0.4$ and 0.6) far below the level at which the collinear magnetic structure inside the sample is destroyed, a macroscopic layer of thickness ~ 300 nm develops on the surface, in which the magnetic moments of the iron ions are oriented noncollinearly with respect to the moments inside the sample. The deviation $\langle\theta\rangle$ of the magnetic moments in $BaFe_{11.6}Sc_{0.4}O_{19}$ was $10° \pm 2°$ for $x=0.4$, and when the Sc concentration was raised to 0.6, the angle $\langle\theta\rangle$ increased to $17° \pm 2°$. The noncollinear magnetic structure in the surface layer in these crystals develops because of further reduction in the energy of the exchange interactions owing to the presence of a ''defect,'' such as the surface. For the first time, therefore, an anisotropic surface layer whose magnetic properties differ from those in the interior of a sample has been observed experimentally in ferromagnetic crystals, as predicted by Néel [L. Néel, Phys. Radium. **15**, 225 (1954)]. © *1999 American Institute of Physics.*
[S1063-7761(99)01111-7]

1. INTRODUCTION

Since the beginning of the 1970's, studies of the magnetic properties of crystalline surfaces have attracted ever increasing attention of researchers. This is because of the need to understand the effect of ''defects'' such as surfaces on the magnetic structure and properties of surface layers and the role of surfaces in the formation of the properties of a material. Studies of processes taking place during phase transitions in the surface layer and of the difference between them and phase transitions inside the crystal and the interrelation between them are also important. At present, these studies are acquiring an ever greater practical significance because, for example, the properties of finely dispersed nanosized powders depend substantially on the properties of the crystallite surface. Thus, an understanding of the mechanism for formation, e.g., of the magnetic properties of nanosized crystallites, will open the path to creating magnetic information carriers with ultrahigh recording densities.

The theoretical description of surface anisotropic layers in ferromagnetic materials was given by Néel in 1954.[1] However, the idea that the surface affects the properties of a material was invoked for interpreting experimental data only much later. Thus, it has been proposed[2] that on thin-film Fe, Co and Ni surfaces there is a nonmagnetic (magnetically dead) layer with a thickness of roughly 6 Å.

Finely dispersed powders and thin films have been extensively used in subsequent research on the properties of surfaces. This is because of the increase in the specific surface of a crystallite as its volume is reduced, as well as the absence of experimental techniques capable of distinguishing the signals from a thin surface layer and those from inside a macroscopic crystal. The widespread use of Mössbauer spectroscopy in studies of surface properties for the case of thin films or finely dispersed powders is explained by the possibility of enhancing (or attenuating) the signal from the surface layer by enriching this layer in the isotope ^{57}Fe (or ^{56}Fe). Thus, the unusual experimental fact[3] that the saturation magnetization of a finely dispersed powder is lower than for a macroscopic crystal of the same material has attracted great interest on the part of researchers. Mössbauer studies have shown that a change in the magnetic structure of the crystallites is also the reason for the reduction in the saturation magnetization of the finely dispersed powder.[4] A ''shell'' model[11,12] has been proposed to explain the experimental data obtained from finely dispersed powders of α-Fe_2O_3,[5] γ-Fe_2O_3,[6,7] $CrFe_2O_4$,[8] CrO_2,[9] $NiFe_2O_4$, $Y_3Fe_5O_{12}$, and $Dy_2BiFe_5O_{12}$,[10] and $BaFe_{12}O_{19}$.[11] According to this model the magnetic structure of the interior of a crystallite is analogous to or, perhaps, completely identical to the structure of a macroscopic crystal, while in a thin surface layer (i.e., in the shell) the magnetic moments are not in collinear alignment with the moments in the interior.

However, the shell model is by no means used everywhere to explain the experimental data; other approaches are used, as well. Thus, it has been assumed[13] that on the surface of a $NiFe_2O_4$ particle there is an angular ordering of the spins with a set of stable configurations, which transforms to a spin-glass state as the temperature is lowered. A disordered shell has been assumed to exist in $CoFe_2O_4$ (Ref. 14) and Fe_3O_4 (Ref. 15) crystallites. Elsewhere,[16] it has been assumed that noncollinear alignment of the spins in γ-Fe_2O_3

particles does not occur only on the surface. It has been conjectured[17] that, in order to overcome the volume anisotropy and complete ordering of the magnetic moments along an external magnetic field, substantially higher fields will have to be applied than those used to prove the shell model. On the other hand, data on textured samples of γ-Fe_2O_3 in strong magnetic fields[18] have been explained by the noncollinear ordering of the spins and it has also been shown[18] that the incomplete ordering of the magnetic moments cannot be explained by a large volume magnetic anisotropy, as was assumed earlier.[17]

The use of finely dispersed powders for studying the properties of surfaces has permitted a major step toward understanding the importance of ''surface'' magnetism. But, it should be noted that the nonuniform sizes of a ensemble of particles, superparamagnetic effects, the strong dependence on fabrication techniques, etc., complicate the study of the properties of surfaces using finely dispersed powders. All these difficulties are eliminated when macroscopic crystals are used.

The predicted[1] surface anisotropy was first observed in 1972 by Krinchik, *et al.*[19] in an antiferromagnetic material with weak ferromagnetism. They[19] proposed that macroscopic anisotropy of a surface layer can occur in magnetic materials where, compared to ferromagnetic materials, the energy of the demagnetizing field is small and there is no magnetic anisotropy in the basal plane, which increses the role of surface anisotropy. It is this situation which made it possible to observe a surface anisotropy for the first time in hematite,[19] an antiferromagnetic material with weak ferromagnetism. Based on the experimental data, it was proposed[19] that, within the confines of this surface anisotropic layer, which was referred[19] to as ''transitional,'' the orientation of the magnetic moments varies smoothly from the direction along which the moments are oriented inside the sample to the direction at the surface. Subsequently, an anisotropic surface layer was observed in macroscopic crystals of $FeBO_3$,[20,21] $ErFeO_3$, and $TbFeO_3$,[22] which also have weak ferromagnetism. It was found experimentally that the thickness of the transition surface layer is ~ 500 nm for $FeBO_3$.[20,21]

Unique possibilities for experimental studies of the properties of the surface of macroscopic crystals are offered by the new technique of simultaneous gamma-ray, x-ray, and electron Mössbauer spectroscopy, which makes it possible to extract information about the state of the surface layer and interior of the crystal simultaneously and to compare the results immediately.[23]

Simultaneous gamma, x-ray, and electron Mössbauer spectroscopy yielded the first direct experimental verification of the existence of a transition surface layer with a thickness[24] of ~ 400 nm in macroscopic crystals of Fe_3BO_6 (which, like hematite, are antiferromagnetic with weak ferromagnetic). Layer-by-layer studies employing simultaneous gamma, x-ray, and electron Mössbauer spectroscopy showed[24] that, as the iron ions move toward the crystal surface, the angle of deflection of their magnetic moments relative to their orientation inside the sample increases smoothly within the transition layer. Simultaneous gamma, x-ray, and electron Mössbauer spectroscopy studies of the surface of $Fe_{3-\delta}Ga_{\delta}BO_6$ crystals have shown that replacement of only 9% of the iron atoms by diamagnetic gallium ions increases the thickness of the transition layer by an order of magnitude.[25]

In studies of hexagonal type-M ferrites ($BaFe_{12}O_{19}$, $SrFe_{12}O_{19}$, and $PbFe_{12}O_{19}$) by simultaneous gamma, x-ray, and electron Mössbauer spectroscopy, no transition surface layer with a magnetic structure different from that of the bulk crystal was observed.[26] An analysis of model Mössbauer spectra showed that if a transition surface layer does exist in these type-M hexaferrites, then it cannot be more than a few nm thick. This is of the same order of magnitude as Néel's theoretical estimates[1] and is substantially smaller than the experimental error, which is roughly 10 nm.

Therefore, the existence of a macroscopic surface layer (referred to in Refs. 21 and 22 as transitional) in antiferromagnetic materials with weak ferromagnetism has been convincingly demonstrated. Observing these layers on the surface of hexagonal type-M ferrites will require[26] methods that can analyze surface layers within a few nm.

In this paper we study the magnetic structure of the surface layer in direct comparison with the structure inside macroscopic crystals of hexagonal type-M ferrites in which part of the iron ions have been replaced by diamagnetic ions. In stating the problem, we assumed that it was possible to observe a transitional surface layer experimentally in the substituted hexaferrites. Here is why: first, it is known[27–29] that replacing a large fraction of the iron ions in type-M hexaferrites by diamagnetic In, Sc, Ga, or Al ions gives rise to a noncollinear magnetic structure inside the crystal. Second, it has been shown[25] that replacing only 9% of the iron ions in Fe_3BO_6 by diamagnetic Ga ions increases the thickness of the transition surface layer by an order of magnitude. This is because the inter-sublattice exchange bonds are weakened, owing both to the introduced diamagnetic ions and to the presence of a surface. For this reason, it was possible to assume that if a small part of the iron ions are replaced by diamagnetic ions (when a collinear magnetic structure is retained inside the sample), then a macroscopic surface layer with noncollinearly ordered magnetic moments can develop on the surface of macrocrystals of the hexagonal ferrites. Preliminary studies[30] showed replacing a small number of the magnetic iron ions in the structure of the hexagonal ferrite Sr–M by diamagnetic Al ions (chemical formula $SrFe_{10.2}Al_{1.8}O_{19}$) gives rise to a layer with a thickness of $\sim 200\,\mu$m on the surface where the magnetic moments are noncollinear with those inside the sample.

For these studies we have chosen a hexagonal Ba-M ferrite in which part of the iron ions are replaced by diamagnetic Sc ions; the chemical formula is $BaFe_{12-\delta}Sc_{\delta}O_{19}$. It has been shown[27–29] that for substitutions $x<1.2$, the collinearity of the moments inside the crystal is preserved, while for $x>1.2$ the collinearity of the moments is disrupted inside the crystal. Thus, for concentrations $x=0.4$ and $x=0.6$ of the diamagnetic ions, i.e., in crystalline $BaFe_{11.4}Sc_{0.4}O_{19}$, and $BaFe_{11.4}Sc_{0.6}O_{19}$, we could be sure that these amounts of scandium ions were well below the amount required to form a noncollinear structure inside the sample so we could study the effect of this substitution on the properties of a thin surface layer.

2. MÖSSBAUER STUDIES OF MAGNETIC STRUCTURE IN CRYSTALS

Mössbauer spectroscopy is widely used in studies of the magnetic structure of materials because the magnetic moment M of the iron ion is always oriented antiparallel to the effective magnetic field H_{eff} at the nucleus. If the energy of the electrical interaction in the crystal is negligibly low and the levels are split owing to the magnetic hyperfine interaction, then the intensities of the components of the Zeeman sextuplet $A_1:A_2:A_3:A_4:A_5:A_6$ in the Mössbauer spectrum depend on the angle θ between the direction of the wave vector of the gamma-rays and the orientation of the effective magnetic field H_{eff} at the nucleus of the iron ion (or the magnetic moment of the iron ion) as follows:

$$3(1+\cos\theta):4\sin\theta:(1+\cos\theta):(1+\cos\theta):4\sin\theta:4(1+\cos\theta). \tag{1}$$

Therefore, the ratio of the second or fifth ($A_{2,5}$) lines of the Zeeman sextuplet corresponding to transitions with $\Delta m=0$ to the intensity of the first or sixth line ($A_{1,6}$) in the Mössbauer spectrum of a single crystal,

$$A_{2,5}/A_{1,6}=4\sin^2\theta/\{3(\cos^2\theta+1)\}, \tag{2}$$

can be used to determine the angle θ which specifies the orientation of the magnetic moments with respect to the gamma ray beam:

$$\theta=\arccos\left(\frac{4A_{1,6}-3A_{2,5}}{4A_{1,6}+3A_{2,5}}\right)^{1/2}=\arcsin\left(\frac{(3/2)A_{2,5}/A_{1,6}}{1+(3/4)A_{2,5}/A_{1,6}}\right)^{1/2}. \tag{3}$$

If the angle θ varies in the sample, then Eqs. (2) and (3) yield an average $\langle\theta\rangle$. Thus, Mössbauer spectroscopy can be used to determine the orientation of the magnetic moments in a crystal to rather high accuracy.

3. SIMULTANEOUS GAMMA, X-RAY, AND ELECTRON MÖSSBAUER SPECTROSCOPY

The simultaneous gamma, x-ray, and electron Mössbauer spectroscopy technique proposed and described in Ref. 23 was used for the measurements. Simultaneous gamma, x-ray, and electron Mössbauer spectroscopy involves simultaneous detection of the Mössbauer spectra of radiation with different mean free paths in the material, specifically, gamma-rays, characteristic x-rays, and conversion and Auger electrons, which carry information on the properties of the crystal within its bulk, in a surface layer several μm thick, and in a surface layer with a thickness of ~ 300 nm, respectively. When a gamma-photon is absorbed an iron atom enters an excited state, and the reverse transition to the ground state is accompanied by the emission of a characteristic x-ray, as well as of conversion and Auger electrons. In simultaneous gamma, x-ray, and electron Mössbauer spectroscopy, the spectra of these x-rays and electrons are measured in a backscattering geometry, together with the spectrum of the gamma-rays transmitted through the sample. The energy of an electron leaving the sample is lower when it leaves an atom that lies deeper in the sample. Simple proportional counters could be used[32–34] for energy resolution of the electrons. Of course, the accuracy of the analysis of the layers in terms of thickness is much poorer than with electric or magnetic separators,[35] but in a number of cases this accuracy is entirely sufficient, so that simultaneous gamma, x-ray, and electron Mössbauer spectroscopy using a proportional detector has been employed for layer-by-layer analysis of the surface layers down to thicknesses of less than 300 nm.[36]

Therefore, simultaneous gamma, x-ray, and electron Mössbauer spectroscopy offers unique prospects for studying the state of the surface layer and interior of macroscopic crystals because of the following. First, the simultaneity of the tests means that the experimental conditions are the same for the surface layers and interior of the crystal. Second, simultaneous gamma, x-ray, and electron Mössbauer spectroscopy uses a single technique (the Mössbauer effect), so the experimental data on the properties of the surface layers and interior of a crystal can be compared directly.

Simultaneous gamma, x-ray, and electron Mössbauer spectroscopy was carried out using the automatic system of which a block diagram is shown in Fig. 1. The various radiations, specifically gamma rays, x rays, and conversion and Auger electrons are detected by the counters Γ, X, and E, respectively in a universal three chamber detector. The signals from the counters were amplified, then discriminators set the thresholds for distinguishing the corresponding energies, and the resulting signals were sent to the corresponding memory buffers for the Mössbauer spectra.

4. EXPERIMENTAL RESULTS

The single crystal $BaFe_{12-\delta}Sc_{\delta}O_{19}$ hexaferrite samples for these studies were synthesized by spontaneous crystallization from a solution in a melt of $NaFeO_2$. The chemical composition and identification of these crystals as ferrites with a type-M hexagonal structure were verified by x-ray analysis, chemical analysis, and the structure of the Mössbauer spectra. The degree of substitution of the iron ions was also determined from the dependence of the Curie temperature on the concentration of Sc. The Curie temperature was determined from the temperature dependences of the effective magnetic fields and by taking temperature scans with a fixed velocity of the gamma-ray source.

For the Mössbauer measurements, slabs in the form of disks with thicknesses of $\sim 80\,\mu$m and diameters ~ 8 mm were cut from the single crystals. The crystallographic C axis was oriented perpendicular to the plane of the slabs. Special attention was devoted to the surface quality of the crystals. Single crystals with a natural mirror facet were selected for studying the surface properties and subjected chemical polishing by etching for one minute in orthophosphoric acid at a temperature of 90 °C. Slabs of the ferrite $BaFe_{12}O_{19}$ without Sc ion substitution were prepared simultaneously by this technique for use as control samples.

Simultaneous gamma, x-ray, and electron Mössbauer

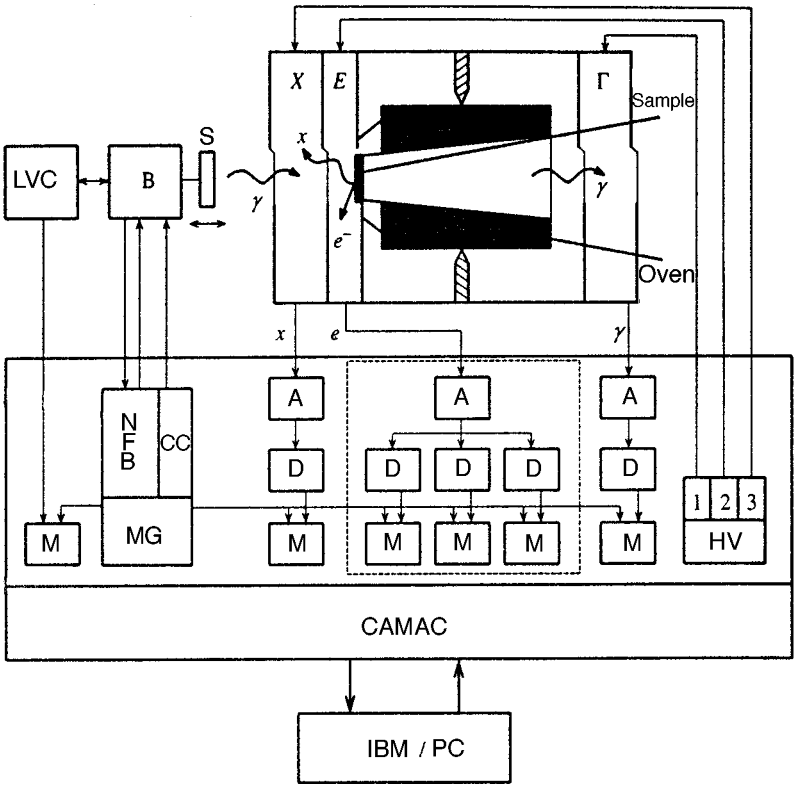

FIG. 1. A block diagram of the automated system for simultaneous gamma-ray, x-ray, and electron Mössbauer spectroscopy. Items indicated in the figure include: Γ, X and E, the detectors for, respectively, gamma rays, characteristic x rays, and secondary electrons, S the gamma-ray source, the sample, oven (heater), B the Doppler modulator for the source motion, MG the motion generator, LVC a laser velocity calibrator, A amplifiers, D discriminators, M memory buffers, HV high voltage supplies, NFB negative feedback channel, and CC computer correction channel.

spectra were obtained experimentally at temperatures from 300 to 750 K. Figures 2, 3, and 4 show examples of spectra of the ferrites $\mathrm{BaFe}_{12-\delta}\mathrm{Sc}_{\delta}\mathrm{O}_{19}$ with $x=0.4$, 0.6, and 0 taken at room temperature with detection of gamma rays, as well as of conversion and Auger electrons. The Mössbauer spectra taken with the x rays carry information from a layer with a thickness of a few μm and are analogous to those taken with the gamma rays. Thus, we do not show the spectra taken in the x ray region in order to make the figures more clearly understandable. As can be seen from Figs. 2–4, the spectral lines are well resolved. This makes it possible to determine the parameters of the hyperfine interactions, as well as the orientations of the magnetic moments in the crystal, with high accuracy.

As Figs. 2(a), 3(a), and 4(a) show, the Mössbauer spectra of the ferrites with added Sc obtained by detecting the gamma rays are similar to the unsubstituted Ba-M ferrite, except that the spectra of $\mathrm{BaFe}_{11.6}\mathrm{Sc}_{0.4}\mathrm{O}_{19}$ [Fig. 2(a)] and $\mathrm{BaFe}_{11.4}\mathrm{Sc}_{0.6}\mathrm{O}_{19}$ [Fig. 3(a)] contain additional well resolved lines corresponding to the $12k'$ sublattice. An analysis of the gamma resonance spectra [Figs. 2(a), 3(a), and 4(a)] showed that in the Zeeman sextuplets of each nonequivalent position, the intensities of the second and fifth lines, corresponding to transitions with $\Delta m=0$, are zero. This means that the angle θ equals zero and, therefore, that the magnetic moments of the iron ions occupying sites within the crystal are collinear with the wave vector of the gamma rays and the crystallographic C axis. This pattern was observed over the entire temperature range studied and is in good agreement with data on the bulk properties of these crystals.[37]

The sections corresponding to velocities of ±4 and ±5 mm/s in the spectra with detection of conversion and Auger electrons [Figs. 2(b) and 3(b)] contain low intensity lines that are clearly absent in the gamma-ray spectra [Figs. 2(a), 3(a), and 4(a)]. It should be noted that the intensities of these lines in the spectra of Fig. 2(b) are roughly a factor of 3 higher than the noise, and for the spectra of Fig. 3(b), the signal-to-noise ratio is still higher. An analysis of the spectra of the conversion and Auger electrons showed that these additional lines are the second and fifth components of the Zeeman sextuplets. This means that the magnetic moments of the iron ions occupying positions in a surface layer of thickness ~ 200 nm are oriented at some angle θ to the direction of the gamma-ray wave vector and, therefore, to the C axis. This pattern is observed up to temperatures of ~ 600 K, above which it is difficult to analyze the spectra because of the poor resolution of the lines. The deviation of the magnetic moments from the wave vector of the gamma rays $\langle\theta\rangle$ calculated using Eq. (3) from the spectra of the $\mathrm{BaFe}_{12-\delta}\mathrm{Sc}_{\delta}O_{19}$ was $10°\pm 2°$ for $x=0.4$. When the concentration of Sc was raised to 0.6, the angle θ increased to 17° ±2°, according to an analysis of the spectrum shown in Fig. 3(b). It should be noted that in these experiments it was not possible to determine the area within which the magnetic moments are deflected. The main result of these studies is the experimental fact that the magnetic moments of the iron ions within a surface layer of thickness ~ 200 nm are noncollinear with the crystallographic C axis, along which the magnetic moments of the iron ions occupying positions in the interior of the crystal are aligned.

In order to verify the correctness of this analysis of the Mössbauer spectra, we took spectra on single crystals that were inclined so that the C axis formed an angle α with the wave vector of the gamma rays. Figures 2(c) and 3(c) show

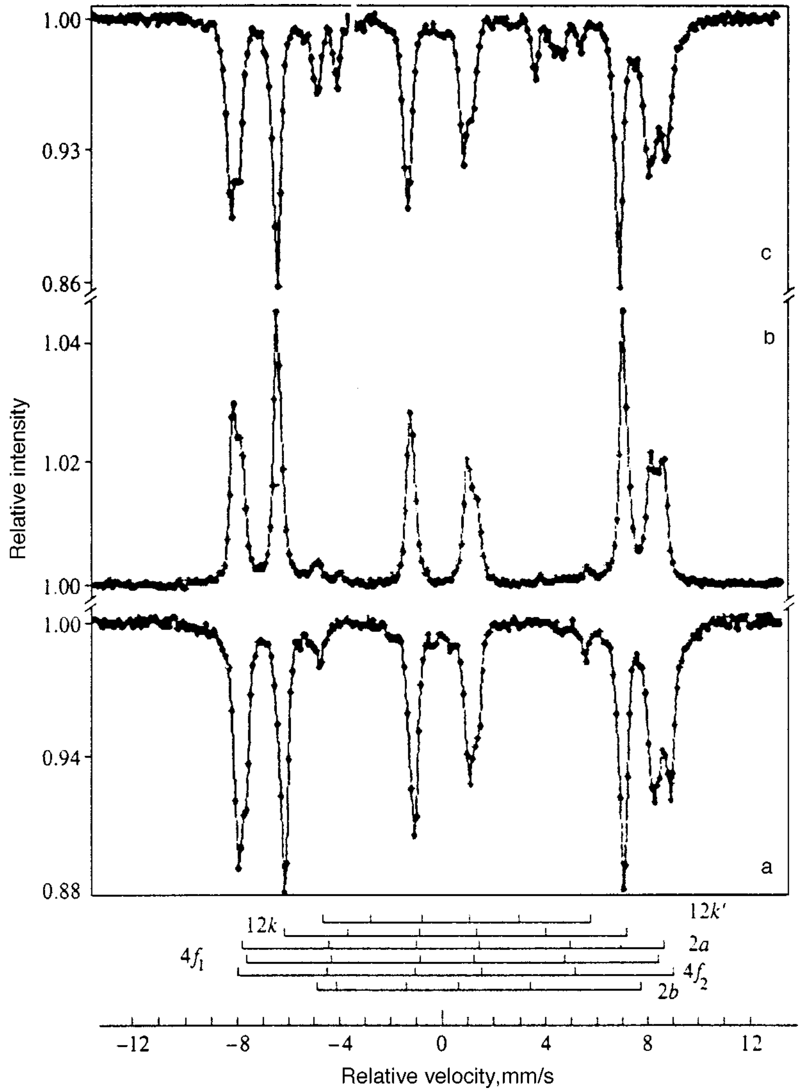

FIG. 2. Room temperature Mössbauer spectra of single crystal hexaferrite $BaFe_{11.6}Sc_{0.4}O_{19}$: a and c with detection of gamma rays carrying information from the volume of the crystal, b with detection of secondary electrons from a surface layer extending from 0 to 200 nm. For a and b the wave vector of the gamma rays is parallel to the crystallographic *C* axis, while for c the wave vector of the gamma rays is directed at an angle of 28° to the crystallographic *C* axis.

examples of gamma resonance spectra obtained for $\alpha = 28° \pm 2°$. These figures show that when the magnetic moments deviate from the propagation direction of the gamma rays, lines of the Zeeman sextuplets corresponding to transitions with $\Delta m = 0$ show up in the spectrum. The angle $\langle \theta \rangle$ calculated from the gamma-ray spectra shown in both Figs. 2(c) and 3(c) using Eq. (3) was $27° \pm 2°$. Therefore, to within the limits of error of the apparatus, the angle of inclination of the crystal, α, and the calculated value of θ were in good agreement. Note that a comparison of the experimental spectra of Figs. 2(b) and 3(b) reveals good agreement between the positions on the velocity scale of the second and fifth lines of the gamma resonance spectra taken with an inclined crystal [Fig. 2(c) and 3(c)] and the second and fifth lines of the spectra obtained with detection of conversion and Auger electrons [Figs. 2(b) and 3(b)].

These results prove convincingly that, when diamagnetic Sc ions are introduced into Ba-M crystals, a macroscopic surface layer with a thickness of ~ 200 nm develops, within which the magnetic moments of the iron ions deviate from the direction of the crystallographic *C* axis and from the direction of the spin moments of the iron ions which occupy sites in the interior of the crystal.

It might be assumed that the observed deviation of the magnetic moments is caused by ‘‘etching’’ of magnetic ions from the surface layer during chemical polishing of the crystals that further reduces the exchange interaction energy in this layer. In order to test this hypothesis on single crystals of $BaFe_{12}O_{19}$ prepared simultaneously with the $BaFe_{12-\delta}Sc_{\delta}O_{19}$ by the same chemical polishing technique, we took the Mössbauer spectra shown in Fig. 4. As can be seen in Fig. 4, when the crystallographic *C* axis is oriented parallel to the wave vector of the gamma rays, the second and fifth lines of the Zeeman sextuplets are absent in the spectra in which either electrons [Fig. 4(b)] or gamma rays [Fig. 4(a)] were detected. As a comparison, Fig. 4(c) also shows the Mössbauer spectrum obtained with detection of gamma radiation on the same single crystal slab of $BaFe_{12}O_{19}$ inclined so that its *C* axis was oriented at an angle of $\alpha = 28° \pm 2°$ to the wave vector of the gamma rays. Figure 4(c) shows that the deviation in the direction of the magnetic moments from the propagation direction of the gamma rays has caused the lines from the Zeeman sextuplet corresponding to transitions with $\Delta m = 0$ to appear. Equation (3) gives $\theta = 30° \pm 2°$, in good agreement with the angle α set in the experiment. Thus, these experiments show convincingly that

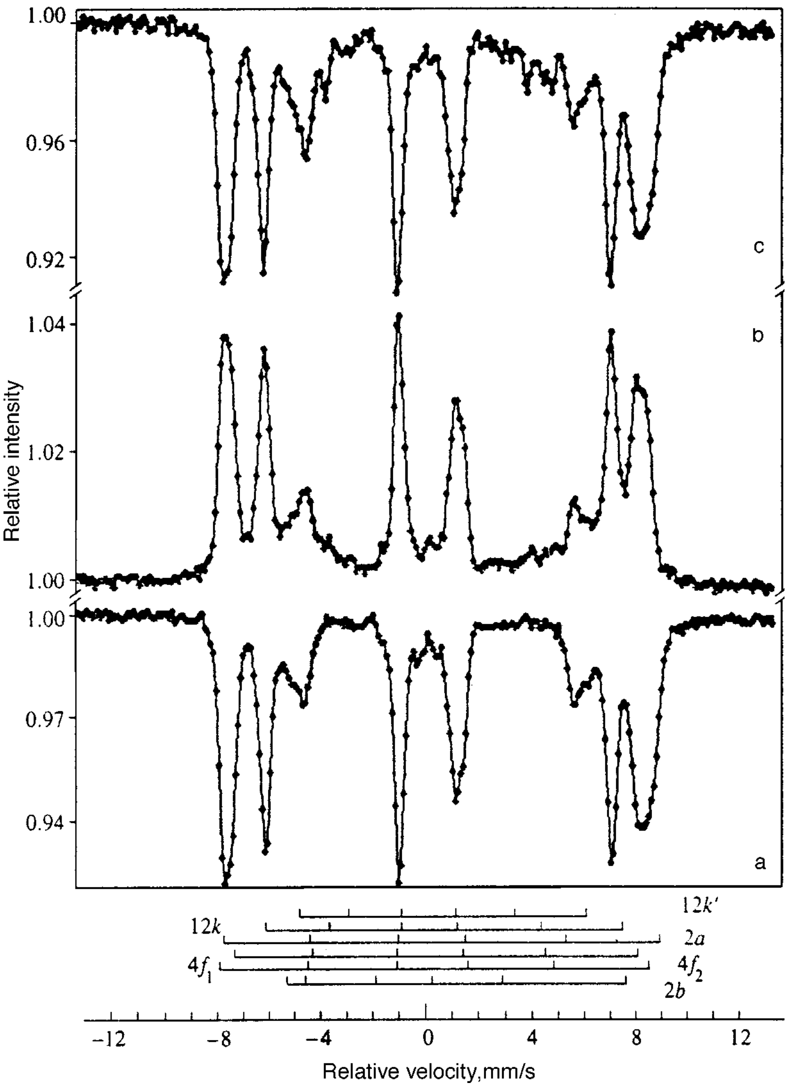

FIG. 3. Room temperature Mössbauer spectra of single crystal hexaferrite $BaFe_{11.6}Sc_{0.6}O_{19}$: a and c with detection of gamma rays carrying information from the volume of the crystal, b with detection of secondary electrons from a surface layer extending from 0 to 200 nm. For a and c the wave vector of the gamma rays is parallel to the crystallographic C axis, while for b the wave vector of the gamma rays is directed at an angle of 28° to the crystallographic C axis.

our technique for preparing the sample surfaces does not cause any changes in the magnetic structure of the surface layer. This experiment also is an additional confirmation of the results of Ref. 26, where it was shown that, to within the experimental error of roughly 10 nm, there is no surface layer on $BaFe_{12}O_{19}$ crystals with a magnetic structure different from that which exists in the interior of the crystal.

Let us consider the reasons for the deviation in the orientation of the magnetic moments in the surface layer from their direction in the interior of the crystal. The main orientational influence on the direction of the magnetic moments of individual sublattices[37] is exerted by exchange interactions in the hexagonal block R between $Fe(2b)$–O–$Fe(4f_2)$ and $Fe(4f_2)$–O–$Fe(12k)$, which have fairly large bond angles ($\sim 140°$ and 130°, respectively). The corresponding Fe–O distances are approximately 1.8 and 1.95 Å. The interactions correspond to the highest values of the exchange integrals. Since the $Fe(2b)$–O–$Fe(4f_2)$–O–$Fe(12k)$ interaction is somewhat stronger the chain consists of two exchange bonds, the spins of the $Fe(2b)$ and $Fe(12k)$ ions are oriented antiparallel to the spin of the $Fe(4f_2)$ ion, despite the strong opposing interaction between the $Fe(2b)$ and $Fe(12k)$ ions, for which the interaction angle is $\sim 125°$, and the Fe–O distances are 2.3 and 1.87 Å, respectively. In the spinel block, the interaction takes place in the usual manner, so that the mutual orientation of the magnetic moments is as shown in Table I. Iron ions in the $2b$-position play a special role in the formation of such an axial structure, as well as of the magnetocrystalline anisotropy. This position has a strong intracrystalline field with a trigonal symmetry where the axis of symmetry coincides with the crystal axis. The importance of the contribution to the magnetic anisotropy of the iron ions occupying $12k$ positions in the low symmetry octahedron (see Table I) has been noted previously.[38]

Neutron diffraction and Mössbauer studies of type-M hexagonal ferrites have shown[27–29] that when iron ions in these ferrites are replaced by diamagnetic Sc ions, there a substantial change in the exchange interaction, even in compounds with a relatively small number of Sc ions. When the amount of Sc ions, $x > 1.2$ a noncollinear magnetic structure appears.[28] In the case of the Sc-substituted ferrite with $x = 1.8$, the magnetic moments of the blocks form a conical block helicoid, within each block of which the magnetic moments of the iron ions are collinear.[28]

The resonant absorption probabilities (Mössbauer effect) listed in Table I were calculated from the area under the lines

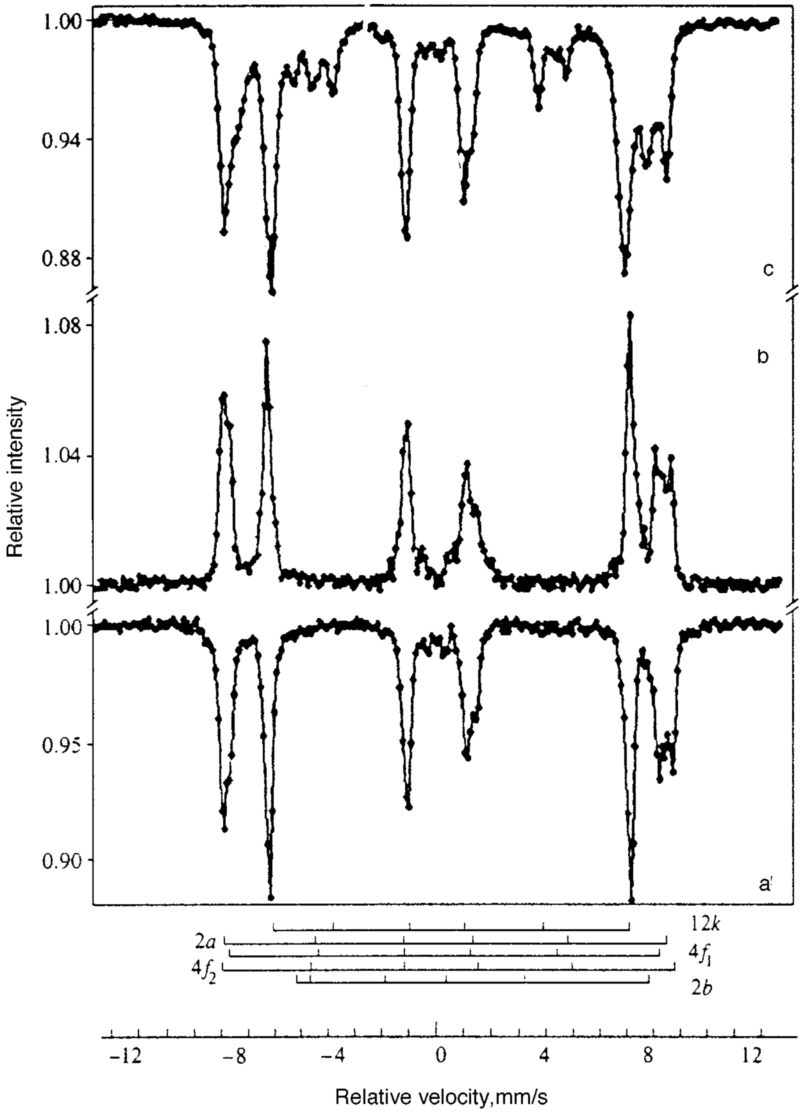

FIG. 4. Room temperature Mössbauer spectra of single crystal hexaferrite $BaFe_{12}O_{19}$: a and c with detection of gamma rays carrying information from the volume of the crystal, b with detection of secondary electrons from a surface layer extending from 0 to 200 nm. For a and b the wave vector of the gamma rays is parallel to the crystallographic *C* axis, while for c the wave vector of the gamma rays is directed at an angle of 28° to the crystallographic *C* axis.

in the experimental Mössbauer spectra. Table I implies that, to within the experimental error, the resonant absorption probabilities for $BaFe_{12}O_{19}$ are in good agreement with the number of ions in the corresponding sublattices of the unit cell of the crystal. The hyperfine interaction parameters obtained from the experimental Mössbauer spectra of

TABLE I. Distribution of Fe^{3+} ions over a unit cell, the surroundings and orientation of their spins in a type-M hexagonal ferrite, and the magnitude of the Mössbauer effect (area under the corresponding spectrum lines) for the hexaferrites $BaFe_{12}O_{19}$, $BaSc_{0.4}Fe_{11.6}O_{19}$, and $BaSc_{0.6}Fe_{11.4}O_{19}$ at room temperature.

	Distribution of Fe^{3+} ions over the unit cell, surroundings, and spin orientations			Mössbauer effect		
				$BaFe_{12}O_{19}$	$BaSc_{0.4}Fe_{11.6}O_{19}$	$BaSc_{0.6}Fe_{11.4}O_{19}$
Sub-lattices	Number of ions	Surroundings	Direction of spins	Area, % (Normalized to 24)	Area, % (Normalized to 23.2)	Area, % (Normalized to 22.8)
12*k*	12	octa	up	12.4±0.2	9.2±0.2	7.3±0.3
12*k*′				-	0.7±0.2	4.4±0.4
$4f_1$	4	tetra	down	4.9±0.4	4.1±0.3	5.4±0.4
$4f_2$	4	octa	down	3.9±0.2	5.8±0.4	3.6±0.4
2*a*	2	octa	up	1.4±0.6	2.2±0.4	1.1±0.5
2*b*	2	trigonal bipyramid	up	1.4±0.2	1.1±0.3	0.9±0.2

TABLE II. Effective magnetic fields H_{eff}, isomeric shifts δ, and quadrupole splittings ΔE for $BaSc_xFe_{12-x}O_{19}$ at room temperature (the isomeric shift δ is taken relative to α-Fe).

Sublattices	H_{eff}, kOe		δ, mm/s		ΔE, mm/s	
$x=0$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$
$12k$	416 ± 1	422 ± 1	0.34 ± 0.01	0.35 ± 0.01	0.42 ± 0.01	0.42 ± 0.02
$4f_1$	493 ± 1	495 ± 1	0.27 ± 0.01	0.29 ± 0.02	0.19 ± 0.02	0.09 ± 0.04
$4f_2$	520 ± 3	523 ± 1	0.38 ± 0.01	0.35 ± 0.02	0.27 ± 0.01	0.20 ± 0.04
$2a$	508 ± 3	508 ± 3	0.33 ± 0.01	0.31 ± 0.02	0.16 ± 0.02	0.11 ± 0.04
$2b$	403 ± 5	401 ± 5	0.30 ± 0.02	0.25 ± 0.10	2.07 ± 0.04	1.84 ± 0.20
$x=0.4$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$
$12k$	416 ± 1	422 ± 1	0.36 ± 0.01	0.35 ± 0.01	0.42 ± 0.02	0.43 ± 0.02
$12k'$	324 ± 1	329 ± 1	0.37 ± 0.01	0.35 ± 0.02	0.42 ± 0.02	0.41 ± 0.04
$4f_1$	495 ± 1	488 ± 1	0.29 ± 0.01	0.30 ± 0.01	0.25 ± 0.02	0.16 ± 0.03
$4f_2$	519 ± 1	519 ± 1	0.46 ± 0.01	0.35 ± 0.01	0.32 ± 0.02	0.28 ± 0.02
$2a$	513 ± 3	514 ± 1	0.31 ± 0.02	0.30 ± 0.03	0.10 ± 0.04	0.18 ± 0.06
$2b$	402 ± 5	-	0.41 ± 0.05	-	2.05 ± 0.10	-
$x=0.6$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$	$[\gamma]$	$[e]$
$12k$	413 ± 1	414 ± 1	0.37 ± 0.01	0.38 ± 0.01	0.41 ± 0.02	0.39 ± 0.03
$12k'$	325 ± 1	327 ± 2	0.38 ± 0.01	0.35 ± 0.03	0.41 ± 0.03	0.36 ± 0.06
$4f_1$	486 ± 1	486 ± 1	0.31 ± 0.01	0.30 ± 0.02	0.24 ± 0.02	0.24 ± 0.04
$4f_2$	512 ± 1	514 ± 1	0.43 ± 0.01	0.33 ± 0.02	0.32 ± 0.02	0.41 ± 0.04
$2a$	505 ± 2	-	0.33 ± 0.01	-	0.12 ± 0.02	-
$2b$	406 ± 3	-	0.43 ± 0.02	-	2.03 ± 0.05	-

$BaFe_{12-\delta}Sc_{\delta}O_{19}$ for $x=0$, 0.4, and 0.6 are shown in Table II. The values listed in Table II are in good agreement with published data.[37,39,40] A comparison of the hyperfine interaction parameters for $BaFe_{12}O_{19}$ and $BaFe_{11.4}Sc_{0.6}O_{19}$ shows that the variation in these parameters for such an apparently negligible replacement of iron ions by Sc ions as $x=0.6$ is substantial, which is indicative of a selective positioning of the Sc ions.

The experimental spectra and the tables show that the behavior of the Sc ions gives rise to a $12k'$ sublattice and a drop in the intensities of the lines corresponding to iron ions in $2b$ and $12k$ positions. The formation of a $12k'$ sublattice can be explained by the fact that Sc ions occupy $2b$ positions. Iron ions at the $12k$ sites have six exchange interactions: three with iron ions at $4f_1$ positions, two with Fe ions at $4f_2$ positions, and one with Fe in the trigonal bipyramid. The appearance of diamagnetic Sc ions in the $2b$ sites means that part of the Fe ions in the $12k$ positions lose the $\mathrm{Fe}(2b)$–O–$\mathrm{Fe}(12k)$ exchange coupling, while another part of the $\mathrm{Fe}(12k)$ ions retains this coupling, so that a nonequivalent situation develops. The ratio of the numbers of ions in the $12k$- and $12k'$-positions will be proportional to the number of magnetic and nonmagnetic ions in $2b$-positions.

Table I implies that the numbers of iron ions in $12k$ and $12k'$ positions for $BaFe_{11.4}Sc_{0.6}O_{19}$ are roughly in the ratio of 2:1. At the same time, if the Sc ions occupy only $2b$-positions, then for the ferrite of this composition with $x=0.6$, in the $2b$ sublattice the ratio of the number of magnetic ions to nonmagnetic ions should, as discussed above, be 2:3, which is not really so. (See Table I.) Therefore, iron ions are not displaced by Sc ions only at $2b$ positions, but also at other positions. As Table I implies, for a given x there is roughly equal replacement of iron ions with opposite orientations of their magnetic moments, i.e., both iron ions at $2b$ sites upward orientations of their spin moments and iron ions with downward spins, which occupy $4f$ sites, are replaced.

These data are consistent with the conclusions of Refs. 27–29, based on neutron diffraction data, that for low concentrations, the Sc ions are located at $2b$ positions and for a degree of substitution of less than 30%, $4f_2$ positions are occupied in addition to $2b$ positions. The increase in the intensity of the lines corresponding to Fe ions in $4f_1$ positions (Table I) should be attributed to the large error in the calculations of the poorly resolved sextuplets for iron ions in $4f_2$ and $2a$ positions. Localization of Sc ions in $2b$ positions weakens the exchange bonds between the S and R blocks, so as the Sc content is changed by $\Delta x=1$, the Curie temperature falls by 160–170°, while when Al or Ga ions are introduced, the Curie temperature (T_C) falls by 50–60° when the substitution is raised by $\Delta x=1$. We have obtained values of the Curie temperature from the experimental data. For the ferrite with $x=0.6$, our value of T_C was 647 K, in good agreement with other published data.[40]

To summarize the above discussion, in the crystals we have chosen, $BaFe_{12-\delta}Sc_{\delta}O_{19}$ for $x=0.4$ or 0.6, diamagnetic Sc ions displace iron ions in $2b$- and $4f_2$-positions and, on forming magnetic bonds, facilitate the formation of a noncollinear magnetic structure. However, the amount of Sc ions ($x=0.4$ or 0.6) is far from high enough to destroy the collinearity in the volume.[27–29] In the surface layer of these crystals, the exchange interaction energy is reduced, both by the diamagnetic ions, and by the presence of the surface. Thus, for example, it has been shown[25] that replacement of only 9% of the iron ions in Fe_3BO_6 by diamagnetic Ga ions increases the thickness of the ''transition'' surface layer by almost a factor of 10.

It has, therefore, been found experimentally for the first time that, when diamagnetic ions are introduced into hexagonal Ba-M ferrite crystals, a macroscopic surface layer with a thickness of ∼200 nm is formed, within which the magnetic moments of the iron ions are noncollinear with those of the iron ions at sites in the interior of the crystal. Thus, for the first time we have experimentally observed a surface layer in ferrites whose magnetic structure differs from that inside the crystal and whose existence was predicted theoretically by Néel in 1954.[1]

The author thanks V. L. Rozenbaum for help in making the measurements.

This work was supported by the Russian Fund for Fundamental Research (Grant Nos. 96-02-10038 and 98-02-18279).

*)E-mail: kamzin@kas.ioffe.rssi.ru

---

[1] L. Néel, J. Phys. Radium **15**, 225 (1954).

[2] L. Liebermann, D. R. Fridkin, and H. B. Shore, Phys. Rev. Lett. **22**, 539 (1969); L. Liebermann, J. Clinton, D. M. Edwards, and J. Mathon, Phys. Rev. Lett. **25**, 232 (1970).

[3] A. E. Berkowitz, W. J. Schuele, and P. J. Flanders, J. Appl. Phys. **39**, 1261 (1968).

[4] J. M. D. Coey, Phys. Rev. Lett. **27**, 1140 (1971); Can. J. Phys. **65**, 1210 (1987).

[5] A. M. van der Kraan, Phys. Status Solidi A **18**, 215 (1973).

[6] A. H. Morrish, K. Haneda, and P. J. Schurer, J. de Phys. Colloque C 6, **37**, C6-301 (1976).

[7] P. M. de Bakker, E. DeGrave, R. E. Vandenberghe, and L. H. Bowen, Hyperfine Interact. **54**, 493 (1990).

[8] A. E. Berkowitz, J. A. Lahut, and C. E. van Buren, IEEE Trans. Magn. **MAG-16**, 184 (1980).

[9] K. Haneda, H. Kojima, A. H. Morrish, P. J. Picone, and K. Wakai, J. Appl. Phys. **53**, 2686 (1982).

[10] A. H. Morrish and K. Haneda, IEEE Trans. Magn. **MAG-25**, 2597 (1989); J. Appl. Phys. **52**, 2496 (1981).

[11] K. Haneda and A. H. Morrish, Nucl. Instrum. Methods, Phys. Res. B **76**, 132 (1993).

[12] K. Haneda, Can. J. Phys. **65**, 1233 (1987).

[13] R. H. Kodama, A. E. Berkowitz, E. J. McNiff Jr.,; and S. Foner, J. Appl. Phys. **81**, 5552 (1997).

[14] D. Lin, A. C. Nunes, C. F. Majkrzak, A. E. Berkowitz, and M. B. Maple, J. Magn. Magn. Mater. **45**, 343 (1995).

[15] S. S. Parkin, R. Sigsbee, R. Felici, and G. P. Felsher, J. Appl. Phys. **57**, 1371 (1985).

[16] F. T. Parker, M. W. Foster, D. Margulis, and A. E. Berkowitz, Phys. Rev. B **47**, 7885 (1993).

[17] Q. A. Pankhurst and P. J. Pollard, Phys. Rev. Lett. **67**, 325 (1991).

[18] P. V. Hendrilsen, S. Linderoth, C. A. Oxborrow, and S. Morup, J. Phys.: Condens. Matter **6**, 3091 (1994).

[19] G. S. Krinchik, A. P. Khrebtov, A. A. Askochenskiĭ, and V. E. Zubov, JETP Lett. **17**, 345 (1973); G. S. Krinchik and V. E. Zubov, Zh. Éksp. Teor. Fiz. **69**, 707 (1975 [Sov. Phys. JETP **42**, 359 (1975)].

[20] V. G. Labushkin, V. V. Rudenko, Yu. R. Sarkisov, V. A. Sarkisyan, and V. N. Seleznev, JETP Lett. **34**, 544 (1981).

[21] V. E. Zubov, G. S. Krinchik, V. N. Selznev, and M. B. Strugatskiĭ, Zh. Éksp. Teor. Fiz. **94**, 290 (1988) [Sov. Phys. JETP **67**, 2122 (1988)]; V. E. Zubov, G. S. Krinchik, V. N. Seleznev, and M. B. Strugatsky, J. Magn. Magn. Mater. **86**, 105 (1990).

[22] E. A. Balykina, E. A. Gan'shina, and G. S. Krinchik, Zh. Éksp. Teor. Fiz. **93** 1879 (1987) [Sov. Phys. JETP **66**, 1053 (1987)]; Fiz. Tverd. Tela **30**, 570 (1988) [Sov. Phys. Solid State **30**, 326 (1988)].

[23] A. S. Kamzin, V. P. Rusakov, and L. A. Grigoriev, *Physics of Transition Metals*, International Conf. USSR Proceed, Part II (1988), p. 271; A. S. Kamzin and L. A. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **16**, *16*, 38 (1990) [Sov. Tech. Phys. Lett. **16**, 616 (1990)].

[24] A. S. Kamzin and L. A. Grigor'ev, [JETP Lett. **57**, 557 (1993)]; A. S. Kamzin and L. A. Grigor'ev, Zh. Éksp. Teor. Fiz. **104**, 3489 (1993) [JETP **77**, 658 (1993)].

[25] A. S. Kamzin, L. A. Grigor'ev, and S. A. Kamzin, Fiz. Tverd. Tela **36**, 1399 (1994) [Phys. Solid State **36**, 765 (1994)]; Fiz. Tverd. Tela **37**, 67 (1995) [Phys. Solid State **37**, 33 (1995)].

[26] A. S. Kamzin, L. P. Ol'khovik, and V. L. Rozenbaum, JETP Lett. **61**, 936 (1965); J. Magn. Magn. Mater. **161**, 139 (1996); Zh. Éksp. Teor. Fiz. **111**, 1426 (1997) [JETP **84**, 788 (1997)].

[27] M. I. Namtalishvili, O. P. Aleshko-Ozhevskiĭ, and I. I. Yamzin, Fiz. Tverd. Tela **13**, 2543 (1971) [Sov. Phys. Solid State **13**, 2137 (1972)].

[28] O. P. Aleshko-Ozhevskiĭ, R. A. Sizov, I. I. Yamzin, and V. A. Lyubimtsev, Zh. Éksp. Teor. Fiz. **55**, 820 (1968) [Sov. Phys. JETP **27**, 425 (1969)].

[29] O. P. Aleshko-Ozhevskiĭ, Ya. Litsevich, A. Murasik, I. I. Yamzin, Kristallografiya, **19**, 331 (1974) [Sov. Phys. Crystallogr. **19**, 201 (1974)].

[30] A. S. Kamzin, L. P. Ol'khovik, and V. L. Rozenbaum, JETP Lett. **67**, 843 (1998).

[31] V. I. Gol'danskii, L. I. Krizhanskii, and V. V. Khrapov, eds., *Chemical Applications of Mössbauer Spectroscopy* [in Russian], Mir, Moscow (1977).

[32] A. P. Kuprin and A. A. Novakova, Nucl. Instrum. Methods Phys. Res. B **62**, 493 (1992).

[33] A. S. Kamzin and L. A. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **19**(8), 38 (1993) [Tech. Phys. Lett. **19**, 517 (1983)].

[34] T. Kobayashi, T. Fukumura, and A. Nakanishi, Nucl. Instrum. Methods Phys. Res. B **76**, 204 (1993).

[35] G. Klingelhofer, U. Imkeller, E. Kankeleit, and B. Stahl, Hyperfine Interact. **69**, 819 (1991).

[36] A. S. Kamzin and L. A. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **19**(8), 50 (1993) [Tech. Phys. Lett. **19**, 245 (1993)].

[37] Sh. Sh. Bashkirov, A. B. Liberman, and V. I. Sinyavskiĭ, *Magnetic Microstructure of Ferrites* [in Russian], Izdatel'stvo Kazanskogo Universiteta, Kazan (1978).

[38] Yu. A. Mamaluĭ, L. P. Ol'khovik, and L. F. Checherskaya, *Physics and Technology of High Pressures* [in Russian], No. 12, 17 (1983).

[39] B. J. Evans, F. Granjean, A. P. Lilot, R. H. Vogel, and A. Gerard, J. Magn. Magn. Mater. **67**, 123 (1987).

[40] G. Albanese, A. Deriu, L. Lucchini, and G. Slokar, Appl. Phys. A: Solids Surf. **26**, 45 (1981); IEEE Trans. Magn. **MAG-17**, 2639 (1981).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Characteristics of magnetic order in perovskite manganites $La_{1-x}Ca_xMnO_3$

T. I. Arbuzova,*) I. B. Smolyak, S. V. Naumov, A. A. Samokhvalov, A. V. Mostovshchikov,
and N. I. Solin

Institute of Metal Physics, Ural Branch of the Russian Academy of Sciences, 620219 Ekaterinburg, Russia

(Submitted 26 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1664–1675 (November 1999)

Solid solutions of the system $La_{1-x}Ca_xMnO_3$ are synthesized, and their magnetic and electrical properties are investigated. As x is increased, the crystal lattice changes symmetrically from an orthorhombic ($x=0$ and $x=0.8$) to a cubic structure ($0<x\leqslant0.6$ and $x=1$). Nonstoichiometric $LaMnO_3$ and compositions with a low Ca content ($0<x<0.4$) are collinear ferromagnets. A metal–insulator transition is observed in them near the Curie temperature. Compositions with $0.6\leqslant x\leqslant0.9$ exhibit a semiconductor-type conductivity and an unusual temperature behavior of the magnetization and the susceptibility with very distinct Curie and Néel temperatures. The magnetic properties of the solid solutions are explained on the basis of the model of ferro–antiferromagnetic phase separation. © *1999 American Institute of Physics.* [S1063-7761(99)01211-1]

1. INTRODUCTION

The perovskite manganites ABO_3 can be classified as magnetic semiconductors characterized by strong coupling between the electronic and magnetic subsystems. This coupling produces a number of unusual properties such as a metal–insulator transition, colossal magnetostriction, magnetic transitions of the antiferromagnetic-ferromagnetic type, charge ordering, and the formation of polarons. The electrical and magnetic properties of manganites are sensitive to changes in the crystal structure and deviations from stoichiometry. An external magnetic field can influence structural phase transitions. Manganites have been of special interest since the discovery of colossal magnetoresistance in them, and for this reason attention has been focused primarily on $La_{1-x}A_xMnO_3$ solid solutions in the composition range $x<0.4$, which exhibit the highest magnetoresistances near the Curie temperature T_C. The range $x>0.5$ has been largely ignored. At the present time there is no consensus as to what causes the onset of colossal magnetoresistance in these compounds. It is often attributed to a double exchange mechanism, but this hypothesis is countered by a mechanism of indirect exchange and the two-phase ferro–antiferromagnetic state of degenerate antiferromagnetic semiconductors. Additional experimental data are needed to explain the entire gamut of physical properties of perovskites within the framework of a single model. The investigation of the magnetic properties of systems of $LaMnO_3$-based solid solutions is also of independent interest in regard to the physics of magnetic semiconductors, in particular, for explaining the exchange interaction mechanisms, the possibility of an inhomogeneous state, the formation of polarons, and the separation of the magnetic phases.

We have investigated the temperature and field dependence of the magnetization, susceptibility, and magnetoresistance of the system $La_{1-x}Ca_xMnO_3$ $(0\leqslant x\leqslant1)$, for which a continuous series of solid solutions exists.

2. SAMPLES AND MEASUREMENT PROCEDURE

The $La_{1-x}Ca_xMnO_3$ samples were prepared by ceramic technology from as-delivered La_2O_3, $CaCO_3$, and Mn_3O_4 powders (all of extreme purity). The samples were synthesized in two stages. The mixture was first annealed in air at 1300 °C for 30 h. It was then ground and compacted into tablets, which were subjected to a second, 50-h anneal. X-ray analysis was performed on a DRON-2.0 x-ray diffractometer in Cr K_α radiation. The samples were single-phase objects. The values of the lattice parameters of the samples are shown in Table I. The unit cell structures were orthorhombic for the $LaMnO_3$ and $La_{0.2}Ca_{0.8}MnO_3$ compositions and were cubic for all others. The technological conditions can influence the crystal structure and physical properties of manganites;[1] the $La_{0.2}Ca_{0.8}MnO_3$ and $La_{0.4}Ca_{0.6}MnO_3$ samples were therefore subjected to another air anneal at 1400 °C for 30 h, followed

TABLE I. Lattice parameters of manganites $La_{1-x}Ca_xMnO_3$.

Composition	$x=0$	$x=0.1$	$x=0.2$	$x=0.3$	$x=0.4$	$x=0.6$	$x=0.8$	$x=1$
Crystal	O	C	C	C	C	C	O	C
Lattice parameters, Å	$a=5.524$ $b=7.788$ $c=5.402$	7.784	7.768	7.708	7.691	7.628	$a=5.334$ $b=7.535$ $c=5.326$	3.731

Legend: O—orthorhombic structure; C—cubic structure.

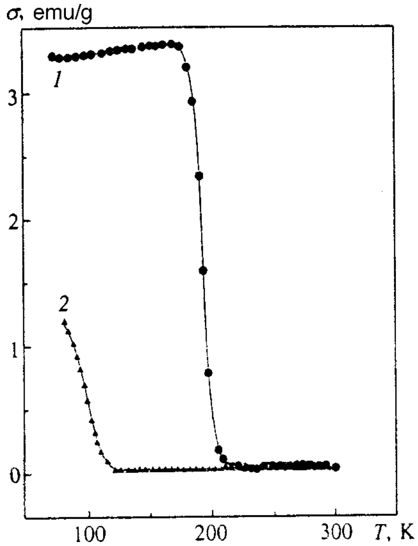

FIG. 1. Temperature dependence of the magnetization in a static field $H=60$ Oe for $LaMnO_3$ (*1*) and $CaMnO_3$ (*2*).

by air hardening. Annealing and hardening did not alter the symmetry of the lattice, but the lattice parameters decreased.

The magnetic properties of the $La_{1-x}Ca_xMnO_3$ solid solutions were measured in the temperature range $T=77-600\,K$ on a vibration magnetometer and a Faraday magnetic balance. The dc electrical conductivity and magnetoresistance were measured by a four-contact procedure on an automated system.

3. EXPERIMENTAL RESULTS

In the system of $La_{1-x}Ca_xMnO_3$ solid solutions the limiting compositions with $x=0$ and $x=1$ must be antiferromagnetic semiconductors with Néel temperatures $T_N=139.5\,K$ Refs. 2 and 3) and $T_N=131\,K$ (Ref. 2), respectively. Deviations from stoichiometry can increase the conductivity and ferromagnetic ordering of the magnetic moments of Mn. Complications would most likely be encountered in trying to obtain samples of stoichiometric composition, because data on the characteristic temperature and field dependence of the magnetization σ and the susceptibility χ of collinear antiferromagnets are all but nonexistent in the literature for $LaMnO_3$ and $CaMnO_3$. We have investigated the magnetic properties of a continuous series of $La_{1-x}Ca_xMnO_3$ solid solutions. The limiting compositions with $x=0$ and $x=1$ can be regarded as reference points in analyzing the influence of Ca substitution on the magnetic properties of $LaMnO_3$.

Figure 1 shows temperature curves of the magnetization in a static magnetic field $H=60\,Oe$ for $LaMnO_3$ and $CaMnO_3$. It is evident that for $LaMnO_3$ the $\sigma(T)$ curve has the typical form for ferromagnets, exhibiting a sharp drop in the magnetization near the Curie temperature T_C. For $CaMnO_3$ the $\sigma(T)$ curve flattens out, suggesting inhomogeneity due to native defects $(T_C=115\,K)$. The magnetic state of a sample at a fixed temperature can be estimated from the field dependence of the magnetization. Figure 2 shows the field dependence of the magnetization for compositions with $x=0$ and $x=1$ at temperatures of 77 K and 300 K. At room temperature the $\sigma(H)$ curves are generally linear for both compounds and extrapolate to zero, indicating the paramagnetic state. At 77 K the field curves have a nonlinear form, as is characteristic of the ferromagnetic ordering region. By extrapolating these $\sigma(H)$ curves to $H=0$ from the region of the para-process, we have estimated the saturation magnetizations: $\sigma_s=73.8\,G\cdot cm^3/g$ for $LaMnO^3$ and $\sigma_s=1.35\,G\cdot cm^3/g$ for $CaMnO_3$. The theoretical values of the saturation magnetization at $T=0$ for ferromagnetically ordered spins of the Mn^{3+} $(S=2)$ in $LaMnO_3$ are $\sigma_{theor}=92.4\,G\cdot cm^3/g$, and for the spins of Mn^{4+} ions $(S=3/2)$ in $CaMnO_3$ we have $\sigma_{theor}=117.2\,G\cdot cm^3/g$. In contrast with $LaMnO_3$, for which the theoretical and experimental values of σ_s are fairly close together (taking into account the Bloch-law decrease of σ_s as T increases), the

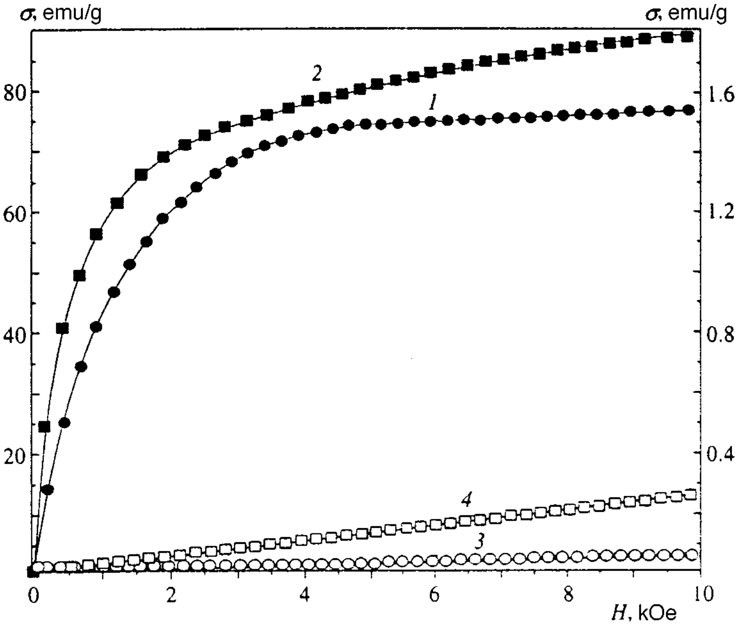

FIG. 2. Field dependence of the magnetization at $T=77$ K (dark symbols) and 300 K (light symbols) for $LaMnO_3$ (curves *1* and *3*, left scale) and $CaMnO_3$ (curves *2* and *4*, right scale).

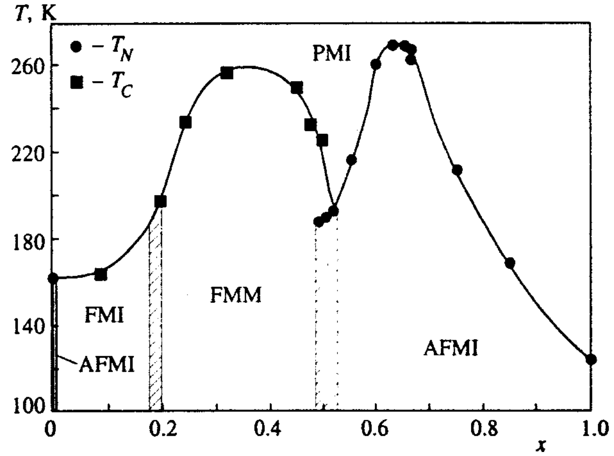

FIG. 3. Phase diagram of the magnetic and electronic states of the system $La_{1-x}Ca_xMnO_3$ (Ref. 4): AFMI=antiferromagnetic insulator; FMI—ferromagnetic insulator; FMM—ferromagnetic metal; PMI—paramagnetic insulator.

experimental value of σ_s for $CaMnO_3$ at 77 K is two orders of magnitude lower than the theoretical value. The reason for this low value of σ_s is that a large part of the sample exists in the antiferromagnetic state, or else the measurement temperature is close to T_C, and long-range ferromagnetic order significantly damages thermal excitations.

Schiffer, Ramirez *et al.*[4,5] have plotted a phase diagram of the magnetic and electronic states of the system $La_{1-x}Ca_xMnO_3$. According to this diagram (Fig. 3), compositions having a low doping level, $x<0.2$, must be ferromagnetic insulators with $T_C = 160$–180 K. As the Ca^{2+} concentration is increased $(0.2<x<0.5)$, the solid solutions become conducting ferromagnets with high $T_C \leqslant 270$ K and almost total saturation of the magnetization at $T=0$. For these compositions we have observed a metal–insulator transition and colossal magnetoresistance in the vicinity of T_C. In the interval $0.5<x<1$ the solid solutions must be nonconducting antiferromagnets.

We have measured the resistivity $\rho(T)$ of $La_{1-x}Ca_xMnO_3$ samples in the temperature range 77–300 K in magnetic fields $H \leqslant 20$ kOe. The temperature dependence of ρ at $H=0$ and $(\rho_H - \rho_0)/\rho_0$ at $H=20$ kOe are shown in Fig. 4. The metal–insulator transition is observed for compositions with $x \leqslant 0.3$. The maximum negative magnetoresistance near the transition point is 55%. The magnetoresistance decreases with increasing distance from the transition temperature. Samples with $x>0.3$ exhibit a semiconductor-type conductivity over the entire temperature range. Their magnetoresistance is substantially lower.

Magnetic measurements have shown that our samples with a low Ca content are ferromagnets. Temperature curves of the magnetization in a field $H=60$ Oe for $La_{0.9}Ca_{0.1}MnO_3$ and $La_{0.7}Ca_{0.3}MnO_3$ are shown in Fig. 5. The Curie temperatures determined by the kink method are $T_C = 170$ K $(x=0.1)$ and $T_C = 205$ K $(x=0.3)$. Field curves of the magnetoresistance for these samples at $T=77$ K and $T=300$ K (above and below T_C) are shown in Fig. 6. Clearly, at 300 K both samples are in the paramagnetic state. At 77 K the $\sigma(H)$ curves exhibit nonlinear behavior. The saturation magnetizations obtained by extrapolation to $H=0$ are $\sigma_s = 74.1\,G\cdot cm^3/g$ for $x=0.1$ and $\sigma_s = 81.9\,G\cdot cm^3/g$ for $x=0.3$, which are lower than the theoretical values of completely ferromagnetically ordered spins of Mn^{3+} and Mn^{4+} ions, $\sigma_s^{\text{theor}} = 93.9\,G\cdot cm^3/g$ $(x=0.1)$ and $\sigma_s^{\text{theor}} = 97.5\,G\cdot cm^3/g$ $(x=0.3)$.

The magnetization exhibits unusual behavior in solid solutions with compositions $0.6 \leqslant x \leqslant 0.9$. Figure 7 shows temperature curves of the magnetization in various magnetic fields for compositions with $x=0.6$ and $x=0.8$. In weak fields the $\sigma(T)$ curves have the usual form for ferromagnets with $T_C = 113$ K $(x=0.6)$ and $T_C = 98$ K $(x=0.8)$. Increasing the magnetic field not only spreads out the phase transition, but also causes the magnetization and the susceptibility to acquire maxima above T_C. Maxima of σ and χ can be observed in antiferromagnets near T_N, in ferromagnets with $T_N > T_C$ (Ref. 7), or in spin glasses near the freezing point T_f. To test the onset of the spin-glass state in these solid solutions, we have measured the dependence $\sigma(T)$ and $\chi(T)$ in two regimes: zero-field cooling and cooling in a magnetic field. None of the characteristic spin-glass hysteresis effects are observed at $T>140$ K.

The temperature dependence of the susceptibility in fields $H=4.5$ kOe and 8.9 kOe for $La_{0.2}Ca_{0.8}MnO_3$ is shown in Fig. 8. We see that the sample exists in the paramagnetic state above 300 K, because the reciprocal susceptibility obeys the Curie–Weiss law with a paramagnetic Curie temperature $\Theta = 142$ K and $\mu_{\text{eff}} = 3.93\mu_B$ $(\mu_{\text{eff}}^{\text{theor}} = 4.1\mu_B)$. The susceptibility maximum near 200 K is probably associated with the Néel temperature. When the temperature is lowered from the high-T range, the sample goes from the paramagnetic to the antiferromagnetic state. In the paramagnetic range the susceptibility of antiferromagnets does not depend on the magnetic field. In the magnetically ordered range $(T<T_N)$ the susceptibility of antiferromagnets depends on the field, and the higher the field H the lower the value of χ (Ref. 8). We have observed a qualitatively similar pattern in $La_{0.2}Ca_{0.8}MnO_3$. It is noteworthy that data on the magnetic properties of $La_{0.2}Ca_{0.8}MnO_3$, obtained on the vibration magnetometer and magnetic balance, are in good agreement [see Figs. 7(b) and 8].

The temperature dependence of the magnetization of manganites is influenced not only by the magnetic field, but also by native defects. For example, the heat treatment of samples with $x=0.6$ and 0.8 at 1400 °C and subsequent air hardening greatly suppress the magnetization in the low-temperature range $T<150$ K. The magnetization maximum in the interval 200–250 K becomes more pronounced, but its position on the T axis stays the same.

4. DISCUSSION

Magnetic oxides with a perovskite structure ABO_3 have important bearing on the experimental investigation of 180° superexchange through anions, because they are devoid of direct cation–cation exchange and 90° superexchange. The interaction between identical magnetic ions other than Jahn–Teller ions (Mn^{3+}, Cr^{3+}, and Cu^{2+}) is antiferromagnetic, so

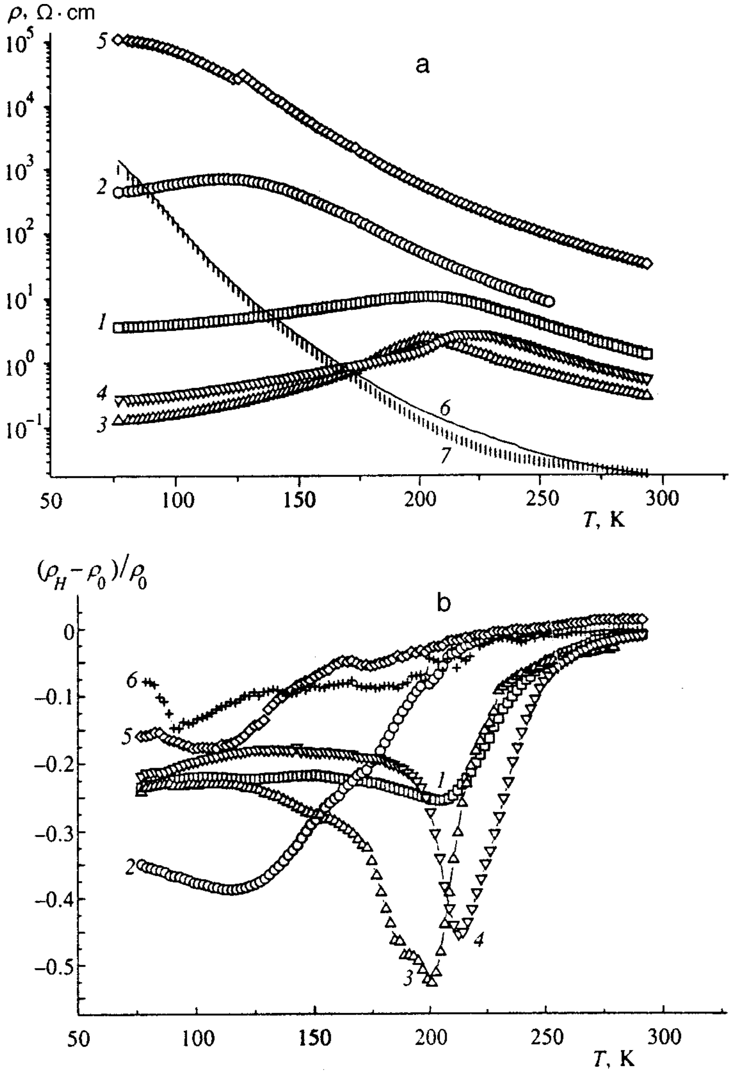

FIG. 4. Temperature dependence of the resistivity at $H=0$ (a) and the magnetoresistance in a field $H=20$ kOe (b) for $La_{1-x}Ca_xMnO_3$: (*1*) $x=0$; (*2*) $x=0.1$; (*3*) $x=0.2$; (*4*) $x=0.3$; (*5*) $x=0.4$; (*6*) $x=0.6$; (*7*) $x=0.8$.

that the majority of $3d$-perovskites are semiconducting antiferromagnets. The ideal ABO_3 structure is cubic ($Pm3m$), but perovskites usually have an orthorhombic or rhombohedral lattice because of Jahn–Teller effects. The symmetry of the lattice is dictated by the relative dimensions of the ions and its tendency toward close-packing. Jahn–Teller effects play an important role in magnetic exchange interactions.

The magnetic properties of $LaMnO_3$-based perovskites were first studied by Jonker and van Santen.[9] They discovered that when La^{3+} ions are replaced by alkaline-earth (A) Ca^{2+}, Sr^{2+}, and Ba^{2+} ions, $La_{1-x}A_xMnO_3$ solid solutions in the composition range $0.2<x<0.4$ become ferromagnets with a high Curie temperature $T_C<280$ K and metallic conductivity. The onset of ferromagnetism in doped manganites, as in pure $LaMnO_3$ with an oxygen excess, was identified with the presence of Mn^{4+} ions and was attributed to the competition of weak ferromagnetic exchange between Mn^{3+} ions, strong antiferromagnetic exchange between Mn^{4+} ions, and strong ferromagnetic exchange between Mn^{4+} and Mn^{3+} ions.[10] Zener[14] has called attention to the high electrical conductivity of high-T_C compositions and, to account for the onset of ferromagnetism, has proposed a mechanism of double exchange with the migration of d-electrons between Mn^{4+} and Mn^{3+} ions. However, it has been shown experimentally[11] that ferromagnetism is possible in nonconducting $LaMnO_3$ for Mn ions of the same valence. The author associates ferromagnetism with the elimination of Jahn–Teller lattice distortion and the tendency of the lattice toward cubic symmetry.

Goodenough[12] has presented a qualitative theory of 180° superexchange with allowance for the covalent bond, which depends on the distance. The degree of covalence exerts a strong influence on exchange interaction and p-type conductivity. The critical distances between magnetic ions for the transition of localized $3d$ states to nonlocalized states have been determined, and semiempirical rules have been formulated for the signs and magnitudes of the superexchange parameters. According to the Goodenough-Kanamori rules, superexchange interaction is antiferromagnetic for small lattice parameters, when the overlap integral of $3d$ orbitals of magnetic ions is a maximum. For large lattice parameters, weak contact of $3d$ orbitals, and a zero overlap integral, superexchange must be ferromagnetic. Consequently, for manganites devoid of Jahn–Teller ordering Mn^{4+}–O_2–Mn^{3+} and Mn^{3+}–O_2–Mn^{3+} interactions have a ferromagnetic charac-

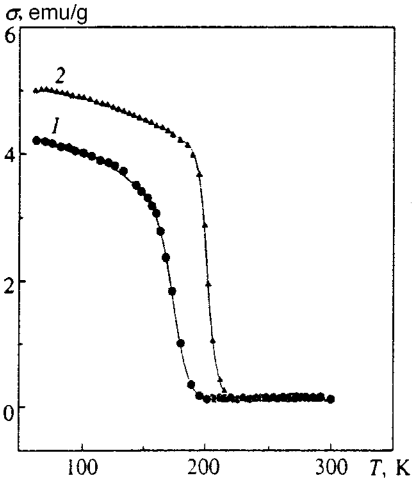

FIG. 5. Temperature dependence of the magnetization for $La_{0.9}Ca_{0.1}MnO_3$ (*1*) and $La_{0.7}Ca_{0.3}MnO_3$ (*2*) in a field $H=60$ Oe.

ter. On the basis of the theory of 180° superexchange and an analysis of experimental data on the magnetic, electrical, and crystallographic properties of perovskites Goodenough has constructed a semiempirical phase diagram for the system $La_{1-x}Ca_xMnO_3$ (Refs. 12 and 13). According to this diagram, compositions with $x<0.2$ must have orthorhombic symmetry and canted (noncollinear) antiferromagnetic order, those with $0.2<x<0.4$ have rhombohedral symmetry and ferromagnetic order, and compositions with $0.5<x<0.9$ have tetragonal symmetry and antiferromagnetic order.

In the nineties manganites aroused interest after colossal magnetoresistance was discovered in them. The magnetoresistance peak in magnetic semiconductors is normally observed near the Curie temperature and can attain large values. At the present time there is no consensus as to what causes the onset of magnetoresistance. Colossal magnetoresistance is attributed either to a double exchange mechanism[14,15] or to indirect exchange and the separation of ferromagnetic and antiferromagnetic phases.[16] It should be noted that the double and indirect exchange mechanisms are qualitatively similar, but differ in the number of free carriers.

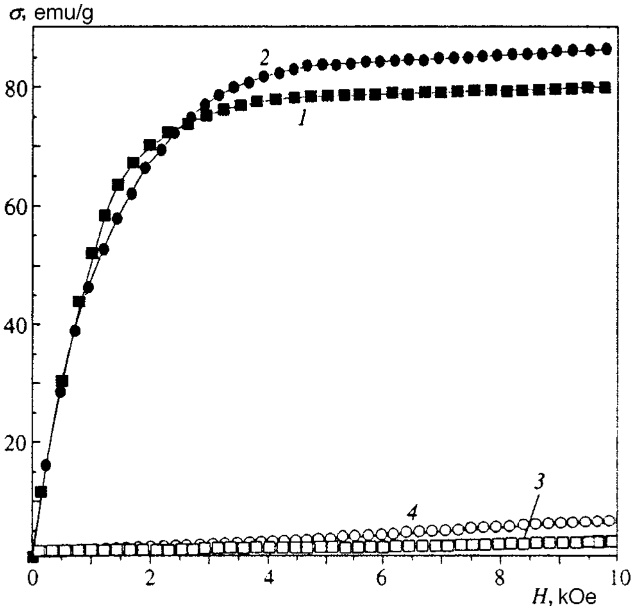

FIG. 6. Field dependence of the magnetization at 77 K (dark symbols) and 300 K (light symbols) for $La_{0.9}Ca_{0.1}MnO_3$ (*1*,*3*) and $La_{0.7}Ca_{0.3}MnO_3$ (*2*,*4*).

The replacement of La^{3+} ions by divalent nonmagnetic ions or an oxygen excess produces Mn^{4+} ions, creating ferromagnetic order and generating free charge carriers (holes). Both mechanisms account for the colossal magnetoresistance and ferromagnetism in manganites, but when double exchange occurs, the magnetic state at $T<T_C$ must be homogeneous (collinear ferromagnetism or canted antiferromagnetism). When the Nagaev mechanism is active, the magnetic state is two-phase in the range $T<T_C$. The spatial separation of the ferromagnetic and antiferromagnetic phases should be manifested in the magnetic properties of manganites.

To date many papers have been published on the magnetic, electrical and crystallographic properties of manganites, but the data can differ even for identical compositions, because native defects have a strong influence on these properties. The investigation of $La_{1-x}Ca_xMnO_3$ samples prepared by the same technology with different compositions x and an

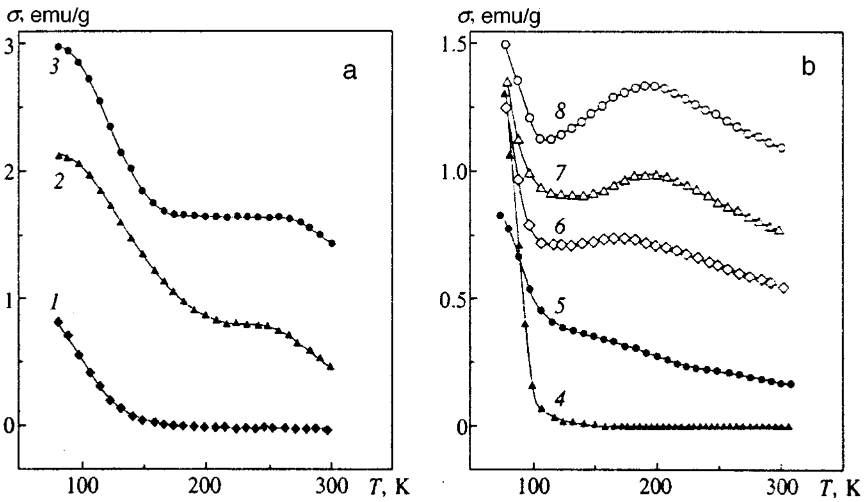

FIG. 7. Temperature dependence of the magnetization for $La_{0.4}Ca_{0.6}MnO_3$ (a) and $La_{0.2}Ca_{0.8}MnO_3$ (b) in various magnetic fields: (*1*) $H=50$ Oe; (*2*) 5 kOe; (*3*) 10 kOe; (*4*) 13 kOe; (*5*) 2 kOe; (*6*) 5 kOe; (*7*) 8 kOe; (*8*) 10 kOe. The scales of σ differ for two curves: 1:3 for curve *1*; 1:10 for curve *4*.

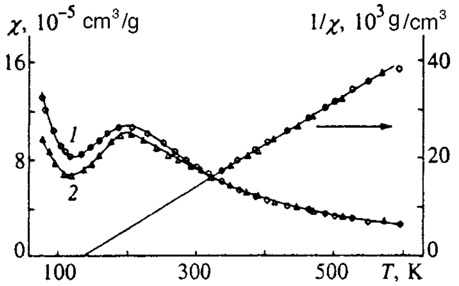

FIG. 8. Temperature dependence of the susceptibility and the reciprocal susceptibility for $La_{0.2}Ca_{0.8}MnO_3$ in fields $H = 4.5$ kOe (*1*) and $H = 8.9$ kOe (*2*).

analysis of published data are helpful in separating the influence of doping from that of native defects.

The magnetic properties of our solid solutions concur qualitatively with the phase diagrams of Goodenough and Ramirez, but a number of anomalies are found. Undoped $LaMnO_3$ exhibits the properties of a collinear homogeneous ferromagnet below $T_C = 193$ K (Figs. 1 and 2. This high value of T_C is probably associated with the presence of Mn^{4+} ions resulting from native defects. The high conductivity of this sample and the transition from conductivity of the semiconductor type to the metallic type near T_C indicate the presence of a fair number of free carriers (Fig. 4). Doped samples in the range $x < 0.4$ are also ferromagnets. They have a cubic structure, high magnetoresistance, and a very distinct metal–insulator transition. The magnetic and electrical properties of $La_{1-x}Ca_xMnO_3$ ($x < 0.4$) can be explained on the basis of a modified double-exchange model.[15] The diffuseness of the magnetic phase transition is attributable to spin polarons. In the high-temperature range the formation of spin polarons is more favorable energetically than the homogeneous paramagnetic state. The polarons are small and do not overlap in space. The system as a whole remains paramagnetic, because the directions of the magnetic moments fluctuate. As the temperature is lowered the dimensions of the spin polarons increase and they begin to overlap. As a result electrons localized near the magnetic moments become delocalized, and the system goes over to the homogeneous ferromagnetic state.

As mentioned above, $La_{1-x}Ca_xMnO_3$ in the range $0.5 < x < 1$ should be antiferromagnetic. Our samples with $0.6 \leqslant x \leqslant 0.9$ exhibit properties of both antiferromagnets and ferromagnets with sharply defined Curie and Néel temperatures (Figs. 7 and 8). We identify the magnetization and susceptibility maxima in the temperature interval $T = 200-250$ K with the Néel temperature, because in the temperature interval $T_C < T < T_N$ the magnetization depends linearly on the field, and the hysteresis effects typical of a spin glass to not occur. Our values of T_N are in good agreement with the data of Ref. 4, which gives $T_N = 255$ K for the composition $x = 0.6$ and $T_N = 205$ K for $x = 0.75$. However, if the entire sample existed in the antiferromagnetic state for $T < T_N$, the susceptibility would either have to decrease or remain constant as the temperature decreased. The abrupt increase in σ and χ for $T < 100$ K, the nonlinear form of the magnetization curve at 77 K, and the nature of the temperature dependence of σ in a weak magnetic field indicate the presence of a ferromagnetic phase at low temperatures. The extrapolation of the reciprocal susceptibility to positive Θ also reveals strong ferromagnetic interaction between the magnetic moments. The magnetic properties of $La_{1-x}Ca_xMnO_3$ ($0.6 \leqslant x \leqslant 0.9$) can be explained on the basis of the Nagaev theory[16] for the two-phase state of degenerate antiferromagnetic semiconductors. The presence of Mn^{4+} and Mn^{3+} ions makes the system amenable energetically to partition into conducting and insulating regions. In the conducting regions the carrier density is high enough to establish magnetic order. However, the entire crystal is still semiconducting (Fig. 4), because a large fraction of the regions exist in the antiferromagnetic insulator state. At low temperatures the magnetization of the crystal is determined by the magnetization of the ferromagnetic regions, because the magnetization of the antiferromagnetic regions is close to zero. In the ferromagnetic regions the total magnetization decreases as the Curie temperature is approached. In the intermediate temperature range $T_C < T < T_N$ the ferromagnetic phase exists in the paramagnetic state, and the antiferromagnetic phase is in the magnetically ordered state. The total susceptibility is the superposition of these contributions. In the high-temperature range $T > T_N$ the entire sample goes over to the paramagnetic state.

The form of the temperature curves of the magnetization and the susceptibility of $La_{1-x}Ca_xMnO_3$ manganites is dictated by the ratio of the volumes of the ferromagnetic and antiferromagnetic phases. Our investigations have shown that reducing the oxygen content of samples with $x = 0.6$ and 0.8 by heat treatment causes the volume of the ferromagnetic phase to decrease, and the ferromagnetic contribution to be suppressed. As a result of heat treatment, these samples acquire the typical dependence $\sigma(T)$ for antiferromagnets, i.e., when the temperature is lowered, $T < T_N$, the magnetization decreases and is not very large in the limit $T \rightarrow 0$, consistent with the data of Ref. 4.

The partition into ferromagnetic and antiferromagnetic phases can also occur in samples having a Ca content $x < 0.5$. The deviation of the saturation magnetic moment of these samples from the theoretical value can be elicited by the presence of the antiferromagnetic phase. However, the advent of antiferromagnetic properties is masked by the ferromagnetic phase. Possible causes of this behavior are the small volume of the antiferromagnetic phase, the hundred- or thousandfold lower magnetization and susceptibility of antiferromagnets relative to ferromagnets, and low Néel temperatures, $T_N < T_C$.

5. CONCLUSION

We have shown that the symmetry of the crystal lattice of the system $La_{1-x}Ca_xMnO_3$ is lowered from orthorhombic to cubic when the concentration of divalent Ca ions in the system is increased. Replacing a portion of the La^{3+} ions by Ca^{2+} ions in the range of low concentrations x, like deviations from stoichiometric composition, increases the conductivity and suppresses ferromagnetism. The maximum values of the Curie temperature T_C, the saturation magnetization, and the magnetoresistance are observed at $x=0.3$. With a further increase in the content of Ca^{2+} ions the Curie temperature, the magnetization, and the resistivity decrease. Especially noteworthy are compositions close to the formula $La_{0.2}Ca_{0.8}MnO_3$, for which maxima of the magnetization and the susceptibility are observed above T_C. The occurrence of coexisting ferromagnetic and antiferromagnetic properties with very distinct Curie and Néel temperatures confirms the existence of two magnetic phases—one ferromagnetic and one antiferromagnetic. The magnetic properties of the complete system of $La_{1-x}Ca_xMnO_3$ solid solutions $(0<x<1)$ can be explained on the basis of the model of indirect exchange and separation of magnetic phases.

This work has received support from the Federal Program ''Surface Atomic Structures'' (Project 2.4.99).

*) E-mail: magsemi@ifm.e-burg.su

---

[1] H. Croft, D. Sills, M. Greenblatt *et al.*, Phys. Rev. B **55**, 8726 (1997).
[2] E. O. Wollan and W. C. Kochler, Phys. Rev. **100**, 545 (1955).
[3] G. J. Matsumoto, J. Phys. Soc. Jpn. **29**, 606 (1970).
[4] P. Schiffer, A. P. Ramirez, W. Bao *et al.*, Phys. Rev. Lett. **75**, 3336 (1995).
[5] A. P. Ramirez, P. Schiffer, S.-W. Cheong *et al.*, Phys. Rev. Lett. **76**, 3188 (1996).
[6] P. J. Woitowicz and M. Reyl, Phys. Rev. Lett. **20**, 1489 (1968).
[7] J. M. Smit and H. P. J. Wijn, *Ferrites*, Wiley, New York (1959).
[8] *Antiferromagnetism* [Russian translation], collected articles, S. V. Vonsovskiĭ (ed.), IIL, Moscow (1956).
[9] G. H. Jonker and J. H. van Santen, Physica (Amsterdam) **16**, 337 (1950).
[10] S. Methfessel and D. C. Mattis, in *Handbuch der Physik*, Vol. 18, Part 1, H. P. J. Wijn (ed.), Springer-Verlag, Berlin (1968), p. 389.
[11] G. H. Jonker, J. Appl. Phys. **37**, 1424 (1966).
[12] J. B. Goodenough, Phys. Rev. **100**, 564 (1955).
[13] J. B. Goodenough, *Magnetism and the Chemical Bond*, Interscience, New York (1963).
[14] C. Zener, Phys. Rev. **82**, 403 (1951).
[15] X. Wang and A. F. Freeman, J. Magn. Magn. Mater. **171**, 103 (1997).
[16] É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996).

Translated by James S. Wood

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Explosive decomposition of heavy-metal azides

B. P. Aduev,*) É. D. Aluker, G. M. Belokurov, Yu. A. Zakharov, and A. G. Krechetov

Kemerovo State University, 650043 Kemerovo, Russia

(Submitted 14 June 1999)
Zh. Éksp. Teor. Fiz. **116**, 1676–1693 (November 1999)

A real-time investigation of the explosive decomposition of heavy-metal azides is reported. A multichannel instrument configuration designed specifically for the goals of the study is described; it is capable of measuring the transient conductivity and the spectral and kinetic characteristics of the luminescence and absorption of exploding samples with nanosecond time resolution. New phenomena are discovered and analyzed in detail: the predetonation conductivity and predetonation luminescence of heavy-metal azides. The conductivity of silver azide in the predetonation state is used to make an experimentally justified decision as to whether the explosion is driven by a thermal or chain mechanism, in favor of the latter. The sum-total of the new data provides the basis for the development of an experimentally justified model of predetonation luminescence and the explosive decomposition process of heavy-metal azides, including the following principal stages: hole trapping by a cation vacancy, reconstruction of the center as a result of chemical reaction with the formation of a quasi-local hole state in the valence band, hole detrapping from the reconstructed center, carrier multiplication as a result of impact ionization by hot holes, and reconstruction of a local state in the bandgap, thereby establishing conditions for repetition of the investigated chain of processes. © *1999 American Institute of Physics.* [S1063-7761(99)01311-6]

1. INTRODUCTION

For close to half a century now scientists have been intrigued by the explosive decomposition of heavy-metal azides (HMAs).[1] The applied aspect of this unflagging interest is related to the fact that the HMAs are the simplest and, hence, the most exhaustively studied representatives of initiating agents and serve as a model object for this class of systems.[2]

The theoretical aspect of the problem runs much deeper. Until recently it has been impossible to reliably detect a single instance of a branched chain reaction involving quasiparticles in a solid. The explosive decomposition of HMAs is indeed of major interest from this perspective.[3,4]

By the early eighties a wealth of experimental data had been accumulated, and basic notions as to the mechanism underlying the explosive decomposition of HMAs had been formulated (see the survey in Ref. 2), laying the foundation for the majority of subsequent research on the problem.[5,6] In our opinion, however, an analysis of these experimental data and the explosive decomposition models developed from them justify a lingering sense of dissatisfaction.

Culling out papers aimed at narrowly confined applications, we discern two experimental approaches to the study of the explosive decomposition of HMAs.

1. The first approach is to investigate slow HMA decomposition processes in exposure to heat, light, ionizing radiations, electric fields, and magnetic fields.[6,7] Obviously, the applicability of data from these studies for the construction of a realistic model of explosive decomposition is highly problematical.[3,4]

2. The second approach involves experiments designed to investigate explosive decomposition directly. These experiments are set up according to the following scheme.[2,8] An explosion is initiated by an impulsive action (shock, a light pulse, an accelerator pulse, etc.). The goal is then to record either the event of an explosion[2] or, better, the time interval (induction period) between the initiating pulse and the instant of explosion as determined from the accompanying flash of light[8] or fragmentation of the sample (using high-speed motion pictures).[2] Clearly, in this experimental setup only certain global parameters of the process are recorded, and the main event is excluded from the researcher's field of vision, viz.: the changes in the characteristics of the sample (or phenomena accompanying these changes) during the evolution of explosive decomposition.

The situation is reminiscent of circumstances in solid-state radiation physics prior to the advent of highly time-resolved spectroscopy techniques, which made feasible the real-time recording of short-lived states; such states are what define, in the final analysis, the physics of radiation-stimulated processes in solids.[9]

The lack of adequate experimental data imparts a speculative quality to existing concepts as to the mechanism of explosive decomposition of HMAs.[2,5] According to these concepts, the energy transformations of an explosion are established by the exothermic reaction

$$2N_3 \rightarrow N_6 \rightarrow 3N_2 + Q. \quad (1)$$

Quantum-mechanical calculations (for isolated N_3 radicals) give $Q \approx 9-12$ eV (Ref. 2). In a solid this reaction is possible if two holes are trapped at contiguous lattice sites. A cation vacancy is regarded as the most likely site for such trapping.[5] The development of an explosion requires that at

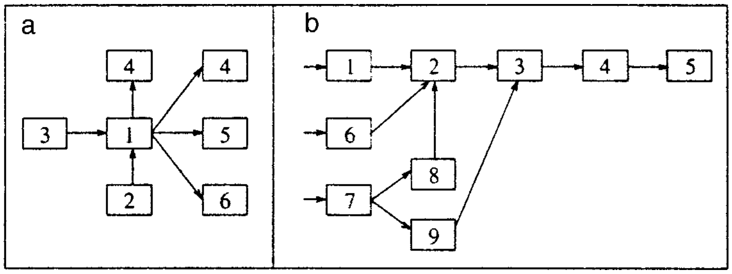

FIG. 1. Block diagram of the experimental configuration for the investigation of explosive decomposition. (a) General block diagram: (*1*) sample; (*2*) excitation source (laser or electron accelerator); (*3*) source of the probe light; (*4, 5, 6*) optical, acoustic, and electrical signal recording channels, respectively. (b) Block diagram of the signal recording channels: (*1*) acoustic sensor; (*2*) oscilloscope; (*3*) television counting device; (*4*) interface; (*5*) computer; (*6*) conductivity measurement cell; (*7*) spectrum instrument (monochromator or spectrograph); (*8*) photomultiplier; (*9*) photoelectron detector.

least some of the energy Q be spent in the reproduction and multiplication of holes. Various multiplication scenarios are considered. Most authors assume that the energy Q is released in the form of phonons and that lattice heating underlies the thermal generation of several electron–hole pairs.[2,8] A so-called ‘‘thermal explosion’’ occurs in this case.[10] A minority of authors[5] advocate the notion of ‘‘chain explosion’’ (Ref. 10). The possible mechanism of hole multiplication is considered to be photomultiplication,[5] or the production of a hot hole directly as a result of the reaction (1) with subsequent multiplication by shock ionization.[11] It must be emphasized that even the very fundamental idea of the reaction (1) as a decisive factor in the explosive decomposition of HMAs, apart from any consideration of the chain or thermal character of the explosion, lacks experimental confirmation.

In our opinion, the entire foregoing discussion indicates that the only road to progress in understanding the mechanism of explosive decomposition of HMAs is the development of new experimental approaches that can be used to record in real time the variations of the characteristics (physical properties) of a sample during an explosion. As the foundation for such an experimental approach we have adopted pulse techniques used in radiation physics,[9] modifying them as necessary in application to the specific attributes of the investigation of explosive samples. The results of the implementation of this approach and an explosive decomposition model proposed on the basis of those results make up the content of the present article.

2. OBJECTS AND PROCEDURE

The objects chosen for the investigation were azides of silver AgN_3, thallium TlN_3, and lead $Pb(N_3)_2$. The azides were prepared in powder form by dual-jet crystallization.[12] The concentrations of the main impurities (Fe, Si, Ca, Mg, Al, Na) were determined polarographically and by complexometric analysis; they were less than 10^{16}–$10^{17}\,\mathrm{cm}^{-3}$.

The investigated thallium and lead azide samples were pressed tablets of diameter 10 mm and thickness 300–400 μm for the thallium azide and 2.5 mm and 30–40μm for the lead azide. The silver azide samples were whiskers with characteristic dimensions $0.1\times0.05\times10$ mm and ‘‘macrocrystals’’ with characteristic dimensions $0.5\times3\times3$ mm. The crystals were grown from solution by a procedure described in Ref. 12. The density of cation vacancies in the crystals was less than $10^{16}\,\mathrm{cm}^{-3}$ (Ref. 12).

The instrument configuration used in the study (Fig. 1) is based on the principles of experimental pulsed radiolysis and photolysis techniques.[9]

The sources of excitation (initiation) are a GIN-600 high-current electron accelerator (effective electron energy 300 keV, current density 1000 A/cm^2, and pulse duration 3 ns) and a YAG:Nd^{3+} laser ($\lambda=1064$ nm, pulse duration 30 ps, and pulse energy 0.5–30 mJ).

The recording apparatus is made up of several synchronized channels.

The *optical channel* consists of an ISP-51 spectrograph or MSD-1 monochromator and a photomultiplier (FÉU-97, FÉU-165-1, 14 ÉLU-FM) or FÉR-7 image-converter streak camera. The photomultiplier output signal is sent to the input of an oscilloscope (S7-8, S7-19); the output of the streak camera is sent through a television counting device utilizing an LI-702 Superkremnikon (a highly sensitive image iconoscope) and then through an interface directly into a computer for processing. The spectral range spanned by the optical channel is 250–1000 nm, and the time resolution is 10 ns or better. By using a spectrograph–chronograph system (instead of the usual monochromator–photomultiplier combination) in the optical channel it is possible to obtain the relaxation pattern of a segment of the spectrum ($\sim$400 nm) within one pulse (‘‘spectrum-in-pulse’’ rather than the customary ‘‘point-by-point spectrum’’ technique[9]), a feature that is essential to the investigation of explosive (‘‘one-time only’’) samples.

In the *conductivity channel* the current through the sample is recorded from the voltage drop across the oscilloscope input resistance, which is connected in series with the sample. When an S8-12 oscilloscope is used, the time resolution is 7 ns and is determined by the oscilloscope; for the S7-19 high-speed oscilloscope the time resolution of the channel is limited by the parameters of the measurement cell and is $\sim$0.3 ns

In the *acoustic signal channel* the sample is attached to the input window of an acoustic sensor in the form of a piezoelectric detector with an intrinsic time resolution $\sim$1 ns. The signal from the acoustic sensor is sent to an oscilloscope.

The channels are synchronized by means of references pulses generated by the input of an initiating pulse to a detector (scattered light from a laser pulse incident on a photodetector, direct action of a laser pulse on the acoustic sensor, etc.). The time-referencing error limits of the signals of the various channels are $\pm$3 ns. A detailed description of the equipment configuration used in the investigation is given in Refs. 3, 4, 13, 14.

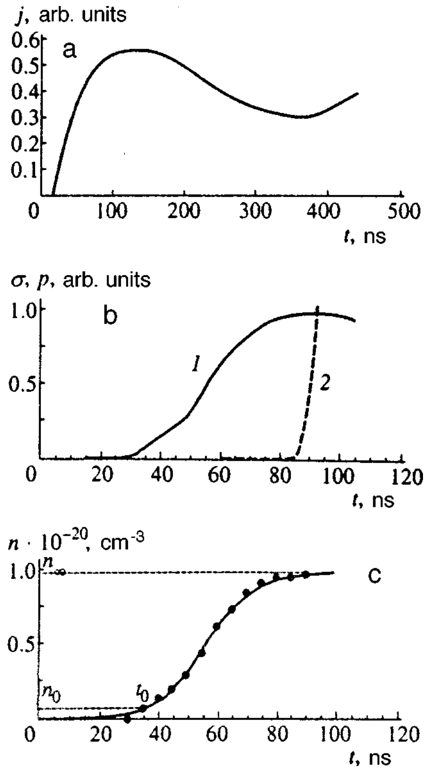

FIG. 2. Kinetics of the explosive conductivity of AgN_3 crystals: (a) complete oscillogram of the current pulse; (b) initial segment of the explosive conductivity pulse (predetonation conductivity): (*1*) conductivity; (*2*) leading edge of the acoustic signal; (c) approximation of the kinetics by Eq. (3): the solid curve is calculated from Eq. (3), and the dots represent the values of n calculated from the experimental values of σ ($\sigma = en\mu$, $\mu = 10\,\mathrm{cm}^2 \cdot \mathrm{V}^{-1} \cdot \mathrm{s}^{-1}$,) and t_0 and n_0 are the time and number density corresponding to the emergence of a current signal that can be reliably detected.

3. EXPLOSIVE CONDUCTIVITY

The first series of experiment to find and investigate the conductivity of HMAs during explosive decomposition (explosive conductivity) was set up as follows. A silver azide whisker with characteristic transverse dimensions $100 \times 150\,\mu\mathrm{m}^2$ was mounted in an air (or vacuum) gap between electrodes of width 2 mm. The explosion was initiated by a laser pulse. Uniform initiation was ensured by making the beam cover the entire length of the interelectrode gap and having the energy of the exciting photons ($\lambda = 1064\,\mathrm{nm}$) fall within the transparency region of the crystal (optical width of the silver azide bandgap $\sim 3.5\,\mathrm{eV}$, thermal width $\sim 1.5\,\mathrm{eV}$, Ref. 15).

3.1. Predetonation conductivity

A typical profile of the explosive conductivity pulse is shown in Fig. 2a. The simplest explanation of the observed kinetics is that the rise of the conductivity in the first peak is associated with the still intact crystal (predetonation conductivity), the decay of the first peak corresponds to rupture of the discontinuity of sample due to growing stresses induced by decomposition, and the next rise is associated with the conductivity of the explosion products (plasma).

The next series of experiments was set up to test this hypothesis. A sample was mounted with its lateral face against the input window of the acoustic sensor to permit synchronous measurement of the conductivity signal and the acoustic signal. The onset of the sample deformation that eventually results in its mechanical fragmentation (explosion) was determined from the leading edge of the acoustic sensor signal.

The sample conductivity preceding the leading edge of the acoustic signal (Fig. 2b) corresponds to the intact sample, i.e., it can be identified with the predetonation conductivity. A detailed analysis of the observed predetonation conductivity reveals the following behavior.

1. As the energy of the initiating pulse is varied in the interval 3–10 mJ, the induction period (the time interval between the initiating pulse and the onset of predetonation conductivity) decreases monotonically and attains 20 ns.

2. The amplitude and shape of the predetonation conductivity pulse do not depend on the energy of the initiating pulse.

3. The maximum values of the recorded current pulses correspond to predetonation conductivities $\sim 1000\,\Omega^{-1} \cdot \mathrm{cm}^{-1}$.

4. The current-voltage curves measured from the maxima of the current pulses are linear.

3.2. Chain-reaction explosive decomposition

The fundamental problem of the nature of the explosive decomposition of HMAs (chain or thermal[16]) can be judged on the basis of the reported data.

We estimate the density of band carriers at the maximum ($\sim 10^3\,\Omega^{-1} \cdot \mathrm{cm}^{-1}$) recorded predetonation conductivities. We invoke the well-known relation $\sigma = en\mu$, where σ is the conductivity, e is the electron charge, n is the density of band carriers (in light of the approximate nature of the estimate, we make no distinction between electrons and holes), and μ is the drift mobility.

For $\mu \approx 10\,\mathrm{cm}^2 \cdot \mathrm{V}^{-1} \cdot \mathrm{s}^{-1}$ (Ref. 7) the conductivity $\sigma = 10^3\,\Omega^{-1} \cdot \mathrm{cm}^{-1}$ corresponds to $n \approx 5 \cdot 10^{20}\,\mathrm{cm}^{-3}$.

The resulting estimate of n as an approximation of the characteristic values for metals indicates the very unusual state of the material in the predetonation phase, which can probably be regarded as a special kind of phase transition. This value of n can serve as a reference point from which to make an experimentally justified decision between the concepts of thermal and chain explosions of HMAs.

We analyze whether the value $n \approx 5 \cdot 10^{20}\,\mathrm{cm}^{-3}$ is realistic from the standpoint of the thermal mechanism of explosion. The thermal width of the silver azide bandgap is $E_g \approx 1.5\,\mathrm{eV}$ (Ref. 15). Assuming the approximations $n \approx n_0 \exp(-E_g/kT)$ and $n_0 \approx 10^{22}\,\mathrm{cm}^{-3}$, we infer that the value $n \approx 5 \cdot 10^{20}\,\mathrm{cm}^{-3}$ corresponds to $T \approx 7000\,\mathrm{K}$. We need not be troubled by the crudeness of the estimate, because attempts to include such factors as the decrease in mobility with increasing temperature, the influence of electron–electron scattering, the true density of states, etc., can only raise the required values of T. The value obtained for T is totally unrealistic inasmuch as the melting point of silver azide is $\sim 523\,\mathrm{K}$ (Ref. 2). Consequently, the experimental values of σ in the predetonation state unequivocally rule out the thermal mechanism of explosive decomposition and can be regarded as direct experimental proof that the explosion of HMAs is a chain reaction.

3.3. Kinetics of the predetonation conductivity

The confirmation of the chain-reaction character of the explosion of HMAs underscores the urgent need to analyze the kinetics of the growth of the predetonation conductivity as a mirror of the kinetics of the evolution of a chain reaction of explosive decomposition.[1)]

The following equation represents the simplest version of a quantitative description of kinetics of the type in Fig. 2(c):

$$dn/dt=\alpha n-\beta n^2, \qquad (2)$$

where n is the density of holes (electrons).

The solution of Eq. (2) subject to the initial condition $n(t_0)=n_0$ is

$$n(t)=\frac{\exp(\alpha(t-t_0))}{n_\infty^{-1}(\exp(\alpha(t-t_0))-1)+n_0^{-1}}, \qquad (3)$$

where n_∞ is the density on the plateau, and t_0 and n_0 are the time and density at which the conductivity attains a value that can be reliably measured. The value of n in Fig. 2(c) is calculated from the relation $\sigma=en\mu$ with $\mu\approx 10\,\mathrm{cm}^2\cdot\mathrm{V}^{-1}\cdot\mathrm{s}^{-1}$ (Ref. 7). We emphasize that the parameters t_0, n_0, and n_∞ in Eq. (3) are evaluated directly from the experimental curve, and the only fitting parameter is α. For all the investigated samples α lies in the interval 10^8–$10^9\,\mathrm{s}^{-1}$, and β lies in the interval 10^{-11}–$10^{-12}\,\mathrm{cm}^3\cdot\mathrm{s}^{-1}$.

The good approximation of the experimental curves [Fig. 2(c)] by Eq. (3) casts doubt on the widely accepted[2,5] (but unproved) opinion that the main exothermic reaction of the HMA decomposition is bimolecular. Indeed, the simplest interpretation of Eq. (2), whose solution is given by Eq. (3), is that the growth (branching) of the chain is governed by a monomolecular process (αn), and its breaking is a bimolecular process (βn^2).

To advance a sensible hypothesis as to the nature of the corresponding processes, it is helpful to analyze the experimental values $\alpha\sim 10^8$–$10^9\,\mathrm{s}^{-1}$ and $\beta\sim 10^{-11}$–$10^{-12}\,\mathrm{cm}^3\cdot\mathrm{s}^{-1}$. The simplest approach is to start with $\beta=vS_r$, where v is the thermal electron (hole) velocity, and S_r is the cross section of the process induced by breaking of the chain. For $v\sim 10^7\,\mathrm{cm}\cdot\mathrm{s}^{-1}$ we have $S_r\approx 10^{-18}$–$10^{-19}\,\mathrm{cm}^2$. These values are typical of indirect interband recombination,[17] i.e., the simplest interpretation of bimolecular breaking of the chain, βn^2, is the interband recombination of electrons and holes.

The simplest interpretation of the linear growth of a chain αn is the trapping of a hole by a point defect. In this case $\alpha=vS_lN$, where v is the thermal velocity, S_l is the capture cross section, and N is the density of defects. For $v\approx 10^7\,\mathrm{cm}\cdot\mathrm{s}^{-1}$ and $N\approx 10^{15}\,\mathrm{cm}^{-3}$ (the usual density of cation vacancies in silver azide whiskers) we obtain $S_l\sim 10^{-14}\,\mathrm{cm}^2$, i.e., the characteristic capture cross section for trapping by an attractive center.[18]

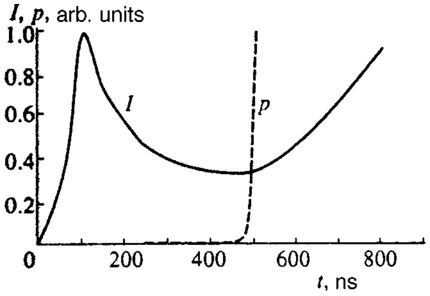

FIG. 3. Kinetics of the explosive luminescence of AgN_3 whiskers ($\lambda=550$ nm, laser pulse initiation, crystal mounted on the input window of the acoustic sensor); the solid curve represents the laser signal, and the dashed curve represents the leading edge of the acoustic signal.

On the basis of these results, therefore, we can advance what seems to us an adequately grounded hypothesis: The growth of the chain reaction of explosive decomposition of HMAs is governed by a monomolecular process: the capture of holes by cation vacancies, and breaking of the chain is governed by the bimolecular interband recombination of electrons and holes.

4. EXPLOSIVE LUMINESCENCE

The shape of the light pulse accompanying the explosive decomposition of a heavy-metal azide (Fig. 3) resembles the shape of the current pulse [Fig. 2(a)] and suggests that this luminescence has two components: predetonation luminescence (the first maximum) and luminescence of the explosion products (the next rise). In any case, the part of the luminescence leading the onset of the acoustic signal (Fig. 3) is unquestionably associated with the intact sample and can be identified as predetonation luminescence.

4.1. Predetonation luminescence and luminescence of the explosion products

Figure 4 shows a streak photograph, taken from the screen of the FÉR camera, of time-resolved spectra of the explosive luminescence of an AgN_3 crystal. Two essentially different types of luminescence are clearly distinguished in this streak photograph: initially wideband luminescence of the sample followed by discrete-line luminescence of the explosion-generated plasma.

In the spectra of the plasma generated by explosion of the HMA we can identify metal lines (silver, thallium, and lead), along with certain nitrogen lines[19] (see Table I). An unidentified line in the vicinity of 770 nm is also observed in all the HMAs.

The wideband luminescence is the most interesting component from our point of view (Fig. 5). The spectrum of this luminescence (or at least part of it) cannot be described by the Planck formula, attesting to the nonthermal character of the process. This conclusion is further corroborated by the qualitative appearance of the spectrum (especially for thallium azide) (Fig. 5) and the nature of the kinetics (decay of the signal intensity before onset of the plasma spectrum; see Fig. 4). All told, these considerations mean that the light radiation preceding the explosion can definitely be identified as predetonation luminescence.[19–21]

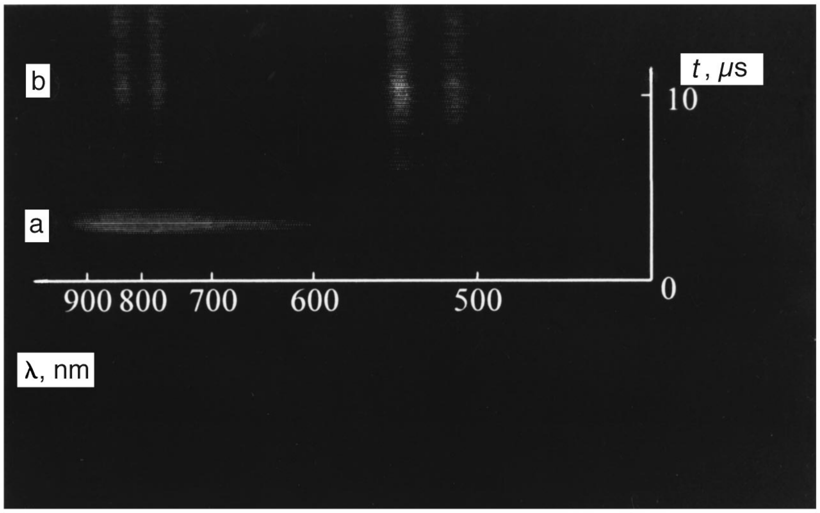

FIG. 4. Streak photograph of the explosive luminescence spectrum of a AgN_3 macrocrystal, reproduced from the screen of the image-converter streak camera: (a) predetonation luminescence region; (b) region of luminescence of the explosion products; the λ and *t* axes indicate the direction of increasing wavelength and time over the screen.

We now discuss certain properties of this luminescence, which are essential to the understanding of its nature and the mechanism of explosive decomposition on the whole.

1. In all the objects the short-wavelength boundary of the luminescence lies in the transparency region. Attempts to detect shorter-wavelength luminescence by means of more sensitive equipment have been in vain.[3,4,16] This result means that the photomultiplication process discussed in Ref. 5 can be eliminated below as a possible mechanism of hole multiplication.

2. A large part of the predetonation luminescence spectrum corresponds to a photon energy greater than the thermal width of the HMA bandgap ($\sim 1-1.5\,\mathrm{eV}$; Ref. 15). We must therefore be dealing with hot luminescence.[22] This is a rigorous assertion for the part of the spectrum with $\hbar\omega > E_g$. For the part of the spectrum with $\hbar\omega < E_g$ the statement is a conjecture, and arguments supporting it will be given below in our discussion of the predetonation luminescence model.

3. Synchronous measurements of the predetonation conductivity and the predetonation luminescence show that this luminescence is observed in the range of very high band carrier densities ($\sim 10^{20}\,\mathrm{cm}^{-3}$).[13,16] Even at far lower densities defect-related luminescence saturation effects are usually observed.[22] (Recall that the concentration of impurities in the as-prepared samples is less than $10^{17}\,\mathrm{cm}^{-3}$.) Moreover, processes of absorption by band carriers become significant at such densities.[22] Essentially structureless predetonation absorption has been observed experimentally in silver azide.[23] It is highly probable that the decay of the predetonation luminescence intensity prior to the onset of plasma luminescence (Fig. 4) is attributable to resorption processes. This conjecture will be proved in Sec. 4.2.

4. It has been shown[2] that the mechanical fragmentation of a sample is preceded by partial melting. Consequently, predetonation luminescence is observed at close to melting-point temperatures, at which luminescence is usually extinguished.[18,22]

TABLE I. Identification of lines in the luminescence spectra of HMA decomposition products.

Material	Band No.	Spectral line, nm	Identification
AgN_3	1	523	Ag (521 nm)
	2	548	Ag (546.5 nm)
	3	595	N (594 nm)
	4	769	Unidentified line
	5	834	N (822 nm)
PbN_6	1	505	N (505 nm)
	2	523	Unidentified line
	3	592	N (593 nm)
	4	679	N (672 nm)
	5	724	Pb (722 nm)
	6	773	Unidentified line
	7	828	N (822 nm)
TlN_3	1	532	Tl (535.046 nm)
	2	588	N^+ (593.179 nm+594.167 nm)
	3	651	Tl (654.977 nm)
	4	747	Unidentified line
	5	815	N (818.816 nm)
	6	870	Unidentified line

4.2. Kinetics of predetonation luminescence

The kinetics of the recorded predetonation luminescence signal (Fig. 3) does not reflect the true kinetics of this luminescence and definitely not the kinetics of explosive decomposition. This disparity is associated with the following considerations.

1. The true luminescence kinetics is distorted by resorption processes.

2. The nonuniform excitation of the macrosamples used in several experiments can have the effect of superimposing the kinetics of propagation of the reaction zone through the sample (detonation wave) on the basic chain-reaction kinetics.

To rule out the influence of these factors, we have performed experiments involving synchronous measurements of the predetonation luminescence and the absorption of silver azide whiskers in the presence of laser pulse initiation (λ

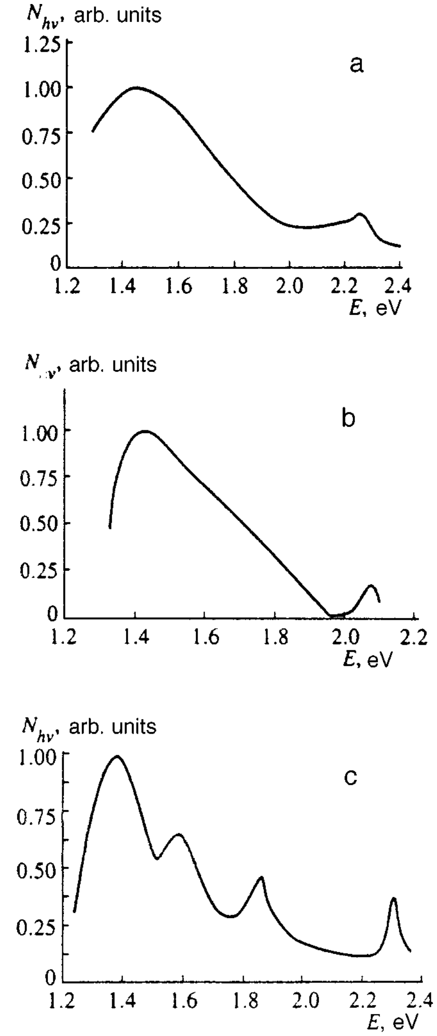

FIG. 5. Predetonation luminescence of HMAs: (a) AgN_3; (b) $Pb(N_3)_2$; (c) TlN_3.

$=1064$ nm). Using this type of excitation and whisker crystals enabled us, as in the measurements of the predetonation conductivity, to make the excitation uniform and thus preclude the influence of detonation processes. The synchronous measurement of luminescence and absorption (at one wavelength) also provided a means for taking into account the distortion of the luminescence signal kinetics due to resorption, i.e., to determine the true luminescence kinetics (Fig. 6).

Resorption is taken into account by the expression[24]

$$I(t)=I_r(t)D(t)(1-\exp(-D(t)))^{-1}, \tag{4}$$

where $I_r(t)$ is the recorded luminescence intensity (luminescence signal), and $D(t)$ is the optical density of the sample.

The resorption-corrected luminescence kinetics is found to be represented by a curve that levels off into a plateau [(Fig. 6(b)] and is very similar to the kinetics of the predetonation conductivity (Fig. 2). The similarity is more than outward appearance. The corrected luminescence kinetics, as in the case of the predetonation conductivity (Fig. 2), is well approximated by the solution of the equation

$$\frac{dI}{dt}=\alpha I-\beta I^2, \tag{5}$$

which has the form

$$I(t)=\frac{\exp(\alpha(t-t_0))}{I_\infty^{-1}(\exp(\alpha(t-t_0))-1)+I_0^{-1}}, \tag{6}$$

where t_0 is the time at which $I(t)$ attains a value I_0 that can be reliably measured, and I_∞ is the value of $I(t)$ on the plateau. As in the case of the predetonation conductivity, the value of the constant α in Eqs. (5) and (6) lie in the interval $\alpha=10^8-10^9\ \mathrm{s}^{-1}$ for different samples.

In our opinion, the agreement of the predetonation conductivity and luminescence kinetics is a very important fact and indicates that they reflect the kinetics of the basic process: explosive decomposition. This fact must be taken into account in constructing a model of predetonation luminescence and explosive decomposition on the whole.

4.3. Model of predetonation luminescence

The predetonation luminescence properties are such that the range of possible models of the phenomenon can be narrowed down considerably.

Above all, the condition $\hbar\omega>E_g$ and the absence of temperature extinction of this luminescence rule out all the kinds of luminescence associated with local centers.[18,22]

Comparison of the spectra of the pre-explosion luminescence with the band structure data[15,23,25] also serves to rule out such forms of fundamental luminescence as edge, exciton, and cross-luminescence.[18,22,26]

Consequently, of all the known types of luminescence from solids, the only remaining candidate is intraband luminescence due to radiative transitions of hot electrons and holes within the conduction or valence band, respectively.[22,27] However, we have not been able to discern any reasonable correlation between the band structure of the in-

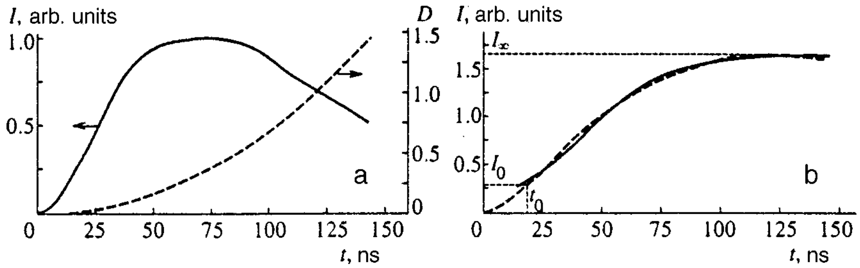

FIG. 6. Predetonation luminescence kinetics of a silver azide whisker at $\lambda=550$ nm (laser pulse initiation): (a) luminescence signal (solid curve) and optical absorption (dashed curve); (b) luminescence kinetics with allowance for resorption, calculated from the experimental values of the luminescence signal and the optical density according to Eq. (4) (dashed curve) and calculated according to Eq. (6) (solid curve); t_0 is the time at which the luminescence signal attains a value I_0 that can be reliably detected.

vestigated objects and the predetonation luminescence spectra, which is one of the basic methods used to identify intraband luminescence.[27]

An altogether different picture emerges when the valence band is assumed to contain quasi-local hole states. If a level corresponding to a quasi-local hole state is present in the interior of the valence band at a distance of 3.2 eV from the top of the valence band in silver azide, at 3.4 eV in thallium azide, or at 3.6 eV in lead azide, a distinct correlation is observed between the luminescence maxima and the density-of-state peaks (Fig. 7).2) However, attempts to obtain an analogous correlation for a level corresponding to a quasi-local hole state in the conduction band have failed.

The sum total of these considerations has led us to propose, as a model of predetonation luminescence, intraband radiative transitions of valence-band electrons into quasi-local hole states situated in the depth of the valence band (Fig. 7). Inasmuch as the lifetime of a hole in a quasi-local hole state is at most $\sim 10^{-14}$ s (Ref. 28), these states must be generated continuously during explosive decomposition.

At this time we do not see any alternative explanation for predetonation luminescence. The proposed model nicely accounts for all the observed properties of this luminescence. It is extremely unlikely that the correlations of the positions of the luminescence maxima for the three objects (two maxima in silver and lead azides, and four maxima in thallium azide) are accidental. We therefore assume that the adequacy of the proposed model, if not proved directly, is highly probable.

One other remark has very important bearing on the what follows. The detrapping of holes from a quasi-local hole state produces hot holes, which can cause hole multiplication as a result of impact ionization[18] and, hence, to the development of a chain reaction.

5. MODEL OF THE EXPLOSIVE DECOMPOSITION OF HEAVY-METAL AZIDES

We first state the main principles on which the model of explosive decomposition must be based.

1. The chain-reaction character of explosive decomposition has been proved experimentally (Sec. 3.2). The decomposition kinetics is described on the assumption that the growth of the chain is linear, and its breaking is a quadratic process (Secs. 3.3 and 4.2).

2. Experimental estimates of the cross sections (Sec. 3.3) support the assumption that the branching of the chain (hole multiplication) is determined by the trapping of holes by a cation vacancy, and breaking of the chain is determined by indirect interband recombination.

3. Photomultiplication of holes is not a contributing factor in the growth of the chain. The principal mechanism of hole multiplication is the impact ionization of hot holes formed in detrapping from a quasi-local hole state (Secs. 4.1 and 4.2).

The chain reaction of the decomposition of HMAs does not cause molecular nitrogen to form in the crystal.[3,6,7]

These experimentally substantiated principles can be used to formulate the main questions that need to be answered (at least at the level of a working hypothesis) by the proposed model.

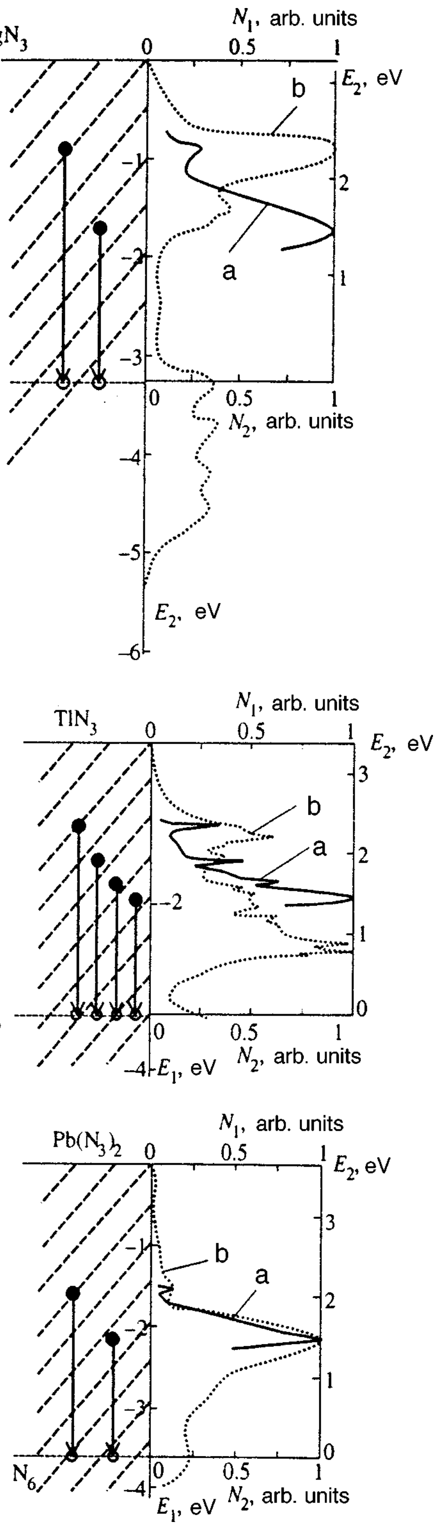

FIG. 7. Comparison of the spectra of predetonation luminescence (a) and the density of states in the valence band (b) for silver, thallium, and lead azides; E_1 and N_1 are the scales for the density of states, E_2 and N_2 are the scales for the predetonation luminescence spectrum, and E_0 is the position of the quasi-local level for best correlation between the maxima of the predetonation luminescence spectrum and the density of states; the arrows indicate optical transitions of valence-band electrons into a quasi-local state.

1. What are the processes by which hole trapping by a

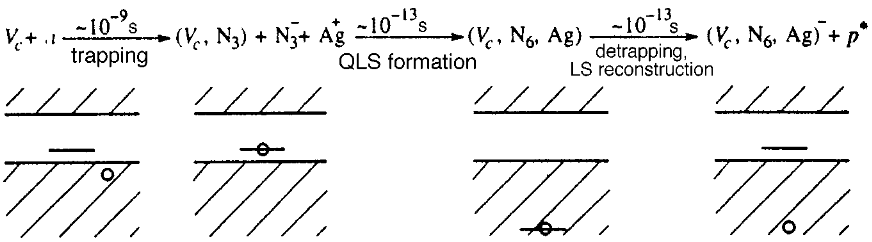

FIG. 8. Main stages of the AgN_3 explosive decomposition reaction.

cation vacancy leads to the formation of an actual quasi-local hole state?

2. How is the reproduction of cation vacancies (or the centers that take their place) maintained during decomposition? What is the necessary condition for the experimentally observed invariance of the quantity α in Eq. (2) as the density of band holes is varies in the interval $10^{18}-10^{20}\,\mathrm{cm}^{-3}$?

A proposed monohole chain-reaction model (in application to silver azide) satisfying the stated requirements is shown schematically in Fig. 8. The figure shows only the main processes underlying chain growth and terminating in the production of a hot hole and the restoration of a local level in the bandgap. The well-known processes of impact ionization by hot holes,[18] which cause hole multiplication, and interband electron–hole recombination,[22] which causes the chain to break, are omitted to avoid cluttering the figure unnecessarily. For the same reason, neither are the radiative transitions responsible for luminescence shown, as they do not affect the energy behavior of the process. The total process in question is conveniently divided into three stages.

1. *Hole trapping at a cation vacancy* (Fig. 8). The parameters characterizing the process have been established in Sec. 3.2: $S_l \sim 10^{-14}\,\mathrm{cm}^2$; $\tau \sim 10^{-9}\,\mathrm{s}$. The end result of this stage is the transformation of an extended (detrapped) state (band hole) into a radical trapped at a cation vacancy, i.e., the formation of a nonequilibrium (V_c, N_3^0) cluster. Inasmuch as the duration of this stage greatly exceeds the duration of the subsequent stages, it is the stage that controls the kinetics of the process.

2. *Formation of quasi-local hole states due to reconstruction of the* (V_c, N_3^0) *cluster in interaction with neighboring* N_3^- *and* Ag^+ *ions*. We an assume that the reconstruction process begins with the formation of a molecular bond between a N_3^0 radical trapped at a vacancy and a neighboring N_3^- ion ($N_3^0 + N_3^- \rightarrow N_6^-$). This process is analogous to the well-studied self-trapping of holes in alkali-halide crystals.[18,29] (Quantum-chemical calculations of the reaction $N_3^0 + N_3^- \rightarrow N_6^-$ for isolated radicals and data on the experimental observation of short-lived N_6^- radicals in solutions by pulsed photolysis are given in Ref. 30). Unlike alkali-halide crystals, however, the completion of this reconstruction process most likely has the effect of ''smearing'' the hole wave function throughout a (V_c, N_6, Ag) cluster. (An analogous process occurs in AgCl, in which the self-trapped hole is a $(AgCl_6)^{4-}$ cluster.[31]

Unfortunately, calculations that could be used to estimate at least approximately the energy released in this stage are nonexistent at the present time. However, making use of the fact that the resultant state is close to N_6^0 and relying on the results of calculations of the reaction (1) in Ref. 2, we can state with some certainty that the process in question is exothermic. In this case, we can assume by reasoning similar to that in Refs. 3 and 4 that the released energy is stored as the potential energy of a hole in a quasi-local hole state (Fig. 8).

According to data obtained in the investigation of predetonation luminescence (Sec. 4.3), the depth of this state is 3–3.5 eV. Multiplying the latter by Avogadro's number, we obtain the total energy that must be released in the complete decomposition of one mole of silver azide if the proposed model is correct. This quantity (~ 3 J/mol) is close to the experimental energy release in an explosion[32]; this proximity can be regarded as evidence of the adequacy of the model.

Consequently, the end result of this stage is the onset of a quasi-local hole state (Fig. 8), which is responsible, in particular, for predetonation luminescence (Sec. 4.3). According to the proposed model, the density of these states is proportional to the density of band holes, thereby reconciling the kinetics of the predetonation conductivity and predetonation luminescence. It must be emphasized that because the intraband luminescence yield is less than 10^{-5} (Ref. 27), it provides us with a good indicator of the multiplication process without influencing its overall kinetics.

The slowest process in this stage is the migration of heavy particles during reconstruction, i.e., the duration of the stage is $\sim 10^{-13}\,\mathrm{s}$.

3. *Hole detrapping* (Fig. 8). The lifetime of a hole in the quasi-local hole is normally no longer than $\sim 10^{-14}\,\mathrm{s}$ (Ref. 28). A hot band hole having an energy of 3–3.5 eV is generated in hole detrapping. This process can be regarded as transfer of the potential energy of a hole in a quasi-local hole state into the kinetic energy of a hot hole. Consequently, hot holes with energies greatly exceeding the width of the HMA bandgap (1–1.5 eV, Ref. 11) are generated. This energy is transformed via two channels: impact ionization, which leads to the multiplication of band holes and electrons, and phonon emission, which causes the sample to heat up, but does not support multiplication.

According to existing notions,[18] at hole energies above the threshold level, which is usually only slightly higher than E_g, the probability of impact ionization is close to unity. Thermalization due to phonon emission becomes predominant at below-threshold energies. In application to the scheme of the model this means that the probability of impact ionization for holes detrapped from the quasi-local hole state is close to unity, i.e., each link of the chain induces doubling of the number of holes or, in reactor terminology, the hole multiplication factor is equal to 2. On the other hand, the sample is heated by phonon emission during the thermalization of holes and electrons whose energies fall below the threshold.

A fundamental question arises: What happens at the local density-of-states peak corresponding to a quasi-local hole state before hole detrapping? In other words, where does the level corresponding to the $(V_c, \mathrm{N_6Ag})^-$ cluster (Fig. 8) go? Above all, we call attention to the fact that the $(V_c\,\mathrm{N_6Ag})$ cluster maintaining the presence of the quasi-local hole state in the depth of the valence band is neutral relative to the lattice. But the $(V_c, \mathrm{N_6Ag})^-$ cluster, which is the result of detrapping, has a negative unit charge relative to the lattice, i.e., this cluster has the same charge as the cation vacancy. As a rule, the position of the energy level of a defect in the band diagram is strictly decided by its charge.[29] We can therefore assume with reasonable certainty that the position of the cluster level does not differ much from the position of the level of the original cation vacancy. Since the transition of the cluster into the equilibrium state is accompanied by the migration of heavy particles, the duration of this stage is also $\sim 10^{-13}$ s.

Consequently, the detrapping of a hole from a quasi-local hole state eradicates the density-of-states peak from the valence band and introduces a local level in the bandgap in a position close to the level of the isolated cation vacancy. The emergence of this level establishes conditions for the above-described chain of processes to be repeated, i.e., for the chain reaction to continue. In other words, from the standpoint of the processes discussed here the system can spontaneously return to its initial state many times. In the proposed model, therefore, the growth of a chain reaction does not require the generation of new vacancies, a result that fits in very well with the experimentally observed invariance of the quantity α in Eq. (2) as the density of band electrons (holes) varies within a range of two orders of magnitude.

Two remarks are called for regarding the model as a whole.

1. The proposed chain-link model is consistent with all currently available experimental evidence. The following principles of the model have been experimentally corroborated to date: the monomolecular character of the growth of the chain process, the role of quasi-local hole states in the generation of hot holes, and the impact mechanism of hole multiplication. Other principles have been established at what we consider to be the level of a reasonable working hypothesis: the role of cation vacancies and the composition of the clusters responsible for the actual local and quasi-local states.

2. The kinetics of explosive decomposition is described by Eq. (2). Therefore the monomolecular growth and bimolecular breaking of the chain are not assumptions of the model, they constitute an experimental fact. However, this experimental fact refers to the already developed explosion process, when the density of band electrons and holes has attained $\sim 10^{18}\,\mathrm{cm}^{-3}$ (Sec. 3.2). The high band-carrier densities lead to saturation of the processes caused by trapping at local centers and, hence, account for the simplicity of the observed kinetics of the process.

In the nascent stage (initiation) of the process the situation can reverse, and processes at preexisting defects can play a decisive role. One of the simplest ways to take this fact into account in the formal kinetics is to write α that appears in (2) in the form $\alpha = f - g(t)$, where f and $g(t)$ are the rate constants of hole trapping by cation vacancies and by competing centers, respectively. The dependence $g(t)$ could be supported, for example, by the ''burnout'' of these centers during initiation. Inasmuch as a chain can grow only under the condition $\alpha > 0$, (i.e., $g < f$), the threshold character of initiation[2] can be attributed specifically to the need to maintain sufficiently ''complete burnout'' $(g < f)$ under the influence of the initiating pulse.

In closing, we wish to call attention to a certain fundamental aspect of the problem that far transcends the issue of exposing the mechanism underlying the explosive decomposition of heavy-metal azides. We feel that the information presented in the article convincingly demonstrates an extremely interesting possibility for the achievement of chemical reactions in solids. A necessary condition for chemical reaction to take place in gases and liquids is that the reacting partners (radicals) must move in such a way as to encounter one another. In a solid, on the other hand, a totally different situation can be realized. Electron excitations are in motion, and their trapping at select sites of the crystal lattice (at structural or impurity defects) causes actual radicals to be formed at the necessary site. Consequently, the long-term migration of heavy particles (usually by diffusion) is superseded by the far more rapid migration of electron excitations. In our opinion, the fascinating prospects for the utilization of this possibility have not yet been fully appreciated in the modern physics and chemistry of the solid state.

The authors are grateful to A. N. Drobchik and A. Yu. Mitrofanov for assistance in the experimental work and to V. I. Krasheninin and V. G. Kriger for productive discussions.

This work has received financial support from the Russian Fund for Fundamental Research (Project 98-03-32001a) and from the Ministry of General and Professional Education of the Russian Federation.

---

*)E-mail: lira@kemsu.ru

1)The explosion (fragmentation) of a sample when the reaction-rate curve reaches the plateau [Figs. 2(b) and 2(c)] is attributable to the fact that the rate at which heat is released exceeds the rate at which heat is removed by heat transfer.[10] In isolated cases sample fragmentation is observed before the rate curve reaches the plateau.[16]

2)Correlation in the intensities of the luminescence and density-of-state peaks might not be observed, because the luminescence intensity also depends on the transition probabilities.

---

[1] F. P. Bowden and A. D. Yoffe, *Fast Reaction in Solids*, Butterworths Scientific Publications, London (1958), p. 242.

[2] *Energetic Materials*, Vol. 1, H. D. Fair and R. F. Walker (eds.), Plenum Press, New York (1987), p. 501.

[3] B. P. Aduev, E. D. Aluker, G. M. Belokurov, and A. G. Krechetov, Russ. Phys. J. **39**, 1135 (1996).

[4] B. P. Aduev, E. D. Aluker, V. G. Kriger, and Yu. A. Zakharov, Solid State Ionics **101–103**, 33 (1997).

[5] V. Kriger and A. Kalensky, Chem. Phys. Rep. **14**, 556 (1995).

[6] V. I. Krasheninin, L. V. Kuz'mina, and V. Yu. Zakharov, Zh. Prikl. Khim. **69**, 21 (1996).

[7] V. I. Krascheninin, L. V. Kuzmina, Yu. A. Zakharov, and A. Yu. Stalinin, Chem. Phys. Rep. **14**, 126 (1995).

[8] E. I. Aleksandrov and A. G. Voznyuk, Fiz. Goreniya Vzryva **19**, 143 (1983).

[9] É. D. Aluker, V. V. Gavrilov, R. G. Deĭch, and S. A. Chernov, *Fast Radiation Processes in Alkali-Halide Crystals* [in Russian], Zinatne, Riga (1987), p. 183.

[10] N. N. Semenov, *Chain Reactions* [in Russian], Nauka, Moscow (1986)., p. 534.

[11] Yu. A. Zakharov, V. A. Nevostruev, S. M. Ryabykh, and Yu. N. Safonov, Khim. Vys. Énerg. **19**, 398 (1985).

[12] S. I. Kurakin, G. M. Diamant, and V. M. Pugachev, Izv. Akad. Nauk SSSR, Neorg. Mater. **26**, 2301 (1990).

[13] B. P. Aduev, E. D. Aluker, G. M. Belokurov, and A. G. Krechetov, Chem. Phys. Rep. **16**, 1479 (1997).

[14] B. P. Aduev, E. D. Aluker, and A. G. Krechetov, Chem. Phys. Rep. **17**, 469 (1998).

[15] A. B. Gordienko, Yu. M. Zhuravlev, and A. S. Poplavnoi, Phys. Status Solidi B **198**, 707 (1996).

[16] B. P. Aduev, É. D. Aluker, G. M. Belokurov, and A. G. Krechetov, JETP Lett. **62**, 215 (1995).

[17] O. V. Bogdankevich, V. S. Letokhov, and A. F. Suchkov, Fiz. Tekh. Poluprovodn. **3**, 665 (1969) [Sov. Phys. Semicond. **3**, 566 (1969)].

[18] É. D. Aluker, D. Yu. Lusis, and S. A. Chernov, *Electronic Excitations and Radioluminescence of Alkali-Halide Crystals* [in Russian], Zinatne, Riga (1979), p. 251.

[19] B. P. Aduev, E. D. Aluker, A. G. Krechetov, and A. Yu. Mitrofanov, Phys. Status Solidi B **207**, 535 (1998).

[20] B. P. Aduev, É. D. Aluker, Yu. A. Zakharov, A. G. Krechetov, and I. V. Chubukin, JETP Lett. **66**, 111 (1997).

[21] B. P. Aduev, É. D. Aluker, A. G. Krechetov, A. Yu. Mitrofanov, A. B. Gordienko, and A. S. Poplavnoĭ, Pis'ma Zh. Tekh. Fiz. **25**(9), 28 (1999) [Tech. Phys. Lett. **25**, 350 (1999)].

[22] V. P. Gribkovskiĭ, *Theory of Light Absorption and Emission in Semiconductors* [in Russian], Nauka i Tekhnika, Minsk (1975), p. 463.

[23] B. P. Aduev, É. D. Aluker, A. G. Krechetov, and Yu. P. Sakharchuk, Pis'ma Zh. Tekh. Fiz. **24**(16), 31 (1998) [Tech. Phys. Lett. **24**, 636 (1998)].

[24] J. Pankove, *Optical Processes in Semiconductors*, Prentice-Hall, Englewood Cliffs, N.J. (1971), p. 456.

[25] E. H. Younk and A. B. Kunz, Int. J. Quantum Chem. **63**, 615 (1997).

[26] P. A. Rodnyĭ, Fiz. Tverd. Tela (Leningrad) **34**, 1975 (1992) [Sov. Phys. Solid State **34**, 1053 (1992)].

[27] B. P. Aduev, É. D. Aluker, V. V. Gavrilov, R. G. Deĭch, and S. A. Chernov, Fiz. Tverd. Tela (St. Petersburg) **38**, 3521 (1996) [Phys. Solid State **38**, 1919 (1996)].

[28] F. Bassani and G. P. Parravicini, *Electronic States and Optical Transitions in Solids*, Pergamon Press, New York-London (1975), p. 391.

[29] A. M. Stoneham, *Theory of Defects in Solids*, Clarendon Press, Oxford (1975), p. 731.

[30] M. S. Workentin, B. D. Wagner, F. Negri *et al.*, J. Phys. Chem. **99**, 94 (1995).

[31] J. P. Spoonhower and A. P. Morcetti, J. Phys. Chem. Solids **51**(7), 793 (1990).

[32] S. Zeman, M. Dimin, S. Truchic, and V. Kabatova, Theor. Chim. Acta **80**, 137 (1984)

Translated by James S. Wood

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Effect of the inner structure of domain walls on the stability of an isolated stripe domain in a pulsed field

Yu. I. Dzhezherya

Institute of Magnetism, National Academy of Sciences, 252680 Kiev, Ukraine

(Submitted 16 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1694–1705 (November 1999)

This paper studies the effect of the inner structure of domain walls on the stability of an isolated stripe domain localized in a thin ferromagnetic film against a pulse of magnetic field applied perpendicularly to the film surface. It is found that the value of the critical amplitude of the pulsed signal strongly depends on the value of the magnetizing field in which the system was initially placed. It is also established that the difference on stability of domains with unipolar and bipolar walls in pulsed fields diminishes as the amplitude of the magnetizing field decreases. Finally, the dependence of the region of stability in a pulse field on the parameters of the system is determined for various domain types. © *1999 American Institute of Physics.* [S1063-7761(99)01411-0]

1. INTRODUCTION

The structure of the magnetization of a nonconducting ferromagnetic system is determined by solving the Landau–Lifshitz equations and the equations of magnetostatics simultaneously. Generally, these equations are nonlinear and contain integral terms, which makes them impossible to solve analytically. A more or less complete analysis can be carried out for simplified models of elements of magnetic structures, among which is the isolated stripe domain and a lattice of stripe domains in a film with perpendicular uniaxial anisotropy. The literature devoted to the investigation of such systems both theoretically and experimentally is vast (see, e.g., Refs. 1–3). There are a number of topics in this field of research whose study is extremely difficult. Among such topics is the allowance for the combined effect on the domain properties of the magnetostatic field of scattering and the exchange interaction. To resolve the problem, one must solve a nonlinear equation containing an integral term of magnetostatic origin.

Several simplified models are used to overcome these difficulties. For instance, in a model widely used in studies of properties of magnetic domains, the domain walls (DW) are interpreted as infinitely thin geometric boundaries with their own surface energy. Within this approximation the dependence of the properties of domain structures on the parameters of the material and the thickness of the ferromagnetic film was found.[4–7] However, this model does not correctly describe the state of the domains when their width becomes comparable to the DW thickness and forces of the exchange interaction come into play.

In another limiting case the solutions of the Landau–Lifshitz equations are determined for the model of a ferromagnet of infinite dimensions. The various properties of an isolated domain in such a system were studied in Refs. 8–12. Among the topics investigated by the researchers were the conditions for stability of an equilibrium system and the spectral characteristics of spin-wave excitations. However, since the size of the ferromagnet was infinite, the contributions of the fields generated by the scattering from the surface inhomogeneities of magnetization had to be ignored. But such fields always exist in real systems, and their effect does not vanish in the limit $L\to\infty$ (here L is the thickness of the film). In this case the value of the scattering field acting on a stripe domain tends to the limit $4\pi M_0$, where M_0 is the saturation magnetization of the material.[1] Hence the research whose results are reported in the papers cited above is of interest from the methodological viewpoint. Only a model that allows for scattering from the surface can be considered meaningful.

The theory of the approximate solution of the singularly perturbed equation describing the structure of magnetization of an isolated stripe domain in a thin ferromagnetic field was developed in Refs. 13 and 14. A unified approach based on a method developed in those papers made it possible to allow for the effect of the magnetostatic and exchange interactions on the domain structure and to determine the limits of applicability of the model of geometric DW.

Here we will discuss the stability of an isolated stripe domain bounded by unipolar and bipolar domain walls against a step magnetic-field pulse. We will also comparatively analyze the stability of the given structures, determine the relationship that exists between the critical amplitude of the pulse and the initial value of the magnetizing field, and analyze the dependence of this relationship on the parameters of the material.

2. DYNAMICAL EQUATION FOR THE PARAMETERS OF A STRIPE DOMAIN IN A THIN FILM

The magnetic energy functional,

$$E=\int_V dV\left\{\frac{\alpha}{2}\left(\frac{\partial \mathbf{M}}{\partial \mathbf{x}}\right)^2-\frac{\beta}{2}M_z^2-H_zM_z\right\}+E_m,$$
$$E_m=\frac{1}{2}\int_V\int_V dV\,dV'\;M_i(\mathbf{r}')M_j(\mathbf{r})\frac{\partial^2}{\partial x_i'\,\partial x_j}\frac{1}{|\mathbf{r}-\mathbf{r}'|},\tag{1}$$

is our starting point in describing the properties of an isolated stripe domain. Here E_m is the energy of the intrinsic magnetic field, repeated indices imply summation, and α and β are the constants of exchange interaction and uniaxial anisotropy, respectively. The anisotropy axis is parallel to the z axis and is perpendicular the xy plane of the film, and H_z is the external field, which is perpendicular to the film plane.

Since the magnetization vector $\mathbf{M}$ in nonconducting ferromagnetic materials meets the condition $\mathbf{M}^2=M_0^2$, we can express its components in terms of the angles in a spherical system of coordinates:

$$\begin{pmatrix} M_x\\ M_y\\ M_z\end{pmatrix}=M_0\begin{pmatrix}\sin\theta\cos\varphi\\ \sin\theta\sin\varphi\\ \cos\theta\end{pmatrix},\tag{2}$$

where θ and φ are the polar and azimuthal angles in a system of coordinates in which the z axis is the polar axis and the angle φ is measured from the x direction.

We will assume that an isolated stripe domain oriented parallel to the y axis, so that the magnetization is uniformly distributed along this direction.

As shown in Ref. 14, in thin ferromagnetic films of thickness L whose value is such that $(L/\Lambda)^2\ll 16$ (with $\Lambda=\sqrt{\alpha/4\pi}$ the magnetic length), the effects of DW twisting are very weak and hence we can assume that the DW structure is approximately uniform in thickness.

Thus, reasoning within these approximations, we conclude that the angles φ and θ depend only on one spatial variable, x, which makes all calculations much simpler.

Integrating in (1) with respect to y and z, we get

$$E=M_0^2S\int dx\,W(\theta,\varphi),\tag{3}$$

where S is the DW area, and

$$W(\theta,\varphi)=\frac{\alpha}{2}\left(\frac{\partial\theta}{\partial x}\right)^2+\frac{\alpha}{2}\sin^2\theta\left(\frac{\partial\varphi}{\partial x}\right)^2-\frac{\beta}{2}\cos^2\theta$$
$$-\left(h_z-\frac{h_z^m(x)}{2}\right)\cos\theta+\frac{h_x^m(x)}{2}\cos\varphi\sin\theta\tag{4}$$

is the magnetic energy density of the system. The components of the magnetostatic field $h_i^m(x)$ are

$$h_x^m(x)=4\pi\cos\varphi\sin\theta-\int dx'\,J(x-x')\cos\varphi(x')\sin\theta(x'),$$

$$h_z^m(x)=\int dx'\,J(x-x')\cos\theta,$$

$$J(x-x')=\frac{2}{L}\ln\left(\frac{L^2+(x-x')^2}{(x-x')^2}\right).$$

To describe the magnetization state of the system with the energy density given by (4), we write the Landau–Lifshitz equations in terms of angular variables:

$$-l^2\frac{\partial^2\theta}{\partial x^2}+\left[1+l^2\left(\frac{\partial\varphi}{\partial x}\right)^2\right]\sin\theta\cos\theta$$
$$=-\varepsilon h_x^m\cos\varphi\cos\theta-\varepsilon\left(h_z-h_z^m-\frac{\partial\varphi}{\partial\tau}\right)\sin\theta,\tag{5a}$$

$$l^2\frac{\partial}{\partial x}\sin^2\theta\frac{\partial\varphi}{\partial x}=-\varepsilon\left(h_x^m\sin\varphi-\frac{\partial\theta}{\partial\tau}\right)\sin\theta,\tag{5b}$$

where $\varepsilon=\beta^{-1}\ll 1$, $l=\sqrt{\alpha/\beta}$ and $\tau=t\omega_0$; here $\omega_0=2\mu_0M_0/\hbar$ and μ_0 is the Bohr magneton.

The right-hand side of Eq. (5b) contains a small parameter, with the result that one of the partial solutions of this equation in the zeroth approximation in ε is independent of the variable x. We examine this solution in greater detail and determine the state of the isolated stripe domain that corresponds to it.

We formally express φ and θ as functions of x and τ:

$$\varphi(x,\tau)=\varphi_0(\tau)+\varepsilon\varphi_1(x,\tau)+\cdots,$$
$$\theta(x,\tau)=\theta_0+\varepsilon\theta_1+\cdots.\tag{6}$$

Here the correction φ_1 can be found by solving the equation

$$l^2\frac{\partial}{\partial x}\sin^2\theta_0\frac{\partial\varphi_1}{\partial x}=-\left(h_{0x}^m\sin\varphi_0-\frac{\partial\theta_0}{\partial\tau}\right)\sin\theta_0.\tag{7}$$

As for Eq. (5a), standard perturbation-theory techniques cannot be used in the solution, since the external field dramatically changes the system state, with the result that the right-hand side of Eq. (5a) is a singular perturbation.

To solve Eq. (5a), we use the perturbation regularization method.[13,14] To this end we add to the right- and left-hand sides of Eq. (5a) an effective operator $\pm\varepsilon h\sin\theta$ (here h is a parameter that is generally time-dependent) and write Eq. (5a) as follows:

$$-l^2\frac{\partial^2\theta}{\partial x^2}+\sin\theta\cos\theta\pm\varepsilon h\sin\theta$$
$$=-\varepsilon h_x^m\cos\varphi\cos\theta+\varepsilon\left(-h_z+h_z^m\pm h+\frac{\partial\varphi}{\partial\tau}\right)\sin\theta.\tag{8}$$

Assuming that the right-hand side of Eq. (8) can be regularized via appropriate selection of the parameter h, we can determine θ_0 by solving the equation

$$-l^2\frac{\partial^2\theta_0}{\partial x^2}+\sin\theta_0\cos\theta_0\pm\varepsilon h\sin\theta_0=0\tag{9}$$

with the boundary conditions $\partial\theta_0/\partial x|_{x\to\pm\infty}=0$.

The correction θ_1 can be found by solving the inhomogeneous linear equation

$$\hat{G}_1(\theta_0)\theta_1=F(\theta_0,\varphi_0),\tag{10}$$

$$\hat{G}_1(\theta_0)=-l^2\frac{\partial^2}{\partial x^2}+\cos 2\theta_0\pm\varepsilon h\cos\theta_0,$$

where

$$F(\theta_0,\varphi_0)=-h_{0x}^{m}\cos\varphi_0\cos\theta_0+\left(-h_z+h_{0z}^{m}\pm h+\frac{\partial\varphi_0}{\partial\tau}\right)\sin\theta_0 .$$

When the parameter h is positive, the boundary-value problem (9) has two different solutions, depending on the sign in front of h:

$$\theta_0(X,h)=\begin{cases}\pi+2\arctan\left\{\sqrt{\dfrac{\varepsilon h}{1+\varepsilon h}}\sinh\left(\sqrt{1+\varepsilon h}\dfrac{X}{l}\right)\right\}\text{ for }+h, & (11a)\\ \pi-2\arctan\left\{\sqrt{\dfrac{\varepsilon h}{1-\varepsilon h}}\cosh\left(\sqrt{1-\varepsilon h}\dfrac{X}{l}\right)\right\}\text{ for }-h, & (11b)\end{cases}$$

with $X=x-x_0$.

An analysis of the results shows that the solution (11a) describes an isolated stripe domain as a bound state of unipolar Bloch walls. Since for the given structure the angle θ_0 changes by 2π when $-\infty$ changes to $+\infty$, from now on we will say that (11a) corresponds to a 2π-DW.

The total variation of the angle θ_0 for (11b) is equal to zero, as a result of which we define the structure of the given isolated stripe domain as that of a $0-$DW. This corresponds to a bound state of bipolar Bloch walls.

The expressions (11a) and (11b) for the ground state of both types of domain structures contain an undefined parameter h. This parameter must be selected in such a way that no secular terms remain on the right-hand side of Eq. (10). This is achieved by satisfying the solvability condition

$$\int dx\,\psi(x)F(\theta_0,\varphi_0)=0, \tag{12}$$

where $\psi(x)$ is the general homogeneous solution of Eq. (10).

Clearly, in the zeroth approximation in ε there are two such solutions:

$$\psi_1(x)=l\,\frac{\partial\theta_0}{\partial x},\quad \psi_2(x)=\sin\theta_0 . \tag{13}$$

If we define the width of an isolated stripe domain as[10]

$$d=\frac{1}{2}\int dx\,(1-\cos\theta_0), \tag{14}$$

there is a one-to-one correspondence between the parameter h and the domain width d, according to which

$$\varepsilon h=\cosh^{-2}\frac{d}{2l}\ \text{ for 0-DW},$$

$$\varepsilon h=\sinh^{-2}\frac{d}{2l}\ \text{ for }2\pi\text{-DW}. \tag{15}$$

Substituting (13) in (12) and employing analytical integration methods, we obtain to within terms proportional to l/d dynamical equations that establish a relationship between the width of an isolated stripe domain and the rate of change of the azimuthal angle:

$$\frac{\partial\varphi_0}{\partial\tau}-h_z-\beta\cosh^{-2}\frac{d}{2l}+8\arctan\frac{L}{d}-4\frac{d}{L}\ln\left[1+\left(\frac{L}{d}\right)^2\right]=0 \tag{16a}$$

for 0-DW, and

$$\frac{\partial\varphi_0}{\partial\tau}-h_z+\beta\sinh^{-2}\frac{d}{2l}+8\arctan\frac{L}{d}-4\frac{d}{L}\ln\left[1+\left(\frac{L}{d}\right)^2\right]=0 \tag{16b}$$

for 2π-DW.

To obtain a closed system of dynamical equations for the parameters of the domain structure, we integrate the right- and left-hand sides of Eq. (7) with respect to x from $-\infty$ to $+\infty$. Since the left-hand side vanishes in the process, the second dynamical equation is determined by the condition

$$\int_{-\infty}^{\infty}dx\left(h_{0x}^{m}\sin\varphi_0-\frac{\partial\theta_0}{\partial\tau}\right)\sin\theta_0=0. \tag{17}$$

Bearing in mind that for $l/d\ll 1$ we have

$$\frac{\partial\theta_0}{\partial\tau}=\frac{\partial\theta_0}{\partial h}\frac{\partial h}{\partial\tau}\approx-\frac{1}{2}\sin\theta_0\frac{\partial(d/l)}{\partial\tau},$$

we find that

$$\frac{1}{l}\frac{\partial d}{\partial\tau}+\Omega_1\sin 2\varphi_0=0, \tag{18}$$

where

$$\Omega_1=4\pi\left\{1-\frac{\pi l}{8L}\left[\ln\frac{L^2}{l^2}\mp\ln\left(1+\frac{L^2}{d^2}\right)-a\right]\right\},$$

$$a=\frac{8}{\pi^2}\int_{-\infty}^{\infty}dx\,\frac{x\ln x}{\sinh x}\approx 0.174.$$

In deriving (18) we discarded the terms proportional to $l/d, l^2/L^2\ll 1$. Here and in what follows the upper signs in the expression with $(\pm,\mp)$ correspond to 2π-DW and the lower signs to 0-DW. Since we have $L^2/l^2\gg L^2/d^2$, the effect of the term following "$\mp$" in the expression for Ω_1 can be ignored. Then the coefficient Ω_1 can be accurately written

$$\Omega_1=4\pi\left(1-\frac{\pi l}{4L}\ln\frac{L}{l}\right). \tag{19}$$

Obviously, in this approximation Eq. (18) is the same for 0-DW and 2π-DW.

On the basis of the earlier remarks we note that the range of applicability of the theory in the thickness of ferromagnetic fields is bounded above and below:

$$l^2\ll L^2\ll 16\Lambda^2. \tag{20}$$

Thus, for both types of isolated stripe domains localized in thin films we have a system of dynamical equations for the parameters of the domain structure [(16) and (18)] that are similar to the Slonczewski equations for 180°-DW (see Ref. 1). Our theory can be used to describe the dynamics of a stripe domain in magnetic bubble materials whose thickness satisfies the condition (20).

Note that for systems of a different type, systems that model the situation in yttrium–iron garnets, the dynamical

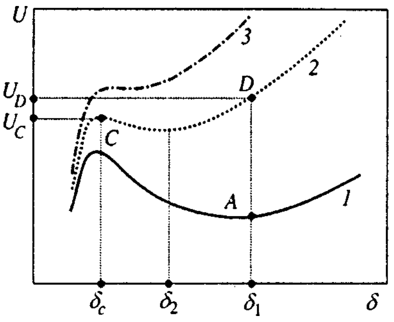

FIG. 1. The δ-dependence of the potential energy of a 0-DW for various values of the magnetic field.

equations in the approximation of ultrathin films ($L \ll l$) with easy-axis magnetization in the film plane were derived by Kovalev *et al.*[15,16] who carried out a full investigation of the dynamics of two DWs coupled by the magnetostatic interaction.

We will now employ the dynamical equations derived in Sec. 2 to study the state of 0-DW and 2π-DW.

Using Eq. (18), we can exclude the variable $\varphi_0(\tau)$ from Eqs. (16a) and (16b). If $h_z = \text{const}$, the first constant of motion can easily be found. Introducing the nation $\delta = d/l$ and $\lambda = L/l$, we can express it as

$$T(\dot{\delta}) + U(\delta, h_z) = E, \tag{21}$$

where

$$T(\dot{\delta}) = \frac{\Omega_1}{2}\left(1 - \sqrt{1 - \left(\frac{\dot{\delta}}{\Omega_1}\right)^2}\right)$$

is the kinetic energy. An analysis of the expression for the kinetic energy shows that there is a limit in the rate of variation of the width of an isolated stripe domain and that this limit is close to twice the Walker velocity. In the adopted notation, this limit is $\dot{\delta}_c = \Omega_1$. The potential energy of an isolated stripe domain, $U(h_z, \delta)$, in dimensionless units is given by the formula

$$U(h_z, \delta) = \delta\left(h_z - 8 \arctan\frac{\lambda}{\delta}\right) - 2\lambda \ln\left(1 + \frac{\delta^2}{\lambda^2}\right) + 2\frac{\delta^2}{\lambda}\ln\left(1 + \frac{\lambda^2}{\delta^2}\right) + \Delta U(\delta), \tag{22}$$

where

$$\Delta U(\delta) = \begin{cases} 2\beta \tanh(\delta/2) & \text{for 0-DW}, \\ 2\beta \coth(\delta/2) & \text{for } 2\pi\text{-DW} \end{cases}$$

is the contribution to the potential energy of an isolated stripe domain that arises due to the allowance for an inner structure of DW.

3. EFFECT OF THE MICROMAGNETIC STRUCTURE ON THE STABILITY OF A 0-DW IN A PULSED FIELD

To be specific we begin with a 0-DW. The δ-dependence of the potential energy of a 0-DW on δ is depicted in Fig. 1 by the curves *1–3*, which correspond to various values of the magnetic field.

At certain values of the magnetic field the potential energy of a 0-DW has two extrema at points that can be found from the condition

$$0 = \frac{\partial U(h_z, \delta)}{\partial \delta} = h_z + \beta \cosh^{-2}\frac{\delta}{2} - 8 \arctan\frac{\lambda}{\delta} + 4\frac{\delta}{\lambda}\ln\left(1 + \frac{\lambda^2}{\delta^2}\right). \tag{23}$$

Obviously, in weak fields, when the width of an isolated stripe domain is large, the second term on the right-hand side of Eq. (23) is negligible, so that (23) corresponds to the result obtained in geometric domain wall models.[4]

We will assume that in the initial stationary state the system is in a magnetizing field h_{0z}. In curve *1* of Fig. 1 this state is denoted by A. As the strength of the external field is increased, the dependence of the potential energy changes (curves *2* and *3*, respectively). If the field amplitude increases slowly, the system has time to reach (due to dissipative processes) a state of equilibrium determined by the new position of minimum in the potential energy. As the field tends to h_c (curve *3*), the minimum in the energy disappears and the 0-DW becomes unstable.

If a step pulse Δh_z of magnetic field is applied to the system, the δ-dependence of the magnetic field suddenly changes (curve *2* now) while the domain width remains the same. Here the new state of the system corresponds to the point D. Then the system begins to move, and the domain width δ changes. If U_D, the coordinate of point D on the energy axis, is smaller than U_C, the potential energy of the 0-DW at the maximum, then the state may remain stable, provided that the rate of variation of the domain width does not reach the critical value Ω_1 and the point δ_2 of the potential-energy minimum. Here the motion of a 0-DW corresponds to periodic pulsations. If $U_D > U_C$ holds (as in the case depicted in Fig. 1), the 0-DW becomes unstable and collapses.

On the basis of the above reasoning we can formulate the conditions for the stability of a 0-DW against a magnetic-field pulse:

$$U(h_{0z} + \Delta h_z, \delta_c) - U(h_{0z} + \Delta h_z, \delta_1) \leqslant 0, \tag{24a}$$

$$U(h_{0z} + \Delta h_z, \delta_1) - U(h_{0z} + \Delta h_z, \delta_2) \leqslant (\Omega_1/2), \tag{24b}$$

where (24a) is the condition for 0-DWstability against collapse, and (24b) is the condition of 0-DW stability against dynamical transformations that appear when the DW reaches the critical rate Ω_1. Note that δ_1 is determined by solving Eq. (23) corresponding to the energy minimum in the field h_{0z}, and the quantities δ_2 and δ_c constitute the solution of the same equation for the field $h_{0z} + \Delta h$ and correspond to the positions of the minimum and maximum in the potential energy of the domain.

Figure 2 illustrates the graphical solution of the system of inequalities (24a) and (24b), which determines the region of 0-DW stability in the $(h_{0z}, \Delta h_z)$ plane, for a film with $\lambda = 10$ and $\beta = 100$.

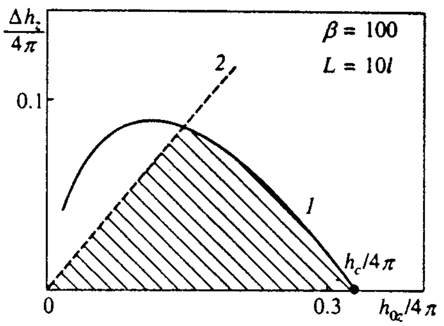

FIG. 2. Graphical solution of the inequalities (24).

Above the curve *2* lies the region corresponding to values of the magnetizing field h_{0z} and the amplitude Δh_z of the pulsed signal for which the rate of variation of the domain width reaches its critical value Ω_1 and an isolated stripe domain becomes unstable against dynamical transformations. We will call h_{0z} and Δh_z the parameters of the field. When the field parameters are below the curve *2*, there is no dynamical 0-DW instability.

The curve *1* divides the plane in Fig. 2 into two parts. The region below *1* corresponds to values of the field parameters at which the 0-DW width does not reach the critical value δ_c (Fig. 1) and the structure of an isolated stripe domain remains stable. Thus, the hatched region in Fig. 2 designates the region where a stable 0-DW exists.

It is clear that the extent of the region of stability depends sensitively on the magnetic film thickness. The traces shown in Fig. 3 give an upper bound for the 0-DW region of stability in films of various thickness L for $\beta=100$.

4. EFFECT OF THE MICROMAGNETIC STRUCTURE OF DOMAIN WALLS ON THE STABILITY OF A 2π-DW

In contrast to the potential energy of a 0-DW, the potential energy of a 2π-DW has only one extremal point, which corresponds to the minimum of the energy (Fig. 4). The coordinate of this point on the δ axis is determined by the condition

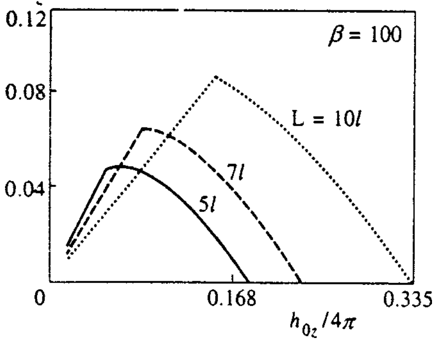

FIG. 3. Regions of 0-DW stability for various values of the thickness of the magnetic film.

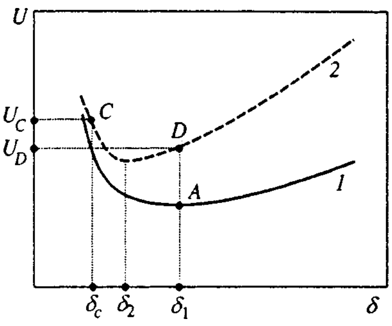

FIG. 4. The δ-dependence of the potential energy of a 2π-DW.

$$0=\frac{\partial U(h_z,\delta)}{\partial\delta}=h_z-\beta\sinh^{-2}\frac{\delta}{2}-8\arctan\frac{\lambda}{\delta}+4\frac{\delta}{\lambda}\ln\left(1+\frac{\lambda^2}{\delta^2}\right). \tag{25}$$

As in the 0-DW case, we will assume that the system in the field h_{0z} is in a stationary state corresponding to point *A* and passes into state *D* when a pulse of magnetic field is applied.

In our case the condition for the stability of the structure of a 2π-DW against dynamical transformations that occur when the critical rate is reached is similar to the condition (24) for a 0-DW. At the same time, as shown in Ref. 13, the decrease in the width of the 2π-DW to a certain critical value δ_c (Fig. 4) induces inhomogeneous longitudinal perturbations of the DW magnetizations. Thus, the 2π-DW becomes unstable if the energy U_D acquired as a result of the action of a field pulse is sufficiently large to reduce the width to δ_c. Here, for an unstable state to be reached, the point δ_c must coincide with the turning point, at which the kinetic energy of the 2π-DW vanishes.

To determine the criterion for the stability of a 2π-DW under compression, we will study the corresponding solution of (11a), (16b), and (18) at the turning point δ_c for stability against small perturbations.

At the turning point $\dot{\delta}=0$ holds, as a result of which Eq. (18) implies that there is a solution that corresponds to $\varphi_0=\pi/2$. Then the system of equations describing the small perturbations $(\delta\theta,\delta\varphi)$ superposed on the ground states (11a) determined at the turning point has the form

$$\begin{aligned}&\hat{G}_1(\theta_0)\,\delta\theta=\varepsilon\frac{\partial}{\partial\tau}\delta u,\\ &\hat{G}_2(\theta_0)\,\delta u=-\varepsilon 4\pi\delta u+\varepsilon\int dx'\,J(x-x')\,\delta u(x')-\varepsilon\frac{\partial}{\partial\tau}\delta\theta,\end{aligned} \tag{26}$$

where

$$\delta u=\sin\theta_0\,\delta\varphi,\quad \hat{G}_2(\theta_0)=\hat{G}_1(\theta_0)+\sin^2\theta_0-l^2\left(\frac{\partial\theta_0}{\partial x}\right)^2.$$

Clearly, to within terms proportional to ε, the solutions of Eqs. (26) can be written[13]

$$\begin{aligned}&\delta\theta=C_1\exp\{i\omega t\}\,\psi_1(x)+C_2\exp\{i\omega t\}\,\psi_2(x),\\ &\delta u=C_3\exp\{i\omega t\}\,\psi_1(x)+C_4\exp\{i\omega t\}\,\psi_2(x),\end{aligned} \tag{27}$$

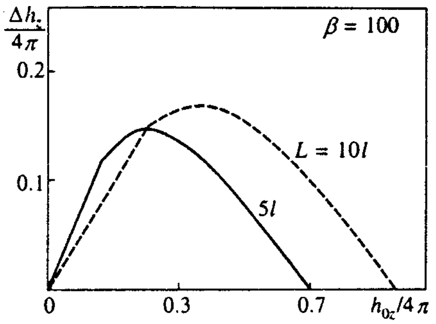

FIG. 5. Graphical solution of the inequalities (30a) and (30b) for films of various thickness.

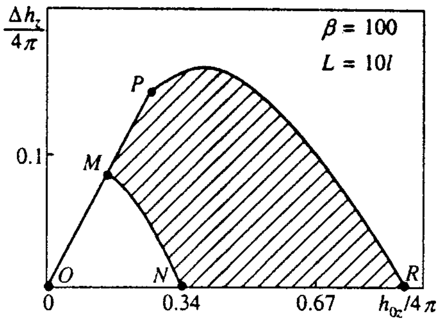

FIG. 6. Regions of 0-DW and 2π-DWstability.

where ψ_1 and ψ_2 are the eigenfunctions (13) of the operators G_1 and G_2, which have eigenvalues of order ε.

Substituting $\delta\theta$ and $\delta\varphi$ in the form (27) into Eqs. (26) and finding the scalar products of the result by the corresponding eigenfunctions (13), we arrive at a system of algebraic equations for the expansion coefficients C_i:

$$\omega C_3 = 0,$$
$$i\omega C_1 + [\Omega_1 - 2\beta \sinh^{-2}(\delta_c/2)]C_3 = 0, \tag{28a}$$
$$\Omega_2 C_2 - i\omega C_4 = 0,$$
$$i\omega C_2 + \Omega_1 C_4 = 0, \tag{28b}$$

where $\Omega_2 = 2\beta \sinh^{-2}(\delta_c/2) + (4/\lambda)\ln(1+\lambda^2/\delta_c^2)$.

The system of equations (28b) has nontrivial solutions for $\omega = \sqrt{\Omega_1\Omega_2}$, a condition that determines the natural frequency of small pulsation perturbations of the 2π-DWat the turning point δ_c. Since the natural frequency is real and positive for all values of δ, the 2π-DW is always stable against such perturbations.

In our case the system of equations (28a) is the most interesting. Clearly, the nontrivial solutions $C_3 \neq 0$ of this system result if

$$\omega = 0, \quad \Omega_1 - 2\beta \sinh^{-2}(\delta_c/2) = 0. \tag{29}$$

An analysis shows that these relationships determine the condition for the occurrence of an instability against deviation from the Bloch orientation of magnetization in the walls bounding the domain.

On the basis of these results we can formulate a criterion for conservation of 2π-DW stability in a pulsed field:

$$U(h_{0z}+\Delta h_z, \delta_c) - U(h_{0z}+\Delta h_z, \delta_1) \leq 0, \tag{30a}$$

$$U(h_{0z}+\Delta h_z, \delta_1) - U(h_{0z}+\Delta h_z, \delta_2) \leq T(\dot{\delta}_c). \tag{30b}$$

Obviously, the systems of inequalities determining the stability regions for 0-DW and 2π-DW are similar. The differences are related to the differences in the dependence of the potential energy of the given systems and in the ways in which the critical widths of domains are determined. For instance, for a 0-DW the critical width is determined by Eq. (23), while for a 2π-DW the critical width is determined by Eq. (29).

Figure 5 depicts the graphical solutions of the inequalities (30a) and (30b), which determine the regions of stability of the 2π-DW structure in the $(h_{0z}, \Delta h_z)$ plane. (The stability regions are bounded above by the corresponding curves.)

5. COMPARATIVE ANALYSIS OF STABILITY OF A 0-DW AND A 2π-DW IN A PULSED FIELD

In Fig. 6 we compare the regions of stability of the structures under investigation for a film with a thickness $\lambda = 10$ and $\beta = 100$. The curve OMN bounds the region of 0-DW stability from above, while the curve OPR bounds the region of 2π-DWstability from above.

Figure 6 shows if the domains of different structure are initially placed in a weak magnetizing field, their response to a magnetic-field pulse is almost the same. Both structures examined in the present paper become unstable against dynamical transformations when the critical rate is reached.

On the other hand, if the magnetizing field is high, the structures examined in the present paper exhibit differences, which are related to the organization of the inner structure of the domain walls. The curve $RPMN$ bounds the region of the field parameters in which a 0-DW disappears but a 2π $=$DW remains stable.

Obviously, in thin films, for certain values of the field parameters h_{0z} and Δh_z a 2π-DW is more stable than a 0-DW. In thick films, due to DW twisting, the states of the structures near the surface are equivalent, so that the differences in the stability of 0-DW and 2π-DW in thin fields manifest themselves less vividly.

The conclusions of the theory developed here were confirmed by the results of experiments[17–19] in which the possibility of using vertical Bloch lines as memory elements was investigated. These experiments revealed that the stability of a 2π-DW is higher than that of a 0-DW.

The author is grateful to Prof. Yu. I. Gorobets for discussing the various aspects of the present research.

[1] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Applied Solid State Science Series, Supplement I, Academic Press, New York (1979).

[2] A. H. Bobeck and E. Della Torre, *Magnetic Bubbles*, North-Holland, Amsterdam (1975).

[3] V. G. Bar'yakhtar and Yu. I. Gorobets, *Magnetic Bubbles and Their Lattices* [in Russian], Naukova Dumka, Kiev (1988).

[4] A. H. Bobeck, Bell Syst. Tech. J. **46**, 1901 (1967).

[5] V. G. Bar'yakhtar, V. V. Gann, and Yu. I. Gorobets, Zh. Tekh. Fiz. **45**, 386 (1975) [Sov. Phys. Tech. Phys. **20**, 240 (1975)].

[6] F. L. Vaĭsman, Yu. I. Gorobets, and S. I. Denisov, Ukr. Fiz. Zh. (Russ. Ed.) **31**, 1234 (1986).

[7] A. P. Babichev, Yu. I. Gorobets, and S. I. Denisov, Ukr. Fiz. Zh. (Russ. Ed.) **30**, 86 (1985).
[8] N. N. Kudel'kin, V. V. Randoshkin, and G. E. Khodenkov, Pis'ma Zh. Tekh. Fiz. **9**, 1357 (1983). [Sov. Tech. Phys. Lett. **9**, 583 (1983)].
[9] G. E. Khodenkov, Fiz. Met. Metalloved. **61**, 850 (1986).
[10] Yu. I. Dzhezherya, Fiz. Tverd. Tela (St. Petersburg) **35**, 2270 (1993) [Phys. Solid State **35**, 1366 (1993)].
[11] Yu. I. Dzhezherya and A. M. Yakovenko, Fiz. Tverd. Tela (St. Petersburg) **37**, 2444 (1995) [Phys. Solid State **37**, 1338 (1995)].
[12] Yu. I. Dzhezherya and I. K. Loktionov, Fiz. Tverd. Tela (St. Petersburg) **39**, 671 (1997) [Phys. Solid State **39**, 586 (1997)].
[13] Yu. I. Gorobets and Yu. I. Dzhezherya, Fiz. Tverd. Tela (St. Petersburg) **40**, 269 (1998) [Phys. Solid State **40**, 243 (1998)].
[14] Yu. I. Dzhezherya, Zh. Éksp. Teor. Fiz. **115**, 1315 (1999) [JETP **88**, 726 (1999)].
[15] A. S. Kovalev, A. M. Kosevich, and I. V. Manzhos, Zh. Éksp. Teor. Fiz. **94**(11), 222 (1988) [Sov. Phys. JETP **67**, 2301 (1988)].
[16] A. S. Kovalev, A. M. Kosevich, I. V. Manzhos, and K. V. Maslov, JETP Lett. **44**, 222 (1986).
[17] F. E. Yurchenko, Zarubezh. Élektron. Tekh. No. 5(336), 5 (1989).
[18] Y. Hidaka *et al.*, IEEE Trans. Magn. **MAG-19**, 1841 (1983).
[19] S. Konishi, N. Shibata, M. Shimaya, Y. Fukuzawa, and K. Narita, J. Appl. Phys. **50**, 7841 (1979).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Electronic properties of C_{60} single crystals doped with lithium by electrodiffusion

A. V. Bazhenov, S. I. Bredikhin, V. V. Kveder,*) Yu. A. Ossipyan, R. K. Nikolaev,
T. N. Fursova, and A. I. Shalynin

Institute for Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 13 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1706–1722 (November 1999)

Diffusion of lithium cations in C_{60} single crystals driven by electric field has been detected and studied. A novel technique for fullerene crystal doping based on injection of ions through a ''superionic crystal/C_{60} single crystal'' heterojunction has been suggested. It has been found that lithium doping of C_{60} single crystals brings about an ESR signal, and this signal as a function of time has been investigated. The electronic conductivity in Li_xC_{60} crystals has a nonmetallic nature. Reflection spectra measured in the IR band have shown that the reflectivity due to free electrons gradually decreases with time, which correlates with the evolution of signals due to ESR and microwave conductivity. Lithium doping of crystals increases the oscillator strength of the $T_{1u}(4)$ vibrational mode and shifts it to lower frequencies (from $1429\,\mathrm{cm}^{-1}$ to $1413\,\mathrm{cm}^{-1}$), which indicates that one electron is present at the C_{60} molecule, and this fact may be treated as evidence that the LiC_{60} phase is generated in a C_{60} crystal. © *1999 American Institute of Physics.* [S1063-7761(99)01511-5]

1. INTRODUCTION

Given the almost spherical shape of the C_{60} molecule and the weak intermolecular interaction, one can classify C_{60} crystals as typical molecular solids, whereas their electronic properties are more like those of semiconductors. Indeed, measurements of their conductivity,[1–3] optical absorption spectra,[4,5] and photoconductivity[6] provide evidence that C_{60} in the crystalline form is a semiconductor with relatively narrow (about 0.5 eV) energy bands and a band-gap width of about 2.2 eV. The valence band of the crystal is formed by A_g orbitals of the C_{60} molecules, which are fully populated by electrons, whereas the conduction band is formed from three almost degenerate, unoccupied T_{1g}, T_{2g}, and G_g orbitals of the C_{60} molecule.

On the basis of the simple one-electron model, one may suppose that the crystalline C_{60} should demonstrate metallic properties at fairly high doping levels (for example, with alkali metals). The small widths of energy bands, the large contribution of Coulomb correlations, and the possibility of strong Jahn–Teller effects caused by molecule deformations, however, can cause a failure of the model of ''rigid'' one-electron bands, which is applied to such systems in many cases. In 1991 thin films of C_{60} fullerenes doped with alkali metals, namely, A_xC_{60}, where A=Li, K, Na, Rb, and Cs,[7] were fabricated for the first time. It was demonstrated later that in A_xC_{60} (with A=K, Rb, Cs and at $x=1$, 3, 4, 6) an electron transfers from the alkali metal to the C_{60} molecule. This means that the triply degenerate unoccupied orbital responsible for formation of the conduction band is populated. In the case of the half-filled conduction band, which occurs at $x=3$, the material is a metal and undergoes a superconducting transition at low temperatures.[8–11] But at $x=4$, probably owing to the Jahn–Teller deformation of molecules,[12] a gap opens around the Fermi level and A_xC_{60} demonstrates dielectric properties. It is also noteworthy that, depending on the fabrication conditions of A_xC_{60}, various stable and metastable phases have been detected.[13] For example, at $x=1$ and with A=K, Rb, and Cs, monomer, dimer, and polymer phases were clearly distinguished using the IR spectroscopy. All this has led us to the conclusion that our understanding of properties of even such extensively investigated materials as A_xC_{60} (A=K, Rb, and Cs and at $x=1,3,4$) is far from clear at the present time.

It is obvious that, for clearer understanding of processes in A_xC_{60} materials, it is advisable to study properties of C_{60} crystals lightly doped with alkali metals. Such samples, however, are difficult to fabricate using conventional diffusion of alkali metals because the system has a tendency to decompose into stable phases with integer x. It is no less interesting to study A_xC_{60} materials where A stands not only for K, Rb, and Cs, but may denote a wide range of other metals. Such systems have been little studied thus far because of difficulties associated with sample fabrication.

In this paper we suggest a nontraditional approach to doping of C_{60} crystals with metal ions by injecting them through a ''superionic crystal/C_{60} crystal'' heterojunction, which we call the electrodiffusion technique. In superionic conductors, the ionic conductivity is usually five to six orders of magnitude higher than their electronic conductivity, so it seemed to us highly probable that an electric current fed through a heterojunction should inject mobile ions into a fullerene crystal. It is known that at $T>260\,\mathrm{K}$ C_{60} crystals have a face-centered cubic (fcc) structure with a lattice constant of 1.4 nm. This structure typically has octahedral voids, which form three-dimensional networks of channels aligned with [111] crystal axes. This was expected to lead to high diffusion coefficients for most metallic atoms, which might permit easy doping of fullerenes with metals when an ionic current is fed through them.

In the presence of electric field $E(r)$, the diffusion equation has the form

$$J=ND(eE+\nabla\mu)/kT, \tag{1}$$

$$dN(r)/dt=-\operatorname{div}J, \tag{2}$$

where $N(r)$ is the ion concentration at a point with coordinate r, D is their diffusion coefficient, $J(r)$ is their current density, T is the temperature, and $\mu(r)$ is the chemical potential of ions at the given point. At low concentrations N, when the interaction among ions can be neglected, $\mu=kT\ln N$, and Eq. (1) reduces to

$$J=NeDE/kT-D\nabla N. \tag{3}$$

If $eEN\gg kT\nabla N$, the electric field notably accelerates the diffusion into a sample and allows one to execute an almost uniform doping of a large sample, even when the diffusion coefficient is small. At large N, the function $\mu(N)$ is nonmonotonic, which leads in the case of conventional diffusion to decomposition of a sample into phases with specific compositions and makes fabrication of homogeneous samples impossible. In this case, the presence of an intense electric field should notably suppress (although not eliminate completely) the decomposition of the material into different phases.

We selected Li ions for our first experiments because the Li_xC_{60} composition was the least studied among all fullerenes doped with alkali metals. The aim of the reported work is investigation of current–voltage characteristics of C_{60}–Li_7SiPO_8 heterojunctions, a feasibility study of C_{60} doping with metal ions using the electrodiffusion technique, and measurements of electronic properties of doped C_{60} crystals.

2. SAMPLES AND EXPERIMENTAL TECHNIQUES

C_{60} single crystals were grown at Institute for Solid State Physics, Russian Academy of Sciences, by the technique of physical vapor-phase transport (sublimation) under a temperature gradient in a sealed-off cell at temperatures ranging from 600 to 640 °C. The starting material C_{60} had a purity of no less than 99.98% after purification by the chromatographic technique and elimination of solvent traces through multiple sublimation in vacuum. For our measurements, we selected single crystals shaped as thin plane-parallel plates with natural faces. The crystal faces were planes of type (111). The sample thickness varied between 0.7 and 2 mm, and the other two dimensions were usually 2 to 4 mm.

In order to inject Li ions, electrodes were mechanically attached to opposite (111) faces of a C_{60} crystals. The cathode was graphite, and the anode was a two-layer $Li_{0.2}WO_3|Li_7SiPO_8$ electrode, which has a high ionic conductivity for lithium cations and cuts off the electronic component of electric current. In accordance with the above, utilization of cells like

$$(+)Li_{0.2}WO_3|Li_7SiPO_8|C_{60}\,\text{crystal}|\text{Graphite}|(-) \tag{4}$$

allowed us to inject lithium cations into C_{60} single crystals and dope samples under investigation in a controlled manner. In the process of doping, cell (4) was placed on a quartz substrate in a sealed-off thermostat. A voltage was fed to a cell in order to measure its current–voltage characteristics and the current as a function of time at various temperatures. The substrate temperature could be tuned over a range of −150 to 500 °C. Most experiments were performed at a voltage across a cell of 10 to 200 V and at temperatures of 400–520 K. The typical current fed across the cell was 2–20 μA. After the process of electrodiffusion (its time varied from one hour to 100 h), the sample was rapidly transported to a quartz ampoule and stored in liquid nitrogen before further experiments.

In order to observe the diffusion process visually, we used an MBS-4 microscope equipped with a silicon CCD camera. In the process of diffusion, a sample was illuminated from below by monochromatic radiation obtained by feeding light from a halogen lamp through a monochromator. The sample image generated by the CCD camera was sent to a computer, and some frames separated by a predetermined time interval (usually 20 to 60 s) were stored on the computer hard disk for further processing. This computer also controlled the temperature and recorded the current fed through the cell as a function of time.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Electrodiffusion of lithium and absorption spectra in the near IR band

An electric field was applied to cell (4) so that the electrode conducting Li ions was positive, and as a result, Li cations were injected into the C_{60} crystal via the $Li_{0.2}WO_3|Li_7SiPO_8|C_{60}$ crystal heterojunction, and electrons via the $C_{60}|$graphite heterojunction. Measurements of current–voltage characteristics at various temperatures have demonstrated that, in the initial stage of C_{60} single crystal doping, the electrodiffusion of Li cations is an activated process with an activation energy $\Delta E\approx 0.93$–0.96 eV. This energy can probably be associated with the activation energy of lithium diffusion in C_{60} single crystals. Saturation of C_{60} crystals with lithium leads to an increase in the Li chemical potential (μ_{Li}) in the fullerene crystal, and the voltage across the cell in the open-circuit mode is of order 2 V ($\mu_{Li}=2$ eV). This means that the lithium concentration generated by the electrodiffusion is notably higher than its equilibrium value, which can be obtained through conventional diffusion. Figure 1 shows typical current–voltage characteristics of cell (4) at $T=478$ K and 508 K after saturation of C_{60} crystals with lithium. It is clear that the shapes of the current–voltage characteristics are typical of the process of charging a chemical battery. When the voltage across the cell is set to zero (its electrodes are short-circuited), this chemical battery is discharged, and some of the lithium is deposited on the surface.

Figure 2 shows absorption spectra of a C_{60} crystal before (curve *1*) and after (curve *2*) electrodiffusion of Li ions. The spectra were recorded at $T=290$ K. The intense absorption beginning in the initial crystal at an energy of 1.6 eV is due to the excitonic absorption in the crystal. One can see in this

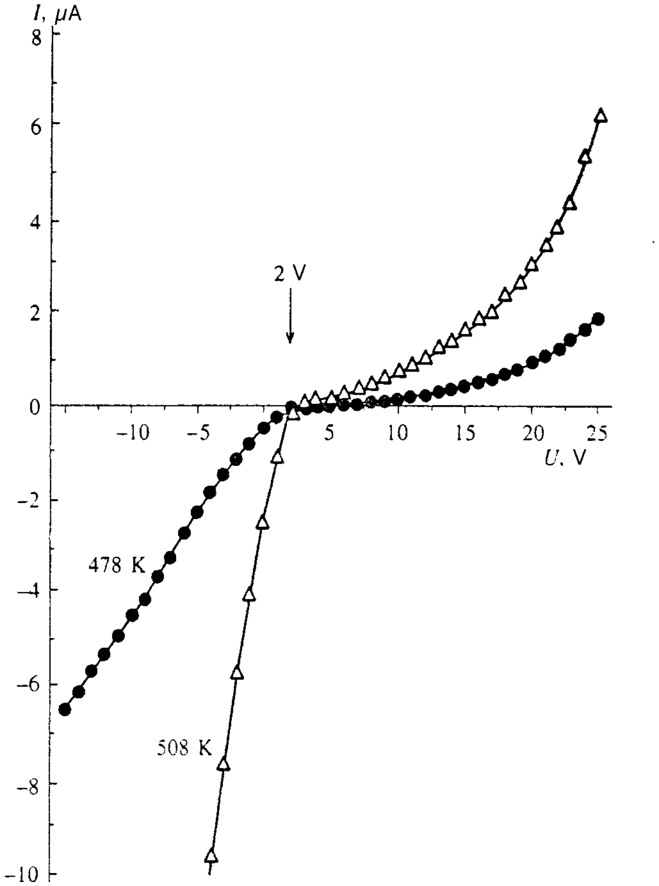

FIG. 1. Current–voltage characteristics of cell (4) at $T = 478$ and 508 K. The positive voltage corresponds to injection of Li^+ ions into the C_{60} crystal.

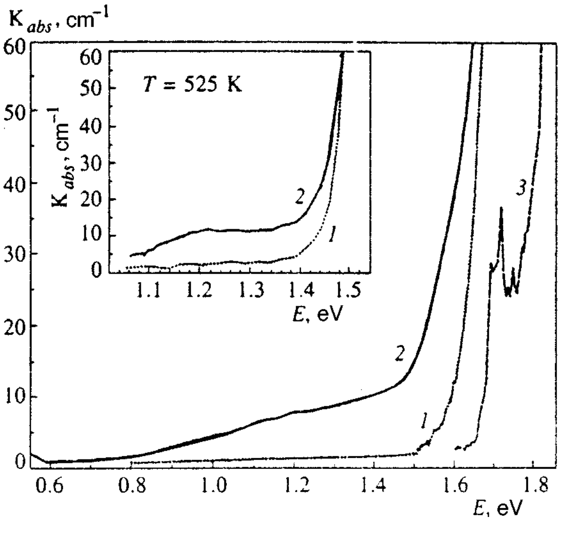

FIG. 2. Absorption spectra of the C_{60} crystal at $T = 290$ K before (curve *1*) and after (curve *2*) Li electrodiffusion. Curve *3* shows the absorption spectrum of the undoped crystal at $T = 10$ K. The inset shows spectra measured *in situ* as Li was injected at $T = 525$ K. Spectrum 1 was taken before applying voltage to the cell, spectrum *2* after 2-h exposure to current.

diagram that after the lithium doping, an additional broad absorption band turns up between 0.9 and 1.5 eV. This band is very similar to spectra of ‘‘photoinduced absorption,’’ i.e., additional optical absorption caused by laser excitation, which was observed in C_{60} crystals.[14,15] By analogy with the earlier investigations,[14,15] we assume that the recorded absorption band is due to electronic transitions from the T_{1g}, T_{2g}, and G_g orbitals of C_{60}, which form the conduction band in the crystal, to the higher band generated by T_{1u} orbitals of the C_{60} molecule. If so, we have to suppose that the electrodiffusion of Li^+ cations and electrons into C_{60} single crystals generates electrons in the conduction band.

The additional absorption of light enables us to visualize the diffusion of Li in the course of experiment. Figure 3 shows photographs of a C_{60} crystals taken at $T = 525$ K in the configuration of light transmission at a photon energy of 1.4 eV (885 nm). Picture 1 was taken at the moment when the current across the cell was turned on, and picture 3 was obtained after 30-min exposure to a current of 10 μA flowing from the anode to cathode. The superionic contact was attached to the left-hand side of the crystal. One can see how a dark ‘‘cloud’’ penetrates into the crystal from the left-hand contact. Under the assumption that the absorption at the wavelength of 885 nm is proportional to the concentration N of Li^+ ions (to be exact, to the concentration of C_{60}^- anions or electrons in the conduction band), one can visualize the process of diffusion in real time. The fast propagating dark ‘‘cloud’’ does not have clearly defined boundaries and, obviously, corresponds to the composition Li_xC_{60}, where $x \ll 1$ and is a slowly changing function of the coordinates within the sample. In the immediate neighborhood of the electrode injecting Li into the sample, an expanding black region with clear-cut edges can be seen, which probably has a composition with $x = 1$. Picture 4 in Fig. 3 illustrates the evolution of the Li concentration after the current direction is reversed. One can see that the dark ‘‘cloud’’ vanishes as lithium is ejected from the sample. Figure 4 shows quantitative characteristics of the process. The current as a function of time before and after its reversal is plotted in Fig. 4a; Fig. 4b shows the charge transmitted through the sample, which approximately reflects the number of Li ions injected into the sample. Figure 4c shows the light absorption at a wavelength of 885 nm (1.4 eV) at two different sites on the sample. The point corresponding to curve *2* is 0.65 mm farther from the Li-injecting contact than the point corresponding to curve *1*. Figure 4 clearly indicates that there is a good correlation between the charge transmitted across the sample and the change in the sample absorption. One can see in Fig. 5 profiles of the absorption coefficient, which is proportional to the Li concentration, as a function of distance x to the lithium-injecting electrode under the forward current (curves *1* and *2*) and after its reversal (curves *3* and *4*).

The reversibility of the lithium injection process in the samples under investigation indicates that a subset of the lithium cations are mobile in C_{60} crystals, and fullerene crystals demonstrate superionic properties. Moreover, the reversibility of the diffusion process with the current reversal provides evidence in favor of the dominant role of electric field in the process of Li diffusion, hence follows the feasibility of this technique for controlled doping of C_{60} crystals with various ions.

After a longer exposure to a direct current across the cell, the lithium concentration in the ‘‘cloud’’ rises to a cer-

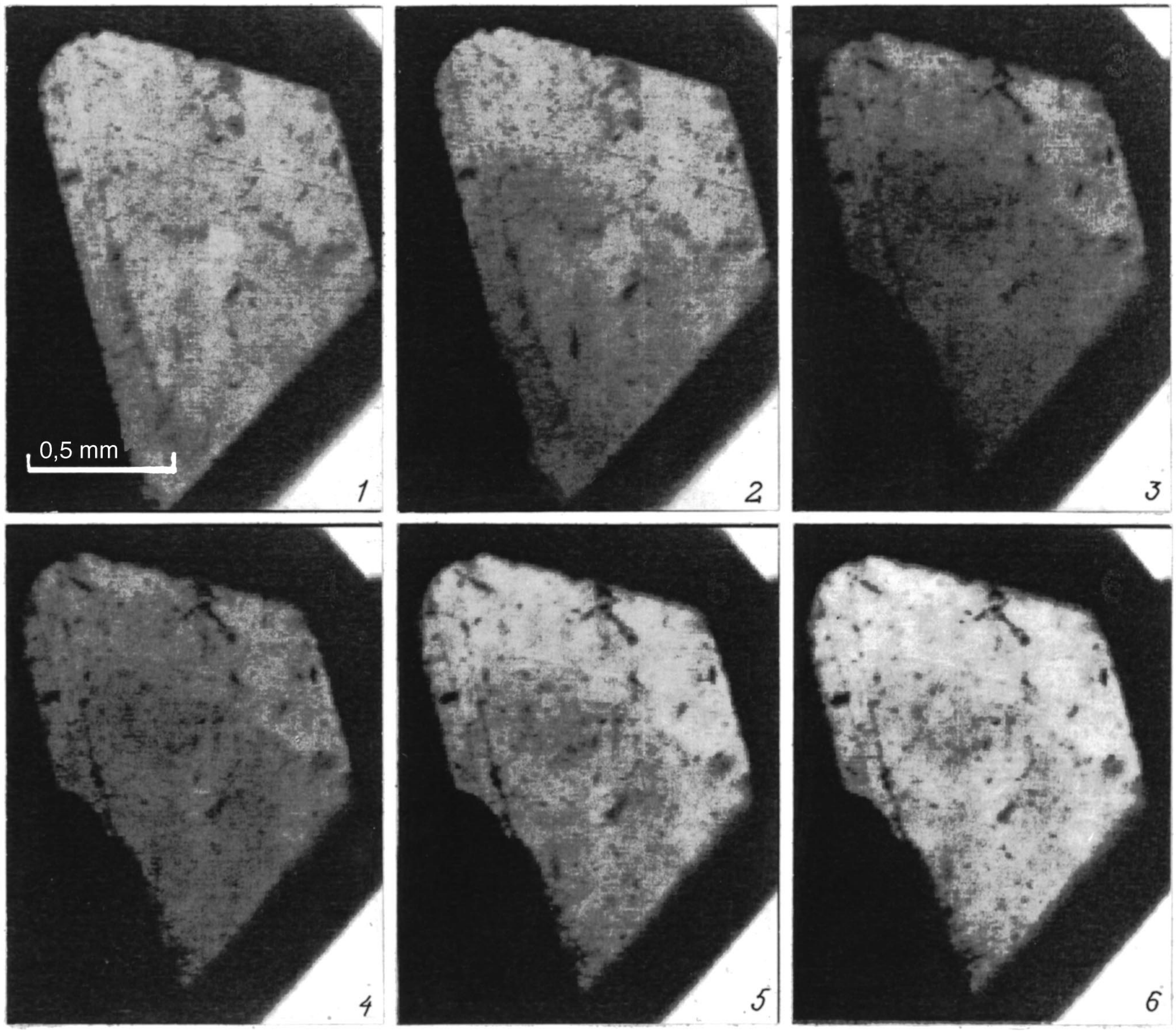

FIG. 3. Photograph of the C_{60} crystal taken *in situ* at $T=525$ K in transmitting light with a photon energy of 1.4 eV (885 nm). The superionic contact can be seen on the left of the sample. Picture *1* was taken before feeding current, pictures *2* and *3* after 10-min and 20-min exposure to a current of order 15 μA in the forward direction. Picture *4* was taken 40 min after switching the current to the opposite direction, picture *5* after 60 min of additional exposure to current in the forward direction, and picture *6* in 60 min after switching off the current.

tain limiting value, which depends on both the temperature and applied voltage, then the evolution of the sample condition proceeds through the expansion of the ''black'' phase, which consists, most probably, of the LiC_{60} $(x=1)$ compound. That the lithium concentration in the sample due to the electrically driven injection is much higher than the equilibrium value is the fact supported by the observation of the growth of isolated LiC_{60} phase clusters in the bulk of the sample (Picture 5 in Fig. 3). If the current is turned off and the temperature is maintained constant, the lithium concentration in the ''cloud'' drops considerably and, at the same time, new clusters of the ''phase'' are generated (Picture 6 in Fig. 3). The kinetic characteristics of this process are strongly dependent on the temperature and degree of oversaturation, and they have not been studied in detail. The only certain observation is that relatively fast cooling of a sample to room temperature (in several minutes) allowed us to preserve a strongly nonequilibrium lithium distribution generated by electrodiffusion.

3.2. ESR and microwave conductivity

Unlike the pristine samples, which were ESR silent, lithium-doped C_{60} crystals demonstrated an intense ESR signal in the form of several lines with *g*-factors close to 2.0 (Figs. 6 and 7). The main lines, which were present in ESR spectra of all lithium-doped samples, can be divided into three groups.

1. A broad isotropic line $(X1)$ with a *g*-factor of 2.004 ± 0.001, whose half-width at $T=290$ K varied from sample to sample between 5 and 8 Oe, and at $T=5$ K dropped to 2–3 Oe.

2. Three narrow anisotropic lines $(X2)$ corresponding to centers with different orientations. These lines have a

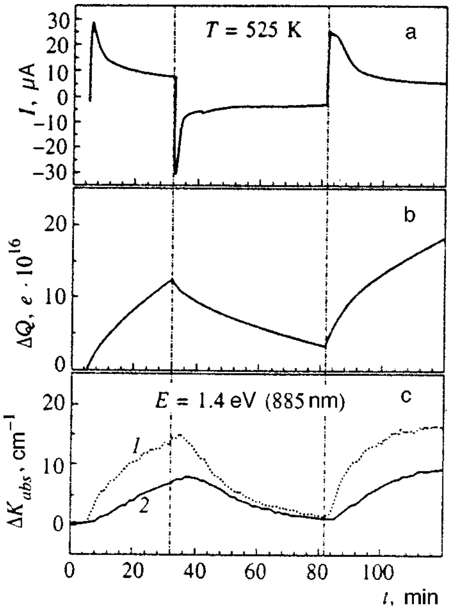

FIG. 4. (a) Current across the cell as a function of time at $T=525$ K. At the moment $t=5$ min a voltage of $+200$ V is switched on, at $t=32$ min its polarity is switched to the opposite direction, at $t=82$ min again a voltage of $+200$ V is applied; (b) current transmitted across the sample versus time; (c) variation of the sample absorption at a wavelength of 885 nm (1.4 eV) at two different points on the sample. Curve *2* corresponds to the region which is 0.65 mm further from the Li contact than the region for curve *1*.

g-factor of 2.002 varying in a range of ± 0.0008 as the sample is rotated. The half-widths of these lines drop from 1.5 Oe at $T=290$ K to 0.35 Oe at 5 K. The anisotropic effect indicates that the symmetry of paramagnetic centers responsible for the $X2$ lines is lower than the cubic one.

3. A pair of $X3$ lines of equal intensities and a half-width of 1.5 Oe at $T=290$ K. The mean g-factor of these lines is 2.003, and the splitting between them is 33 Oe. The intensity of these lines is very weak in comparison with other lines and poorly reproducible from sample to sample. These lines are probably due to electrons localized at C_{60} molecules which have chemical impurities with nuclear spin $I=1/2$ in their immediate neighborhood, and the splitting of 33 Oe is due to the hyperfine interaction. Since the intensity of these lines is very low, they can be ascribed to an uncontrolled impurity which is present in the samples.

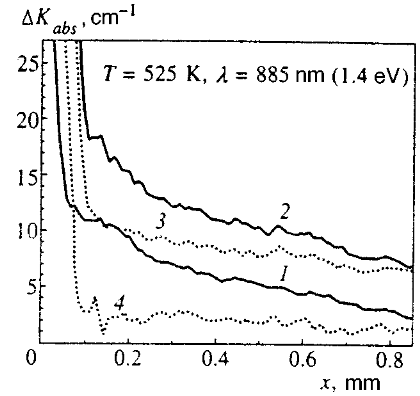

FIG. 5. Profiles of absorption coefficient in C_{60} crystals at a wavelength of 885 nm at $T=525$ K as a function of distance x from the Li contact. Curves *1* and *2* were recorded 12 and 29 min after feeding current in the forward direction. Curves *3* and *4* were recorded 11 and 41 min after switching the current polarity.

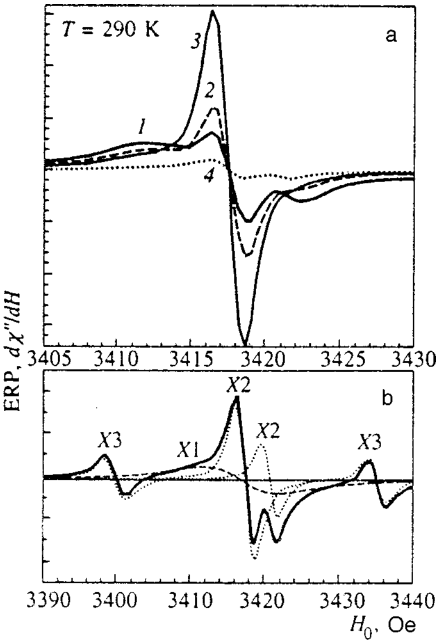

FIG. 6. (a) ESR spectra recorded at $T=290$ K in a C_{60} crystal after Li electrodiffusion at 520 K. Spectrum *1* was recorded immediately after cooling, spectra *2–4* after regular intervals of several hours of exposure to the temperature of 290 K; (b) spectrum *4* on an extended scale and its decomposition into separate lines.

The total intensity of the ESR spectrum at sufficiently small doping times (far from C_{60} saturation with lithium) is approximately proportional to the total charge transmitted through the cell and the amount of additional optical absorption in the sample. This allows us to associate $X1$ and $X2$ ESR-active centers with the lithium doping of the samples. The proportion of the Li^7 isotope with nuclear spin 3/2 in natural lithium is 92.5%, so if the paramagnetic electron were localized close to the lithium nucleus, we would observe splitting of each line into four hyperfine components. The absence of this hyperfine splitting supports our hypothesis that unpaired electron spins are located on C_{60} molecules, and the negative charge of the molecule is canceled by Li^+ ions located in the voids of the crystal.

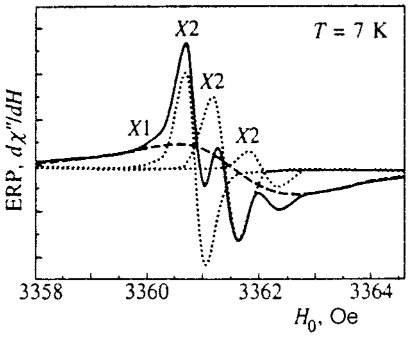

FIG. 7. Typical ESR spectrum of a Li_xC_{60} crystal recorded at $T=7$ K and its decomposition into separate lines.

At temperatures below 250 K, we did not detect any changes in the ESR intensity with time. At temperatures higher than 280 K, however, the evolution of intensities of the $X1$ and $X2$ lines was quite considerable. This is illustrated by Fig. 6a, where spectrum *1* was recorded immediately after a rapid (during several minutes) cooling of the sample from $T=520$ K, at which it was saturated with lithium, to room temperature. Spectra *2–4* were recorded sequentially with a time interval between them of several hours. In this process, the sample was kept at room temperature. In order to calculate the concentration of different paramagnetic centers, we normalized the ESR signal due to the sample to the ESR signal of a paramagnetic reference sample, then decomposed the ESR spectrum into separate lines (via a least-squares fit) and calculated total intensities of separate lines, proportional to the line intensity times its half-width squared. We managed to account for the evolution of ESR spectra on the basis of an assumption that two independent processes proceed concurrently.

1. A drop in the concentration of $X1$ centers (the broader ESR line) owing to transformation of $X1$ centers to $X2$ (narrower ESR line).

2. A drop in the concentration of $X2$ centers owing to their transformation to ESR-silent complexes.

If the initial concentration of $X1$ centers is much higher than that of $X2$ centers, the first process leads to an increase in the concentration of $X2$ centers, notwithstanding the second process. In fact, spectrum *1* in Fig. 6a yielded a concentration of $X1$ centers of $6.7\times10^{17}\,\mathrm{cm}^{-3}$ and that of $X2$ centers $0.4\times10^{17}\,\mathrm{cm}^{-3}$, whereas in spectrum *3* the concentrations of $X1$ and $X2$ centers were already $2.5\times10^{17}\,\mathrm{cm}^{-3}$ and $1.5\times10^{17}\,\mathrm{cm}^{-3}$, respectively. Thus, the concentration of $X1$ centers had dropped by $4.2\times10^{17}\,\mathrm{cm}^{-3}$ in several hours, whereas the concentration of $X2$ centers had increased by $1.1\times10^{17}\,\mathrm{cm}^{-3}$. When the $X1$ center concentration had become too small, process 2 dominated, which led to a drop in the concentrations of both centers, as is shown by spectrum *4* in Fig. 6a. The kinetics of processes 1 and 2 are strongly affected by the temperature and conditions of sample manufacture (local concentration of Li). These processes are nonexponential and deserve a dedicated investigation.

It is probable that $X1$ centers (the broader ESR line) correspond to conduction electrons in the weakly doped region of the $\mathrm{Li}_x\mathrm{C}_{60}$ crystal with $x\ll1$, whereas $X2$ centers correspond to the electrons localized on C_{60} molecules in the phase with $x=1$ formed in the process of self-organization. Then process 1, which transforms $X1$ to $X2$, corresponds to the decomposition of the ‘‘cloud’’ of uniformly distributed Li ions into clusters of the phase with $x=1$, which was observed visually (pictures 5 and 6 in Fig. 3).

Process 2, which is responsible for the disappearance of the $X2$ ESR signal, is probably more complicated: it probably leads to formation of monomer or polymer species including even numbers of Li atoms, which form covalent bonds with C_{60} molecules within the phase clusters. Additional arguments supporting this hypothesis will be given in Sec. 4, which deals with measurements of light absorption by vibrational modes in doped $\mathrm{Li}_x\mathrm{C}_{60}$ samples.

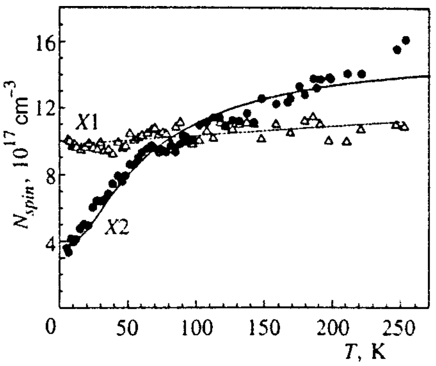

FIG. 8. Effective numbers of ESR-active spins corresponding to $X1$ and $X2$ centers whose ESR spectra are shown in Fig. 7 as functions of temperature. The numbers of spins were calculated as ratios between integrated ESR signals of $X1$ and $X2$ centers on one side and the integrated ESR intensity of a paramagnetic reference sample on the other.

Effective numbers of paramagnetic spins in $X1$ and $X2$ centers (whose spectra are given in Fig. 7) in one of the samples are plotted versus temperature in Fig. 8. The spin numbers were calculated by dividing the total intensities of ESR signals due to these centers by the total ESR intensity due to the paramagnetic reference sample with a known number of spins. It is clear that $X1$ centers correspond to nondegenerate electrons weakly interacting with each other. This supports our assumption that $X1$ centers correspond to electrons on C_{60} molecules in weakly doped regions of the sample. In contrast, the effective number of paramagnetic $X2$ centers drops considerably with decreasing temperature, approximately following the function

$$N=N_{\mathrm{para}}+N_{\mathrm{af}}\exp\left(-\frac{E_{\mathrm{af}}}{kT}\right), \tag{5}$$

where $E_{\mathrm{af}}=4.9$ meV, $N_{\mathrm{para}}/N_{\mathrm{af}}=0.3$, $N_{\mathrm{af}}=1.28\times10^{18}\,\mathrm{cm}^{-3}$. Thus, either the electrons are localized and the antiferromagnetic interaction between them is of order 5 meV, or this behavior is caused by Fermi degeneracy in the electronic system.

Figure 9 shows the conductivity of the same sample measured at a frequency of 9300 MHz. The conductivity

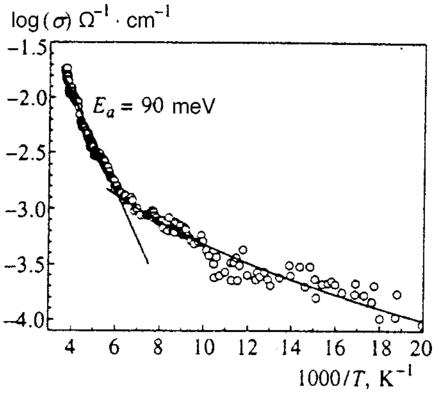

FIG. 9. Conductivity of the $\mathrm{Li}_x\mathrm{C}_{60}$ crystal versus temperature measured at a frequency of 9300 MHz. The conductivity $\sigma(T)$ was derived from measurements of the FWHM ΔF of the resonant curve of a rectangular microwave cavity loaded with the sample.

$\sigma(T)$ was derived from the change in the FWHM ΔF of the resonant curve of a rectangular microwave cavity after placing the sample inside it. It is clear that the sample conductivity is not of metallic character. In the temperature range $T>170\,\mathrm{K}$ the curve of $\sigma(T)$ is in good agreement with the activation temperature dependence $\sigma(T)=\sigma_0\exp(-E_a/kT)$ with an activation energy $E_a\approx 90\,\mathrm{meV}$, whereas at $T<160\,\mathrm{K}$ the curve is in fairly good agreement with the function $\sigma(T)=\sigma_{0H}\exp[(-A/T)^{1/4}]$, which characterizes hopping conductivity with variable hop length.

Let us discuss these results. Above all, note that after a sufficiently long electrodiffusion and slow cooling of the sample, the intensity of the $X2$ signal is usually much lower than that corresponding to the Li concentration in the dark regions with phase $x=1$. The cause is process 2. In our experiments, the intensity of the $X2$ signal was usually 0.1 to 1% of the number of Li atoms in the sample estimated on the basis of the electric charge transmitted through it and the visually detected volume of the dark phase (Figs. 3 and 4). The data given above were taken from a sample through which a charge of $10^{21}e/\mathrm{cm}^3$ was transmitted, which is approximately equal to the total number of C_{60} molecules in the sample. As a result, the sample was almost entirely opaque at the wavelength of 885 nm (1.4 eV), and we could suppose that a notable fraction of its volume was occupied by the phase $x=1$. The total ESR signal from this sample at $T=290\,\mathrm{K}$, however, corresponded to only $1.7\times10^{18}\,\mathrm{cm}^{-3}$ electrons. This means that the major fraction of the electrons had been knocked out of the game owing to formation of new compounds containing one or several C_{60} molecules connected by covalent bonds and even numbers of Li atoms in covalent bonds with C_{60}. Thus, we have a system with a high concentration of defects containing few free electrons. In such a system, the Fermi level is probably below the mobility edge of the conduction band, the electrons are localized at low temperatures, and hopping conductivity takes place.

4. OPTICAL SPECTRA IN THE REGION OF VIBRATIONAL MODES

Reflection spectra of Li_xC_{60} single crystals were measured on a Brücker 113v Fourier spectrometer at room temperature in a spectral range of 80 to $5000\,\mathrm{cm}^{-1}$ ($\sim$0.1–0.6 eV) with a resolution of $1\,\mathrm{cm}^{-1}$. The pristine single crystals had mirror-reflecting faces, and their reflection spectrum is labeled by number *1* in Fig. 10a. After lithium doping, the faces of the single crystal became dull. At the same time, the crystal surface, which had been black in visible optical range, was covered with yellow-green spots located close to the Li electrode. Reflection spectra from different areas on the Li_xC_{60} crystal surface recorded using the IR microscope of the Fourier spectrometer are given in Fig. 10. Note that the roughness of the crystal surface has a considerable effect on the intensity of reflected light, so spectra of Li doped samples are plotted in arbitrary units. The reflection spectra from both yellow-green (Fig. 10a) and dark (Fig. 10b) areas of the surface through which lithium was introduced into the crystal are different from those of the pristine C_{60} crystal since they have a background monotonically rising as the wave number drops. This spectral component indicates the presence of either free or quasi-free carriers in the surface regions of the lithium-doped crystal, which is in accord with measurements of the microwave conductivity. At room temperature, the background intensity drops with time: the spectra shown by dotted lines in Fig. 10 were taken 40 min after the spectra plotted by solid lines. This fact indicates that the density of quasi-free carriers drops, which is also consistent with the drop in the microwave conductivity and ESR intensity with time at $T=300\,\mathrm{K}$. Moreover, one can see in Fig. 10 that the dropping rate of the background intensity is different on different areas of the surface through which lithium was introduced: in the yellow-green areas (Fig. 10a) this dropping rate is notably lower than in the dark areas (Fig. 10b). Thus, the kinetics of the density of quasi-free carriers in the entire sample volume has complex characteristics.

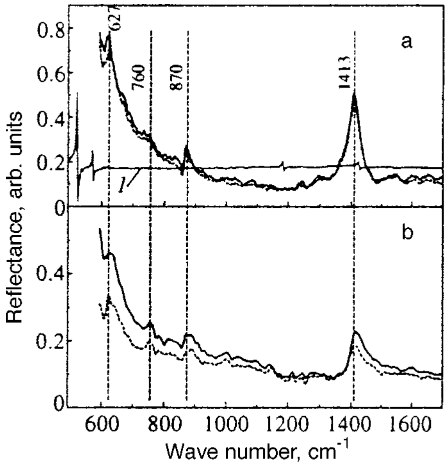

FIG. 10. Reflection spectra taken from the surface of a fresh as-grown Li_xC_{60} sample at $T=290\,\mathrm{K}$: (a) regions of increased Li concentration (the surface is tinted yellow-green in the visible light); (b) regions with a lower concentration (dark areas of the surface). The solid lines show spectra taken in one hour after Li diffusion, the dotted lines show spectra recorded after another 40 min of exposure to a temperature of 290 K. Spectrum *1* in Fig. 10a was taken from the pristine (undoped) sample.

The spectra of the Li_xC_{60} samples also contain at least four intense lines associated with different vibrational modes: 1415, 870, 760, and $627\,\mathrm{cm}^{-1}$. The intense line centered at $1415\,\mathrm{cm}^{-1}$ is an analog of the $T_{1u}(4)$ ($1429\,\mathrm{cm}^{-1}$) vibrational mode of the initial C_{60} crystal. The accuracy of the center frequency of the vibrational mode manifesting in the peak at $1415\,\mathrm{cm}^{-1}$ is very important for calculating the density of electrons in the C_{60} molecule. Unfortunately, we could not make use of the Kramers–Kronig technique for calculating the mode frequency because of the effect of surface roughness on the light scattering. The method of dispersion analysis allowed us to approximate the reflection spectrum with poor accuracy and yielded a center frequency of $1410\,\mathrm{cm}^{-1}$. The actual frequency of this mode is probably $1413\pm2\,\mathrm{cm}^{-1}$.

In the reflection spectrum of the initial crystal, the intensities of lines associated with the $T_{1u}(4)=1429\,\mathrm{cm}^{-1}$ and $T_{1u}(3)=1183\,\mathrm{cm}^{-1}$ modes are approximately equal (spectrum *1* in Fig. 10a). After the lithium doping, firstly, the oscillator strength of the $T_{1u}(4)$ mode becomes significantly higher than that of the $T_{1u}(3)$ mode, secondly, the former mode shifts to the low-energy side by $16\,\mathrm{cm}^{-1}$. A similar red shift and an increase in the oscillator strength of the $T_{1u}(4)$ mode were detected previously[16] in C_{60} films doped with K and Rb and interpreted in terms of the ''charged phonon'' model.[17] In accordance with this model, introduction of electrons to the lower unpopulated orbitals of the C_{60} molecule, which form the conduction band of the crystal, leads to interaction of the vibrational modes of C_{60} through a virtual electronic transition from these states to the band which is about 1.2 eV higher. Depending on the electron population x on the C_{60} molecule, the vibrational modes shift to the low side, and this shift is proportional to x, and their oscillator strengths increase as x^2. Among the four dipole-active vibrational modes detected in C_{60}, the $T_{1u}(4)$ mode has the highest constant λ of electron–phonon coupling. The coupling constant λ of the $T_{1u}(3)=1183\,\mathrm{cm}^{-1}$ mode is small, so lithium doping has little effect on this mode.

Investigations[16] of C_{60} films doped with K and Rb led the researchers to a conclusion that the $T_{1u}(4)$ mode shifts to the low-energy side by about $15\,\mathrm{cm}^{-1}$ when one electron is placed on the C_{60} molecule. Using the frequency of the $T_{1u}(4)$ mode measured as a function of the number x of electrons per one molecule,[16] we have inferred that, in the case under investigation, there is one electron per one C_{60} molecule near the crystal surface through which Li was introduced. Therefore, our assumption that the LiC_{60} phase $(x=1)$ is generated in the doped crystal seems likely.

Unlike the spectrum of the undoped C_{60} crystal, that of LiC_{60} contains a large set of relatively weak lines over a broad spectral range. The most intense among them are the lines at 870, 760, and $627\,\mathrm{cm}^{-1}$ (Fig. 10), which have not been detected in C_{60} crystals doped with K, Rb, and Cs (Refs. 16 and 18). Previously[19] spectral lines at 880, 670, and $450\,\mathrm{cm}^{-1}$ were detected after lithium doping of graphite: the line at $880\,\mathrm{cm}^{-1}$ was ascribed to a skeletal oscillation of the Li–Li bond in seven- and six-atomic lithium clusters, and the line at $670\,\mathrm{cm}^{-1}$ was attributed to valence oscillations of the Li–C bond. In the case under study, the line at $870\,\mathrm{cm}^{-1}$ can also be put down to a skeletal oscillation of the Li–Li bond, and the line at $627\,\mathrm{cm}^{-1}$ probably corresponds to a valence oscillation of the Li–C_{60} bond. The notable shift of its frequency from that measured in graphite intercalated with lithium $(670\,\mathrm{cm}^{-1})$ can be ascribed to the difference between the masses of the carbon atom and C_{60} molecule. Identification of the line at $760\,\mathrm{cm}^{-1}$ seems more difficult. Probably, the lithium can have two different states in the C_{60} crystal and form both the ionic Li^{+}–C_{60} and covalent Li–C_{60} bonds, which manifest as two valence oscillations at 627 and $760\,\mathrm{cm}^{-1}$. Note that the possibility of forming the Li–C_{60} covalent bond accounts for the disappearance of $X2$ ESR-active centers with time.

As was noted above, the exposure of samples to room temperature leads to the considerable drops in the ESR signal, microwave conductivity, and IR reflection due to quasi-free electrons. Probably, a fraction of Li is ejected from the sample bulk to the surface. This process can cause formation of cracks in the sample after a time. Studies of degraded samples, however, revealed that their surfaces had large areas which looked black under light of wavelengths less than 1000 nm. This observation indicates that their degradation cannot be fully interpreted in terms of the lithium ejection to the surface, and this fact impels us to assume that other phases (or compounds) are formed in doped samples, which contribute to absorption in the spectral range under investigation and have a considerable (more than 3 to 5%) volume deficiency, which leads to cracking. Moreover, such phases should not contribute to the ESR signal, neither can they have a high conductivity.

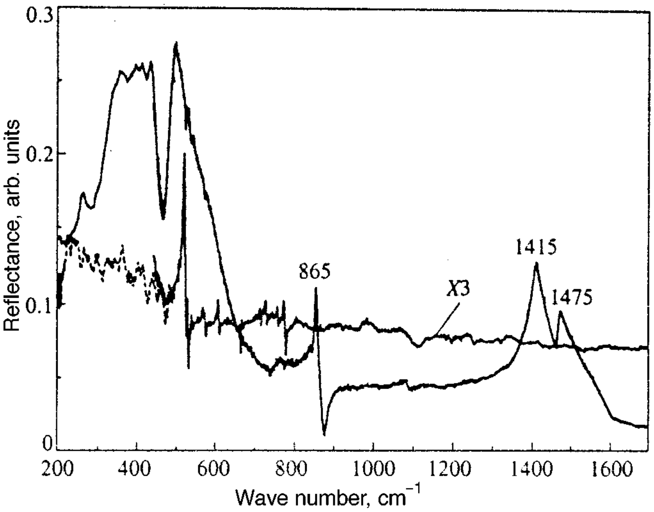

FIG. 11. Reflection spectra of the surface (dashed line) and cleaved surface (solid curve) of an ''aged'' sample, in which the conductivity and ESR signal are almost zero, at $T=290$ K.

This assumption is supported by IR reflection spectra of degraded samples shown in Fig. 11. The reflection spectrum of the sample surface shown by the dashed line in Fig. 11 is very similar to the $(C_{60})_2$ dimer spectrum recorded previously,[18] when transmission spectra of C_{60} monomer, dimer, and polymer phases in RbC_{60} were studied. It was shown[18] that the transition from the initial C_{60} phase to dimers, for example, leads to splitting of the $T_{1u}(4)$ (1429 cm^{-1}) mode into several lines and generates a set of lines at frequencies around $700\,\mathrm{cm}^{-1}$. This was ascribed to the lower symmetry of the C_{60} molecule in the dimer phase.

The spectrum shown by the solid line in Fig. 11 was taken from a cleaved surface of a degraded sample. Its comparison to the data reported earlier[18] indicates that this spectrum can be attributed to the polymer phase. The splitting of the $T_{1u}(4)$ $(1429\,\mathrm{cm}^{-1})$ mode of the initial C_{60} crystal into two lines at 1415 and $1480\,\mathrm{cm}^{-1}$ and the smaller number of spectral lines than in the dimer phase provide evidence in favor of this interpretation. The latter effect is due to the higher symmetry of the C_{60} molecule in the polymer phase than in the dimer phase. The plot also clearly shows lines that indicate the presence of lithium clusters $(865\,\mathrm{cm}^{-1})$ and the Li–C_{60} bond $(760\,\mathrm{cm}^{-1})$. Moreover, as in the case of the monomer phase observed on the surface of a pristine sample (Fig. 10a), the oscillator strengths of vibrational modes at frequencies of 1415 and $1480\,\mathrm{cm}^{-1}$ are significantly higher than that of the 1183 cm^{-1} mode. A third fact that is noteworthy is the difference between the spectral positions of the 1415 cm^{-1} and 1480 cm^{-1} lines on one side and those detected[20] in the photopolymerized phase (1424 and 1460 cm^{-1}), i.e., the 1415 cm^{-1} line is shifted to the low-energy side by 9 cm^{-1} as compared to those in the spectrum of the photopolymerized phase. This shift is less than 15 cm^{-1}, which is typical of the case when one electron is added to a C_{60} molecule in the monomer phase.[16] These three facts has led us to a conclusion that lithium is incorporated in the structure of the polymer phase synthesized in our experiment, where the number of electrons per one C_{60} molecule is probably much less than unity.

The main difference between the LiC_{60} polymer phase and RbC_{60} polymer reported earlier[18] is associated with the presence of free (or quasi-free) carriers. It was found previously[18] that the optical conductivity of RbC_{60} increases in the sequence of the dimer–monomer–polymer phases. In the case under study, the polymer phase of LiC_{60} is not conducting. This statement is supported by both microwave conductivity data and the reflection spectrum in Fig. 11, which does not show a monotonic increase in the reflectivity with decreasing wave number, characteristic of the contribution of free carriers. Thus, in contrast to the case of RbC_{60}, the Li–C_{60} covalent bond is formed in the polymer phase of LiC_{60}, whose valence oscillation manifests in the line at 760 cm^{-1} in the reflection spectrum. The reflection line at 627 cm^{-1}, which is also attributed to the valence oscillation of the Li–C_{60} bond in pristine samples, but is absent in the polymer phase of degraded Li_xC_{60}, should probably be associated with the oscillation of the Li^+–C_{60} ionic bond. Thus, our investigation of reflection spectra indicates that the assumption about formation of dimer and polymer phases in degraded Li–C_{60} crystals is true. Moreover, the absence of an ESR signal in the Li_xC_{60} polymer allows us to draw the conclusion that an even number of Li atoms are bound to the C_{60} molecule by covalent bonds.

5. CONCLUSIONS

We have suggested and implemented a novel technique for controlled doping of fullerene crystals through diffusion of ions driven by an electric field from a superionic contact. We have studied features of electrodiffusion of Li ions in C_{60} crystals and electronic properties of Li_xC_{60} single crystals fabricated by this method.

In lithium-doped C_{60} crystals, we have detected an ESR signal in the form of several lines with g-factors around 2.0. We have observed and studied evolution of ESR spectra with time and the effective numbers of unpaired spins corresponding to the $X1$ and $X2$ centers as functions of temperature.

The temperature dependence of the microwave conductivity of lithium-doped C_{60} crystals has been investigated. The sample conductivity is of a nonmetallic nature. In the temperature range $T>170\,\mathrm{K}$ the curve of $\sigma(T)$ is well approximated by the function $\sigma(T)=\sigma_0\exp(-E_a/kT)$ with an activation energy E_a of about 90 meV, whereas at $T<160\,\mathrm{K}$ it is approximated by the function $\sigma(T)=\sigma_{0H}\times\exp[(-(A/T)^{1/4}]$, which is typical of the hopping conductivity with variable hop length. Thus, Li_xC_{60} single crystals contain a lot of defects, but the concentration of free electrons is low. In such a system, the Fermi level should be located below the edge of the conduction band, electrons should be localized at low temperatures, and hopping conductivity should take place.

In IR reflection spectra of the Li_xC_{60} samples, we have detected lines associated with vibrational modes at 1415, 870, 760, and 627 cm^{-1}, and a background whose intensity increases monotonically as the wave number drops and which is due to free or weakly localized electrons. The spectral position of the 1415 cm^{-1} line indicates that one electron is bound to the C_{60} molecule. The lines at 627 and 760 cm^{-1} are put down to valence oscillations of the Li^+–C_{60} ionic bond and Li–C_{60} covalent bond, respectively. The former decays with time as crystals degrade, and only lithium in covalent bonds (alongside lithium clusters) is detected in degraded samples. The shapes of reflection spectra of degraded samples indicates that the conducting monomer phase in pristine samples transforms to nonconducting phases of dimers on the surface and polymers in the bulk of degraded samples. This supports our assumption that a subset of the Li atoms form covalent bonds with C_{60} molecules, which leads to the decay of both ESR and microwave conductivity signals.

The work was part of Task 2-1-98 Fullerenes and Atomic Clusters of the State R&D Program Physical Properties, Phase Transitions, and Metastable Phases of Fullerene Crystals and Their Derivatives, and was partly supported by the HTSC State R&D Program, Project 96031.

*)E-mail: kveder@issp.ac.ru

---

[1] T. Asakawa, M. Sasaki, T. Shiraishi, and H. Koinuma, J. Appl. Phys. **34**, 1958 (1995).

[2] M. Hosoya, K. Ichimura, Z. H. Wang, G. Dresselhaus, M. S. Dresselhaus, and P. C. Eklund, Phys. Rev. B **49**, 4981 (1994).

[3] T. Rabenau, S. Roth, and R. K. Kremer, Acta Phys. Pol. A **87**, 881 (1995).

[4] V. V. Kveder, V. D. Negrii, E. A. Shteinman, A. N. Izotov, Yu. A. Ossipyan, and R. K. Nikolaev, Zh. Éksp. Teor. Fiz. **113**, 734 (1998) [JETP **86**, 405 (1998)].

[5] A. N. Izotov, V. V. Kveder, Yu. A. Ossipyan, E. A. Shteinman, R. K. Nikolaev, and N. S. Sidorov, Zh. Éksp. Teor. Fiz. **114**, 2211 (1998) [JETP **87**, 1205 (1998)].

[6] S. Matsuura, T. Ishiguro, K. Kikuchi, and Y. Achiba, Phys. Rev. B **51**, 10217 (1995).

[7] R. C. Haddon, A. F. Hebard, M. J. Rosseinsky *et al.*, Nature (London) **350**, 320 (1991).

[8] A. F. Hebard, M. J. Rosseinsky, R. C. Haddon *et al.*, Nature (London) **350**, 600 (1991).

[9] W. Kratschmer, L. D. Lamb, K. Fostiropulos, and D. R. Huffman, Nature (London) **347**, 354 (1990).

[10] S. H. Irons, J. Z. Liu, P. Klavins, and R. N. Shelton, Phys. Rev. **52**, 15517 (1995).

[11] H. Funasaka, K. Sakurai, K. Sugiyama, K. Yamamoto, and T. Takahashi, Chem. Phys. Lett. **241**, 154 (1995).

[12] R. Kerkoud, P. Auban-Senzier, D. Jerome, S. Brazovskii, N. Kirova, I. Luk'yanchuk, F. Rachdi, and C. Goze, Synth. Met. **77**, 205 (1996).

[13] *Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials,* Vol. 2, K. M. Kadish and R. S. Ruoff (Eds.), The Electrochem. Soc., Inc., Pennington (1995).

[14] D. Dick, X. Wei, S. Jeglinski, R. E. Benner, and Z. V. Vardeny, Phys. Rev. Lett. **73(20)**, 2760 (1994).

[15] A. V. Bazhenov, A. V. Gorbunov, M. Yu. Maksimuk, T. N. Fursova, Zh. Éksp. Teor. Fiz. **112**, 246 (1997) [JETP **85**, 135 (1997)]; A. V. Bazhenov, A. V. Gorbunov, M. Yu. Maksimuk, and T. N. Fursova, Mol. Mater. **7**, 191 (1996).

[16] T. Pichler, R. Winkler, and Kuzmany, Phys. Rev. B **49**, 15879 (1994).

[17] M. J. Rice and H. Y. Choi, Phys. Rev. B **45**, 10173 (1992).

[18] K. Kamaras, D. B. Tanner, and L. Forro, Fullerene Sci. Technol. **5**(2), 465 (1997).

[19] D. E. Sklovsky, H. Gaucher, G. N. Bondarenko *et al.*, *High Pressure Lithium Intercalation in Catalytic Carbon Nanotubes*, Intern. Symp. of Intercal. Comp., France (1997).

[20] A. M. Rao, P. Zhou, K.-A. Wang *et al.*, Science **259**, 955 (1993).

Translation provided by the Russian Editorial office

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Two-dimensional weak localization effects in high temperature superconductor $Nd_{2-x}Ce_xCuO_{4-\delta}$

G. I. Harus, A. N. Ignatenkov, A. I. Ponomarev,*) L. D. Sabirzyanova,
and N. G. Shelushinina

Institute of Metal Physics, 620219 Ekaterinburg, Russia

A. A. Ivanov

Moscow Engineering Physics Institute, 115409 Moscow, Russia

(Submitted 16 October 1998)
Zh. Éksp. Teor. Fiz. **116**, 1723–1734 (November 1999)

A systematic study of the resistivity and Hall effect in single-crystal films $Nd_{2-x}Ce_xCuO_{4-\delta}$ ($0.12 \leqslant x \leqslant 0.20$) is presented, with special emphasis on the low-temperature dependence of the normal state conductance. Two-dimensional weak localization effects are found both in a normally conducting underdoped sample ($x=0.12$) and *in situ* superconducting optimally doped ($x=0.15$) or overdoped ($x=0.18$) samples in a high magnetic field $B>B_{c2}$. The phase coherence time τ_φ ($5.4\cdot 10^{-11}$ s at 2 K) and the effective thickness of a CuO_2 conducting layer d ($\simeq 1.5$ Å) have been estimated by fitting 2D weak localization theory expressions to magnetoresistivity data for magnetic fields perpendicular to the *ab* plane and in plane. Estimates of the parameter d ensure strong carrier confinement and justify a model consisting of almost decoupled 2D metallic sheets for the $Nd_{2-x}Ce_xCuO_{4-\delta}$ single crystal. © *1999 American Institute of Physics.* [S1063-7761(99)01611-X]

1. INTRODUCTION

The field of high-transition-temperature (high-T_c) superconductivity has generated several thousand publications in the last few years. For a short overview of the lattice structure and phase diagram of the most widely studied copper oxide compounds, such as hole-doped $La_{2-x}Sr_xCuO_4$ and $YBa_2Cu_3O_{6+x}$ or electron-doped $L_{2-x}Ce_xCuO_4$ (L=Nd or Pr), one can consult, e.g., the review in Ref. 1 or book by Plakida.[2] The copper oxide high-T_c materials are basically tetragonal, and all of them have one or more CuO_2 planes in their structure, which are separated by layers of other atoms (Ba–O, La–O, Nd–O,...). Most researchers empirically and theoretically have come to a consensus that superconductivity is related to processes occurring solely in the conducting CuO_2 planes, with the other layers simply providing the carriers (they are therefore called charge reservoirs). In the CuO_2 planes, each copper ion is strongly bonded to four oxygen ions separated by approximately 1.9 Å.

Due to the layered character of the crystal structures, the high-T_c copper oxide compounds are highly anisotropic in their normal-state electrical properties. Although the resistivity in the CuO_2 planes, ρ_{ab}, shows metallic temperature dependence, the temperature behavior and the magnitude of the resistivity parallel to the c axis, ρ_c, are strongly dependent on crystal structure, and on the concentration of charge carriers.

Systematic data for ρ_c in a number of high-T_c materials obtained on well characterized single crystals are presented by Ito *et al.*[3] For hole-doped systems $YBa_2Cu_3O_{6+x}$ and $La_{2-x}Sr_xCuO_4$ ρ_c exhibits a marked change in magnitude as well as in temperature dependence with changing hole concentration (i.e., changing x). For the underdoped (low x) and optimally doped (superconducting) compounds ρ_c is nonmetallic ($d\rho_c/dT<0$) at low enough temperatures. In both systems the anisotropy coefficient, ρ_c/ρ_{ab}, decreases noticeably with doping, being $\sim 10^2$ for the superconducting compounds.

The crystal structure T' of the electron doped $Nd_{2-x}Ce_xCuO_{4-\delta}$ system is the simplest among the superconducting cuprates with the perovskite structure.[4] The Cu atoms in the CuO_2 layers of hole-doped $La_{2-x}Sr_xCuO_4$ or $Nd_{2-x-y}Ce_xSr_yCuO_{4-\delta}$ ($y>x$) superconductors are surrounded by apical O atoms, forming octahedrons (T structure) or semioctahedrons (T^* structure). The most important difference in the crystal structures of $Nd_2(Ce)CuO_4$ and $La_2(Sr)CuO_4$ is that the apical oxygen atoms in the T' structure are displaced so as to make an isolated CuO_2 plane (Fig. 1).

The undoped system Nd_2CuO_4 is an insulator, with the valence band mainly of O $2p$ character, and the empty conduction band being the upper Hubbard Cu $3d$ band. The Coulomb $3d-3d$ repulsion at the Cu site U ($\simeq 6-7$ eV) is strong, and it is greater than the oxygen to metal charge-transfer energy Δ ($\simeq 1-2$ eV). As the gap between the conduction and valence bands is determined just by the energy Δ, these cuprates are classified as charge-transfer semiconductors.[5]

The combination of Ce doping and O reduction results in n-type conduction in CuO_2 layers in $Nd_{2-x}Ce_xCuO_{4-\delta}$ single crystals.[4,6] An energy band structure calculation[7] shows that the Fermi level is located in a band of $pd\sigma$-type formed by $3d(x^2-y^2)$ orbitals of Cu and $p_\sigma(x,y)$ orbitals of

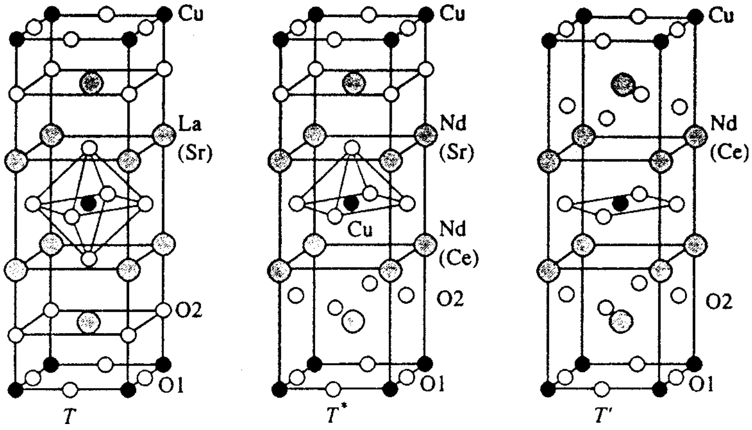

FIG. 1. Crystal structure of three types of copper oxides (Ref. 4).

oxygen. The $pd\sigma$ band appears to be of highly two-dimensional (2D) character, with almost no dispersion in the z-direction normal to CuO_2 planes. The electrons are concentrated within the confines of conducting CuO_2 layers, separated from each other by a distance $c \simeq 6$ Å.

In accordance with such a structure $Nd_{2-x}Ce_xCuO_{4-\delta}$, single crystals have a significantly higher resistive anisotropy than Y- or La-systems: $\rho_c/\rho_{ab} \simeq 10^4$ for $x=0.15$[8,9] and for $x=0.16-0.20$ with different values of oxygen deficiency δ[10] at room temperature, it increases with decreasing temperature.[10] Measurements by Ito *et al.*[3] for another electron-doped compound with the same T' structure, $Pr_{2-x}Ce_xCuO_{4-\delta}$, show that for $x=0.15$ the anisotropy is very large, $\rho_c/\rho_{ab} \geqslant 10^4$, and nonmetallic ρ_c is observed. Preliminary measurements on a Pr-system with different x indicated that, as in the case of Y- and La-systems, ρ_c decreases with increasing carrier concentration much more rapidly than ρ_{ab}.

The larger anisotropy in Nd- or Pr-systems compared with La-or Y-systems would imply that fluorite-type Nd_2O_2 or Pr_2O_2 layers block out-of-plane conduction more effectively than NaCl-type La_2O_2 or Ba_2O_2 layers.[3]

The nonmetallic behavior of out-of-plane conductance suggests that the carriers are confined tightly in the CuO_2 plane.[3] It is thus of interest to search for two-dimensional effects in the in-plane conductance of the layered cuprates. There are several previous reports on the manifestation of 2D weak localization effects in the in-plane conductance of $Nd_{2-x}Ce_xCuO_{4-\delta}$ single crystals or films. Thus a linear dependence of resistivity on $\ln T$ comes about at $T<T_c$ for samples with $x \simeq 0.15$, in which the superconducting state is destroyed by a magnetic field.[11] Furthermore, a highly anisotropic negative magnetoresistance, predicted for 2D weak localization, has been observed in the nonsuperconducting state at low temperatures: in samples with $x=0.11$[12] and in unreduced samples with $x=0.15$[13] or $x=0.18$.[14] We report here a systematic study of 2D weak localization effects for a number of optimally reduced samples of $Nd_{2-x}Ce_xCuO_{4-\delta}$ with $0.12 \leqslant x \leqslant 0.20$.

TABLE I.

x	t, Å	T_c, K	$\rho_0 \cdot 10^5$, $\Omega \cdot$cm	$\rho_{300\ K} \cdot 10^5$, $\Omega \cdot$cm	$\rho_{300\ K}/\rho_0$	$A \cdot 10^9$, $\Omega \cdot$cm/K^2
0.12	5500	-	-	102	-	-
0.15	5000	20	8.2	42.4	5.2	4.0
0.17	5700	12	8.6	29.6	3.4	2.7
0.18	5000	6.0	6.0	23.5	3.9	2.2
0.20	4000	<1.3	1.1	10.0	9.1	1.1

2. EXPERIMENTAL PROCEDURE

The flux separation technique was used for $Nd_{2-x}Ce_xCuO_{4-\delta}$ film deposition.[15] High-quality c-axis oriented single crystal films with thickness around 5000 Å and $0.12 \leqslant x \leqslant 0.20$ were grown. The values of T_c after sample reduction are shown in Table I.

Figure 2 demonstrates that T_c of the film with $x=0.15$ is in agreement with previously published data for bulk single crystals.[4] The values of T_c for overdoped films with $x \geqslant 0.17$ are higher than for corresponding bulk crystals in accordance with the information of Xu *et al.*[16] that superconductivity survives up to $x=0.22$ in $Nd_{2-x}Ce_xCuO_{4-\delta}$ films.

Standard four-terminal measurements of the resistivity ρ and Hall effect R ($\mathbf{j} \| ab, \mathbf{B} \| c$) in the dc regime were carried out in the temperature range $T_c < T \leqslant 300$ K without a magnetic field B, and in magnetic fields up to $B=12$ T at temperature down to 0.2 K. The electrical contacts were prepared by evaporating thin silver strips onto the sample, and attaching silver wires to these with conducting glue.

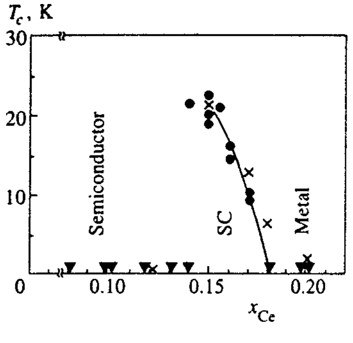

FIG. 2. Phase diagram of $Nd_{2-x}Ce_xCuO_{4-\delta}$. Notation: triangles and circles—data of Ref. 4 (triangles—$T_c=0$); crosses—our data.

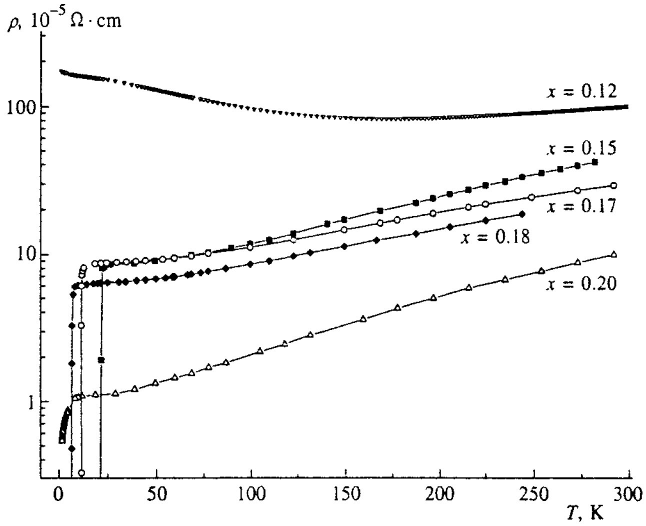

FIG. 3. Temperature dependence of in-plane resistivity for the samples investigated $Nd_{2-x}Ce_xCuO_{4-\delta}$.

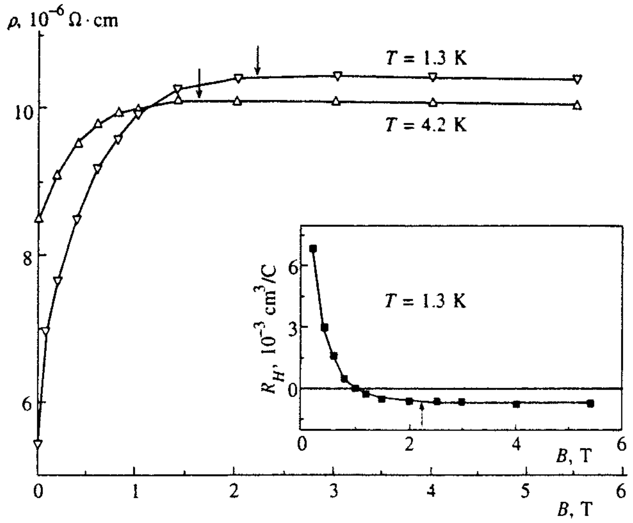

FIG. 4. In-plane resistivity ($\mathbf{j}\perp\mathbf{B}$) of the sample with $x=0.20$ as a function of magnetic field $\mathbf{B}\perp ab$ at two different temperatures. Arrows indicate values of the second critical field. Inset: Hall coefficient ($\mathbf{j}\| ab$; $\mathbf{B}\perp ab$) as a function of magnetic field at $T=1.3$ K.

3. RESULTS

The temperature dependence of the zero-field in-plane resistivity for the investigated samples at T up to 300 K is shown in Fig. 3. A clear resistance minimum is observed at $T\simeq 150$ K for the nonsuperconducting sample with $x=0.12$. The $\rho(T)$ dependence is described by $\rho=\rho_0+AT^2$ at $T=20-180$ K for $x=0.15$, 0.17, and 0.18, and over the wider $T=10-300$ K for $x=0.20$. The values of ρ_0 and A are presented in Table I.

We describe here the magnetoresistance measurements only for the underdoped nonsuperconducting sample ($x=0.12$) and for two overdoped superconducting samples ($x=0.18$ and $x=0.20$). Detailed investigations of $\rho(B,T)$ dependencies for the optimally doped sample with $x=0.15$ were presented earlier,[17] as were some results on the $x=0.18$ sample.[18]

In the superconducting samples, normal-state transport at low T is hidden unless the magnetic field stronger than the second critical field B_{c2} is applied. As we are interested in the low-temperature $\rho(T)$ dependence, we have destroyed the superconductivity with a magnetic field $B_\perp$ perpendicular to the CuO_2 planes. In Fig. 4, the $\rho(B_\perp)$ dependence for $x=0.20$ at $T=1.3$ K and $T=4.2$ K in a magnetic fields up to $B=5.5$ T are presented. In the inset of Fig. 4, the dependence of the Hall coefficient R on magnetic field $B_\perp$ at $T=1.3$ K is also shown. On the assumption that $B_{c2}^{\perp}(T)$ is a field in which $\rho(B_\perp)$ and $R(B_\perp)$ at given T come up to their normal-state value, we have $B_{c2}^{\perp}=2.2$ T at $T=1.3$ K and $B_{c2}^{\perp}=1.5$ T at $T=4.2$ K.

In our previous investigation[18] of the sample with $x=0.18$, negative magnetoresistance was observed after the destruction of superconductivity by a magnetic field up to 5.5 T at $T\geqslant 1.4$ K. In Fig. 5, $\rho(B_\perp)$ is shown for this sample at much lower temperatures (down to 0.2 K) and in fields up

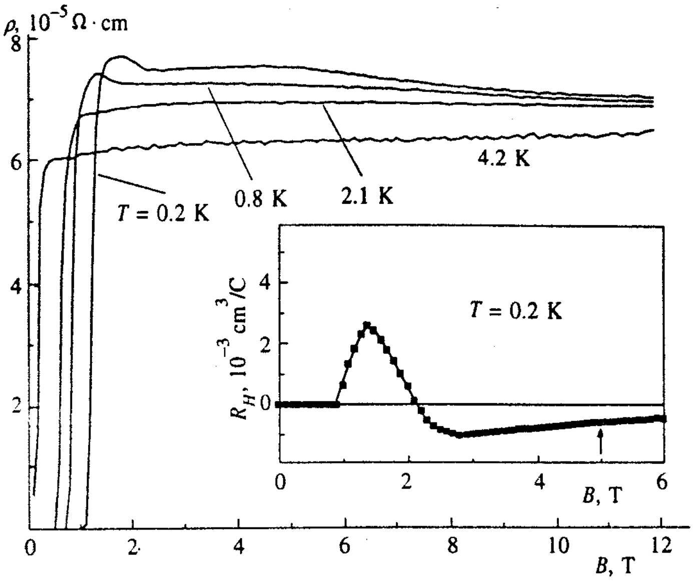

FIG. 5. In-plane resistivity ($\mathbf{j}\perp\mathbf{B}$) of the sample with $x=0.18$ as a function of magnetic field $\mathbf{B}\perp ab$ at different temperatures. Inset: Hall coefficient ($\mathbf{j}\| ab$;$\mathbf{B}\perp ab$) as a function of magnetic field at $T=0.2$ K. The arrow shows the estimate for the second critical field.

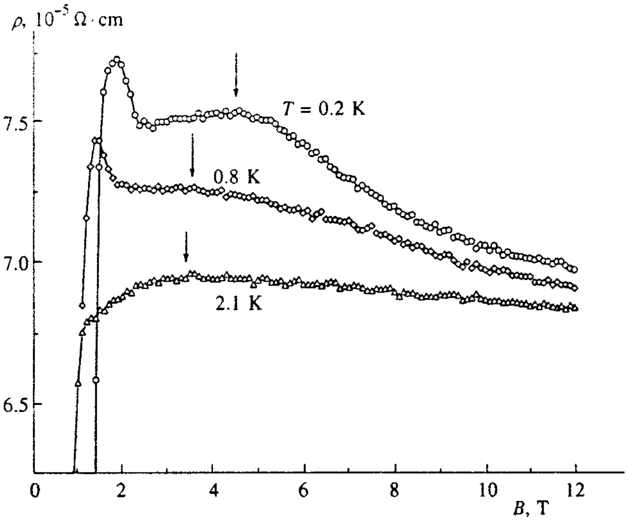

FIG. 6. Negative magnetoresistance at $B>B_{c2}$ in the sample with $x=0.18$. Arrows indicate values of the second critical field B_{c2} at different temperatures.

to 12 T. The inset of Fig. 5 shows an example of $R(B_\perp)$ at given $T<T_c$. The nonmonotonic $R(B_\perp)$ behavior, with reversal of the sign of the Hall effect, is usually observed in the mixed state of the superconductor.[18,19] The transition to the normal state is completed at $B=B_{c2}$, where the Hall coefficient becomes nearly constant with the same value as in the normal state at $T>T_c$ ($B_{c2}\simeq 5$ T at $T=0.2$ K). Values of B_{c2} for $x=0.18$ estimated in this way at different temperatures are marked by the arrows in Fig. 6. This figure also clearly demonstrates the transition from positive to negative magnetoresistance after the suppression of superconductivity.

In Fig. 7, the results of the theoretical description of the magnetoconductivity at $B>B_{c2}$ are presented. Figure 8 demonstrates that the resistivity of the sample with $x=0.18$ is a linear function of $\ln T$ in magnetic fields $B>B_{c2}$. The experimental points for $B=1.5$ T are also shown. If the logarithmic temperature dependence of the resistivity is typical of the normal state, then the discrepancy between the experimental points and the logarithmic law indicates that the normal state has not yet been attained at $B=1.5$ T.

We have also measured the in-plane conductivity in a nonsuperconducting sample with $x=0.12$ for magnetic fields perpendicular and parallel to the CuO_2 planes up to 5.5 T at $T=1.9$ K and 4.2 K (Fig. 9). The positive magnetoconductivity (negative magnetoresistance) observed in this sample is obviously anisotropic relative to the direction of the magnetic field.

4. DISCUSSION

A logarithmic temperature dependence of the conductivity is one indication of the interference quantum correction due to 2D weak localization. A magnetic field normal to the diffusion path of a carrier destroys the interference. In a two-dimensional system, it causes negative magnetoresistance for the field perpendicular to the plane, but no effect for the parallel configuration. Weak localization effects are almost totally suppressed for $B_\perp>B_{tr}$,[20] where the so called "transport field" is defined as the field at which

$$2\pi B_{tr} l^2=\Phi_0. \tag{1}$$

Here l is the elastic mean free path and $\Phi_0=\pi c\hbar/e$ is the elementary flux quantum.

Let us compare Eq. (1) with the relations between the coherence length ξ and the second critical field in the pure superconductor ($\xi\ll l$),

$$2\pi B_{c2}\xi^2=\Phi_0, \tag{2}$$

or in the so-called "dirty limit" ($\xi\gg l$):

$$2\pi B_{c2}\xi l=\Phi_0. \tag{3}$$

From Eqs. (1) and (2) we have $B_{tr}/B_{c2}=(\xi/l)^2$, so $B_{tr}\ll B_{c2}$, and it is impossible to observe weak localization effects in the pure case. In contrast, from Eqs. (1) and (3) one

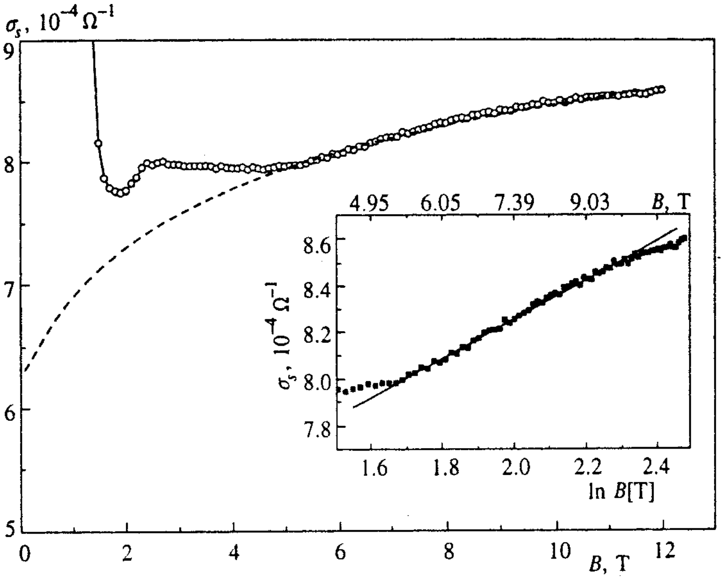

FIG. 7. Fit of the expression (4) to experimental data on the surface conductivity of the sample with $x=0.18$ at $T=0.2$ K. Fit parameters of the broken line are: $B_\varphi=0.1$ T, $\alpha=6.6$. Inset: Surface conductivity as a function of $\ln B$.

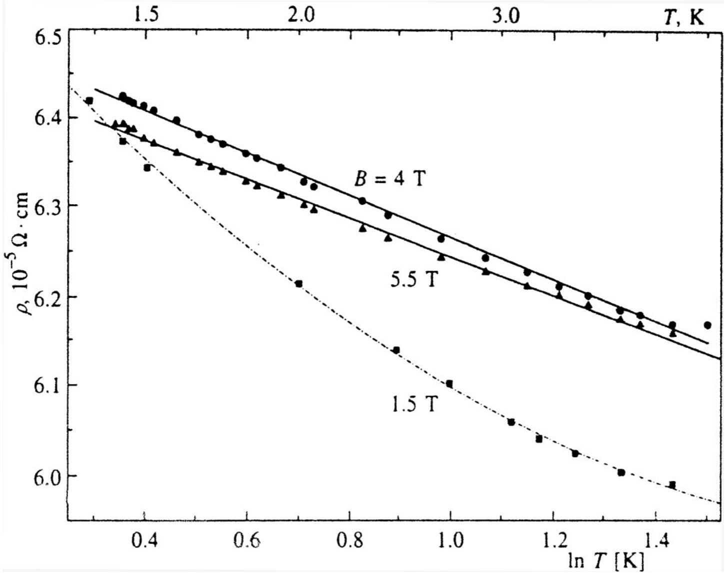

FIG. 8. In-plane resistivity of the sample with $x=0.18$ as a function of ln T in different magnetic fields.

has $B_{tr}/B_{c2}=(\xi/l)$, $B_{tr}>B_{c2}$, and weak localization should survive in the normal state ($B>B_{c2}$) of a dirty superconductor.

In Table II, the parameters of investigated samples essential to a description of localization are presented. From the experimental values of the in-plane resistivity ρ and Hall constant R in the normal state, we have obtained the surface resistance $R_s=\rho/c$ per CuO$_2$ layer and the bulk and surface electron densities $n=(eR)^{-1}$ and $n_s=nc$ ($c=6$ Å is the distance between CuO$_2$ layers). Using the relations[21] $\sigma_s=(e^2/\hbar)k_F l$ for the 2D conductance $\sigma_s=1/R_s$, and $k_F=(2\pi n_s)^{1/2}$ for the Fermi wave vector, we have estimated the important parameter $k_F l$, the mean free path l, and then according to Eq. (1) the characteristic field B_{tr}. For the sample with $x=0.15$ we use the data of Ref. 17.

In a random two-dimensional system, the parameter $k_F l$ can serve as a measure of disorder.[21] It is seen from Table II that we have a wide range $\sim 1-10^2$ of $k_F l$ in the investigated series of samples. For $k_F l \gg 1$, a true metallic conduction takes place in CuO$_2$ layers. Thus, we have a rather pure 2D system with $k_F l \sim 10$ for $x=0.15$ or $x=0.18$, and an extremely pure system with $k_F l \sim 10^2$ for $x=0.20$. It is quite remarkable that even at such high values of the parameter $k_F l$, a trace of localization comes to light: for $B_\perp > B_{c2}$, ρ is greater at 1.3 K than at 4.2 K (see Fig. 4). As for the sample with $x=0.12$, where $k_F l$ is of the order of unity, this system is in close proximity to transition from weak logarithmic to strong exponential localization as disorder increases ($k_F l$ decreases).

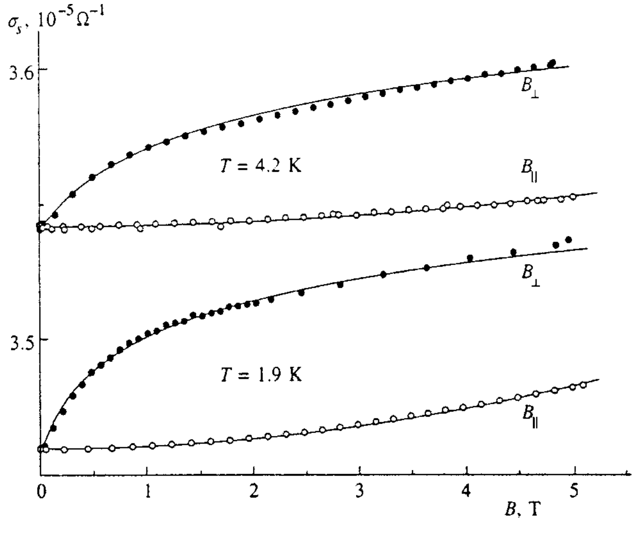

FIG. 9. Surface conductivity of the sample with $x=0.12$ as a function of magnetic field $B_\perp$ ($\mathbf{B}\perp ab$) or $B_\parallel$ ($\mathbf{B}\parallel ab$) at different temperatures.

The second critical field $B_{c2}^{\perp}$ at temperatures around $T \simeq 1.4$ K (see Figs. 4 and 5) and values of ξ estimated according to Eqs. (2) or (3) are also shown in Table II. In the pure system with $x=0.20$, $\xi<l$, and negative magnetoresistance is not detected at $B>B_{c2}$, at least for $T \geq 1.3$ K (see Fig. 4). Systems with $x=0.15$ and 0.18 are situated close to the dirty limit $\xi \gg l$, and there exist appreciable field ranges $B_{c2}<B<B_{tr}$ where negative magnetoresistance due to 2D weak localization is actually observed (see Ref. 17 and Fig. 6).

In 2D weak localization theory, the quantum correction to the Drude conductivity in a perpendicular magnetic field is[22]

$$\Delta\sigma_s(B_\perp)=\alpha\frac{e^2}{2\pi^2\hbar}\left\{\Psi\left(\frac{1}{2}+\frac{B_\varphi}{B_\perp}\right)-\Psi\left(\frac{1}{2}+\frac{B_{tr}}{B_\perp}\right)\right\}, \quad (4)$$

TABLE II.

x	$n\cdot 10^{-22}$, cm^{-3}	$k_F l$	l, Å	B_{tr}, T	$B_{c2}^{\perp}$, T ($T=1.4$ K)	ξ, Å
0.12	0.2*	2	≈10	≈270	-	-
0.15	2.0*	18	≈30	≈30	5.5	180
0.18	1.1	25	40	22.5	4.0	200
0.20	1.0	150	240	0.6	2.2	150

*Data from Ref. 1 at $T=80$ K.

where α is a factor of the order of unity, Ψ is the digamma function, and $B_\varphi = c\hbar/4eL_\varphi^2$. Here $L_\varphi = \sqrt{D\tau_\varphi}$ is the phase coherence length, D is the diffusion coefficient, and τ_φ is the phase breaking time.

The fit of (4) to the experimental $\sigma_s(B_\perp)$ data for $x=0.18$ at $B_\perp > B_{c2}^\perp$ is shown in Fig. 7. For each temperature there are two fit parameters: the characteristic field B_φ (or L_φ) and the factor α. The widest range of requisite magnetic fields and thus the most accurate fit results are obtained for the lowest temperature $T=0.2$ K. With $B_{tr}=22.5$ T, the best fit is obtained for $B_\varphi \simeq 0.1$ T ($L_\varphi = 560$ Å) and $\alpha = 6.6$. The fitting procedure is highly sensitive to the value of the parameter α. In contrast, the value of B_φ is obtained only to order of magnitude, as we have no zero-field and weak-field data. Nevertheless, there is no doubt that the inequality $B_\varphi \ll B_{tr}$ is valid.

In the field range $B_\varphi \ll B \ll B_{tr}$, the expression (4) can be written

$$\Delta\sigma_s(B_\perp) = \alpha\frac{e^2}{2\pi^2\hbar}\left\{-\Psi\left(\frac{1}{2}\right) - \ln\frac{B_\perp}{B_{tr}}\right\}. \tag{5}$$

The inset of Fig. 7 shows that the experimental data at $T=0.2$ K can be fitted rather closely by this simple formula over a wide range of fields, $5 \leqslant B \leqslant 11$ T. But as we have the factor $\alpha = 6.6$, negative magnetoresistance is too large to be due to the destruction of weak localization only. Thus we conclude that some additional mechanism of negative magnetoresistance must be at work.

There exists another quantum correction to the normal-state conductivity with a logarithmic dependence of magnetoresistivity on B, namely the correction due to disorder-modified electron–electron interaction (EEI) in the Cooper channel.[23] In the range of magnetic fields $B_T \ll B \ll B_{tr}$, we have

$$\Delta\sigma_s^{\mathrm{EEI}}(B_\perp) = -\frac{e^2}{2\pi^2\hbar}g(T)\ln\left(\frac{B_\perp}{B_{tr}}\right), \tag{6}$$

where $B_T = \pi c\hbar/2eL_T^2$, $L_T = \sqrt{\hbar D/kT}$ is the thermal coherence length, and $g(T)$ is the effective interaction constant of two electrons with opposite momenta. For the attractive electron–electron interaction due to virtual phonon exchange, $g > 0$, and according to (6) the magnetoresistance should be negative.

As we have dealt with *in situ* superconducting samples, so that $g > 0$, the contribution due to EEI is most probably the reason for the extra negative magnetoresistance at very low temperatures ($B_T = 0.02$ T at $T = 0.2$ K). With increasing temperature, the magnitude of the EEI contribution decreases rapidly ($\alpha = 2.5$ at $T = 0.8$ K), and at $T \geqslant 1$ K the estimated value of the factor is close to unity ($\alpha = 0.77$ at $T = 2.1$ K), as it should be for weak localization.

It should be noted that pronounced negative magnetoresistance due to the suppression of weak electron localization is observed in ordinary superconductors as well. The electron transport properties of three-dimensional (3D) amorphous α-Mo_3Si and α-Nb_3Ge superconducting films have been investigated in magnetic fields up to $B = 30$ T at temperatures down to $T = 0.35$ K.[24] The authors have found that both the temperature and field dependence of the resistivity ρ can be qualitatively described by weak localization theory. At low temperatures and in magnetic fields above the upper critical field, $B > B_{c2}$, magnetoconductivity is proportional to $B^{1/2}$. This field dependence is consistent with weak localization in the high-field limit ($B \gg B_\varphi$) for 3D disordered systems, in contrast to a 2D system with $\Delta\sigma(B) \propto \ln B$.

One important indication of the 2D character of a system is the strong dependence of magnetoresistance on magnetic field orientation. Highly anisotropic (negative) magnetoresistance is actually observed in a nonsuperconducting sample with $x = 0.12$ (see Fig. 9). From the fit to $\sigma_s(B_\perp)$ by the functional form (4) with $\alpha = 1$, we find $L_\varphi = 770$ Å at $T = 1.9$ K and $L_\varphi = 550$ Å at $T = 4.2$ K, so that the phase coherence time $\tau_\varphi = 5.4 \cdot 10^{-11}$ s at $T = 1.9$ K and $\tau_\varphi = 2.7 \cdot 10^{-11}$ s at $T = 4.2$ K.

We explain the much weaker negative magnetoresistance for the parallel configuration $\mathbf{B} \| ab$ by incorporating finite-thickness (d) corrections into the strictly 2D theory:[25]

$$\Delta\sigma_s(B_\|) = \frac{e^2}{2\pi^2\hbar}\ln\left(1 + \frac{d^2L_\varphi^2}{3\lambda_\|^4}\right), \quad \lambda_\|^2 = \frac{c\hbar}{eB_\|}. \tag{7}$$

By fitting the theoretical expression (7) to the curves for $\sigma(B_\|)$ (see Fig. 9), we have found the effective thickness of a conducting CuO_2 layer, $d \simeq 1.5$ Å. This value yields an estimate for the extent of the electron wave function in the normal direction, and ensures strong carrier confinement ($d < c$). The single crystal NdCeCuO can therefore be regarded as an analog of an ultra-short-period superlattice (1.5 Å wells /4.5 Å barriers).

As the 2D version of weak localization theory is able to describe the behavior of $\sigma(B,T)$ in our sample, the inequality $\tau_{\mathrm{esc}} > \tau_\varphi$ should be valid for the escape time of an electron from one CuO_2 plane to another. Then we have $\tau_{\mathrm{esc}} \geqslant 5 \cdot 10^{-11}$ s. The escape time between adjacent wells in a superlattice can also be estimated from the value of the normal diffusion constant, $\tau_{\mathrm{esc}} = c^2/D_\perp$. For the parameters of our sample at 300 K, we have[9] $D_\| / D_\perp = 1.7 \cdot 10^4$ with the in-plane diffusion constant $D_\| = 1.2\ \mathrm{cm^2 s^{-1}}$. Then $\tau_{\mathrm{esc}} \simeq 5 \cdot 10^{-11}$ s even at room temperature, so $\tau_{\mathrm{esc}} > \tau_\varphi$ with certainty at low temperatures.

5. CONCLUSION

We have investigated the low-temperature and magnetic field dependence of the normal state in-plane resistivity, ρ_{ab}, in a layered copper oxide single crystal $Nd_{2-x}Ce_xCuO_{4-\delta}$. The material is regarded as an intrinsic two-dimensional conduction system (a collection of 2D conducting CuO_2 planes), and the results are interpreted in terms of the 2D weak localization model. Three indications of 2D weak localization have been displayed: logarithmic temperature dependence of the resistivity, significant negative magnetoresistance for a field normal to the *ab*-plane, and pronounced magnetoresistance anisotropy (much weaker effect for a parallel configuration). A strong dependence of the magnitude of magnetoresistance on the direction of the magnetic field is the most important experimental test for the two-dimensional character of a conducting system.

In a series of samples with $x=0.12-0.20$, we have a full range of disorder parameter values, $k_F l=2-150$. Estimates of essential microscopic parameters, such as the elastic mean free path l, the inelastic scattering length L_φ, and the effective thickness of a conducting layer d, have shown that in accordance with the adopted model, $d \ll l < L_\varphi \ll t$ (t is the geometrical thickness of a sample). Moreover, our estimates show that the thickness of the conducting layer is less than the distance between CuO_2 layers, $d<c$, and this favors carrier confinement within a separate CuO_2 layer. Thus, the NdCeCuO single crystal can be described as a natural superlattice with a confining potential induced both by the specific $pd\sigma$ symmetry of the electron wave function and strong Coulomb correlation effects.

This research was supported by the Russian Program ‘‘Current Problems in Condensed Matter Physics’’ (Grant No. 98004) and the Russian Fund for Fundamental Research (Grant No. 99-02-17343).

*⁾E-mail: ponomarev@imp.uran.ru

---

[1] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[2] N. M. Plakida, *High-Temperature Superconductors*, International Program of Education, Moscow (1996).
[3] T. Ito, Y. Nakamura, H. Takagi, and S. Uchida, Physica C **185–189**, 1267 (1991); T. Ito, H. Takagi, S. Ishibashi, T. Ido, and S. Uchida, Nature (London) **350**, 596 (1991).
[4] S. Uchida, H. Takaji, and Y. Tokura, ISEC-89, Tokyo, 306 (1989).
[5] J. Zaanen, G. A. Sawatsky, and J. W. Allen, Phys. Rev. Lett. **55**, 418 (1985).
[6] H. Takagi, S. Uchida, and Y. Tokura, Phys. Rev. Lett. **62**, 1197 (1989).
[7] S. Massidda, N. Hamada, J. Yu, and A. J. Freeman, Physica C **157**, 571 (1989).
[8] Z. Z. Wang, T. R. Chien, N. R. Ong *et al.*, Phys. Rev. B **43**, 3020 (1991).
[9] A. I. Ponomarev, V. I. Tsidilkovski, K. R. Krylov *et al.*, J. Supercond. **9**, 27 (1996).
[10] Beom-hoan O and J. T. Markert, Phys. Rev. B **47**, 8373 (1993).
[11] Y. Hidaka, Y. Tajima, K. Sugiama *et al.*, J. Phys. Soc. Jpn. **60**, 1185 (1991).
[12] S. J. Hagen, X. Q. Xu, W. Jiang *et al.*, Phys. Rev. B **45**, 515 (1992).
[13] A. Kussmaul, J. S. Moodera, P. M. Tedrow *et al.*, Physica C **177**, 415 (1991).
[14] S. Tanda, M. Honma, and T. Nakayama, Phys. Rev. B **43**, 8725 (1991).
[15] A. A. Ivanov, S. G. Galkin, A. V. Kuznetsov *et al.*, Physica C **180**, 69 (1991).
[16] X. Q. Xu, S. N. Mao, Wu Jiang *et al.*, Phys. Rev. B **53**, 871 (1996).
[17] A. I. Ponomarev, K. R. Krylov, G. I. Harus *et al.*, JETP Lett. **62**, 516 (1995).
[18] G. I. Harus, A. N. Ignatenkov, N. K. Lerinman *et al.*, JETP Lett. **64**, 444 (1996).
[19] Y. X. Jia, J. Z. Liu, M. D. Lan *et al.*, Phys. Rev. B **47**, 6043 (1993).
[20] M. I. Dyakonov, Solid State Commun. **92**, 711 (1994).
[21] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[22] S. Hikami, A. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).
[23] B. L. Altshuler, A. G. Aronov, A. P. Larkin, and D. E. Khmelnitskii, Zh. Éksp. Teor. Fiz. **81**, 768 (1981) [Sov. Phys. JETP **54**, 411 (1981)].
[24] A. V. Samoilov, N.-C. Yeh, and C. C. Tsuei, Phys. Rev. B **57**, 1206 (1998).
[25] B. L. Altshuler and A. G. Aronov, JETP Lett. **33**, 499 (1981).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Structure of the mixed state induced in thin YBaCuO superconducting films by the field of a small ferromagnetic particle

A. Yu. Aladyshkin,*⁾ A. K. Vorob'ev, P. P. Vysheslavtsev, E. B. Klyuenkov,
A. S. Mel'nikov, Yu. N. Nozdrin, and I. D. Tokman

Institute of the Physics of Microstructures, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

(Submitted 2 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1735–1749 (November 1999)

The temperature dependence of the local energy barrier to formation of the mixed state in YBaCuO thin-film superconducting samples has been determined. The measurement technique is based on use of a small ferromagnetic particle as the magnetic field source. It is found that the energy barrier to creation of vortices (for the field oriented parallel to the CuO planes) is anomalously small while the dependence of the corresponding threshold current $j_c(T)$ differs substantially from the temperature dependence of the pair-breaking current. The experimental results are interpreted in terms of the model of a Josephson medium. The observed temperature dependence of j_c points to a strong suppression of the superconducting order parameter at the intergrain boundaries, which for the most probable type of boundaries: superconductor–insulator–superconductor, is evidence of anisotropic pairing. © *1999 American Institute of Physics.* [S1063-7761(99)01711-4]

1. INTRODUCTION

In recent years a great deal of attention has been given to experimental and theoretical studies of the mixed state in thin films of high-temperature superconductors (HTSC's) (see, e.g., the review in Ref. 1 and the references cited therein). The results of measurements of the magnetic properties can be interpreted on the basis of the well-known Bean model. Such calculations turn out to be quite complicated for samples of arbitrary shape and can be performed with relative ease only for certain particular cases (e.g., for samples of cylindrical or ellipsoidal shape). Thus, an analysis of the experimental data on the magnetization of thin films, especially in a field perpendicular to the surface, is a very difficult task. The large demagnetizing factor of such samples leads to a high density of the screening current at the edges of the film, as a result of which the poorly controlled edge structure of the samples to a significant degree determines the experimentally measurable magnetic characteristics.

In the present paper we propose an original method (preliminary results were presented in Ref. 2) for directly measuring local the characteristics of the mixed state (the Bean–Livingston energy barrier for entry of magnetic lines of force **B** from the sources, the critical current of barrier suppression j_c, and the depinning current j_p), allowing one to neglect the influence of edge effects, which are substantial for measurements in uniform fields. The given method is based on an experimental analysis (with the aid of a Hall sensor) of the spatial distribution of the residual magnetization created by the vortices pinned at the pinning centers and penetrating into the film under the action of the field of a small ferromagnetic particle (micromagnet). The use of such a micromagnet, located a small distance a from the film surface ($a \ll L$, L is the lateral dimension of the film), as the field source allows one to neglect edge effects since the currents at the edges of the film in this case are negligibly small. Note that systems of this type (a magnetic dipole above the surface of a superconductor) are being actively investigated at the present time, in particular, as applied to problems of the magnetic force microscope, levitation, etc.[3–5] As a rule, such studies assume that the magnetic dipole moment is quite small and only weakly perturbs the structure of the mixed state. In the present paper we present an experimental and theoretical study of a qualitatively different situation, where the mixed state itself is created by the field of the magnetic dipole (micromagnet). The scenario of the appearance of such a vortex state depends substantially on how the film goes over to the superconducting state, in the presence or in the absence of the field of the micromagnet. Freezing of the vortices in the field of the micromagnet was investigated experimentally in Ref. 6. This case corresponds to the theoretical analysis of Refs. 5, 7–9, based on a comparison of the free energy of a superconductor without vortices and with a single vortex. As will be shown in this paper, a study of the formation of a vortex state (in a dipole field) in samples cooled to $T < T_c$ in the absence of a field is also of great interest since it would yield important information about local characteristics of the sample. In this context we propose the following scenario of the appearance of the mixed state in our experiments. As the distance a between the film and the micromagnet is decreased, the local Meissner current on the surface of the film exceeds a critical value, which leads to the generation near the surface of vortex half-loops. Such half-loops, with increase of their radius, reach the opposite surface of the film, where they split up into a vortex–antivortex pair. These newly formed vortices become pinned to pinning centers and create a residual field. In this scenario it is possible, in particular, to determine the local energy barrier to entry of the vortex lines into the thin superconducting film through its surface. This problem reduces to measuring the threshold distance $a_{\rm cr}$, starting from which destruction of the Meissner state occurs. Note that the problem of determining the local surface energy barrier is of interest in connection with results[10] which indicate a substantial dependence of the barrier on the orientation of the surface relative to the crystal axes, which cannot be explained in terms of the anisotropic Ginzburg–Landau theory.[11,12]

Using this method we have obtained the temperature dependence of the surface energy barrier suppressing current density j_c in thin YBaCuO films in the temperature range 77–90 K. The measured value of j_c turns out to be extremely small in comparison with the theoretical value of the pair-breaking current for homogeneous single-crystal superconductors $j_{\rm GL}$ ($j_c/j_{\rm GL}\sim 10^{-2}$), which is in good agreement with the results of an analysis of the hysteretic dependence of $M(H)$ (Ref. 10). Our experiments have shown that the temperature dependence of $j_c \propto \tau^2$ differs substantially from that of $j_{\rm GL}\propto \tau^{1.5}$, where $\tau=(T_c-T)/T_c$. In our opinion, the observed effects are evidence for the existence in the sample of Josephson-coupled grains, and the temperature dependence points to strong suppression of the magnitude of the order parameter at the grain boundaries. Note that the existence of grains is confirmed by the results of tunnel scanning microscope studies. To interpret our experimental data, we carried out a theoretical analysis of the spatial distribution of the vortices in a thin film in the Bean model, which takes account of the magnetostatics of thin superconducting films in the mixed state.

The paper is organized as follows. Section 2 describes the characteristics of the samples and the experimental setup. Section 3.1 presents results of measurements of the depinning current j_p in a uniform magnetic field. Section 3.2 describes a technique for measuring the critical distance $a_{\rm cr}$ from the micromagnet to the film surface and present results of such measurements. Section 4 analyzes the temperature dependence of the critical distance $a_{\rm cr}$ in order to obtain information about the temperature dependence of the current j_c. Section 5 considers a theoretical model which describes the resulting vortex state and compares the calculated characteristics of the vortex structure with the experimental results.

2. SAMPLE CHARACTERISTICS AND EXPERIMENTAL SETUP

Experiments on measuring the energy barrier at the fixed temperature $T=77$ K were performed on a large number of YBaCuO films differing in thickness (from 850 Å to 3000 Å), preparation technique (magnetron sputtering or laser ablation), and type of substrate (sapphire with yttria-stabilized zirconia (YSZ) sublayer, $NdGaO_3$.) The results of all these experiments convincingly demonstrate the existence of a low energy barrier to entry of vortex lines from the superconductor surface.

To check the temperature measurements of the characteristics of the mixed state, we used four c-oriented YBaCuO films ($M1$, $M2$, $M3$, $M4$) with dimensions 20×20 mm^2 and thickness 850 Å, which were grown in a low-temperature, single-stage process *in situ* by the method of reverse magnetron sputtering. An annular target, prepared from a presynthesized and pressed powder of the phase $YBa_2Cu_3O_x$ sputtered in an argon–oxygen mixture at the optimal ratio Ar:O$=1:1$ and optimal pressure 50 Pa. The temperature was measured and maintained with an accuracy of ± 4 °C in the range 600–750 °C. The substrate was $NdGaO_3$ with orientation ⟨100⟩. The films differed in their conditions of preparation. The temperature of the condensation surface was the same for $M3$ and $M4$ and approximately equal to 700 °C, and for $M1$ and $M2$ it was 20 °C and 40 °C higher, respectively. The deposition rate for $M1$, $M2$, and $M3$ was 4.7 Å/min and the grain size was 5000 Å, while for $M4$ it was two times lower and the grain size was 1.4 times greater. The half-width at half-maximum of the rocking curve for ω-scanning of the (005) reflection of YBaCuO was 1°. The films had the following parameters: $T_c\approx 84-86$ K, width of the resistive junction $\approx 1-2$ K, low microwave (10 GHz) resistance $\sim 10^{-4}-10^{-3}\,\Omega$ at 77 K and high depinning current density $j_p(77\,{\rm K})=10^5-10^6$ A/cm^2, and normal-state resistance (at 90 K) $\sim 100\,\mu\Omega\cdot$ cm. Details of sample preparation and sample characteristics are described more fully in Refs. 13–15.

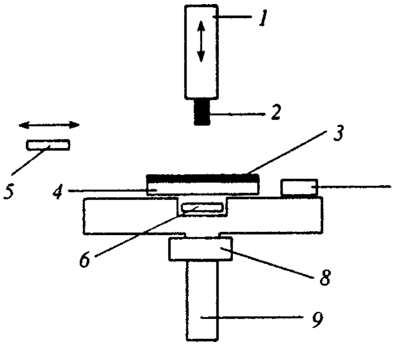

FIG. 1. Diagram of the experimental setup: *1*—bar; *2*—micromagnet; *3*—YBaCuO film; *4*—substrate; *5*—scanning Hall sensor $P1$; *6*—fixed Hall sensor $P2$; *7*—temperature-sensitive element; *8*—heater; *9*—heat sink.

The setup for performing the temperature measurements was a copper table with a massive base immersed in liquid nitrogen (Fig. 1). The investigated film was placed on the table, under which a heater was mounted on a rod. The temperature was determined from the change in the resistance of a copper wire which played the role of a temperature sensor and was situated in contact with the film on the surface of the table. A prescribed temperature was maintained by a thermostat, which created a balance between the delivered heat and the removed heat by controlling the current flowing through the heater.

A particle of $SmCo_5$ with dimensions $300\times 300\times 500\,\mu$m^3 and magnetic moment 6.7×10^{-3} G·cm^3 served as the micromagnet. Characteristics of the vortex distributions were measured by two identical Hall sensors based on InSb films with dimensions of the working region $50\times 100\,\mu$m^2. A scanning sensor $P1$ was situated above the film surface at a distance of 100 μm and was used to measure the spatial dependence of the z component of the magnetic field $B_z(x,y)$ at $T=77$ K. The fixed Hall sensor $P2$ was located under the film precisely under the micromagnet

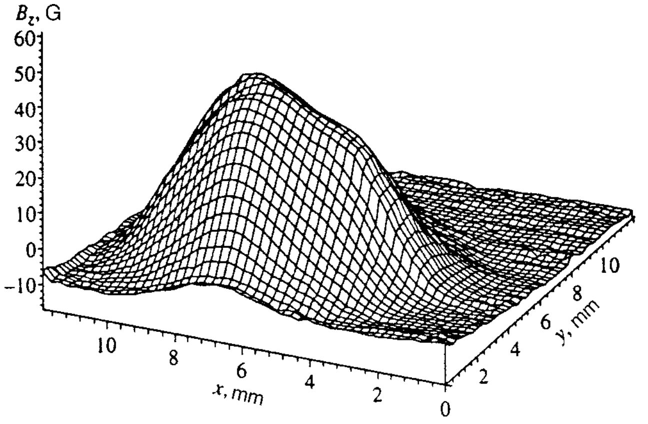

FIG. 2. Spatial distribution of the residual magnetic field B_z in the film, measured by the scanning sensor $P1$ after switching the external uniform magnetic field on and off ($H=500$ G, $T=77$ K).

and was used for temperature measurements of the residual magnetization. The sensors in the xy plane had spatial resolution $R_{H1} \approx 100\,\mu$m and $R_{H2} \approx 700\,\mu$m, respectively.

3. EXPERIMENTAL RESULTS

3.1. Film in a homogeneous magnetic field

To determine the depinning current density j_p and its temperature dependence, we performed experiments in a uniform magnetic field. The film, located in a solenoid, was cooled in zero field down to temperatures $77\,\mathrm{K} \leqslant T \leqslant T_c$. After increasing the field to values $B \geqslant 600$ G, thereby ensuring complete penetration of the flux into the film and subsequently lowering the field to zero, we used the scanning Hall sensor $P1$ to record the spatial distribution of the trapped field $B_z(x,y)$ (Fig. 2). The $B_z(x,y)$ measurements are found to be in agreement with typical experimental data for YBaCuO films.[16] In Fig. 2 it can be seen that the region where the vortices are pinned is not localized; this indicates penetration of vortex lines from the edges.[1] It should be noted that some of the behavior (logarithmic growth of the filed near the center and change of sign of the second derivative) predicted by theoretical calculation in the model of the critical state with uniform depinning current density $j=j_p=\mathrm{const}$ are not observed in the experiment. In our view, this may be due to the finite spatial resolution R of the experimental setup (in our case $R=R_{H1} \approx 100\,\mu$m, and in Ref. 16 $R=1.6$ mm) or to large-scale inhomogeneities which lead to the dependence $j_p(x,y)$. Nevertheless, in the critical-state model with $j=j_p=\mathrm{const}$ (Ref. 17) (i.e., ignoring the inhomogeneities) it is possible to estimate the mean depinning current density j_p, averaged over the surface of the film, using the formula

$$j_p = \frac{c \max[B_z]}{2\pi d \ln(L/R_{H2})}, \tag{1}$$

where d is the film thickness and $d=850$ Å. The temperature dependence $j_p(T)$ obtained in this way is plotted in Fig. 3. Note that this dependence is nearly linear at temperatures far from T_c.

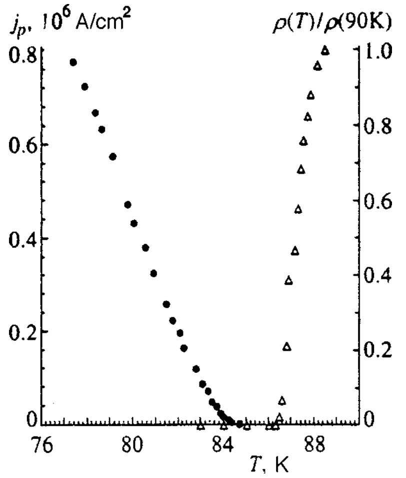

FIG. 3. Typical temperature dependence of the depinning current j_p (●) and the resistivity ρ (△).

3.2. Film in the field of the micromagnet

As was indicated in the Introduction, the use of a micromagnet makes it possible to neglect the influence of edge effects, which are important in uniform fields. The field of a micromagnet at distances $r \gg l$ (where l is the characteristic dimension of the micromagnet) falls off with increasing distance as $1/r^3$. Therefore, despite the increase in the current at the edge (by virtue of the large demagnetizing factor), the current density near the edges turns out to be negligibly small in comparison with the pair-breaking current and insufficient for vortex creation. The form of the spatial distribution of B_z on the surface of the thin film magnetized in the field of the micromagnet (Fig. 4) confirms that vortex lines cannot penetrate the film from the edges: the vortices are localized in a bounded region of the film under the micromagnet. As a result, in experiments on samples with finite dimensions the presence of an edge does not have an effect,

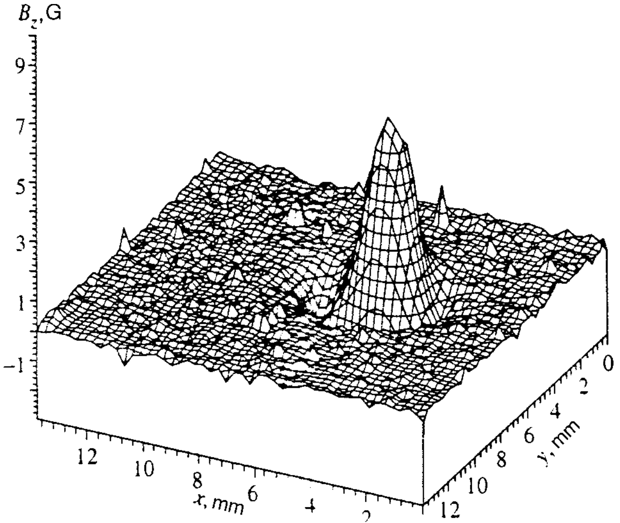

FIG. 4. Spatial distribution of the residual magnetic field B_z in the film, measured by the scanning sensor $P1$ after lowering the micromagnetic to a height $a < a_{\mathrm{cr}}$ and its subsequent removal from the film ($T=77$ K).

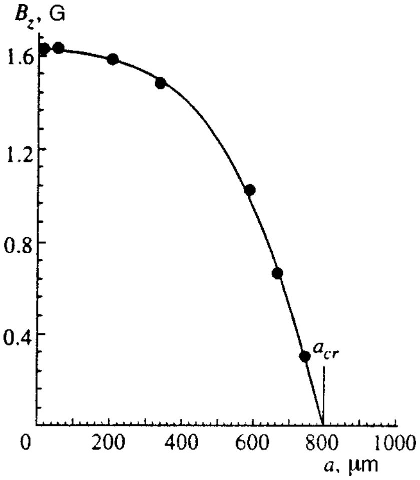

FIG. 5. Typical dependence of the residual field B_z in the center of the film on the distance a at $T=77$ K.

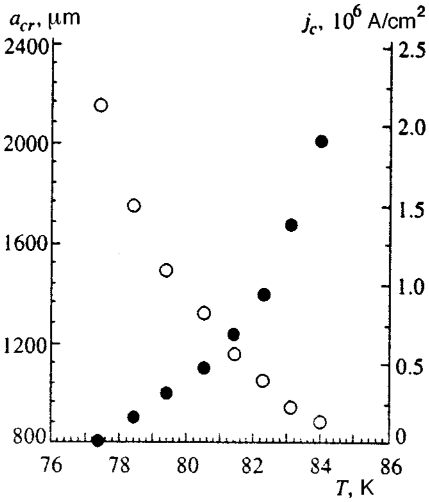

FIG. 6. Typical dependence of $a_{\rm cr}$ (●) and j_c (○) on the temperature T.

and the possibility exists of carrying out local measurements. In the present section we show how the technique of measuring local characteristics of the mixed state is used to find the temperature dependence of the surface energy barrier suppressing current density j_c.

The experiment was performed in the temperature interval $77\,{\rm K}<T<T_c$ and in the range of distances between the center of the micromagnet and the surface $150\,\mu{\rm m}<a<4000\,\mu{\rm m}$ for each sample. The procedure for measuring the residual magnetization arising in the film in the field of the micromagnet was as follows. The film with frozen-in zero flux was cooled to a fixed temperature which was maintained with an accuracy $\Delta T=0.01$ K. The micromagnet, originally located far from the film, was lowered to some height a, which was measured with a micrometer, and was then raised to its original height. After this, the magnetic field of the film was measured with the sensor $P2$. Using this technique, we obtained the dependence of the residual magnetization on the distance a to the micromagnet for each sample at various temperatures. Let us enumerate our main results:

a) the vortex lines penetrate into the thin-film sample through its surface and not from its edges;

b) the resulting vortex distribution has regions with positive and negative components of B_z; the maximum of the absolute value of B_z grows with decreasing a;

c) the existence of a critical distance $a_{\rm cr}$ is observed, which corresponds to onset of entry of vortices into the superconductor (Fig. 5); for $a>a_{\rm cr}$ residual magnetization is absent, which corresponds to the Meissner state (the vortex state is formed only for $a<a_{\rm cr}$);

d) the critical distance $a_{\rm cr}$ increases as the temperature T is increased (Fig. 6).

To increase the accuracy of the measurements, the critical distance $a_{\rm cr}$ (in the presence of noise corresponding to a magnetic field ≈ 0.1 G) was determined by extrapolation of the dependence of the signal at the sensor $P2$ on the distance a (Fig. 5). The character of the observed experimental distributions $B_z(x,y)$ was the same for all samples. The temperature dependence of $a_{\rm cr}$ (Fig. 6) is well approximated by a power law $a_{\rm cr}\propto\tau^n$. For all samples the exponent $n\simeq-0.7$. Using these data, we will find the value of j_c using the model described above.

4. DISCUSSION OF RESULTS

4.1. Temperature dependence of the critical current j_c

The relation between the critical distance $a_{\rm cr}$ and the current density j_c needed for analysis of the experimental data can be easily obtained by solving the problem of the distribution of the Meissner current induced by the field of the micromagnet. For the case of an infinite superconducting film the initial equation has the form

$$\nabla\times\nabla\times\mathbf{A}+\frac{f(z)}{\lambda_{ab}^2}\mathbf{A}=4\pi\delta(z+a)\nabla\times\{\mathbf{m}\delta(x,y)\}, \tag{2}$$

where x,y,z is the coordinate system with origin on the film surface, λ_{ab} is the effective penetration depth for currents parallel to the xy plane; and the micromagnet (magnetic dipole) is located at the point $z=-a$, $x=y=0$ with $\mathbf{m}\|\mathbf{z}$. The function $f(z)$ is defined as follows: $f(z)=1$ for $0<z<d$ and $f(z)=0$ for $z<0$ and $z>d$.

Note that λ_{ab} depends strongly on the structure of the samples: for uniform single-crystal films λ_{ab} coincides with the London penetration depth λ_L while for grainy media this length depends on the intergrain Josephson interaction.[18] From Eq. (2) with the help of the relation

$$j_\varphi=-\frac{c}{4\pi\lambda_{ab}^2}A_\varphi \tag{3}$$

we obtain the following expression for the current density:

$$j_\varphi=-\frac{cm}{2\pi\lambda_{ab}^2}\int_0^\infty \exp(-qa)\times\frac{k\cosh[k(d-z)]+q\sinh[k(d-z)]}{(k^2+q^2)\sinh(kd)+2kd\cosh(kd)}q^2J_1(q\rho)dq, \tag{4}$$

where r,φ,z is the cylindrical system of coordinates, $k^2=q^2+\lambda_{ab}^{-2}$, and J_1 is the first-order Bessel function.

TABLE I. Main parameters of the investigated samples.

Film	T_c, K	j_c(77 K), A/cm^2	j_p(77 K), A/cm^2	p
$M1$	86.4	2.2×10^6	0.8×10^6	2.1 ± 0.1
$M2$	86.9	2.6×10^6	0.85×10^6	2.0 ± 0.1
$M3$	85.5	2.0×10^6	0.95×10^6	2.0 ± 0.1
$M4$	86.6	1.0×10^6	0.7×10^6	1.9 ± 0.1

Clearly, Eq. (4) can also be used for a film of finite dimensions in the limits $a \ll L$ and $r \ll L$. To analyze the experimental data it is convenient to use a simplified expression that follows from Eq. (4) for $a \gg \max[\lambda_{ab}, \lambda_{ab}^2/d]$:

$$j_\varphi(r) = -\frac{3cmra}{2\pi\lambda_{ab}(r^2+a^2)^{5/2}}\frac{\cosh[(z-d)/\lambda_{ab}]}{\sinh(d/\lambda_{ab})}. \qquad (5)$$

Employing Eq. (5) and the experimental dependence of $a_{\mathrm{cr}}(T)$, we can determine the value of the critical current density $j_c(T) = \max|j_\varphi(r,\varphi)| = |j_\varphi(a_{\mathrm{cr}}/2,0)|$ corresponding to the onset of vortex penetration. Here we have also made use of the condition $\lambda_{ab} \gg d$, which turns out to hold in the temperature range under consideration (see the estimates derived below). Figure 6 displays a typical dependence of $j_c(T)$. Let us turn our attention now to some important features of the obtained results:

a) for the entire temperature range $j_c(T)$ is significantly smaller than j_{GL}, the Ginzburg–Landau critical current density needed to suppress the surface barrier in ideal samples (at $T = 77$ K we have $j/j_{\mathrm{GL}} \sim 10^{-2}$); thus, our results are in agreement with the results of Refs. 2 and 10, where the smallness of the Bean–Livingston barrier to entry of vortex lines parallel to the CuO planes was demonstrated;

b) for temperatures near T_c the experimental data are well approximated by the dependence $j_c \propto \tau^p$, where $p \simeq 2$ (see also Table I, which lists the main parameters of the samples).

Let us now consider possible mechanisms of the observed strong suppression of the Bean–Livingston barrier. First, suppression of the barrier can clearly be connected with surface roughness and surface defects. This mechanism lowers the critical current of penetration of the first vortices: $j \sim \gamma j_{\mathrm{GL}}$, where $\gamma < 1$. Nevertheless, it is improbable that such a mechanism should give the small value $\gamma \sim 10^{-2}$ corresponding to the experimental data. Besides, even if we assume that surface roughness leads to the appearance of vortex half-loops near the surface, breaking of these half-loops into vortex–antivortex pairs would still require a rather strong current density $j \sim j_{\mathrm{GL}}$. It should also be noted that barrier suppression due to surface roughness, in our view, cannot explain the observed temperature dependence $j_c \propto \tau^p$ with exponent $p \simeq 2$, which differ substantially from the temperature dependence of the Ginzburg–Landau current density $j_{\mathrm{GL}} \propto \tau^{1.5}$.

Second, a low barrier can be easily explained in terms of the model of a grainy Josephson medium. Here the role of Josephson contacts can be played, for example, by grain boundaries. The effective critical pair-breaking current in such a model is equal in order of magnitude to the Josephson critical current between the grains, which is much less than j_{GL}. According to Ref. 18, the effective penetration depth can be written as follows:

$$\lambda_{ab}^2 = \frac{c\Phi_0}{16\pi^2\mu l j_c}, \qquad (6)$$

where μ is the effective permeability of the medium, associated with the ratio λ_L/l (for $\lambda_L/l \sim 1$ we have $\mu < 1$.) For $\lambda_L(77\,\mathrm{K}) = 3000$ Å, $l \approx 0.5\,\mu$m, and $j_c = 2\times10^6$ A/cm^2 we obtain $\lambda_{ab} \sim l$. It can be seen that at all temperatures $T > 77$ K the condition $\lambda_{ab} > l$, which was used above in the analysis of the experimental data, is valid. At temperatures $T < 77$ K the penetration depth λ_{ab} can be less than the grain size l and the simple model (see Ref. 18) based on the free energy functional averaged over the scale l is no longer valid. The temperature dependence of j_c observed in our experiment differs significantly from the temperature dependence characteristic for Josephson contacts of the form superconductor–insulator–superconductor ($j_c \propto \tau$; $\tau \ll 1$) (Ref. 19), and points to a substantial suppression of the order parameter at the grain boundaries. This conclusion is found to be in agreement with the results of Ref. 20, which measured the magnetic susceptibility of polycrystalline YBaCuO films with resistivity $\rho_{ab}(100\,\mathrm{K}) = 800\,\mu\Omega\cdot$cm and $T_c = 80$ K. If we ignore the possible existence of superconductor–normal metal–superconductor contacts in YBaCuO films, then our results point to an anisotropic type of pairing in the grains.

4.2. Model of the critical state in thin films

To analyze the characteristics of the mixed state formed in the field of a micromagnet, we considered a simple model of the critical state in thin films with thicknesses $d \ll \lambda_{ab}$, which allowed us to find the steady-state vortex distribution in an external magnetic field.

As was stated above, vortex–antivortex pairs are created in a bounded region near the maximum of the Meissner current $r_{\max} = a/2$. The vortices that have entered the film under the action of the Lorentz force will move to the opposite side, redistributing the residual magnetization as a result. A steady-state vortex distribution in the film is possible provided the following conditions hold (see, e.g., Ref. 21). First, the current density should not exceed the critical value j_c (otherwise additional vortices will be generated). Second, in order that the vortex structure be fixed, the current density in the region of the vortices should not be greater than the depinning current density j_p. A simple variant of the nascent vortex structure is a set of two nonoverlapping regions in which either the vortex density or the antivortex density is nonzero.

Let $\mathbf{j}_1(\mathbf{r},\mathbf{r}')$ be the current density at the point $\mathbf{r}$ created by a vortex located at the point $\mathbf{r}'$:

$$\mathbf{j}_1(\mathbf{r},\mathbf{r}') = j_1(|\mathbf{r}-\mathbf{r}'|)\mathbf{z}\times\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}.$$

Then the expression for the current density $\mathbf{j}_{\mathrm{tot}}(\mathbf{r})$ created by the fixed vortices distributed with concentration $n(\mathbf{r})$ has the form

$$\mathbf{j}_{\mathrm{tot}} = \int n(\mathbf{r}')\mathbf{j}_1(|\mathbf{r}-\mathbf{r}'|)d\mathbf{r}'. \tag{7}$$

For the case of an axially symmetric distribution $n(\mathbf{r}) = n(r)$, Eq. (7) reduces to the form

$$j_{\mathrm{tot},\varphi} = \frac{1}{N}\int_0^\infty 2\pi n(r')G(r,r')r'dr', \tag{8}$$

where $G(r,r')$ is the current density created by a vortex ring with concentration of vortices $n(r) = N\delta(r-r')/2\pi r$, where N is the total number of vortices.

An analysis of the creation of a vortex structure can be performed at the threshold of vortex-pair generation $[(a_{\mathrm{cr}}-a)/a_{\mathrm{cr}} \ll 1]$. In this case, the steady-state vortex distribution consists of two narrow rings located a great distance apart: an inner vortex ring of radius r_1 and half-width θ_1 and an outer vortex ring of radius $r_2 \gg r_1$ and half-width θ_2 ($\theta_{1,2} \ll r_{1,2}$). We also assume that $j_p \ll j_c$. This condition is in agreement with our experimental results for T close to T_c (see Sec. 3.1 and 4.1). The indicated conditions allow us to simplify the expression for $G(r,r')$ in Eq. (8) (see Appendix A) and obtain the following system of equations:

$$\fint_a^b \frac{c\Phi_0 n_+(r')dr'}{2\pi^2 d(r-r')} + j_\varphi(r) + \frac{3Nc\Phi_0 r_1}{8\pi^2 d r_2^3} = -j_p, \quad a \leqslant r \leqslant b, \tag{9}$$

$$-\fint_c^d \frac{c\Phi_0 n_-(r')r'dr'}{2\pi^2 dr(r-r')} + j_\varphi(r) + \frac{Nc\Phi_0}{4\pi^2 d r_2^2} = -j_p,$$

$$c \leqslant r \leqslant d, \tag{10}$$

$$j_\varphi(\tilde{r}) + \frac{c\Phi_0 N}{4\pi^2 d\tilde{r}^2} + \frac{3c\Phi_0 N r}{4\pi^2 d\tilde{r}^2} = -j_c, \tag{11}$$

$$\left.\frac{d}{dr}j_\varphi(r)\right|_{\tilde{r}} + \left.\frac{d}{dr}\left(\frac{c\Phi_0 N}{4\pi^2 dr^2}\right)\right|_{\tilde{r}} + \left.\frac{d}{dr}\left(\frac{3c\Phi_0 N r}{4\pi^2 dr^2}\right)\right|_{\tilde{r}} = 0, \tag{12}$$

where $a \leqslant r \leqslant b$ and $c \leqslant r \leqslant d$ are respectively the regions of existence of the vortices and antivortices; $\tilde{r}$ is the point at which the total current reaches its maximum; n_+ is the vortex density; and n_- is the antivortex density. Equations (9) and (10) describe the conditions of immobility of the vortex structure $[j_{\mathrm{tot}}(r) = j_p$ for $a \leqslant r \leqslant b$ and $c \leqslant r \leqslant d]$, while Eqs. (11) and (12) correspond to the situation where the maximum value of the current density is equal to j_c.

Taking the approach employed, for example, in Ref. 21, it is possible to invert the integral equations (9)–(12) (see Appendix B). After solving the resulting system we obtain the following expressions for the vortex and antivortex densities:

$$n_+ = \frac{3m}{\Phi_0 a_{\mathrm{cr}}^4}\sqrt{\theta_1^2-(r-r_1)^2}, \quad n_- = \frac{12 m a_{\mathrm{cr}}}{\Phi_0 r_2^5}\sqrt{\theta_2^2-(r-r_2)^2}, \tag{13}$$

where

$$r_1 = \frac{2\pi j_p d a_{\mathrm{cr}}^4}{3mc}, \quad r_1 r_2^4 = a_{\mathrm{cr}}^5, \tag{14}$$

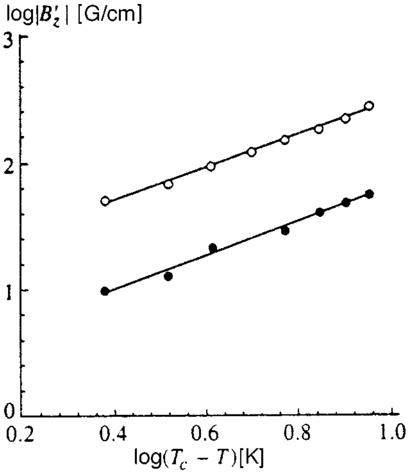

FIG. 7. Temperature dependence of $B_z' = dB_z/da$ (● — experiment, ○ — theory).

$$\theta_1^2 = \frac{Nc\Phi_0}{2\pi^3 d j_p}, \quad \theta_1 = \frac{1}{2}\theta_2. \tag{15}$$

Then

$$N = 2\pi r_1 \int_{r_1-\theta_1}^{r_1+\theta_1} n_+(r)dr = 2\pi r_2 \int_{r_2-\theta_2}^{r_2+\theta_2} n_-(r)dr$$

$$= \frac{9m\pi}{1.25^{5/2}\Phi_0 \cdot 4a_{\mathrm{cr}}}\left(1-\frac{a}{a_{\mathrm{cr}}}\right). \tag{16}$$

To compare the results obtained using the considered theoretical model with the experimental data, it is necessary to find a relation linking the experimentally measured value of B_z with the parameters of the vortex structure. Because two narrow vortex rings ($\theta_{1,2} \ll r_{1,2}$) are formed near the generation threshold, to find the magnetic field created by the film we can take the vortex density to be equal to

$$n(r) = \frac{N}{2\pi r}[\delta(r-r_1)-\delta(r-r_2)]. \tag{17}$$

The desired expression for the z component of the magnetic field B_z measured by the Hall sensor $P2$ (located under the film at the distance $H = 700\,\mu$m) has the form

$$B_z = \int_0^\infty \frac{H}{(r'^2+H^2)^{3/2}}\Phi_0 n(r')r'dr'$$

$$= \frac{NH\Phi_0}{2\pi}\left[\frac{1}{(r_1^2+H^2)^{3/2}} - \frac{1}{(r_2^2+H^2)^{3/2}}\right]. \tag{18}$$

It is interesting to compare the slope of the experimental curve $B_z(a)$ for $a = a_{\mathrm{cr}}$ with the theoretical value

$$B_z' \equiv \left.\left(\frac{dB_z}{da}\right)\right|_{a=a_{\mathrm{cr}}} = -\frac{9mH}{1.25^{5/2}\cdot 8a_{\mathrm{cr}}^2}\left[\frac{1}{(r_1^2+H^2)^{3/2}} - \frac{1}{(r_2^2+H^2)^{3/2}}\right]. \tag{19}$$

As can be seen from Fig. 7, the temperature dependence $B_z'(T)$ demonstrates good agreement between the theoretical calculations and the experimental data.

5. CONCLUSION

In the present paper we have proposed a new experimental method for determining the local characteristics of superconducting films, in particular, the surface energy barrier for entry of vortex lines into the superconductor. The measurement technique is based on the use of a small ferromagnetic particle as the magnetic-field source and is therefore free of the drawbacks characteristic of methods employing uniform external fields to determine the characteristics of samples with a large demagnetizing factor. The primary virtue of the method is that it allows one to obtain direct information about important parameters of a superconductor regardless of the geometry of the sample. The experiments conducted on YBaCuO films have shown that the energy barrier to creation of vortices (for the field oriented parallel to the CuO planes) is anomalously small. The temperature dependence of the corresponding threshold current j_c (induced by the micromagnet) differs substantially from the temperature dependence of the pair-breaking current responsible for overcoming the Bean–Livingston barrier in ideal samples. These facts can be interpreted in terms of the model of a Josephson medium. In this case, the value that we have obtained for $j_c(T)$ is the intergrain critical current density. The observed temperature dependence of j_c is indicative of a strong suppression of the superconducting order parameter at the intergrain boundaries, which for the most probable type of boundaries (superconductor–insulator–superconductor) is evidence of anisotropic pairing. Further analysis of the temperature dependence of $j_c(T)$ in samples with different microstructure is thus very important for answering the question of the type of pairing in high-temperature superconductors.

In conclusion, we would like to express our gratitude to A. A. Andronov for valuable remarks and discussion of the results. This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grant No. 97-02-17437) and the International Center–Foundation for Promising Research in Nizhniĭ Novgorod (Grant No. 99-2-03).

APPENDIX A

Here we obtain the asymptotic limits for the Green's function $G(r,r')$ in Eq. (8). We find the current density created by a ring of vortices with concentration

$$n(r)=N\delta(r-r')/(2\pi r) \tag{A1}$$

(r' is the radius of the ring). It is well known that the expression for $j_1(\mathbf{r},\mathbf{r}')$ for $|\mathbf{r}-\mathbf{r}'|\gg\lambda_{\mathrm{eff}}$ ($\lambda_{\mathrm{eff}}=\lambda_{ab}^2/d$) has the form

$$j_1(\mathbf{r},\mathbf{r}')=\frac{c\Phi_0}{4\pi^2|\mathbf{r}-\mathbf{r}'|^2 d}. \tag{A2}$$

After substituting Eqs. (A1) and (A2) into expression (7) and integrating over $r'=|\mathbf{r}'|$ we obtain

$$G(r,r')=\frac{c\Phi_0 N}{8\pi^3 d}\int_{-\pi}^{\pi}\frac{(r-r'\cos\varphi)\,d\varphi}{(r^2+r'^2-2rr'\cos\varphi)^{3/2}}. \tag{A3}$$

After some mathematical manipulations, expression (A3) can be reduced to the form

$$G(r,r')=\frac{c\Phi_0 N}{4\pi^3 d}\left[\frac{E(k)}{r(r-r')}+\frac{K(k)}{r(r+r')}\right], \tag{A4}$$

where $K(k)$ and $E(k)$ are the complete elliptical functions of the first and second kind and $k^2=4rr'/(r^2+r'^2)$.

We write down the asymptotic limits of the Green's function $G(r,r')$ [Eq. (A4)]:

1) for $|r-r'|\ll r'$

$$G(r,r')=\frac{c\Phi_0 N}{4\pi^3 d}\,\frac{1}{r'(r-r')}; \tag{A5}$$

2) for $|r-r'|\gg r'$ and $r>r'$

$$G(r,r')=\frac{c\Phi_0 N}{4\pi^2 d}\,\frac{1}{r^2} \tag{A6}$$

(this form of the Green's function is used to calculate the current density created by a vortex ring in the region of an antivortex ring);

3) for $|r-r'|\gg r_0$ and $r<r'$

$$G(r,r')=\frac{3c\Phi_0 N}{8\pi^2 d}\,\frac{r}{r'^3} \tag{A7}$$

(this form of the Green's function is used to calculate the current density created by an antivortex ring in the region of a vortex ring).

APPENDIX B

To solve the system of integral equations, we used the approach developed in Ref. 22. In this situation, according to Ref. 22, for the singular integral equation

$$\frac{1}{\pi i}\int_a^b\frac{\varphi(t)\,dt}{t-t_0}=f(t_0)+C,\quad a\le t_0\le b, \tag{B1}$$

(where $\varphi(t)$ is an unknown function satisfying the Helder condition, and C is an unknown constant) the unique solution has the form

$$\varphi(t_0)=\frac{\sqrt{(t_0-a)(t_0-b)}}{\pi i}\int_a^b\frac{f(x)\,dx}{\sqrt{(x-a)(x-b)}(x-t_0)}, \tag{B2}$$

$$C=\frac{1}{\pi i}\int_a^b\frac{f(x)\,dx}{\sqrt{(x-a)(x-b)}}. \tag{B3}$$

*)E-mail: alay@ipm.sci-nnov.ru

---

[1] E. H. Brandt, Rep. Prog. Phys. **58**, 1465 (1995).
[2] A. S. Mel'nikov, Yu. N. Nozdrin, I. D. Tokman, and P. P. Vysheslavtsev, Phys. Rev. B **58**, 11672 (1998).
[3] M. W. Coffey, Phys. Rev. B **52**, R9851 (1995).
[4] J. H. Xu, J. H. Miller, Jr., and C. S. Ting, Phys. Rev. B **51**, 424 (1995).
[5] J. C. Wei, J. L. Chen, L. Horng, and T. J. Yang, Phys. Rev. B **54**, 15429 (1995).
[6] H. J. Hug, A. Moser, I. Parashikov, B. Stiefel, O. Fritz, H.-J. Güntherodt, and H. Thomas, Physica C **235–240**, 2695 (1995).
[7] I. D. Tokman, Phys. Lett. A **166**, 412 (1992).

[8] G. M. Genkin, V. V. Skuzovatkin, and I. D. Tokman, J. Magn. Magn. Mater. **130**, 51 (1994).
[9] Y. N. Nozdrin, P. P. Vysheslavtsev, I. D. Tokman, and I. M. Gordion, IEEE Trans. Appl. Supercond. **5**, 1424 (1995).
[10] D.-X. Chen, R. B. Goldfarb, R. W. Cross, and A. Sanchez, Phys. Rev. B **48**, 6426 (1993).
[11] V. P. Damjanovic and A. Yu. Simonov, J. Phys. I **1**, 1639 (1991).
[12] V. P. Dam'yanovich and A. Yu. Simonov, Sverkhprovodimost: Fiz., Khim., Tekh. **4**, 1512 (1991).
[13] Yu. N. Drozdov, S. V. Gaponov, S. A. Gusev *et al.*, Supercond. Sci. Technol. **9**, A166 (1996).
[14] R. K. Belov, Yu. N. Drozdov, S. V. Gaponov *et al.*, IEEE Trans. Appl. Supercond. **5**, 1797 (1995).
[15] Yu. N. Drozdov, S. V. Gaponov, S. A. Gusev *et al.*, Supercond. Sci. Technol. **7**, 1642 (1997).
[16] H. Darhmaoui, J. Jung, J. Talvacchio, M. A.-K. Mohamed, and L. Friedrich, Phys. Rev. B **53**, 12330 (1996).
[17] P. N. Mikheenko and Yu. E. Kuzovlev, Physica C **204**, 229 (1993).
[18] É. B. Sonin, JETP Lett. **47**, 496 (1988).
[19] V. Ambegaokar and A. Baratoff, Phys. Rev. Lett. **10**, 486 (1963).
[20] W. Widder, L. Bauernfeind, H. F. Braun, H. Burkhardt, D. Rainer, M. Bauer, and H. Kinder, Phys. Rev. **55**, 1254 (1997).
[21] I. L. Maksimov, A. A. Elistratov, JETP Lett. **61**, 208 (1994).
[22] N. I. Muskhelishvili, *Singular Integral Equations*, P. Noordhoff, Groningen (1953) [Russian orig., 2nd ed., Fizmatgiz, Moscow, 1963].

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Spectroscopy of electroreflection, the electron band structure, and the mechanism of visible photoluminescence of anisotropically etched silicon

E. F. Venger, T. Ya. Gorbach, L. A. Matveeva,*) and S. V. Svechnikov

Institute of Semiconductor Physics, National Academy of Sciences, 252028 Kiev, Ukraine

(Submitted 13 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1750–1761 (November 1999)

We present the results of studies of electroreflection in the 1.1–4.4 eV spectral range, of electron Auger spectroscopy, and of electron diffraction involving the photoluminescent Si–SiO_2 system prepared via anisotropic chemical etching of the Si(100) surface. These results are explained on the basis of a four-layer model of the band structure and energy transition diagram for a system with a quantum well at the silicon surface, surface electron states at the boundary, and a gradient of the band potential in the transition layer. We find that light-emitting silicon remains an indirect-gap semiconductor and that the visible photoluminescence is due to direct recombinations of the light-excited electrons and holes in the quantum well at the center of the Brillouin zone with the participation of the band of deep localized states, which is due to the presence of oxygen at the silicon surface. © *1999 American Institute of Physics.* [S1063-7761(99)01811-9]

1. INTRODUCTION

Since visible photoluminescence was first detected in silicon at room temperature,[1] much attention has been paid to this phenomenon, which is unexpected in indirect-gap semiconductors. The unflagging fundamental and practical interest in the emission of radiation by silicon[2–5] is due to the specific optical and electronic properties of this material, properties that also make it possible to record electroluminescence at room temperature.[6,7] This broadens the possibility of using one of the main materials of electronics not only in microelectronics but also in optoelectronics.

In silicon direct radiative recombination of electrons at a side minimum in the conduction band and a hole at the top of the valence band is usually impossible without the participation of phonons or violation of selection rules, so the photoluminescence efficiency of silicon is extremely low. To increase this efficiency, the electrons and holes must be localized at the same point of the Brillouin zone (in *k*-space). However, the results of numerous studies indicate that intense photoluminescence at room temperature can be detected after the silicon surface has been treated appropriately by one of the following methods: anodic electrolytic etching,[1] chemical etching,[8] photochemical etching,[9] and anisotropic chemical etching.[10] It was also found that silicon nanospheres deposited on quartz or germanium substrates and covered by a layer of silicon dioxide emit light at room temperature when excited by laser light,[11] and so do Si–SiO_2 superlattices.[12]

It has been established without doubt that irrespective of the method used in preparing light-emitting silicon, the energy of the emitted photons exceeds the energy of the indirect transition (1.12 eV) and does not coincide with the known energies of direct optical transitions for silicon at the critical points of the Brillouin band. Therefore, it is quite natural that most studies of light-emitting silicon have been aimed at establishing the nature and mechanism of light emission in silicon, which were usually found from the spectral dependence of the photoluminescence intensity and the effects of various factors, such as the nanocrystallite size,[13] thermal annealing,[14] ultraviolet,[15] laser,[16] electron,[17] and gamma[18] radiation, chemical etching,[19] electric field,[20] and the surroundings.[21] However, despite the substantial advances in the studies of properties of light-emitting silicon, the physical mechanism of visible photoluminescence of silicon is still unclear.

Although excitonic annihilation at the surface states of quantum-size structures is the generally accepted model, a standard mechanism of visible photoluminescence in silicon has yet to be developed. Neither has it been established without doubt that light-emitting silicon is an indirect-gap semiconductor, although there are indications of this in the results of studies of light-induced absorption in porous silicon.[22] According to theoretical calculations,[23–26] light-emitting silicon becomes a quasidirect-gap semiconductor with an increased band gap and less stringent selection rules due to the effect quantum-size confinement in nanostructures. The latter is important if we want to increase the photoluminescence efficiency.

Studies of the electron band structure of light-emitting silicon by the modulation spectroscopy method should help in understanding the nature of efficient visible photoluminescence of indirect-gap semiconductors. The high sensitivity of this method to details of the band structure is well known.[27] The advantage of modulation spectra over classical spectra lies in the fact that they allow the fine structure, usually hidden by the structureless background, to be established. Here the electroreflection method is the most sensitive, since the electroreflection signal is determined by the third energy derivative of the optical constants (and hence of the reflection). The signal exists only at the critical points of the

Brillouin zone in the direct-transition region, vanishes as we move away from these point, and is determined by both optical and electron properties of the semiconductor surface under investigation.[28] No studies of light-emitting silicon by this method have been conducted so far, and there are no data on the electroreflection and changes in the electron structure of silicon initiated by the transition to the light-emitting state in the literature.

This paper is a report on the studies of the energy spectrum of light-emitting silicon, prepared by anisotropic etching, The results were extracted by the electroreflection method and compared with the spectrum of ordinary (non-light-emitting) bulk silicon. We also examined the results of electron diffraction and electron Auger spectroscopy. On the basis of the data, we interpret the nature of photoluminescence in way that differs from the common approach.

2. EXPERIMENT

Layers of photoluminescent silicon with a thickness of up to 0.3 μm each were formed by anisotropic chemical etching of a mirror-smooth n-Si(100) surface in a $\mathrm{HF:HNO_3=20:1}$ mixture in the course of 20–30 min. Photoluminescence was recorded at room temperature by a standard device and was excited by the 514.5-nm light of an argon laser with a flux of less than 0.1 W cm^{-2} (see Ref. 10).

The essence of the electroreflection method consists in recording the modulation of the reflectivity of the sample that appears when an external modulating electric field is applied to the sample. The quantity that is measured is the relative reflectivity modulation, $\Delta R/R$. The electroreflection spectra were measured in an electrolytic cell with a 0.1-normal aqueous solution of KCl at room temperature in the 1.1–4.4 eV spectral range, which included the photon-emission range for silicon and the direct transitions at the center of the Brillouin zone and at the edge of the zone in the direction $\langle 100\rangle$ of the wave vector, where the absolute minimum X_1 of the conduction band of silicon is located.

The measurements were done by a device based on the DMR-4 monochromator with automatic recording of the spectrum on the display with a linear energy scale and a 0.003-eV resolution. The sensitivity of the device in measuring $\Delta R/R$ reached one part in a million, and the accuracy of measuring the signal strength was 2%. The electroreflection signal was recorded at 0.2–0.7 eVmodulating voltages to ensure the weak-field measuring regime.

An analysis of the electroreflection spectra in a specific region in the k-space of the Brillouin zone for a specific energy transition made it possible to determine the transition energy E_g, the phenomenological Lorentz broadening parameter Γ, which allows for dissipation processes in the electron transition, and the energy relaxation time τ of the light-excited charge carriers. The calculations were done by the three-point method[29] using the energy position of the dominating peaks in the electroreflection spectrum with allowance for the peak intensities.

The morphology of the surface of the silicon plates and its modification as a result of anisotropic etching were studied with the JEM-100 CX scanning electron microscope. The atomic composition of the surface was determined from the Auger spectra for low-energy (LVV) and high-energy (KLL) transitions obtained in the JEOL JAMP-10S microanalyzer with 10-keV electrons and a 3-Vmodulating voltage. To analyze the chemical composition of light-emitting silicon by depth we use *in situ* layer-by-layer etching of the silicon surface by 3-keV Ar^+ ions with a current density of 1–2 μA mm^{-2} and a spot diameter of 100 μm. The etching rate was 3 nm per minute. The perfection of the crystal structure at the plate surface was estimated from the electron-diffraction patterns produced by the ÉMR-100 electron-diffraction camera with an accelerating voltage of 75 keV. If necessary, the surface layer of silicon was removed in an aqueous solution of hydrofluoric acid, $\mathrm{HF:H_2O=1:1}$.

3. RESULTS OF THE EXPERIMENT

In the process of anisotropic etching, the surface of the silicon plate dimmed, acquired a microrelief, and became opalescent. Under laser excitation, such silicon samples were found to emit light at room temperature in the 600–800 nm spectral range with a maximum at 720 nm, and the halfwidth of the photoluminescence band amounted to 0.25 eV. The shape of the spectrum was found to be the same as that obtained in other studies of porous silicon. A specific feature of the luminescence is the small halfwidth of the emission band in comparison to the values for ordinary silicon (0.32–0.58 eV)[2–4,18–21] and the fact that luminescence did no degrade after long storage periods (more than one year). A detailed report on the results of studies of photoluminescence, Raman scattering of light, and the surface morphology of anisotropically etched silicon can be found in Ref. 10. After light-emitting silicon was treated by an aqueous solution of hydrofluoric acid, the opalescence disappeared and the surface acquired a mat finish, but the microrelief remained. We were unable to record any photoluminescence of such a sample (just as with the initial bulk silicon sample).

An analysis of the electron-diffraction patterns produced by the electron-diffraction camera suggested that the initial and anisotropically etched surfaces differ in crystal structure. The electron-diffraction pattern of the initial surface exhibited distinct point reflections. The electron-diffraction pattern of the anisotropically etched surface contained rings, which is an indication that a polycrystalline structure is present. After the light-emitting layer was removed, the electron-diffraction camera produced patterns with Kikuchi lines, which indicates that the crystal structure of the microrelief surface of the silicon is nearly perfect. According to Refs. 10 and 30, as a result of anisotropic etching of Si(100) plates, a layer of the crystal modification of the β-cristobalite of $\mathrm{SiO_2}$ forms at the surface. In our research we used the data on electron-diffraction patterns to show that this layer has a polycrystalline structure with a tetragonal lattice.

A detailed analysis of Auger spectra showed that there was no oxygen in the initial plate and confirmed that an oxide phase is formed at the silicon surface as a result of anisotropic etching. The intensities of the high-energy KLL-peaks in the Auger spectra of oxygen (500 eV) and

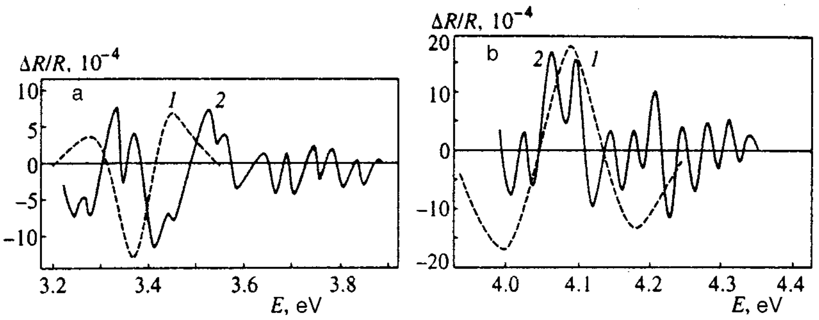

FIG. 1. Electroreflection spectra of the initial (curve *1*) and photoluminescent (curve *2*) silicon surfaces in the regions of the E_0'- and E_0-transitions (point Γ, Fig. a) and of E_2-transitions (point *X*, Fig. b).

silicon (1590 eV) were used to determine the O-to-Si ratio and its changes as we approach the SiO_2–Si boundary. First this ratio did not change, but then, at a distance of 100 nm from the silicon surface, a composition gradient was observed with a smooth transformation from SiO_2 to Si through the SiO_x phases (the incompletely oxidized states Si_2O_3, SiO, and Si_2O). For the low-energy silicon *LVV*-transition, the Auger spectra in the transition layer revealed large shifts of Auger electrons (-9 eV), which is characteristic of the Si bond in SiO_2.

The results of measurements of the electroreflection of the initial and light-emitting silicon surfaces are depicted in Fig. 1. Electroreflection signals from the initial surface were recorded in the 3.2–3.55 eV(curve *1* in Fig. 1a) and 3.9–4.25 eV(curve *1* in Fig. 1b) spectral ranges. According to electron band structure of silicon,[31] the electroreflection spectrum in Fig. 1a corresponds to the direct transitions Γ_{25}'–Γ_{15} and Γ_{25}'–Γ_2, which occur between the valence band and the conduction band at the center of the Brillouin zone. In silicon at $k=0$, the orbitally degenerate band Γ_{15} lies below the nondegenerate conduction band Γ_2', since the order in which the structures E_0 and E_0' appear in the spectrum is opposite the sequence observed in other semiconductors (Fig. 2). The low-energy structure E_0' corresponds to Γ_{25}'–Γ_{15} transitions. In view of broadening, the signals from the E_0'- and E_0-transitions overlap and usually are not resolved at room temperature. In the 3.9–4.25 eV spectral range (curve *1* in Fig. 1b), the electroreflection signal is due to the X_4–X_1 transitions at the edge of the zone (the E_2-transition). In agreement with selection rules, the indirect transition Γ_{25}'–$X_1$with an energy of 1.12 eV did not manifest itself in the electroreflection spectrum. The electroreflection spectra of the photoluminescent silicon surface are depicted in Figs. 1 (curve *2*) and 3. In comparison to the spectra of the initial surface, these spectra showed that

1. There is inversion of the polarity of electroreflection signal in the vicinity of the E_0' and E_0-transitions, but there is no inversion of the polarity for the E_2-transition;

2. The value of the phenomenological broadening parameter decreases, which made it possible to allow for E_0'- and E_0-transitions;

3. There is a signal in the spectral range where photoluminescence is observed (Fig. 3);

4. Periodic alternating-sign oscillations appear in the short-wavelength region of each transition;

5. The dominating peaks separated by 40 meV are doublets; and

6. The removal of the oxide layer causes the disappearance of the electroreflection effect in the photoluminescence region, the disappearance of short-wavelength oscillations and peak splitting, and the restoration of the polarity of the initial signal.

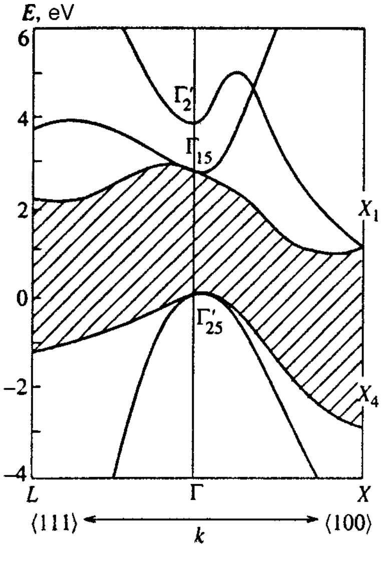

FIG. 2. Electron band structure of silicon.

4. DISCUSSION

In interpreting the experimental data we assumed that measurements of the reflection of photoluminescent silicon samples without a modulating field do not reveal the pres-

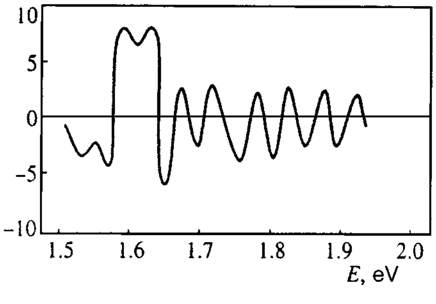

FIG. 3. The electroreflection spectrum of anisotropically etched silicon in the photoluminescence region.

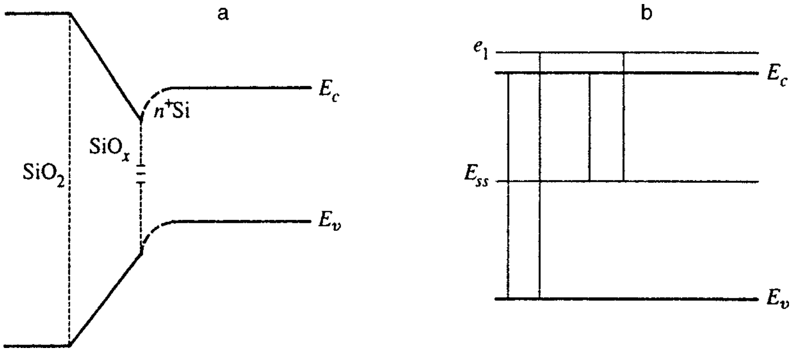

FIG. 4. Schematic of the band diagram of the photoluminescent Si–SiO$_2$ structure (a), and the diagram of energy transitions that form the electroreflection and photoluminescence spectra (b).

ence of interference effects, and electroreflection spectroscopy provides information about the presence of direct transitions in the band structure of the semiconductor and establishes the energy distribution of the electron in the structure. Hence the differences in the behavior of the electroreflection spectra that manifest themselves when we go from bulk silicon with a surface prepared in the standard way to light-emitting silicon are due to the change in the electron properties and the structure of the energy bands of the silicon surface brought on by anisotropic etching. The shape of the electroreflection spectra and the occurrence of photoluminescence in anisotropically etched silicon can be explained by the four-layer model of the band diagram, which does not contradict the experimental data on electron Auger spectroscopy either. The model presupposes the existence in anisotropically etched silicon of interphase surface electron states situated deep inside the forbidden band of silicon, a transition layer of SiO$_x$ at the Si–SiO$_2$ boundary, and surface size quantization of the energy of the electrons belonging to the conduction band. A schematic of the proposed band diagram is depicted in Fig. 4a.

The results of electroreflection studies have shown that at the initial surface of the silicon substrate, and also after the oxide layer is removed from the microrelief surface, the bands undergo depletion bending, i.e., the surface is depleted of electrons and the bands are bent upward. In the process of anisotropic etching, silicon atoms are forced to leave the substrate surface and are replaced by oxygen atoms, as shown by Auger spectroscopy. This process is accompanied by ordered growth of crystalline SiO$_2$, which changes the concentration and spectrum of the surface electron states in the surface layer. On the surface silicon under a layer of SiO$_2$ (Fig. 4a) the potential is of enrichment type and the bands are bent downward. The surface is enriched with electrons due to the appearance of interphase surface electron states at the boundary and a built-in positive change. This causes a change in the phase of the electroreflection signal. Since a change from a depletion potential to an enrichment potential was observed only for Γ-transitions and was absent for X-transitions (the phase of the electroreflection signal was retain), we can state that the interphase surface electron states lie in k-spacewhich are not near the minimum of the silicon conduction band, but at the center of the Brillouin zone.

According to the selection rules, the electroreflection effect is observed if[27]

$$\nabla k[E_c(k)-E_v(k)]=0, \tag{1}$$

where E_c and E_v are electron energies in the conduction and valence bands, respectively. It is likely that the condition (1) is met also for the electron transition between the main band and the band of surface electron states lying at a depth E_{ss}:

$$\nabla k[E_c(k)-E_{ss}(k)]=0. \tag{2}$$

If we allow for (2), the occurrence of an electroreflection signal from anisotropically etched silicon in the spectral range depicted in Fig. 3 can be explained by the presence of the direct transitions E_{ss}–Γ_{15} at the center of the Brillouin zone, in the same way as the presence of the Γ'_{25}–Γ_{15} transitions explains the origin of an electroreflection signal (for the latter transitions the electroreflection spectrum is given by curve *2* in Fig. 1a).

The doublet nature of the peaks of electroreflection in light-emitting silicon can be explained by the effect of surface size quantization of the electron energy in the enriched surface layer of silicon. The band diagram in Fig. 4a suggests that between the SiO$_x$ layer and silicon there is a quantum well, i.e., the lateral motion of the electrons is free, while in the direction perpendicular to the boundary the system is quantized. The presence of 2D electrons in the conduction band leads to a situation in which in addition to electron transitions between the main bands there are transitions between the valence band or the band of surface electron states and the first quantized level e_1 in the 2D quantum well (Fig. 4b). The presence of two transitions with energies E_g and E_g+e_1 splits the peaks in the electroreflection spectrum.

Gushchina *et al.*[32] were the first to detect a discrete energy spectrum related to size quantization in photoluminescent silicon. Their experiment revealed that such a spectrum appears in differential transmission spectra and was attributed to the presence of quantized energy levels. The energy of optical transitions in the quantum-size system for the first quantization level is given by the expression

$$E=E_g+\frac{\hbar^2\pi^2}{2\mu L^2}, \tag{3}$$

where E_g is the band-to-band transition energy, μ is the band-to-band effective mass, and L is the width of the quantum well. The value $e_1=0.04$ eV found from the experimen-

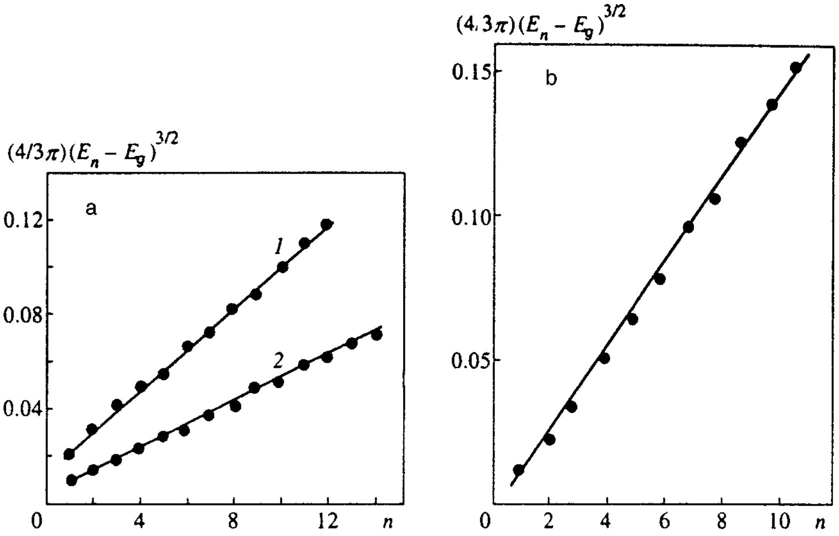

FIG. 5. Experimental curves (4) for determining the electro-optical energy: (a) in the region of the E_0- (straight line *1*) and E_2-transitions (straight line *2*), and (b) in the photoluminescence region.

tal electroreflection spectra under the assumption that the effective hole mass is $m_p = 0.49m_0$ at the point Γ'_{25} (here m_0 is the free electron mass) and that the effective electron mass is $m_e = 0.156m_0$ at the point Γ'_2 (Ref. 31), and with allowance for the expression (3) the value of the quantum-well width L is 4 nm.

In the presence of a quantum well and a SiO_x layer, the presence of alternating-sign oscillations observed in the electroreflection spectra above the optical absorption edge can be explained by the Franz–Keldysh effect involving direct transitions in a semiconducting structure with a built-in electric field. Franz–Keldysh oscillations in a uniform electric field are usually observed if[33]

$$n\pi = \Phi + \frac{4}{3}\left(\frac{E_n - E_g}{\hbar\theta}\right)^{3/2}, \tag{4}$$

where n and E_n are the number of an extremum in the oscillations and the position of this extremum on the energy scale, Φ is a phase factor, and $h\theta$ is the electro-optical energy. For the energy transitions E_0 ($E_g = 3.42$ eV) and E_2 ($E_g = 4.2$ eV), the experimental curves are depicted in Fig. 5a, and for the transition with the participation of surface electron states ($E_g = 1.6$ eV), in Fig. 5b. The values for the E_0-, E_2-, and E_{ss}-states, which we found from the slopes of the curves specified by (4) and depicted in Fig. 5, amounted to 41, 27, and 58 meV.

The electro-optical energy is determined by the field-induced broadening of the spectrum and depends in the following way on the band-to-band effective masses of the charge carriers and the electric field strength:[28]

$$(\hbar\theta)^3 = e^2\hbar^2F_s^2/\mu, \tag{5}$$

where e is the electron charge, and F_s is the surface electric field. The value $\hbar\theta = 41$ meV and the value of $\mu = 0.119m_0$, determined from the given values of m_e and m_p for the Γ'_{25}–Γ'_2 with allowance for (5), were employed to find the value of the built-in electric field, $F_s = 1.5\times10^5$ V cm^{-1}. Allowing for (5) and the value $\hbar\theta = 27$ meV for the E_2-transition, we calculated the value of the band-to-band effective mass $\mu = 0.432m_0$. With the effective electron mass $m_e = 0.98m_0$ at point X_1 (Ref. 31), the value of the effective hole mass at point X_4 proved to be $m_p = 0.824m_0$.

The Franz–Keldysh effect with the participation of localized states was studied theoretically by Vinogradov,[34] while we were the first to observe it in experiments. For an energy transition with the participation of a miniband, the value $\hbar\theta = 58$ meV at $F_s = 1.5\times10^5$ V cm^{-1} was used together with (5) to determine the value of the band-to-band effective mass for the given transition, $\mu = 0.416m_0$. The large number of oscillations in the electroreflection spectra and the fact that the condition (4) holds point to the uniformity of the built-in electric field, which is due to the variable energy gap of the SiO_x layer. For a layer of thickness $d = 100$ nm, the pinning of the Fermi level at the silicon surface at the boundary, $E_{\text{pin}} = eF_sd = 1.5$ eV, is close to the energy ''depth'' of the miniband of the surface electron states (1.6 eV).

Note that the electro-optical Franz–Keldysh effect in a built-in electric field was also observed in photoreflection spectra.[35–37] In variable-gap GaAlAs–GaAs structures[35] and in insulator–n^+n–semiconductor structures,[36] oscillations was observed at $F_s = 3\times10^5$ V cm^{-1}. In CdTe/CdMnTe structures with single quantum wells the built-infield was found be as high as 7.5×10^4 V cm^{-1}, with the width of the quantum well ranging from 5 to 15 nm (see Ref. 37).

The width of the enrichment layer, $l = 2kT/eF_s$, at room temperature ($kT = 25$ meV) amounted to 3.3 nm in a field $F_s = 1.5\times10^5$ V cm^{-1}, which is close to the calculated value $L = 4$ nm of the width of the quantum well, obtained from the data on peak splitting in the electroreflection spectrum. The closeness of the values of l and L suggests that there is quantum well at the silicon surface, in accordance with the proposed model of the band diagram of the photoluminescent $Si–SiO_2$ structure.

The model made it possible to explain the effect of electroreflection in the photoluminescence region, signal inversion, the presence of periodic oscillations, and peak splitting in the electroreflection spectra. The diagram of the energy transitions at the center of the Brillouin band that are responsible for the spectra of electroreflection and photoluminescence in anisotropically etched silicon is depicted in Fig. 4b. All processes included in the diagram participate in the electroreflection, while the photoluminescence is due solely to the transitions in which the levels E_{ss} take part. Here the photoluminescent silicon remains an indirect-gap semiconductor, since, according to electroreflection data, the energies of the direct transitions at the singular points Γ and X remained almost unchanged when the initial silicon was replaced by anisotropically etched silicon. Only a slight variation in the energy was observed, which was due to surface quantization (40 meV) and the appearance of internal mechanical stresses associated with compression. Estimates of the values of these stresses made on the basis of the data on Raman scattering of light taken from Ref. 10 yielded 5×10^8 N/m^2, which changes E_g by 50 meV.

Thus, in anisotropically etched silicon, no displacement of the absolute minimum of the conduction band from point X_1 to point Γ and no strong increase in the indirect gap due to quantum-size confinement are observed. Hence the onset of photoluminescence in such silicon is caused neither by a quantum-size increase in the band gap nor by transitions between the tails of the density states at the edges of the mobility edges nor by excitonic annihilation at the surface states of quantum size structures.

A comparison of the spectral photoluminescence and electroreflection curves suggests that the carriers are generated as a result of excitation of a transition in which the band of surface electron states at point $k=0$ participates. The energy of this transition exceeds the energy E_g of the indirect transition. Transitions with radiative recombination of electrons and holes take place at the same point of k-space of the Brillouin zone, which requires neither the violation of selection rules nor the participation of phonons, and this raises quantum efficiency substantially.

The surface radiative recombination of electrons and holes excited by laser light is confined to the quantum well, where the probability of finding a nonradiative center is low. Indeed, the Lorentz broadening parameter Γ decreases from 132 meV to 80 meV for the E_0-transition when the initial silicon is replaced by light-emitting silicon, while for the E_{ss}-transition responsible for photoluminescence we have $\Gamma=40$ meV.

As is known, Γ is the dissipative part of the self-energy of an electron excited by light to the conduction band and is determined by scattering of the electron energy in an optical transition. The presence of a quantum well at the photoluminescent surface of silicon makes Γ smaller and increases the energy relaxation time $\tau=\hbar/\Gamma$. The value of the relaxation time was found to increase from 5×10^{-15} s to 8×10^{-15} s, while for the E_{ss}-transition $\tau=1.6\times10^{-14}$ s. The increase in τ causes an increase in the mean free path of an electron–hole pair and to a corresponding increase in the mobility of such pairs. Note that an increase in the energy relaxation time for electrons was also observed by Gushchina *et al.*[32] in photoluminescent silicon with the quantum size effect (compared to the value in the initial sample).

The following factors also contribute to the increase in photoluminescence intensity: the passivation by oxygen of broken bonds at the light-emitting surface (the decrease in the number of nonradiative recombination centers), the decrease in the rate of nonradiative surface recombination in the built-in electric field, and the increase in the absorption of the exciting light by the well-developed microrelief surface of anisotropically etched silicon. The high stability of photoluminescence with the passage of time can be explained by the passivation of the photoluminescent surface by a thick oxide layer. Fauchet *et al.*[38] also reported an increase in intensity and stability of photoluminescence in oxidized porous silicon in comparison to the values in the initial sample.

5. CONCLUSIONS

On the basis of a comparison of the results of experimental investigations of the electroreflection of the initial surface and the anisotropically etched photoluminescent surface of silicon we have developed a four-layer model of the band structure of Si–SiO$_2$, determined the parameters of the structure, and examined a mechanism of visible photoluminescence that differs from standard mechanisms. Our findings yield the following conclusions.

1. Light-emitting, anisotropically etched silicon is an indirect-gap semiconductor. The surface photoluminescence that arises in this semiconductor at room temperature is due to the emergence of a miniband of deeply-lying levels of surface electron states in the forbidden band of anisotropically etched silicon, levels formed by adsorbed oxygen atoms. Direct radiative recombination of electrons and holes takes place in a quantum well near the Si–SiO$_2$ boundary during the energy transition at the center of the Brillouin zone between the conduction band and the miniband of surface electron states.

2. The increase of quantum efficiency is promoted by the increase in the absorption of the exciting radiation by the well-developed surface of the boundary and by the redistribution of the radiative and nonradiative recombination centers at the silicon surface after anisotropic etching is completed. The decrease in the number of nonradiative recombination centers in the region where electron–hole pairs are produced is due to the high perfection of the Si–SiO$_2$ boundary.

3. Passivation of the photoluminescent boundary by a thick oxide layer makes the structure of Si–SiO$_2$ prepared by anisotropic etching extremely stable and inert to the surroundings. The stability of photoluminescence in this structure is not inferior to that in porous silicon prepared by special methods by Fauchet *et al.*[38] The simple technology used in manufacturing the photoluminescent layer and the high efficiency and stability of photoluminescence require further investigation in order to obtain electroluminescence in anisotropically etched silicon.

*⁾E-mail: matveeva@class.semicond.kiev.ua

---

[1] L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).

[2] M. S. Bresler and I. N. Yasievich, Fiz. Tekh. Poluprovodn. **27**, 873 (1993) [Semiconductors **27**, 475 (1993)].

[3] P. M. Fauchet, L. Tsybeskov, G. Peng *et al.*, IEEE J. Sel. Top. Quantum Electron. **1**, 126 (1995).

[4] R. T. Collins, P. M. Fauchet, M. A. Tischler, Phys. Today **50**(1), 24 (1997).

[5] V. Ya. Degoda, V. M. Ogenko, G. V. Vesna, and S. N. Naumenko, Zh. Prikl. Spektrosk. **65**, 247 (1998).

[6] S. V. Belov, Pis'ma Zh. Tekh. Fiz. **18**(24), 16 (1992) [Sov. Tech. Phys. Lett. **18**, 800 (1992)].

[7] L. V. Belyakov, D. N. Goryachev, D. I. Kovalev, O. M. Sreseli, I. D. Yaroshetskiĭ, F. Koch, and V. Petrova-Koch, Fiz. Tekh. Poluprovodn. **29**, 1288 (1995) [Semiconductors **29**, 667 (1995)].

[8] R. W. Fathauer, T. George, A. Ksendzov, and R. P. Vasquez, Appl. Phys. Lett. **60**, 995 (1992).

[9] F. M. Qureshi, I. C. Barnard, and R. E. Palmer, in *Book of Abstracts EMRS-1998 Spring Meeting*, France, Strasbourg (1998), p. B-22.

[10] T. Ya. Gorbach, G. Yu. Rudko, P. S. Smertenko *et al.*, Appl. Phys. A: Solids Surf. **58**, 183 (1994).

[11] Y. Kanemitsu, T. Ogawa, K. Shiraishi, and K. Takeda, Phys. Rev. B **48**, 4883 (1993).

[12] D. J. Lockwood, Z. H. Lu, and J.-M. Baribean, Phys. Rev. Lett. **76**, 539 (1996).

[13] K. Behrenmeier, Fereydoon Nomavar, G. B. Amisola *et al.*, Appl. Phys. Lett. **62**, 2408 (1993).

[14] S. M. Prokes, J. Appl. Phys. **73**, 407 (1993); Appl. Phys. Lett. **62**, 3244 (1993).

[15] A. V. Petrov and A. G. Petrukhin, Fiz. Tekh. Poluprovodn. **28**, 82 (1994) [Semiconductors **28**, 49 (1994)].

[16] B. M. Kostishko, A. M. Orlov, and T. G. Emel'yanova, Pis'ma Zh. Tekh. Fiz. **21**(19), 32 (1995) [Tech. Phys. Lett. **21**, 782 (1995)].

[17] B. M. Kostishko, A. M. Orlov, and T. G. Emel'yanova, Izv. Ross. Acad. Nauk, Neorg. Mater. **32**, 1432 (1996).

[18] E. V. Astrova, V. V. Emtsev, A. A. Lebedev, D. I. Poloskin, A. D. Remenyuk, Yu. V. Rud', and V. E. Khartsiev, Fiz. Tekh. Poluprovodn. **29**, 1301 (1995) [Semiconductors **29**, 674 (1995)].

[19] T. Ya. Gorbach, G. Yu. Rudko, P. S. Smertenko *et al.*, Semicond. Sci. Technol. **11**, 601 (1996).

[20] B. M. Kostishko and L. I. Gonchar, JETP Lett. **66**, 382 (1997).

[21] T. V. Torchinskaya, N. E. Korsunskaya, L. Yu. Komenkova *et al.*, Semicond. Phys. Quantum Electron. Optoelectron. **1**, 61 (1998).

[22] P. D. J. Calcott, K. J. Nash, L. T. Canham, M. J. Kane, and D. Brumhead, J. Lumin. **57**, 257 (1993).

[23] G. D. Sanders and Yia-Chung Chang, Phys. Rev. B **45**, 9202 (1992); Appl. Phys. Lett. **60**, 2525 (1992).

[24] J. P. Proot, C. Delerue, and G. Allan, Appl. Phys. Lett. **61**, 1948 (1992).

[25] Yu. V. Kopaev, S. N. Molotkov, and S. S. Nazin, JETP Lett. **55**, 728 (1992).

[26] C. Delerue, G. Allan, and M. Lamroo, Phys. Rev. B **48**, 11 024 (1994).

[27] M. Cardona, *Modulation Spectroscopy*, Mir Publishers, Moscow (1973), p. 17 [English orig.: Academic Press, New York (1969)].

[28] V. A. Tyagaĭ and O. V. Snitko, *Electroreflection of Light in Semiconductors* [in Russian], Naukova dumka, Kiev (1980), p. 47.

[29] D. Aspnes, Surf. Sci. **37**, 418 (1973).

[30] E. F. Venger, T. Ya. Gorbach, L. A. Matveeva, and S. V. Svechnikov, Pis'ma Zh. Tekh. Fiz. **20**, 60 (1994) [Tech. Phys. Lett. **20**, 287 (1994)].

[31] I. M. Tsidil'kovskiĭ, *Band Structure of Semiconductors*, Nauka, Moscow (1978), p. 65. [English transl.: Pergamon Press, New York (1982)].

[32] N. V. Gushchina, V. S. Dneprovskiĭ, E. Yu. Dovydenko, V. A. Karavanskiĭ, and D. K. Okorokov, Zh. Éksp. Teor. Fiz. **106**, 1830 (1994) [JETP **79**, 994 (1994)].

[33] R. N. Bhattacharya, H. Shen, P. Parayanthal, F. H. Pollak, T. Coutts, and H. Aharoni, Phys. Rev. B **37**, 4044 (1988).

[34] V. S. Vinogradov, Fiz. Tverd. Tela (St. Petersburg) **15**, 395 (1973) [Sov. Phys. Solid State **15**, 285 (1973)].

[35] Xu Hondwei, Zhou Xiachuan, Xu Guichang *et al.*, Appl. Phys. Lett. **61**, 2193 (1992).

[36] H. Shen, M. Dutta, L. Fotiadis, P. G. Newman, R. P. Moerkirk, W. H. Chang, and R. N. Sacks, Appl. Phys. Lett. **57**, 2188 (1990).

[37] E. V. Goncharova, V. P. Kochereshko, and M. A. Yakobson, J. Cibert, and Le Si Dang, JETP Lett. **61**, 894 (1995).

[38] P. M. Fauchet, L. Tsybeskov, S. P. Duttagupta, and K. D. Hirschman, Thin Solid Films **297**, 254 (1997).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Optical and magnetooptical properties of granular alloys with giant magnetoresistance in the IR region of the spectrum

A. B. Granovskiĭ*) and M. V. Kuz'michev

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

J. P. Clerc

Laboratoire de l'Institut Universitaire des Systèmes Thermiques Industriels (IUSTI), Université de Provence, 13453 Marseille Cedex 13, France

(Submitted 26 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1762–1769 (November 1999)

The features of the optical and magnetooptical properties of granular alloys with giant magnetoresistance in the IR region are examined in reference to the magnetorefractive effect and the equatorial Kerr effect. Calculations are performed within the semiclassical approximation with consideration of spin-dependent scattering in the bulk of the granules and on their surfaces (interfaces). The expressions obtained for $\sigma_{xx}(\omega)$ and $\sigma_{xy}(\omega)$ are found to be sensitive to scattering on the surfaces and in the bulk of the granules, as well as to granule size, the type of impurities trapped on the interfaces, the frequency of the incident light, and the external magnetic field. For granular thin films exhibiting giant magnetoresistance, the theory predicts significant relative changes in the optical reflection and transmission coefficients when the sample is magnetized to saturation (0.02% and 20%, respectively, for giant magnetoresistance of the order of 20%), as well as Kerr and Faraday effects that are nonlinear with respect to magnetization. © *1999 American Institute of Physics.* [S1063-7761(99)01911-3]

1. INTRODUCTION

It has essentially been proven that the giant magnetoresistance in magnetic multilayers and granular alloys is associated with the spin-dependent scattering of conduction electrons.[1,2] The occurrence of spin-dependent scattering in such systems is also manifested in other transport phenomena, for example, in the thermal conductivity and thermopower,[3] the anomalous Hall effect,[4,5] the optical properties,[6–9] and the high-frequency impedance.[10]

The reflection (R), transmission (T), and absorption (k) coefficients of metals in the IR region of the spectrum are determined by the frequency dependence of the diagonal component of the intraband conductivity tensor $\sigma_{xx}(\omega)$. Since the static conductivity $\sigma_{xx}(\omega\rightarrow 0)$ of materials with giant magnetoresistance depends strongly on the magnetization of the sample,[1,2] it is clear that $\sigma_{xx}(\omega)$ and, consequently, all the optical properties should also depend on magnetization. This phenomenon, which is called the magnetorefractive effect, was first discovered experimentally and explained theoretically for multilayers by Jacquet and Valet.[6] Thus, the magnetorefractive effect consists of changes in the optical properties of systems with giant magnetoresistance when they are magnetized. This phenomenon can be observed only in the IR region of the spectrum, where the optical properties are determined by intraband scattering processes. The theory for this effect in multilayers has been thoroughly developed.[6–9] At the same time, this effect has a fairly general character and should be observed in all systems characterized by significant magnetoresistance, particularly in granular metal-metal alloys with giant magnetoresistance or granular metal-dielectric alloys with tunneling magnetoresistance. The magnetorefractive effect was recently discovered[11] in reflection from granular $Co_x(Al_2O_3)_{1-x}$ films, and its magnitude in the near-IR region of the spectrum was smaller than might have been expected from the analogy to multilayers. Therefore, a theory for the magnetorefractive effect in granular films is developed in the first part of this paper, and the high sensitivity of this effect to the microstructure of the alloy is demonstrated.

The magnetooptical properties of ferromagnets in the IR region of the spectrum are determined by the off-diagonal part of the conductivity tensor $\sigma_{xy}(\omega)$, which is the dynamic analog of the anomalous Hall conductivity $\sigma_{xy}(0)$. In homogeneous ferromagnets both $\sigma_{xy}(0)$ and $\sigma_{xy}(\omega)$ depend linearly on magnetization; therefore, the magnetooptical Kerr and Faraday effects are linear with respect to magnetization. However, in the case of materials with giant magnetoresistance, $\sigma_{xy}(0)$ exhibits a nonlinear dependence on magnetization, which is caused by spin-dependent scattering.[12] It is shown in Sec. 3 of this paper that $\sigma_{xy}(\omega)$ can display very complex behavior (nonlinear with respect to magnetization and nonmonotonic with respect to external magnetic field) in granular alloys and that it leads to magnetooptical effects that are nonlinear with respect to magnetization. As far as we know, such effects have not been discussed in the literature. To stress that such nonlinear behavior appears as a result of spin-dependent scattering and is not related to the nonlinear magnetooptical effects caused by high-intensity light beams, we call it the nonlinear field-dependent magnetooptical effect.

2. MAGNETOREFRACTIVE EFFECT

The optical and magnetooptical spectra of a granular film are calculated below in the Zhang–Levy model[13] using the self-averaging limit of the scattering probabilities[1,13] and by considering the intraband dynamics of the conduction electrons in the IR energy range of the incident light in the normal skin-effect regime.

The optical response of a metal is determined by the diagonal part of the dielectric tensor $\varepsilon_{xx}(\omega)$, which can be expressed in terms of the corresponding component of the conductivity tensor:

$$\varepsilon_{xx}(\omega)=\varepsilon'_{xx}-i\varepsilon''_{xx}=1-i\,\frac{4\pi\sigma_{xx}(\omega)}{\omega}. \tag{1}$$

For example, the reflection coefficient R of a metal with normally incident light is

$$R=\frac{k^2+(1-n)^2}{k^2+(1+n)^2}, \tag{2}$$

where n is the refractive index. In addition, n and k are related to the dielectric tensor: $\varepsilon'_{xx}=n^2-k^2$, $\varepsilon''_{xx}=2nk$.

The conductivity $\sigma_{xx}(\omega)$ of a granular alloy in formula (1) is determined by the contributions of electrons with spins parallel and antiparallel to the magnetization [spin up (↑) and spin down (↓)]. Then[14]

$$\sigma_{xx}(\omega)=\sigma^{\uparrow}_{xx}(\omega)+\sigma^{\downarrow}_{xx}(\omega)=\frac{\sigma^{\uparrow}_{xx}(\omega=0)}{1+i\omega\tau^{\uparrow}}+\frac{\sigma^{\downarrow}_{xx}(\omega=0)}{1+i\omega\tau^{\downarrow}}, \tag{3}$$

where $\tau^{\uparrow(\downarrow)}$ is the electron mean free time between collisions, and in the Zhang–Levy model[13]

$$\sigma^{\uparrow(\downarrow)}_{xx}(\omega=0)=\frac{ne^2\hbar}{2m}\,\frac{1}{\Delta^{\uparrow(\downarrow)}}, \tag{4}$$

$$\Delta^{\uparrow(\downarrow)}\equiv\frac{\hbar}{2\tau^{\uparrow(\downarrow)}},\qquad \Delta^{\uparrow(\downarrow)}=\frac{\varepsilon_F}{k_F}\left(\xi_0\pm\xi_1\frac{\langle M_z\rangle}{M_s}\right), \tag{5}$$

$$\xi_0=\frac{1-x}{l_{nm}}+\frac{x(1+p_b^2)}{l_m}+\frac{3x(1+p_s^2)}{r_0 l_s/a_0}, \tag{6}$$

$$\xi_1=\frac{2xp_b}{l_m}+\frac{6xp_s}{r_0 l_s/a_0}. \tag{7}$$

Here x is the concentration of ferromagnetic granules in the nonmagnetic matrix; the mean free paths l_{nm}, l_m, and l_s characterize the scattering on impurities in the nonmagnetic matrix, in the bulk of a granule, and on a granule surface, respectively; p_b and p_s are the ratios of the spin-dependent scattering potential to the spin-independent scattering potential in the bulk and on the surface of a granule; r_0 is the granule radius (it is assumed that all the granules are identical single-domain spheres); a_0 is the lattice parameter; $\langle M_z\rangle$ is the mean magnetic moment of the granules in the direction of the magnetic field H_z; and M_s is its value at saturation. Following the Zhang–Levy model, we assume that electron transfer is of the s type. Since $\langle M_z\rangle$ depends on magnetic field, $\sigma_{xx}(\omega=0)$ and $\sigma_{xx}(\omega)$ are functions of magnetic field. The giant magnetoresistance is defined as

$$\frac{\Delta\rho}{\rho}=\frac{\rho(H_c)-\rho(H)}{\rho(H_c)}=\frac{\xi_1^2}{\xi_0^2}\left(\frac{\langle M_z\rangle}{M_s}\right)^2, \tag{8}$$

where $\rho(H_c)=\rho_c$ is the resistivity of the alloy in the demagnetized state (in a field equal to the coercive force H_c when $\langle M_z\rangle=0$). When the field is equal to the saturation field $H_s(\rho(H_s)=\rho_s)$, the giant magnetoresistance takes its maximum value

$$\frac{\Delta\rho_s}{\rho}=\frac{\xi_1^2}{\xi_0^2}.$$

The parameters describing the magnetorefractive effect in optical reflection and absorption can be defined in the following manner:

$$\frac{\Delta R}{R}=\frac{R(H_c)-R(H_s)}{R(H_c)}=\frac{R(\omega,H_c)-R(\omega,H_s)}{R(\omega,H_c)}, \tag{9}$$

$$\frac{\Delta T}{T}=\frac{T(H_c)-T(H_s)}{T(H_c)}=\frac{T(\omega,H_c)-T(\omega,H_s)}{T(\omega,H_c)}. \tag{10}$$

In the general case they can be calculated using Fresnel coefficients on the basis of Eqs. (1) and (3)–(7). In the Hagen–Rubens limit $\omega\tau\ll 1$ we have

$$R=1-2\sqrt{\frac{\omega}{2\pi\sigma}}=1-2\sqrt{\frac{\omega\rho}{2\pi}}, \tag{11}$$

$$T=\left(\frac{\omega\rho}{2\pi}\right)e^{-\alpha t}, \tag{12}$$

where $\alpha=2k\omega/c$ and t is the film thickness. Then, for $\omega\tau\ll 1$ we can write

$$\begin{aligned}\frac{\Delta R}{R}&=\frac{R(H_c)-R(H_s)}{R(H_c)}=2\sqrt{\frac{\omega}{2\pi}}(\sqrt{\rho_s}-\sqrt{\rho_c})\\&=2\sqrt{\frac{\omega\rho_c}{2\pi}}\left(\sqrt{\frac{\rho_s}{\rho_c}}-1\right)=2\sqrt{\frac{\omega\rho_c}{2\pi}}\left(\sqrt{1-\frac{\Delta\rho_s}{\rho}}-1\right)\\&\approx-\sqrt{\frac{\omega\rho_c}{2\pi}}\frac{\Delta\rho_s}{\rho},\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\Delta T}{T}&=\frac{T(H_c)-T(H_s)}{T(H_c)}=1-\left(1-\frac{\Delta\rho_s}{\rho}\right)\\&\quad\times\exp\left[-t\,\frac{2\omega}{c}\sqrt{\frac{2\pi}{\omega}}\left(\sqrt{\frac{1}{\rho_s}}-\sqrt{\frac{1}{\rho_c}}\right)\right]\\&\approx 1-\left(1-\frac{\Delta\rho_s}{\rho}\right)\exp\left[-\frac{t}{c}\sqrt{\frac{2\pi\omega}{\rho_s}}\frac{\Delta\rho_s}{\rho}\right],\end{aligned} \tag{14}$$

where it is assumed in the final approximate expressions that $\Delta\rho_s/\rho\ll 1$. It follows from (13) and (14) that the magnetorefractive effect depends strongly on the magnitude of the giant magnetoresistance and that it is preferable to observe it in optical transmission.

Figure 1 presents the results of a calculation of the frequency dependence of the magnetorefractive effect in the reflection and transmission coefficients, which was performed using Fresnel coefficients. The parameters of the granular alloy were chosen to correspond to a typical alloy

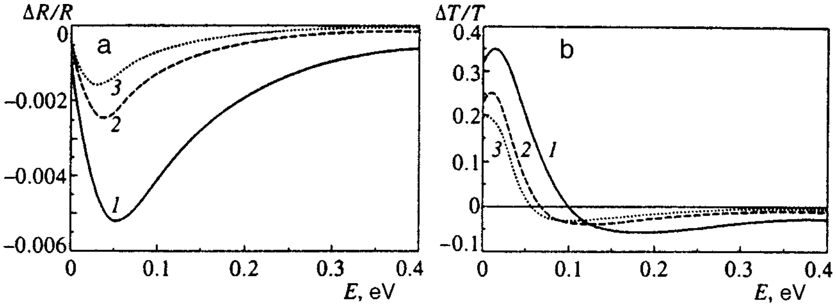

FIG. 1. Spectral dependences of the relative change in the reflection coefficient (a) and the transmission coefficient (b) when the sample is magnetized for granular alloys with giant magnetoresistance of various strength (normal light incidence): *1*—$\Delta\rho_s/\rho=30\%$, $r_0=20$ Å; *2*—$\Delta\rho_s/\rho=22\%$, $r_0=40$ Å; *3*—$\Delta\rho_s/\rho=17\%$, $r_0=60$ Å. The film thickness $t=200$ Å, $P_b=0.2$, $P_s=0.42$, $c=0.2$, $l_{nm}=200$ Å, $l_m=50$ Å, and $l_s/a_0=2$.

with giant magnetoresistance. The granule size, which is one of the parameters determining the magnitude of the magnetoresistance [see formulas (6)–(8)], was varied in the calculations. Similar spectral dependences were obtained by varying other microscopic parameters that influence the magnitude of the giant magnetoresistance [see (6)–(8)]. Several conclusions can be drawn from the data in Fig. 1. First, the amplitude of the reflective magnetorefractive effect in the near-IR region of the spectrum and on the edge of the visible range is comparatively small ($\Delta R/R<0$) and does not exceed 0.1%, in agreement with the experimental data in Ref. 11. However, even in that case the amplitude of the effect is two orders of magnitude greater than the amplitude of the magnetooptical Kerr effects. Second, $\Delta R/R$ increases in the mid-IR region of the spectrum, and, as follows from formulas (1)–(3) and (9), the maximum value of $\Delta R/R$ is achieved at the frequency $\omega=1/\tau$. Thus, experimental investigations of the magnetorefractive effect permit determination of the relaxation time of the electrons responsible for the optical properties. Third, as in the limiting case of $\omega\tau\ll 1$ considered above, it is seen that the transmission effect can reach values comparable to the giant magnetoresistance; however, the transmission effect has strong dispersion and changes sign, in agreement with the calculations for multilayers.[6] Fourth, the magnetorefractive effect is highly sensitive to the microstructure of the alloy, specifically to any variation of the parameters characterizing the spin-dependent scattering or the granule size. This finding also corresponds to the experimental data in Ref. 11.

It is noteworthy that, strictly speaking, the calculations performed are applicable only to granular metal–metal alloys; therefore, a comparison with the data in Ref. 11 has only a qualitative character.

We also stress that the magnetorefractive effect is isotropic, i.e., it does not depend on the orientation of the magnetization in the film plane; therefore, it can easily be distinguished from the classical magnetooptical effects.

3. NONLINEAR FIELD-DEPENDENT MAGNETOOPTICAL EFFECT

The magnetooptical response is determined by both the diagonal component $\varepsilon_{xx}(\omega)=\varepsilon'_{xx}-i\varepsilon''_{xx}$ and the off-diagonal component

$$\varepsilon_{xy}(\omega)=\varepsilon'_{xy}-i\varepsilon''_{xy}=\frac{4\pi\sigma_{xy}(\omega)}{\omega} \tag{15}$$

of the dielectric tensor. In particular, the parameter $\delta(\omega,H)$ of the equatorial Kerr effect, which characterizes the relative change in the intensity I of reflected light when the sample is magnetized, can be expressed in the case of p polarization in the form

$$\delta(\omega,H)=\frac{I(H)-I(0)}{I(0)}=a\varepsilon'_{xy}+b\varepsilon''_{xy}, \tag{16}$$

$$a=2\,\frac{A}{A^2+B^2}\sin 2\varphi,\quad b=2\,\frac{B}{A^2+B^2}\sin 2\varphi, \tag{17}$$

$$A=\varepsilon''_{xx}(2\varepsilon'_{xx}\cos^2\varphi-1),$$

$$B=(\varepsilon''^2_{xx}-\varepsilon'^2_{xx})\cos^2\varphi+\varepsilon'_{xx}-\sin^2\varphi,$$

where φ is the angle of light incidence.

In the case of ferromagnets $\varepsilon_{xy}\propto\sigma_{xy}\propto\langle M_z\rangle$, and $\varepsilon_{xx}\propto\sigma_{xx}$ and does not depend or depends weakly on $\langle M_z\rangle$ or magnetic field. Therefore, as a result, $\delta(\omega)\propto\langle M_z\rangle$. However, this is not true in the case of granular alloys with giant magnetoresistance: σ_{xx} depends on $\langle M_z\rangle$, and $\sigma_{xy}(\omega)$, as will be shown below, is nonlinear with respect to $\langle M_z\rangle$.

In the case of weak scattering of the conduction electrons, the off-diagonal component σ_{xy} of the conductivity tensor is described by the following expression, which resembles the Drude formula:[14]

$$\sigma_{xy}(\omega)=\frac{\sigma^{\uparrow}_{xy}(\omega=0)}{[1+i\omega\tau^{\uparrow}]^2}+\frac{\sigma^{\downarrow}_{xy}(\omega=0)}{[1+i\omega\tau^{\downarrow}]^2}. \tag{18}$$

The static off-diagonal conductivity $\sigma_{xy}(\omega=0)$ is linear with respect to the spin-orbit coupling and can be expressed in terms of the coefficient R_s of the anomalous Hall effect:[14]

$$R_s=R_s^{\uparrow}+R_s^{\downarrow},\quad R_s^{\uparrow(\downarrow)}=\frac{\sigma^{\uparrow(\downarrow)}_{xy}(\omega=0)}{4\pi\langle M_z\rangle}\rho^2. \tag{19}$$

Then

$$\frac{\sigma^{\uparrow(\downarrow)}_{xy}(\omega)}{\langle M_z\rangle}=\frac{4\pi R_s^{\uparrow}(H)}{\rho^2(H)[1+i\omega\tau_d^{\uparrow}(H)]^2}+\frac{4\pi R_s^{\downarrow}(H)}{\rho^2(H)[1+i\omega\tau_d^{\downarrow}(H)]^2}. \tag{20}$$

It is generally assumed that d electrons are responsible for determining $\sigma_{xy}(\omega=0)$, R_s, and $\sigma_{xy}(\omega)$ (Ref. 14). Thus, the relaxation time in formulas (18) and (20) can differ from the relaxation time in formula (3). Therefore, we introduced the corresponding notation $\tau_d^{\uparrow(\downarrow)}$ in (20). Because of the spin-

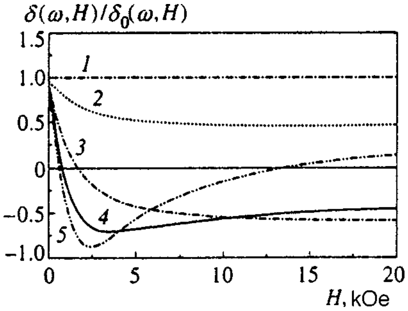

FIG. 2. Field-dependent equatorial Kerr effect in granular alloys for various levels of spin-dependent scattering: curve *1*—$p_b=0$, $p_s=0$, $\Delta\rho_s/\rho=0\%$; curve *2*—$p_b=0.2$, $p_s=0$, $\Delta\rho_s/\rho=0.5\%$; curve *3*—$p_b=p_s=0.2$, $\Delta\rho_s/\rho=10\%$; curve *4*—$p_b=0.2$, $p_s=0.42$, $\Delta\rho_s/\rho=30\%$; curve *5*—$p_b=0.2$, $p_s=0.52$, $\Delta\rho_s/\rho=40\%$. Parameters: $r_0=20$ Å, $c=0.2$, $l_{nm}=200$ Å, $l_m=150$ Å for s electrons and $l_m=50$ Å for d electrons, $l_s/a_0=2$, $R_s^s/R_s^b=1.5$, $\hbar\omega=0.01$ eV, and $\varphi=70°$.

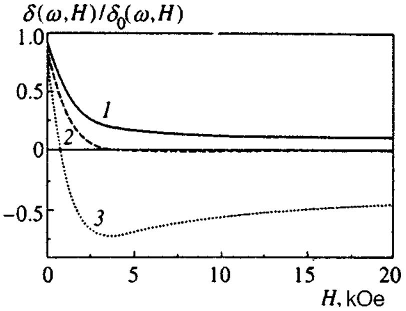

FIG. 3. Field-dependent equatorial Kerr effect in granular alloys with giant magnetoresistance for various granule sizes: curve *1*—$r_0=60$ Å, $\Delta\rho_s/\rho=30\%$; curve *2*—$r_0=40$ Å, $\Delta\rho_s/\rho=22\%$; curve *3*—$r_0=20$ Å, $\Delta\rho_s/\rho=17\%$. Parameters: $p_b=0.2$, $p_s=0.42$, $c=0.2$, $l_{nm}=200$ Å, $l_m=150$ Å for s electrons and $l_m=50$ Å for d electrons, $l_s/a_0=2$, $R_s^s/R_s^b=1.5$, $\hbar\omega=0.01$ eV, and $\varphi=70°$.

dependent scattering, the values of ρ and $\tau_d^{\uparrow(\downarrow)}$ for granular alloys depend on magnetic field. It was recently shown (see Ref. 12 and the references therein) that R_s in granular alloys with giant magnetoresistance depends strongly on magnetic field, more strongly than does the resistivity $\rho(H)$. If the main contribution to R_s is made by electrons with spin down and $\langle M_z\rangle/M_s$ is described by the Langevin function $L(H)$, the coefficient of the anomalous Hall effect of a granular alloy in the Zhang–Levy model can be written[12]

$$R_s^{\downarrow}(H)=\left\{R_s^{b\downarrow}\frac{(1+p_b^2)^2[1-2p_bL(H)+p_b^2]}{(1+p_b)^4(1-p_b)^2}\right.$$
$$\left.+R_s^{b\downarrow}\frac{(1+p_x^2)^2[1-2p_sL(H)+p_s^2]}{(1+p_s)^4(1-p_s)^2}\right\}$$
$$\times\frac{[\xi_0+\xi_1L(H)]^2}{\xi_0^2}, \qquad (21)$$

where R_s^b and R_s^s are the values of the coefficients of the anomalous Hall effect in the bulk and on the surface of a granule. Therefore, three factors can lead to dependences of the Kerr effect that are nonlinear with respect to magnetization: first, the dependence of the optical properties on magnetization, i.e., the parameters a and b in (16), which, according to (17), depend on $\sigma_{xx}(\omega)$; second, the dependence of $\sigma_{xy}(\omega)$ on $\rho(H)$ and $\tau_d^{\uparrow(\downarrow)}(H)$; and, finally, the strong field dependence of R_s, which largely determines the behavior of $\sigma_{xy}(\omega)$. All three factors are related to spin-dependent scattering.

The results of the calculation of the field dependences of the equatorial Kerr effect using (16)–(21) are presented in Figs. 2–4. All the figures present the normalized ratio $\delta(\omega,H)/\delta_0(\omega,H)$, where $\delta_0(\omega,H)$ is the parameter of the equatorial Kerr effect in the absence of spin-dependent scattering ($p_b=0$, $p_s=0$, and consequently $\Delta\rho_s/\rho=0$). Clearly, $\delta(\omega,H)/\delta_0(\omega)=1$ for a homogeneous ferromagnet, as is shown by curve *1* in Fig. 2. In all the other cases there is a strong field dependence of this ratio up to the saturation field. As follows from these figures, the field dependence of the equatorial Kerr effect does not correlate with the field dependence of the magnetization. The nonmonotonic course of the field dependences and possible sign reversal are also noteworthy. It follows from Fig. 2 that there is a nonlinear field dependence of the equatorial Kerr effect even in the presence of weak spin-dependent scattering, although, of course, it becomes strong when the giant magnetoresistance increases, the parameters of the spin-dependent scattering vary (Fig. 2), or the granule size varies (Fig. 3).

Figure 4 demonstrates the importance of scattering on the granule surfaces in shaping the field dependence of the equatorial Kerr effect. The type of impurities and their concentration on the granule surfaces alter the magnitude and sign of the corresponding Hall coefficient[4,5] and thus have a strong influence on the equatorial Kerr effect. Hence it also follows that the field dependence of $R_s(H)$ can be the main factor determining the nonlinearity of the equatorial Kerr effect.

Clearly, the other magnetooptical Kerr and Faraday effects will also have similar features, viz., a nonlinear dependence on magnetization, and this nonlinearity can exceed 100%.

Thus, it has been shown in this work that the spin-dependent scattering in granular alloys leads to two features in the IR region of the spectrum: the magnetorefractive effect, which consists of changes in the optical reflection, transmission, and absorption coefficients when the alloy is

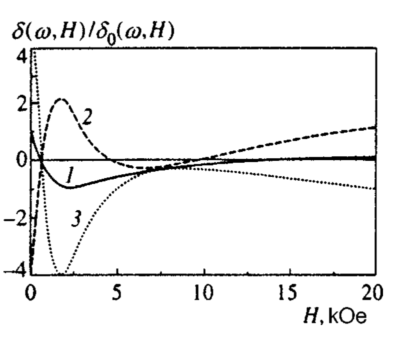

FIG. 4. Field-dependent equatorial Kerr effect in granular alloys with giant magnetoresistance for $R_s^s/R_s^b=10$ (curve *1*) -0.9 (*2*), and -1.1 (*3*). Parameters: $p_b=0.2$, $p_s=0.42$, $\Delta\rho_s/\rho=30\%$, $c=0.2$, $l_{nm}=200$ Å, $l_m=150$ Å for s electrons and $l_m=50$ Å for d electrons, $l_s/a_0=2$, $r_0=20$ Å, $\hbar\omega=0.01$ eV, $\varphi=70°$.

magnetized, and a strong nonlinear field dependence of the parameters of the magnetooptical effects, which is nonmonotonic in the general case.

This work was carried out with partial support from the Russian Fund for Fundamental Research. One of us (A.B.G.) thanks the Université de Provence (Marseille, France) for its hospitality.

[a] E-mail: granov@magn.phys.msu.su

---

[1] P. M. Levy, *Solid State Physics: Advances in Research and Applications*, Academic Press, New York–London (1994), Vol. 47, p. 367.
[2] M. A. Gijs and E. W. Bauer, Adv. Phys. **46**, 285 (1997).
[3] H. Sato, Y. Kobayashi, Y. Aoki *et al.*, Phys. Rev. B **52**, 9823 (1995).
[4] A. Granovsky, F. Brouers, A. Kalitsov, and M. Chiev, J. Magn. Magn. Mater. **166**, 193 (1997).
[5] A. B. Vedyaev, A. B. Granovskiĭ, A. V. Kalitsev, and F. Brauérs, Zh. Éksp. Teor. Fiz. **112**, 2198 (1997) [JETP **85**, 1204 (1997)].
[6] J. C. Jacquet and T. Valet, Mater. Res. Soc. Symp. Proc. **384**, 477 (1995).
[7] R. Atkinson, P. M. Dodd, N. F. Kubrakov *et al.*, J. Magn. Magn. Mater. **156**, 169 (1996).
[8] N. F. Kubrakov, A. K. Zvezdin, K. A. Zvezdin, and V. A. Kotov, Zh. Éksp. Teor. Fiz. **114**, 1101 (1998) [JETP **87**, 600 (1998)].
[9] J. M. Genkin, Phys. Lett. A **241**, 293 (1998).
[10] V. V. Ustinov, J. Magn. Magn. Mater. **165**, 125 (1997).
[11] I. V. Bykov, E. A. Gan'shina, A. B. Granovskiĭ *et al.*, Fiz. Tverd. Tela (St. Petersburg), **41** (1999) (in press) [Phys. Solid State **41**, (1999) (in press)].
[12] A. B. Granovskiĭ, A. V. Kalitsov, and F. Brauérs, JETP Lett. **65**, 509 (1997).
[13] S. Zhang and P. M. Levy, J. Appl. Phys. **73**, 5315 (1993).
[14] A. B. Vedyaev, A. B. Granovskiĭ, and O. A. Kotel'nikova, *Kinetic Phenomena in Disordered Ferromagnetic Alloys* [in Russian], Izd. MGU, Moscow (1992), p. 59.

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Semiconductor–metal transition in FeSi in an ultrahigh magnetic field

Yu. B. Kudasov,*) A. I. Bykov, M. I. Dolotenko, N. P. Kolokol'chikov, M. P. Monakhov,
I. M. Markevtsev, V. V. Platonov, V. D. Selemir, O. M. Tatsenko, A. V. Filippov

*Russian Federal Nuclear Center, National Research Institute of Experimental Physics, 607190 Sarov,
Nizhniĭ Novgorod Region, Russia*

A. G. Volkov, A. A. Povzner, P. V. Bayankin, V. G. Guk, V. V. Kryuk

Ural State Technical University, 620002 Ekaterinburg, Russia

(Submitted 21 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1770–1780 (November 1999)

We study the conductivity and magnetic susceptibility of single-crystal iron monosilicide in ultrahigh magnetic fields (up to 500 T) at low temperatures. The experimental methods used in measuring the conductivity and magnetic susceptibility are discussed. At 77 K we detect a gradual increase in the conductivity of iron monosilicide by more than a factor of 100 as the magnetic field gets stronger. At 4.2 K we detect a first-order phase transition in a field of 355 T accompanied by a sudden change in the value of the magnetic moment by $0.95\mu_B$ per iron atom and a transition to a phase with high conductivity. The results are discussed within the scope of the spin-fluctuation theory. © *1999 American Institute of Physics.*
[S1063-7761(99)02011-9]

1. INTRODUCTION

In the family of narrow-gap semiconductors, iron monosilicide is famous for its unusual properties. For instance, the violation of sum rules in FeSi has attracted wide attention in optical spectroscopy,[1] and metallization with increasing temperature occurs much faster than could be expected from band-structure calculations.[2] Near the Fermi level in the energy spectrum of FeSi there lies a narrow threefold degenerate band formed by the *d*-electrons of iron, with the result that various theoretical models[3–7] in which the strong electron correlations of the *d*-electrons are taken into account have been used to explain the observed anomalies. Although the effective electron masses obtained from band calculations[8] are large, they are still much smaller than the experimental values, which suggests strong mass renormalization due to electron–electron interactions.

Since the Zeeman splitting in magnetic fields on the order of several hundred teslas is comparable to the width δ of the gap in the spectrum of the *s*-, *p*-, and *d*-electronsof FeSi ($\delta = 0.11$ eV; see Ref. 9), it can give rise to a dramatic transformation of the electron spectrum. Hence ultrahigh magnetic fields constitute a powerful instrument for studying the electronic structure of narrow-gap semiconductors. Recently the behavior of FeSi in strong magnetic fields has been studied theoretically.[10,11] The study predicted the existence of a first-order phase transition into a metallic phase at absolute zero and a magnetic induction $B = 170$ T accompanied by a sudden jump in the magnetic order of order of two-tenths of a Bohr magneton,[11] or one Bohr magneton[10] per iron atom.

Some of the results obtained at $T = 77$ K can be found in the brief communication in Ref. 12.

2. EXPERIMENTAL TECHNIQUES AND PREPARATION OF SAMPLES

To generate ultrahigh magnetic fields with an induction of up to 450 T we used an MK-1 magnetocumulative generator.[13] At present this is the only device that allows reproducible uniform magnetic fields with B up to 1000 T to be created in large effective volumes.[13] In this device, the initial magnetic field ($B \approx 16$ T), generated in a thin-walled multilayer multifilar solenoid by a discharge of a powerful capacitor bank, is amplified through compression of a conducting cylinder by products of the blast of a cylindrical charge of an explosive. The time of compression of the initial magnetic flux is about 16 μs. The MK-1 generator was used in a single-cascade variant, i.e., without intermediate internal cascades,[13] which, on the one hand, somewhat reduces the peak value of the magnetic-field induction (450–550 T), but, on the other, makes it possible to obtain a smooth magnetic-field pulse (Fig. 1a). The effective volume when the magnetic field is at its maximum is a cylinder with the following nominal dimensions: 10 mm in diameter and 100-mm long. The large effective volume of the MK-1 generator allowed using several samples in a single experiment.

The samples and magnetic-field pickups were placed on a glass–textolite plate. In the experiments in which the initial temperature was $T = 77$ K the plate was submerged in a foam-plastic vessel filled with liquid nitrogen. When the initial temperature was $T = 4.2$ K, the plate was placed in a helium cryostat with a vacuumized double glass wall (Fig. 2). The cooling of the sample started 30 to 40 minutes before the beginning of the experiment. To this end the helium from Dewar vessel was transferred, via a heater, to the helium cryostat. The temperature in the cryostat was monitored by a carbon thermometer manufactured from the calibrated resis-

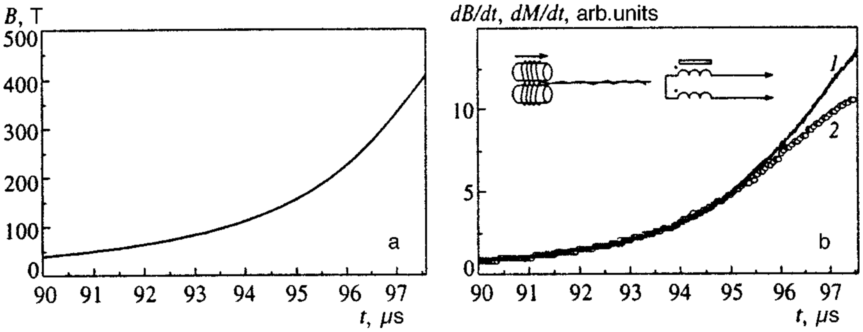

FIG. 1. Time dependence of the magnetic field at the final stage of the pulse (a), the time derivative of the magnetic field from the induction pickup (curve *1* in b)), and the signal from the compensation pickup (curve *2* in (b)) at $T=77$ K. The inset depicts the schematic of the compensation pickup (the arrow depicts the direction of the magnetic field) and the respective electric circuit.

tor TVO-0.125. The thermometer was in the upper part of the cryostat.

Since the rate at which the magnetic field increases changes substantially in the process of field generation (from approximately $0.5\,\mathrm{T}\,\mu\mathrm{s}^{-1}$ when the initial magnetic field is generated to approximately $10^2\,\mathrm{T}\,\mu\mathrm{s}^{-1}$ at the end of the pulse), it is difficult to measure the magnetic field accurately with a single induction pickup over the entire length of the pulse. Hence in our experiments we used a set of single-coil induction pickups with diameters ranging from 0.8 mm to 14.0 mm in which the PÉTV-2 brand of wire with diameters ranging from 71 μm to 0.25 mm was used. Earlier studies (see, e.g., Refs. 13 and 14) revealed that such an approach can guarantee a 5 to 10% accuracy of measurements in fields of up to 500 T. All signals were recorded using four-channel Tektronix-784 and Tektronix-744 oscilloscopes.

The single crystals of FeSi used in the experiments were grown by the Czochralski method in an atmosphere of spectrally pure helium under a pressure of 0.4 atm in the ⟨100⟩ direction. The rate of crystal growth amounted to 0.4 mm per minute when the crucible with the melt and the crystal were rotated in opposite direction at a rate of three rotations per minute. The initial components were 99.98%-pure carbonyl iron and band-purified semiconducting silicon. The preparation of single-crystal FeSi is described in detail in Ref. 15.

The induction measurements of the conductivity and magnetic susceptibility were done in the compensation pickup (see the inset in Fig. 1b).[16] The PÉTV-2 wire 71 mm in diameter was placed in the spiral grooves of two frames 2 mm in diameter fabricated of Caprolan, with $N=9$ turns on each frame. The degree of compensation of the coils was tested in a hf magnet. The total areas NS of the coils, where S is the area of one turn, differed by less than 2%. A hole was drilled (1.6 mm in diameter) in one of the frames to attach the sample under investigation.

The effective signal from an induction pickup can be related to the conductivity and magnetization of the sample. To separate these factors, we used two types of samples. A FeSi single crystal was crushed in a porcelain mortar to a powder with a granule size of about 100 μm. The powder was packed under pressure into the opening of the pickup. The density of the ''filling'' was monitored by weighing and on the average was $k=0.41$. The signal from this sample contained contributions from the conductivity and magnetization.

For the second type of sample we used a single-crystal powder of FeSi mixed with polymethyl methacrylate. The mixture was polymerized by dichloroethane. As a result we obtained single-crystal FeSi granules in an insulator matrix. This made it possible, in measuring magnetization, to avoid the appearance of an additional diamagnetic moment due to induction currents.

In the experiment at $T=77$ K with FeSi powder in an insulator matrix no peaks in the signal from the compensation pickup were observed and the shape of the signal coincided with that of the signal from the magnetic-field induction pickup. Figure 1b depicts the signals from the magnetic-field induction pickup and from the compensation pickup with the pure FeSi powder. Here additional induction emf is clearly visible against the background of pickup decompensation emf (the time derivative of the magnetic field). Thus, in a pulsed magnetic field, an additional diamagnetic moment arises in the powder, and it is related to the conductivity of the powder.

In our experiment the magnetic field was axial and uniform and λ was always much larger than R, where λ is the thickness of the skin layer, and R is the sample radius. Then the induction current density j in the sample is proportional to r (the radius), and the magnetic-field induction B is proportional to r^2. The diffusion equation for the magnetic field,

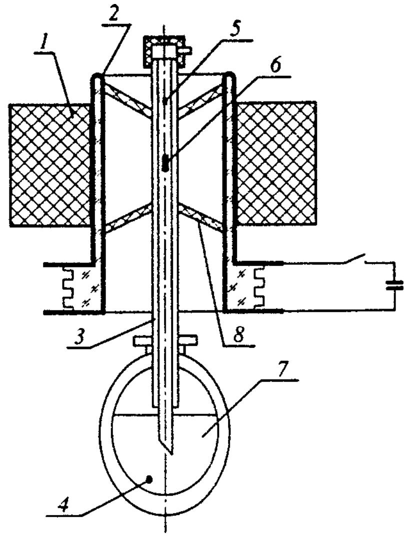

FIG. 2. The schematic of the device used in the experiment when the initial temperature was $T=4.2$ K: *1*, explosive charge; *2*, solenoid; *3*, glass vacuumized cryostat; *4*, heater for raising the helium; *5*, carbon thermostat; *6*, the measuring unit (samples and the magnetic-field pickups); *7*, helium cryostat; and *8*, foam-plastic mounts.

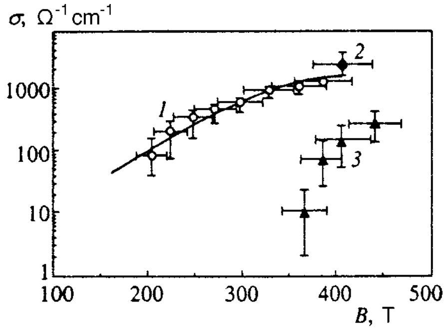

FIG. 3. Diagrams representing the dependence of the conductivity of FeSi powder (*1*) and a single-crystal plate of FeSi (*2*) in a magnetic field at 77 K (the solid curve corresponds to the dependence that follows from spin-fluctuation theory) and of the conductivity in the experiment with an initial temperature of 4.2 K (*3*).

$\Delta B = \mu_0 \sigma\, \partial B/\partial t$ (where μ_0 is the permeability of free space, or the magnetic field constant, and σ the conductivity of the powder) yields the following distribution of the magnetic field in the sample:

$$B(r) = B_0 + \frac{\dot{B}\mu_0\sigma}{4}(r^2 - R^2), \tag{1}$$

where B_0 is the magnetic-field induction outside the sample. This yields a formula for calculating the time dependence of the conductivity:

$$\sigma(t) = \frac{8}{\pi R^4 \mu_0 \dot{B} N} \int_0^t \Delta U(\tau)\, d\tau, \tag{2}$$

where ΔU is an additional emf related to the induction currents in the powder. The conductivity of the powder calculated by using the data in Fig. 1b and formula (2) is depicted in Fig. 3 (small open circles).

The experiments involving FeSi samples whose temperature was $T = 4.2$ K revealed a sharp peak in a field of 355±20 T. As is known, the appearance of sharp peaks in the signal is due to discontinuities in the magnetic moment when the samples undergo metamagnetic transitions.[16] Moreover, the readings from the pickup with pure FeSi powder demonstrated that after a peak a diamagnetic moment develops, which implies a transition to the conducting phase. Figure 4 depicts the signal from the pure FeSi powder, where both features are clearly exhibited.

Using the signal depicted in Fig. 4, we can calculate the conductivity using the method described above (see Fig. 3) and the discontinuity in the magnetic moment. We eliminate the background signal (see the inset in Fig. 4). The expression for the discontinuity in the specific magnetization has the form

$$\Delta I = \frac{1}{\pi R^2 \mu_0 N k} \int_{t_1}^{t_2} \Delta U(\tau)\, d\tau. \tag{3}$$

Integration is done between the points *1* and *2* (see the inset in Fig. 4). Using this expression, we can find the discontinuity in the magnetic moment in Bohr magnetons μ_B per iron atom:

$$\Delta P = \frac{M}{\pi R^2 \mu_0 \mu_B N \rho k N_A} \int_{t_1}^{t_2} \Delta U(\tau)\, d\tau, \tag{4}$$

where N_A is the Avogadro number, ρ is the density, and M is the molar mass. Using Eq. (4) and the data in Fig. 4, we find that for an initial temperature of 4.2 K the size of the discontinuity in the magnetic moment is $(0.95 \pm 0.2)\mu_B$/at. Fe.

We also measured the conductivity of a $4 \times 4 \times 0.3$ mm single-crystal plate of FeSi at $T = 77$ K. To do this, we employed a high-frequency method. The measuring circuit is shown in Fig. 5. The source *G* of high-frequency oscillations was a G4-154 generator. The oscillation frequency was about 49 MHz. The signal travelled through a valve *V*, a band-pass filter *F*, and an isolation transformer and was fed to a 30-m cable. The other end of the cable was attached to a flat coil 3 mm in diameter consisting of five turns (the wire used in the coil was of the PÉTV-2 brand with a diameter of 71 μm). The coil was placed flat on the FeSi plate in the cavity of an MK-1 generator, and the two ends of the coil were led out as a twisted pair. The axis of the coil was perpendicular to the direction of the external magnetic field, which reduced the emf induced by the pulses of this field. Here the plane of the plate was parallel to the external magnetic field, which prevented the sample from being heating

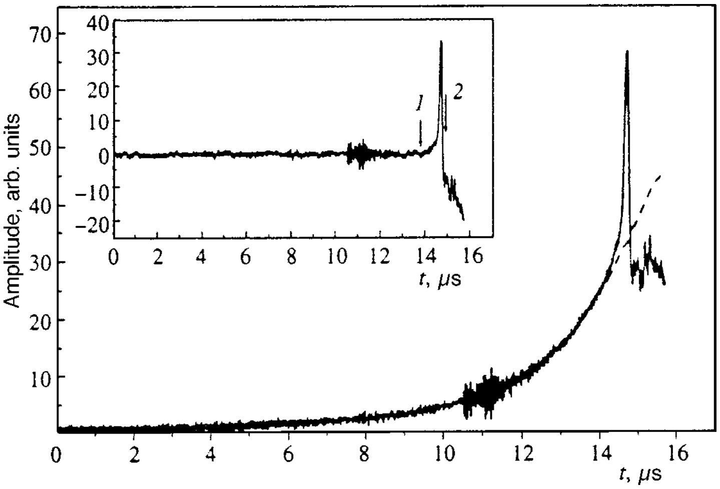

FIG. 4. Signal from the induction pickup with pure FeSi powder at an initial temperature of 4.2 K (the solid curve), and the scaled signal from the induction field pickup (the dashed curve). The inset shows the signal after the background proportional to $\partial B/\partial t$ is eliminated. The arrows indicate the limits of integration in Eq. (4).

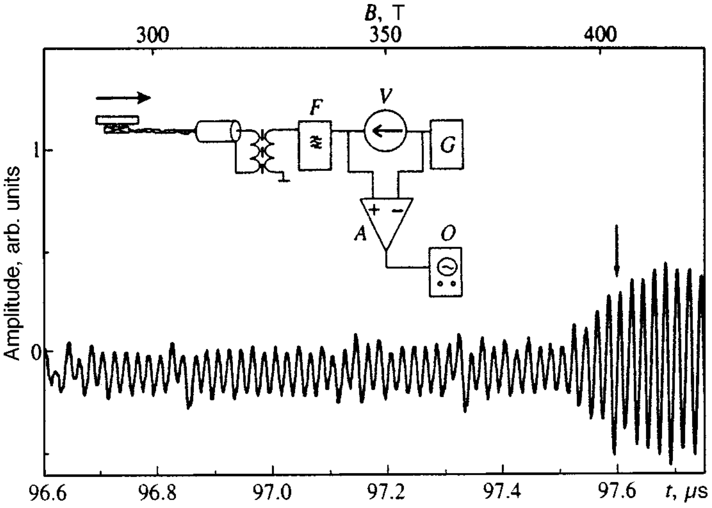

FIG. 5. High-frequency circuit for measuring conductivity and the hf signal obtained in the experiment at an initial temperature $T = 77\,\mathrm{K}$.

by induction currents (according to our estimates, the heating of the sample did not exceed 4 K). The incident hf wave was reflected at the end of the cable to which the coil was attached, with the reflectivity depending on the conductivity of the plate. The subtraction circuit A made it possible to discriminate the reflected signal, which was then recorded by an oscillograph O. The electrodynamics of the given circuit is similar to that used by Sakakibara *et al.*[17] but is more noiseproof.

The dynamic range of the hf method of measuring conductivity is limited from below. Hence in the experiment with the initial temperature $T = 77\,K$ this method makes it possibly only to establish the presence of conductivity of the FeSi plate at the level of $2\times 10^4\,\Omega^{-1}\mathrm{cm}^{-1}$ (the conductivity was calculated by a method similar to the one used by Sakakibara *et al.*[17]). Figure 5 depicts the hf signal, where the signal of the sample conductivity at approximately 400 T is clearly visible, which suggests that the conductivity is higher than $2\times 10^4\,\Omega^{-1}\,\mathrm{cm}^{-1}$.

3. STATE OF THE SAMPLES IN ULTRAHIGH MAGNETIC FIELDS

When the sample is placed in a pulsed magnetic field, a solenoidal electric field is generated in the sample, and in the case of a metal or semiconductor there are induction currents. Although the vortex field and the currents can be large (the maximum electric field strength on the surface of a cylindrical sample 1.6 mm in diameter is $400\,\mathrm{V\,cm^{-1}}$), still they are not large enough to dramatically alter the electronic structure of the substance. Hence secondary factors prove to be more important, i.e., factors that arise in a conducting medium due to the pulses of the magnetic field: magnetic pressure and Joule heating.

The magnetic pressure at the center of the sample, where it is at its maximum, can be calculated by the formula

$$P = \frac{(\delta B)^2}{2\mu_0}, \tag{5}$$

where δB is the difference between the magnetic-field inductions outside the sample and at the center of the sample, and μ_0 is the permeability of free space. Here it is assumed that the permeability of the medium is close to unity. Using (5) we find that

$$P = \frac{R^4 \dot{B}^2 \sigma^2}{32\mu_0}. \tag{6}$$

In all cases the magnetic pressure estimated by (6) was found to be lower than 50 bar. Such pressure cannot dramatically alter the electronic structure or, with allowance for the shortmagnetic-field pulse, deform the sample.

To estimate to what extent the sample is heated, we assume that at low temperatures the heat capacity of FeSi is a cubic function of temperature: $C = \beta T^3$, where $\beta = 0.6\,\mathrm{J\,m^{-3}\,K^{-1}}$ (see Ref. 18). Since we are interested in the upper bound on the heating of the sample, we ignore all types of heat transfer and calculate the temperature at the sample surface, where the release of heat is greatest. Then, using the expressions (1) and (2), we arrive at the following formula for the time dependence of the heat released by a cylindrical sample per unit time:

$$Q(\tau) = \frac{1}{4}\int_0^{\tau} \sigma(\lambda)\dot{B}^2 R^2\, d\lambda. \tag{7}$$

Hence the upper bound on the temperature as a function of time is

$$T(\tau) = \left[\frac{4}{k\beta}\,Q(\tau) + T_0\right]^{1/4}. \tag{8}$$

Figure 6 depicts the field dependence of the upper bound on the temperature of a powder sample in the experiment in which the initial temperature was 4.2 K. Clearly, after the sample passes to the conducting state, it heats up very strongly. In the experiment with an initial temperature of 77 K, the heating of the powder sample is not so intense: in a field of 400 T the upper bound of the sample temperature is 90 K. The dynamics of the temperature of the single-crystal plate used in the hf method can also be estimated by (7) and (8) if instead of R we take one-half of the plate thickness. At an initial temperature of $77\,K$, the upper bound on the plate temperature at 400 T was found to be 81 K.

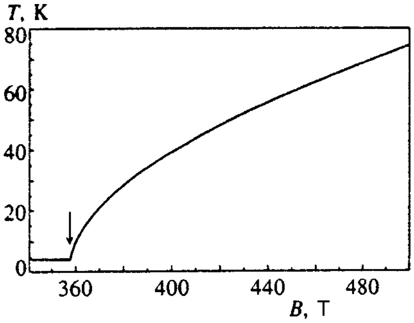

FIG. 6. Upper bound on the temperature of the FeSi powder in the experiment with an initial temperature of 4.2 K. The arrow indicates the phase transition at 355 T.

4. DISCUSSION AND CONCLUSIONS

The results of experiments conducted in magnetic fields with an induction of up to 450 T at $T=77$ K suggests that at this temperature there is neither magnetic ordering nor a discontinuous semiconductor–metal transition in FeSi. Instead of such a transition, we observed a gradual increase of the conductivity of FeSi with B by a factor larger than 100 (at $B=450$ T) in comparison to the conductivity in a zero field. Since the effective masses of the mobile charge carriers in FeSi are much larger than the mass of a free electron, the behavior of this semiconductor in a magnetic field differs dramatically from that of ordinary narrow-gap semiconductors and semimetals, e.g., InSb, PbTe, and Bi. The distance between the Landau levels in FeSi is very small, with the result that orbital quantization can be ignored. Then, with a band gap in the spectra of *s*-, *p*-, and *d*-electrons equal to 0.11–0.12 eV ($=2\delta$) (see Refs. 10, 11, and 19), the ''collapse'' of the energy gap due to the Zeeman splitting of the energies of (both *s*- and *p*-, and *d*-) electrons must occur at $B\approx 400$ T. Hence $\ln\sigma$ increases linearly with field induction for $B<400$ T. However, we found (see Fig. 3) that the nonlinear nature of the field dependence of the logarithm of the conductivity of FeSi manifests itself much earlier, starting at $B=250$ T. This is an indication that the value of the band gap found from band calculations in Refs. 10, 11, and 19 and from experiments on the temperature dependence of conductivity[20] is overestimated at the given temperature.

The reason for this discrepancy in nearly ferromagnetic semiconductors is the splitting of the energy spectrum due to spin-density fluctuations (SDF) of *d*-electrons,[21,22] which ensures the narrowing of the energy gap in their spectra with increasing temperature:

$$\delta(B,T)=\delta(0,0)-2U_m(B,T). \tag{9}$$

Here $2\delta(0,0)$ and $2\delta(B,T)$ is the gap without allowance for SDF and the gap renormalized by SDF as a function of magnetic field and temperature, $U_m(B,T)=[\xi_0^2+U^2\langle m^2\rangle]^{1/2}$ is the effective Zeeman splitting of the energies of *d*-electrons in a magnetic field in the presence of SDF, with $\xi_0=UM_0+2H$ (where H is the magnetic field strength (in units of $2\mu_B$), M_0 is the uniform magnetization of the d subsystem (in units of $2\mu_B$), and U is the parameter of the intratomic Coulomb interaction), and $\langle m^2\rangle^{1/2}$ the SDF amplitude determined by the fluctuation–dissipation theorem and incorporating contributions from thermal and zero-point spin-density fluctuations.[22] Note that the splitting caused by SDF does not give rise to magnetization. Using the approach developed in Refs. 12, 21, and 22, we see that magnetization develops because of a change in the capacitances of the subbands corresponding to opposite directions of spins. With allowance for charge-density fluctuations, the magnetization is described by the equation[12,22]

$$M_0=\frac{1}{4N_0}\sum_{\sigma=\pm 1}\sigma\int d\varepsilon\left\langle \sum_{\nu,\alpha} g^0(\varepsilon+\alpha\sqrt{\xi_{\boldsymbol{\nu}}^2(\tau)}\,)\right.$$
$$\left.\times\left(1+\frac{\alpha\sigma\xi_{\boldsymbol{\nu}}^z(\tau)}{\sqrt{\xi_{\boldsymbol{\nu}}^2(\tau)}}\right)f(\varepsilon-\mu-\eta_{\boldsymbol{\nu}}(\tau))\right\rangle. \tag{10}$$

Here $g^0(\varepsilon)$ is the density of states of noninteracting d-electrons, $\xi_{\boldsymbol{\nu}}(\tau)$ is the vector of the fluctuating exchange field, ν is the crystal lattice vector, $\nu=(\boldsymbol{\nu},\tau)$, τ is the imaginary Matsubara time,

$$\sum_{\nu}(\cdots)=T\sum_{\boldsymbol{\nu}}\int_0^{1/T}(\cdots)\,d\tau,$$

N_0 is number of lattice sites, the Fourier transform of the local exchange field $\xi_\nu(\tau)$ has the form $\xi_q^{(\gamma)}=r_{q,\gamma}\exp\{i\vartheta_{q,\gamma}\}$ (at $\mathbf{q}=0$ we have $\vartheta_{q,\gamma}=0$, and $r_{q,\gamma}$ is real),

$$\langle m^2\rangle=\frac{1}{U^2N_0}\sum_{\nu}\xi_{\boldsymbol{\nu}}^2(\tau),$$

$\langle m_\gamma^2\rangle$ is the mean-square amplitude of the γth component of the SDF, with $\gamma=x,y,z$, $f(\varepsilon-\mu)$ is the Fermi function, with μ the chemical potential,

$$\langle(\cdots)\rangle=\int_0^\infty\int_0^{2\pi}\prod_{\gamma,q(\neq 0)}dr_{q,\gamma}d\vartheta_{q,\gamma}(\cdots),$$

and $\eta_{\boldsymbol{\nu}}(\tau)$ is the random charge field that fluctuates in both space and time and is related to the longitudinal fluctuations of the local magnetic moment:

$$\eta_{\boldsymbol{\nu}}^2(\tau)=\frac{\sum_{\alpha=\pm 1}\alpha g^0(\mu+\alpha U_m)(\xi_{\boldsymbol{\nu}}(\tau)-U_m^2)}{\sum_{\alpha=\pm 1}g^0(\mu+\alpha U_m)}.$$

For the semiconductor state of the *d*-electrons of FeSi ($5<T<100$ K) (see Ref. 22), the SDF amplitude can be calculated from the formula $\langle m^2\rangle^{1/2}=bT$, where $b=7/U$, with $U=0.8$ eV. Numerical estimates of the value of the magnetic-field induction corresponding to the collapse of the gap in the spectrum of *d*-electrons at an initial temperature of 77 K yield $B_d=275$ T. Thus, the observed deviation of the experimental dependence $\ln\sigma(B)$ from the linear law above 250 T can be explained by the SDF narrowing of the gap in the *d*-spectrum. Here, in view of the extreme smallness of the SDF amplitude in the system of *s*- and *p*-electrons and the smallness of parameter of $(s,p-d)$-exchange interaction

($I=0.1U$; see Ref. 3), the energy gap in the spectrum of *s*- and *p*-electrons disappears at $B_s=389$ K. Hence the contribution of this group of carriers to the conductivity of FeSi is extremely small at low magnetic-field induction ($B<40$ T) and increases significantly only when $B>400$ T. The field dependence of the conductivity of FeSi calculated in spin-fluctuation theory with an initial temperature $T=77$ K is depicted in Fig. 2.

Concluding the analysis of this part of the experiment, we would like to mention the work of Anisimov *et al.*,[10] who showed that there can be a first-order phase transition between a singlet semiconductor and a ferromagnetic metal with a critical point $(B_c, T_c)=(170\,\mathrm{T}, 280\,\mathrm{K})$ at $U=3.4$ eV and $\delta=0.02$ eV. However, according to Anisimov *et al.*,[10], the emergence in FeSi of a metallic state is accompanied by ferromagnetic ordering, a phenomenon not observed by us at $T=77$ K (<280 K).

At the initial temperature 4.2 K of the experiment we clearly detected a discontinuity in the magnetic moment of FeSi amounting to $0.95\mu_B$/at. Fe at $B=355$ T. It must also be noted that no sample heating or magnetic pressure was observed (see Fig. 6).

Comparing our results with the theoretical estimates made in Ref. 19, we can conclude that the value of the field induction found in that paper is underestimated by a factor of approximately two (as in Ref. 10). More than that, the size of the magnetization discontinuity obtained in Refs. 11 and 19 smaller than the value observed in the experiment by a factor of approximately four. This shows that the density of states of FeSi differs from the value established in Refs. 11 and 19 (to make their estimates, Anisimov *et al.*[10] used the model of a rectangular band). Note that Anisimov *et al.*[10] mentioned the possibility of observing a discontinuity in the magnetic moment of order $1\mu_B$/at. Fe, but only in very high magnetic field, higher than 1000 T.

Our calculations of the field dependence of magnetization, based on spin-fluctuation theory and the data on the density of states of *d*-electrons taken from Refs. 11 and 19, done at the initial temperature $T=4.2$ K, revealed that at low temperatures the amplitude of thermal SDF is extremely small (in comparison to δ). Hence if we allow only for thermal SDF the value of the critical field is of order 1000 T. At the same time, according to Refs. 21 and 22, at low temperatures there are zero-point SDFs in iron monosilicide with an amplitude $\langle m^2\rangle_0^{1/2}=0.14\mu_B$/at. Fe. The resulting value of magnetic induction corresponding to the peak in the field dependence of the response function (see Fig. 3) proves to be equal to 320 T, which agrees with the experimental value. According to the magnetic state equation (10), the same magnetic field, $B_c=320$ T, corresponds to the magnetization discontinuity of $1.02\mu_B$/at. Fe.

Thus, the metamagnetic transition detected at $T=4.2$ K and the absence of such a transition at $T=77$ K in FeSi and the established value of the magnetic field strength at which the energy gap in the spectrum of *d*-electrons do not agree (when considered in the mean-field approximation) with the results of band calculations. One of the possible reasons for this discrepancy is the renormalization of the energy spectrum of *d*-electrons due to SDF (including zero-point SDF) accompanied by a redistribution of the number of states among the states with opposite directions of spins.

We believe that it is very important to study the temperature dependence of the first-order phase transition at low temperature and to determine the critical point from experiments.

Our research in ultrahigh magnetic fields was carried out in the ''Kapitza experiments'' program supported by the Russian Ministry of Atomic Energy and the Russian Ministry of Science and Technology.

*⁾E-mail: kudasov@ntc.vniief.ru

---

[1] D. van der Marel, A. Damascelli, and K. Schulte, E-prints archive cond-mat/9701005 (1997).

[2] V. Jaccarino, G. R. Wertheim, and J. H. Wernic, Phys. Rev. **160**, 476 (1967).

[3] P. V. Gel'd, A. G. Volkov, S. V. Kortov, A. A. Povzner, and V. Yu. Ivanov, Dokl. Akad. Nauk SSSR **320**, 1097 (1991) [Sov. Phys. Dokl. **36**, 698 (1991).

[4] M. B. Hunt, M. A. Chernikov, E. Felder, and H. R. Ott, Phys. Rev. B **50**, 14 933 (1994).

[5] C. M. Varma, Phys. Rev. B **50**, 9952 (1994).

[6] P. V. Gel'd, A. A. Povzner, and A. G. Volkov, Dokl. Akad. Nauk SSSR **283**, 358 (1985) [Sov. Phys. Dokl. **30**, 585 (1985).

[7] C. Fu, M. P. C. M. Krijn, and S. Doniach, Phys. Rev. B **49**, 2219 (1994).

[8] L. F. Mattheiss and D. R. Hamann, Phys. Rev. B **47**, 13 114 (1993).

[9] L. I. Vinokurova, A. V. Vlasov, and É. T. Kulatov, Proc. (Trudy) of the Institute of General Physics, USSR Academy of Sciences, **32**, 26 (1991).

[10] V. I. Anisimov, S. Yu. Ezhov, and I. S. Alfimov, Phys. Rev. Lett. **76**, 1735 (1996).

[11] E. Kulatov, H. Ohta, T. Arioka *et al.*, in *Proceedings of the 5th International Symposium on Research in High Magnetic Fields* (Sidney, Australia, 1977).

[12] Yu. B. Kudasov, A. I. Bykov, M. I. Dolotenko, N. P. Kolokol'chikov, M. P. Monakhov, I. M. Markevtsev, V. V. Platonov, V. D. Selemir, O. M. Tatsenko, A. V. Filippov, A. G. Volkov, A. A. Povzner, P. V. Bayankin, and V. G. Guk, JETP Lett. **68**, 350 (1998).

[13] A. I. Pavlovskiĭ and R. Z. Lyudaev, in *Current Aspects of Experimental Science and Technology* [in Russian], A. P. Aleksandrov (Ed.), Nauka, Leningrad (1984), p. 206.

[14] B. A. Boiko, G. V. Boriskov, A. I. Bykov *et al.*, in *Proceedings of the 11th IEEE International Pulsed Power Conference*, G. Gooperstein and I. Vitkovsky (Eds.), IEEE Inc. (1997), p. 1486.

[15] A. A. Frolov, Proc. (Trudy) of the Institute of General Physics, USSR Academy of Sciences, **32**, 3 (1991).

[16] I. S. Dubenko, A. K. Zvezdin, A. S. Lagutin, R. Z. Levitin, A. S. Markosyan, V. V. Platonov, and O. M. Tatsenko, JETP Lett. **64**, 202 (1996).

[17] T. Sakakibara, T. Goto, and N. Miura, Rev. Sci. Instrum. **60**, 444 (1989).

[18] T. Arioka, E. Kulatov, H. Ohta *et al.*, Physica B **246–247**, 541 (1998).

[19] P. V. Gel'd, A. A. Povzner, Sh. Sh. Abel'skiĭ, and L. F. Romasheva, Dokl. Akad. Nauk SSSR **313**, 1107 (1990) [Sov. Phys. Dokl. **35**, 758 (1990).

[20] A. G. Volkov, A. A. Povzner, and P. V. Gel'd, Fiz. Tverd. Tela (St. Petersburg) **26**, 1675 (1984) [Sov. Phys. Solid State **26**, 1015 (1984)].

[21] A. A. Povzner, A. G. Volkov, and P. V. Bayankin, Fiz. Tverd. Tela (St. Petersburg) **40**, 1437 (1998) [Phys. Solid State **40**, 1305 (1998)].

[22] A. A. Povzner, A. G. Volkov, and P. V. Bayankin, Fiz. Nizk. Temp. **23**, 1054 (1997) [Low Temp. Phys. **23**, 792 (1997)].

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Mechanism of Fermi acceleration in dispersing billiards with time-dependent boundaries

A. Yu. Loskutov,*⁾ A. B. Ryabov, and L. G. Akinshin

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 17 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1781–1797 (November 1999)

The paper is devoted to the problem of Fermi acceleration in Lorentz-type dispersing billiards whose boundaries depend on time in a certain way. Two cases of boundary oscillations are considered: the stochastic case, when a boundary changes following a random function, and a regular case with a boundary varied according to a harmonic law. Analytic calculations show that the Fermi acceleration takes place in such systems. The first and second moments of the velocity increment of a billiard particle, alongside the mean velocity in a particle ensemble as a function of time and number of collisions, have been investigated. Velocity distributions of particles have been obtained. Analytic and numerical calculations have been compared. © *1999 American Institute of Physics.* [S1063-7761(99)02111-3]

1. INTRODUCTION

The term *billiard* is applied to a dynamic system in which a point-like particle moves within a certain region Q with a piecewise smooth boundary ∂Q under the condition that the law of equality between the angles of incidence and reflection applies. Depending on the billiard boundary configuration, the motion of the particle (billiard ball) can be regular, ergodic, or mixing. The term *dispersing billiard*[1] applies to a system whose boundary ∂Q is convex inside the region Q. It is well known that such a billiard has a mixing property, and the billiard ball dynamics in this case is chaotic.

If the set ∂Q is constant with time, the system is called a billiard with a constant (fixed) boundary, but if $\partial Q = \partial Q(t)$, this is a billiard with a perturbed (moving) boundary. Billiards with fixed boundaries have been well studied (see Refs. 1–7 and references therein). At the same time, there have been very few publications devoted to billiards with perturbed boundaries,[7–11] although their studies are of great interest from the viewpoints of both solutions of some problems of statistical mechanics and the feasibility of an unbounded increase in a ball velocity, the latter problem originating from that of the so-called Fermi acceleration.[12,13]

Fermi acceleration is the phenomenon of infinite acceleration of particles of various nature owing to their scattering by moving massive scatterers. This mechanism of acceleration was first suggested by Fermi[12] to account for the origin of cosmic rays of very high energies. Later various models were suggested,[14–21] which described this phenomenon with a lesser or greater degree of success. For example, Ulam[14] demonstrated that, if a particle moves between an oscillating and a fixed wall, and the oscillation phase of the former at the moment of collision is a random value, the particle can acquire an infinitely high velocity. A more detailed investigation of Ulam's model was conducted by Lieberman and Lichtenberg,[16] who showed that, in the case of a smooth time dependence of the wall velocity, stochastic layers are separated by invariant curves. These curves set limits on the energy acquired by the particle. If this dependence is not sufficiently smooth, there are no invariant curves, and the particle velocity can increase without bound. Later investigations (see Refs. 16,19–21 and references therein) of various versions of Ulam's model revealed some relation between the law governing wall oscillations (i.e., the smoothness of the wall velocity as a function of time and the degree of its randomness) and the presence of the Fermi acceleration.

In chaotic billiards, even if the boundary velocity is a smooth function of time, the incidence angle of a particle can be treated as a random parameter. Consequently, the normal velocity component at the collision point (this is the component that changes as a result of scattering, whereas the tangential component is constant) is a stochastic value. Obviously, changes in the velocity are also random in this case.

The paper is devoted to the problem of Fermi acceleration treated on the example of a generalized billiard, namely, a Lorentz gas with an open horizon and a perturbed boundary. We focus attention on two different cases of stochastic and regular (harmonic) oscillations of the boundary. Note that in all publications on this topic of which we are aware, the problem of Fermi acceleration was investigated in integrable or almost integrable systems. In view of this, our paper presents the first investigation of chaotic billiards with perturbed boundaries.

The paper comprises three main sections. The first of them is devoted to the basic concepts and derivation of maps that describe the dynamics of a billiard. The second describes the analytic and numerical study of the feasibility of Fermi acceleration. The third presents numerical calculations of the particle velocity as a function of time and number of scattering events, and compares them to analytic results.

1063-7761/99/89(11)/9/$15.00

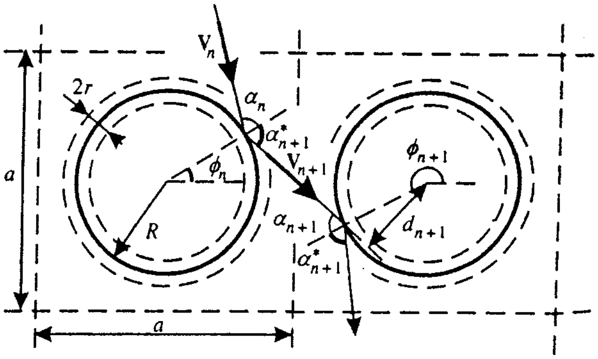

FIG. 1. Configuration of Lorentz gas model. The scatterers (circles of radius R) are located at sites of a periodic lattice with period a.

2. LORENTZ GAS

This section is devoted to the main concepts and derivation of mappings that determine the dynamics of a two-dimensional Lorentz gas.

Consider a plane area Q with a piecewise smooth boundary ∂Q. The dispersing billiard[1,4] is a system composed of neutral ∂Q_i^0 and dispersing ∂Q_i^+ (i.e., convex in the region Q) sections of the boundary ∂Q. One representative of such billiards is a system defined in an unbounded domain D and composed of a set of round infinitely heavy scatterers B_i with boundaries ∂Q_i and of radii R located at sites of an infinite periodic lattice with period a (Fig. 1). Given that B_i are fixed, the billiard in the region $Q = D \backslash \cup_{i=1}^{r} B_i$ is called a Lorentz gas. A particle moves among the scatterers and reflects from them in accordance with the mirror reflection law. Such a billiard has been studied in detail in the case of $\partial Q = \text{const}$ (see Refs. 1, 3, 4, 6 and references therein).

The ratio $(a/R)^2$ is one of the main parameters of the Lorentz gas. Depending on this parameter, we distinguish Lorentz gases with a bounded horizon $[(a/R)^2 < 4]$, with an open horizon $[4 < (a/R)^2 < 8]$, and with an infinite horizon $[(a/R)^2 > 8]$. In the first case, the particle motion is limited to one lattice cell, in the second and third cases it can travel throughout the entire space. In the case of an infinite horizon, statistical properties of a billiard change because of higher probabilities of long free paths,[3,4,22–24] whereas in Lorentz gases with bounded and open horizons correlations decay exponentially. The mean free path is defined as $l = \pi A/P$, where A is the area of a billiard where a particle can go and P is the scatterer perimeter. For a system with an open horizon $l = (a^2 - \pi R^2)/2R$, and for a billiard with an infinite horizon l has no upper bound.

Suppose that the radii of scatterers B_i in a Lorentz gas are perturbed in accordance with a certain law, i.e., all boundaries ∂Q_i perform small oscillations in the normal direction. In this paper we consider two different cases: periodic (and phase-synchronized) oscillations, and random changes in scatterer radii. The first case corresponds to the situation when all boundaries oscillate in phase following the same law. The second case describes oscillations of many scatterer boundaries with the initial phases distributed randomly.

2.1. Lorentz gas with a fixed boundary

It is known that one can select as canonical variables for billiards with unperturbed boundaries the azimuthal angle ϕ and incidence angle α between the interior normal to the surface and particle velocity before the collision. Let us introduce the reflection angle α^* between the exterior normal and velocity after the collision (Fig. 1). It is obvious that $\phi \in [0, 2\pi]$, and the angles α and α^* vary over the interval $[-\pi/2, \pi/2]$. In order to describe the dynamics of an unperturbed billiard, one has to calculate a mapping $(\alpha_n, \phi_n) \to (\alpha_{n+1}, \phi_{n+1})$ which transforms the variables (α, ϕ) at the moment before the nth collision with ∂Q to their values at the moment before the $(n+1)$th collision. It clearly follows from geometrical considerations (Fig. 1) that

$$\phi_n + \alpha_n^* + \pi = \phi_{n+1} + \alpha_{n+1}. \tag{1}$$

Moreover, $\alpha_n^* = -\alpha_n$ since these angles are measured in opposite directions.

Let us introduce a reference frame with its origin at the center of a circle on which the latest scattering event has taken place and determine the equation of the straight line along which the particle travels after the collision. Then one can easily calculate the distance at which the particle passes another center at a distance of p cells along the horizontal axis and q cell along the vertical axis:

$$d_{n+1} = a[p \sin(\phi_n + \alpha_n^*) - q \cos(\phi_n + \alpha_n^*)] - R \sin \alpha_n^*. \tag{2}$$

The parameter p is assumed to be positive if the particle moves on the right of the center and negative if the particle moves on the left. Accordingly, q is positive if the particle moves upwards and negative if it moves downwards. The values p and q are determined using the scattering condition, i.e., these are integers with the smallest absolute values at which the condition $|d_{n+1}| \leqslant R$ is satisfied. After calculating the impact parameter d_{n+1}, one can easily calculate the angle at which the collision with the next scatterer will take place:

$$\alpha_{n+1} = \sin^{-1} \frac{d_{n+1}}{R}. \tag{3}$$

The Jacobian of the resulting mapping defined by Eqs. (1)–(3) is

$$\frac{\partial(\phi_{n+1}, \alpha_{n+1})}{\partial(\phi_n, \alpha_n)} = \frac{\cos \alpha_n}{\cos \alpha_{n+1}}.$$

Thus, the mapping preserves the phase volume $\cos \alpha d\alpha d\phi$. Hence follows, in particular, that if the billiard is ergodic, the distribution with respect to α_n is described by the formula

$$\rho_\alpha(\alpha) = \frac{1}{2} \cos \alpha, \tag{4}$$

where 1/2 is the normalization factor.

2.2. Lorentz gas with oscillating scatterer boundaries

Now we can easily obtain a mapping that describes the dynamics of a billiard with a perturbed boundary under the assumption that the boundary oscillation amplitude is much smaller than its radius, i.e., we can neglect geometrical changes in its boundaries.

Suppose that the dispersing component ∂Q^+ of boundary ∂Q contracts and expands (Fig. 1), so that its radius varies following the law

$$R=R(t)=R+r(t), \quad \text{where} \quad \max|r(t)| \ll R.$$

Then the boundary velocity is a function of time, $u(t) = \dot{r}(t)$. Further, we assume for definiteness that $u(t) = u_0 \cos(\omega t)$, where $u_0=\omega r_0$. In this case, in addition to parameters α and ϕ, we have to introduce another two variables, namely, the particle velocity $\mathbf{v}$ and collision time t. Given that only the normal (radial) component of the velocity changes in the process of scattering, and the tangential component remains unchanged, we obtain a mapping for the absolute value of particle velocity after the collision:

$$v_{n+1}=\sqrt{v_n^2-4u_n v_n \cos\alpha_n+4u_n^2}. \tag{5}$$

Here $u_n \equiv u_0 \cos\omega t_n$ is the boundary velocity at the moment of the nth scattering event. The relation between the angles of incidence and reflection, in its turn, can be expressed as

$$\alpha_n^* = -\sin^{-1}\left(\frac{v_n}{v_{n+1}}\sin\alpha_n\right).$$

Now, by calculating the separation between sequential scattering events, one can easily obtain a mapping for the collision time t_n:

$$\begin{aligned} t_{n+1}&=t_n+\frac{l_{n+1}}{v_{n+1}}, \\ l_{n+1}&=\sqrt{[R(\cos\phi_{n+1}-\cos\phi_n)-pa]^2+[R(\sin\phi_{n+1}-\sin\phi_n)-qa]^2}. \end{aligned} \tag{6}$$

Here l_n is the free path. Under the assumption that $r \ll R$, the mappings for variable ϕ and impact parameter d are the same as for the unperturbed billiard [Eqs. (1) and (2)].

3. FERMI ACCELERATION

As a result of impacts with a perturbed boundary, the billiard ball velocity always changes. As earlier research has shown,[11] these changes in the velocity are random. Therefore, let us consider an ensemble of particles and calculate their velocity distribution and average velocity as a function of time t and number of collisions n (the number of collisions and time are not directly proportional because a faster particle undergoes more impacts during a time interval than a slower one). In this section, we will first consider the issue of the mean change in the velocity in billiards with arbitrary shapes and perturbed boundaries, then we will discuss the problem of Fermi acceleration in a Lorentz gas with randomly and regularly oscillating scatterers.

3.1. Average change in the velocity in the general case

Consider two sequential collisions of a ball hitting against a wall in a billiard of an arbitrary configuration (Fig. 2). Denote by α_0 the particle incidence angle in the first collision, and by α_1 this angle in the second collision (they are introduced as in Sec. 2.1). Further, denote by v_0 and v_1 the absolute values of the ball velocity before the first and second collisions, respectively. The velocity components are labeled by the following indices: the superscripts τ and n denote the tangential and normal velocity components, respectively, the first subscript is the velocity index, the second is set to unity if the velocity component is considered before the collision and to zero after the collision. Thus, v_{10}^τ denotes the tangential component of velocity v_1 at the point of the first collision, and v_{11}^τ is the tangential component at the point of the second collision. In the general case, they are not equal (Fig. 2). Let $u(t)$ be the boundary velocity. The following relation should, obviously, hold:

$$\langle u(t)\rangle_t=0, \tag{7}$$

which means that the boundary remains, on average, at its place.

Consider a single collision between a particle and a wall. The tangential velocity component in this case is, obviously, constant, whereas the change in the normal component can be easily calculated in the reference frame connected to the wall. Thus, we can write for the first collision

$$\begin{aligned} v_{10}^n&=-v_{01}^n+2u(t_n)=-v_0\cos\alpha_0+2u(t_n), \\ v_{10}^\tau&=v_{01}^\tau=v_0\sin\alpha_0, \\ v_1&=\sqrt{v_0^2-4v_{01}^n u(t_n)+4u^2(t_n)}. \end{aligned} \tag{8}$$

It is clear that, if only one collision is considered, $\langle \Delta v_{10}^\tau\rangle = 0$ and $\langle \Delta v_{10}^n\rangle=0$ for a billiard of an arbitrary configuration. Moreover, changes in the velocity are associated only

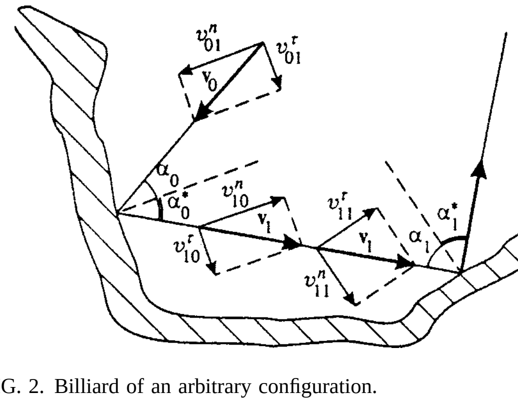

FIG. 2. Billiard of an arbitrary configuration.

with the normal component since the tangential component is unchanged after the reflection. Consequently, the average growth in the velocity depends on the normal velocity component in the next collision. In the general case, however, one can consider only the average velocity component, i.e.,

$$\langle v_{11}^n\rangle=\langle v_1\cos\alpha_1\rangle$$
$$=\langle\sqrt{v_0^2-4v_0u(t)\cos\alpha_0+4u^2(t_n)}\cos\alpha_1\rangle, \tag{9}$$

where averaging is performed over angles α_0 and α_1 and time t.

It seems appropriate to illustrate Eq. (9) on two examples.

Ulam's model.[14,16–21] Two parallel heavy walls are placed at distance L between them, and a ball moves between these walls. One wall oscillates periodically with amplitude δ such that $L\gg\delta$. The specific time dependence of these oscillations is unimportant for our analysis, the only important point is that the wall motion should satisfy condition (7). Since the tangential velocity component in this model is constant, $v^\tau=\mathrm{const}$, the velocity and incidence angle are related by the formula

$$v\sin\alpha=v^\tau=\mathrm{const}. \tag{10}$$

The normal velocity component, in its turn, has the same absolute value before the first collision, v_{10}^n, and before the second, v_{11}^n. Consequently,

$$\langle v_{11}^n\rangle=\langle v_{10}^n\rangle=\langle -v_{01}^n+2u(t_n)\rangle=v_{01}^n.$$

Thus, there is no particle acceleration on average in this model.

Lorentz gas. Owing to the strong mixing in this model, we can assume that angles α_0 and α_1 are mutually independent, hence

$$\langle v_{11}^n\rangle=\langle\cos\alpha_1 v_1\rangle_{\alpha_0,\alpha_1,t}=\langle\cos\alpha_1\rangle_{\alpha_1}\langle v_1\rangle_{\alpha_0,t}.$$

Therefore, fluctuations in the velocity (increase and decrease) due to collisions are associated with changes in its absolute value, but not in its normal component, as was in the previous case. As will be shown below, the value $\langle v_n\rangle$ increases with n, therefore Fermi acceleration is feasible in the Lorentz gas.

This conclusion can probably be extended to other systems in which the incidence angle and velocity are not directly related by Eq. (10), as in Ulam's model. An intermediate configuration between Ulam's model and a scattering billiard is the ''stadium-shaped'' billiard, in which the feasibility of Fermi acceleration was studied numerically.[11]

3.2. Stochastically perturbed scatterer boundary

Let the boundary velocity of a scatterer at which the nth collision takes place be

$$u_n(t)=u_0\cos\varphi_n, \tag{11}$$

where u_0 is the boundary velocity amplitude, and $\{\varphi_n\}$ is a set of uncorrelated random values uniformly distributed over the interval $[0,2\pi)$. Let us calculate the velocity distribution function and the average velocity in an ensemble of particles as a function of the number n of scattering events and time t. In the case of a low particle velocity, $v\ll u_0$, the major contribution to velocity given by Eq. (5) is due to the last term on the right-hand side, hence

$$v_{n+1}\approx 2|u(t_n)|.$$

If boundary oscillations are determined by Eq. (11),

$$\langle v_{n+1}\rangle\approx 2\langle|u(t_n)|\rangle_t=4\frac{u_0}{\pi}.$$

Thus, even after the first collision, the average velocity becomes larger than u_0.

Now let us calculate the change in the velocity at $v\gg u_0$. By expanding the right-hand side of Eq. (5) in powers of u/v, we obtain an expression for the velocity change:

$$\Delta v_n=v_{n+1}-v_n=-2u_n\cos\alpha_n+2\frac{u_n^2}{v_n}$$
$$\times\sin^2\alpha_n+v_nO\left(\left(\frac{u_n}{v_n}\right)^3\right), \tag{12}$$

where u_n is the scatterer boundary velocity during the nth collision.

Using Eq. (4) and the uniformity of the phase distribution at the moment of collision, we obtain $\langle\Delta v_n\rangle$ and $\langle(\Delta v_n)^2\rangle$:

$$\mu_s\equiv\langle\Delta v_n\rangle=\frac{M_s}{v},\quad \sigma_s^2\equiv\langle(\Delta v_n)^2\rangle=\frac{4}{3}u_0^2. \tag{13}$$

Here we have introduced for simplicity of further calculations the parameter $M_s\equiv u_0^2/3$, where subscript s denotes the stochastic effect. After averaging, only the second term on the right of Eq. (12) contributes to the velocity increase, and in calculations of the variance the first term is sufficient.

If the number n of scattering events is sufficiently large, we can replace the first equation in (13) with a differential equation

$$\frac{\partial v(n)}{\partial n}=\frac{M_s}{v(n)}. \tag{14}$$

Its solution with the initial condition $v(0)=v_0$ yields the most probable velocity as a function of the number of collisions:

$$v(n)=\sqrt{2M_sn+v_0^2}. \tag{15}$$

Since the particle velocity is expressed as a sum of independent random quantities Δv_n with known mean and variance, it follows from Lyapunov's central limit theorem that the distribution function of the random value $v_n=v_0+\Sigma_{i=1}^n\Delta v_i$ tends to a normal distribution with mean $v(n)$ and variance $n\sigma_s^2$. Thus, the velocity distribution has the shape of a spreading Gaussian. The position of the distribution peak is at the most probable velocity $v(n)$, proportional to the square root of n.

This reasoning applies only to the case of a sufficiently high particle velocity, $v\gg u_0$. In order to describe the distribution at lower velocities, let us introduce an additional condition, namely, that there is no flow of particles to the region of negative velocities: $(v\partial\rho/\partial v)_{v=0}=0$. It is well known that the Gaussian distribution that satisfies this condition has the form

$$\rho(v,n)=\frac{1}{\sigma_s\sqrt{2\pi n}}\left[\exp\left(-\frac{[v-v(n)]^2}{2\sigma_s^2 n}\right)+\exp\left(-\frac{[v+v(n)]^2}{2\sigma_s^2 n}\right)\right]. \tag{16}$$

This allows us to calculate the mean velocity in the particle ensemble as a function of the number of scattering events:

$$V(n)=\sigma_s\sqrt{\frac{2n}{\pi}}\exp\left(-\frac{v^2(n)}{2\sigma_s^2 n}\right)+v(n)\Phi\left(\frac{v(n)}{\sigma_s\sqrt{2n}}\right), \tag{17}$$

where $\Phi(x)=(2/\sqrt{\pi})\int_0^x\exp(-x^2)dx$ is the error function. Hereafter V denotes the mean velocity in the particle ensemble. By substituting all coefficients and expanding the expression for the velocity, we obtain

$$V(n)=C\sqrt{n}+O\left(\frac{1}{\sqrt{n}}\right), \tag{18}$$

where the constant $C=\sqrt{2}[\sigma_s\exp(-M_s/\sigma_s^2)/\sqrt{\pi}+\Phi(\sqrt{M_s}/\sigma_s)\sqrt{M_s}]\approx 1.143u_0$.

Thus, Eqs. (16) and (18) determine the velocity distribution and the mean velocity in the ensemble as functions of the number of scattering events.

To calculate the mean velocity versus time we use the Fokker–Planck equation:

$$\frac{\partial\rho(v,t)}{\partial t}=-\frac{\partial}{\partial v}[A\rho(v,t)]+\frac{1}{2}\frac{\partial^2}{\partial v^2}[B\rho(v,t)],$$

where the factors A and B are given by the expressions

$$A\equiv\left\langle\frac{\Delta v}{\tau}\right\rangle=\frac{M_s}{l},\quad B\equiv\left\langle\frac{\Delta v^2}{\tau}\right\rangle=\frac{\sigma_s^2 v}{l}.$$

Here the mean time between collisions $\tau=l/v$, l is the mean free path, and Δv and Δv^2 are defined by Eq. (13). By substituting the resulting coefficients in the equation, we obtain

$$\frac{\partial\rho(v,t)}{\partial t}=-\frac{M_s}{l}\frac{\partial}{\partial v}\rho(v,t)+\frac{1}{2}\frac{\sigma_s^2}{l}\frac{\partial^2}{\partial v^2}[v\rho(v,t)]. \tag{19}$$

If parameters M_s and σ_s are determined in accordance with Eq. (13), the solution of this equation in the limit of high velocities much larger than the initial value, i.e., after a sufficiently long time interval, tends to

$$\rho(v,t)=\frac{1}{\sqrt{2tA\pi v}}\exp\left(-\frac{v}{2tA}\right),$$

where $A=M_s/l$. The latter expression yields the mean particle velocity:

$$V(t)=\frac{M_s}{l}t+v_0=\frac{1}{3}\frac{u_0^2}{l}t+v_0. \tag{20}$$

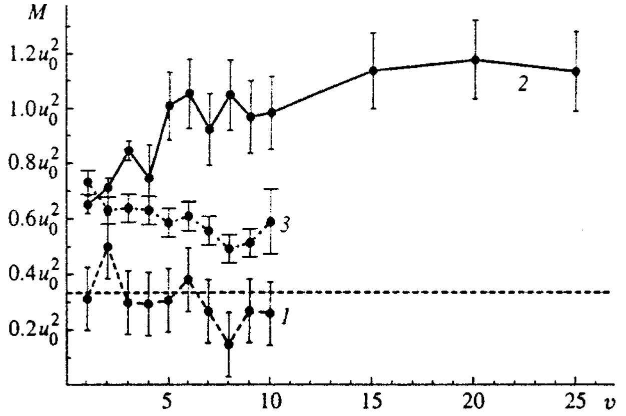

FIG. 3. Parameter $M\equiv v\langle\Delta v\rangle$ as a function of particle velocity. Curves *1* and *2* are calculated by the Lorentz gas model for the cases of random and regular boundary oscillations, respectively. Curve *3* is calculated using the simplified mapping (22). The dashed line shows M_s in the Lorentz gas in the case of stochastic boundary oscillations calculated by Eq. (13). Results are obtained at $u_0=0.01$, $a=1$, and $R=0.4$.

Thus, the system under investigation gives rise to the Fermi acceleration, with the particle velocity increasing as a linear function of time.

3.3. Periodically perturbed scatterer boundaries

Suppose that all scatterer boundaries contract and expand following a certain periodic law with a constant initial phase. Then, during one half of the period, the particle velocity should increase as a result of collisions and decrease during the other half. If the particle velocity is high enough, the time τ_s between scattering events is longer than the period T of scatterer oscillations. This leads to correlations in particle velocity variations, so the sequential increments in the velocity defined by Eq. (12) can no longer be treated as independent.

This section presents numerical calculations of the velocity variance and its average increase, alongside the decay rate of the correlation function $R(m)=(\Delta v_n\Delta v_{n+m})$. They indicate, in particular, that correlations can lead to larger first and second moments of velocity distributions. The calculations were performed on the basis of the Lorentz gas model with the following parameters: the scatterer radius $R=0.4$; the separation between their centers $a=1$ [thus, the basic model parameter $(a/R)^2=6.25$]; the amplitude of the scatterer surface velocity $u_0=0.01$; the oscillation frequency $\omega=1$.

It follows from the analysis of the previous subsection that at high particle velocities $\langle\Delta v\rangle\sim 1/v$. Therefore, the variable most convenient for the analysis and graphic representation is $M\equiv\langle\Delta v\rangle v$. Figure 3 shows M plotted against the particle velocity in the case of stochastic (curve *1*) and periodic (curve *2*) boundary oscillations. One can see that, in the case of stochastic oscillations, the variable $M_s\approx u_0^2/3$ coincides with to the result of the previous subsection. In the case of regular oscillations, M_r first increases, and then most likely tends to a constant $M_r^{\max}=(1.15\pm 0.10)u_0^2$ at $v\geqslant 15$, which corresponds in this specific billiard configuration to $n\geqslant 150$ particle collisions with the boundary during one oscillation period on average. In addition, it is clear that the particle acceleration in the case of regular boundary oscillations is a factor of three to four higher than in the case of stochastic oscillations.

For the analysis of velocity changes in chaotic billiards with periodically oscillating boundaries, the following approach can be suggested. Consider an approximate mapping for velocity (12). Since correlations between parameters α_n decay exponentially (as follows from the billiard configuration), the formulas can be averaged over α using Eq. (4). Then

$$\langle \Delta v\rangle_\alpha = -\frac{\pi}{2}u_0\cos\omega t_n + \frac{u_0^2\cos\omega t_n}{v_n}. \tag{21}$$

During the oscillation period, the largest contribution to changes in the velocity is due to the first term on the right. Therefore, it is sufficient in the first approximation to take into account only the changes in the velocity due to the first term, and the second can be neglected. On the other hand, correlational corrections to the second term generate terms of higher orders than that of its average. Therefore, correlation effects in the second term can be neglected. For this reason, the two values related to the first and second terms can be calculated independently:

$$\langle \Delta v\rangle = \langle \Delta v\rangle_{\mathrm{I}} + \langle \Delta v\rangle_{\mathrm{II}},$$

where $\langle v\rangle_{\mathrm{II}} = u_0^2/(3v)$, which coincides with μ_s in the stochastic case [Eq. (13)], and $\langle \Delta v\rangle_{\mathrm{I}}$ is the correction due to correlations. Discarding the second term on the right of Eq. (21), we have the following mapping for calculating $\langle \Delta v\rangle_{\mathrm{I}}$:

$$v_{n+1} = v_n + \gamma\cos\theta_n, \quad \theta_{n+1} = \theta_n + \frac{l_{n+1}\omega}{v_{n+1}}. \tag{22}$$

Here $\gamma = -\pi u_0/2$, and the collision phase $\theta_n \equiv \omega t_n$ is substituted for time. This mapping is exactly equivalent to Ulam's well-known mapping,[14–21] the only difference being that in this case the free path l_n is a random parameter distributed over a certain interval.

Let us analyze numerically this mapping at the same values of u_0 and ω as those selected in our analysis of the Lorentz gas. Suppose that the free path l_n has a normal distribution with mean $l = 0.62$ and variance $\sigma_l^2 = 0.657$. This corresponds to the variance and mean free path calculated numerically at $R = 0.4$ and $a = 1$ (see the previous subsection). Figure 3 shows $\langle \Delta v\rangle_{\mathrm{I}} v + u_0^2/3$ (curve *3*) derived from mapping (22). One can see in the graph that the first moment of the velocity distribution defined by this mapping becomes positive, but it is still smaller than the observed velocity increase in the Lorentz gas. Nonetheless, this mapping is easier for analysis than Eq. (21).

Now let us estimate the variance and decay rate of correlations in the velocity change. Suppose that the particle velocity is so high that its change after n scattering events is negligible. It is clear that, in order to satisfy this condition, one can choose v and u_0 in a proper manner. Let us calculate correlations between velocity increments Δv_m and Δv_{m+n} [Eq. (12)] for $n \to \infty$. Taking into account in the first approximation only the first terms on the right of Eq. (12), we obtain

$$R(n) \equiv \langle \Delta v_m \Delta v_{m+n}\rangle = u_0^2\frac{\pi^2}{4}\langle \cos\omega t_m \cos\omega t_{m+n}\rangle,$$

which takes into account, as follows from Eq. (4), that $\langle \cos\alpha_n\rangle = \pi/4$. By setting the oscillation frequency to unity and introducing the notation $S_n \equiv \Sigma_{i=1}^n \tau_{m+i}$, where $\tau_i = t_i - t_{i-1}$, we obtain

$$\langle \cos t_m \cos t_{m+n}\rangle = \langle \cos t_m \cos(t_m + S_n)\rangle.$$

The variable S_n can be expressed as

$$S_n = \sum_{i=1}^{n} (l + \Delta l_i)/v,$$

where Δl_i is the deviation from the mean free path on the ith collision. Since S_n is the sum of independent random quantities, its distribution at large n tends to the normal distribution with mean nl and variance $n\sigma_l^2$, where σ_l^2 is the mean free path variance. By expanding the cosine of the sum and averaging over S_n, we obtain the following expression for the correlation function of velocity increments:

$$R(n) \simeq \frac{\pi^2}{8}u_0^2\cos(\omega n\tau)\exp\left(-\frac{n}{N}\right), \tag{23}$$

where ω is the frequency of scatterer oscillations, $N = v^2/(\omega^2\sigma_l^2)$. Thus, correlations between sequential changes in the particle velocity are the stronger, the higher the velocity, and their "half-life" N, i.e., the number of collisions after which correlations drop by factor e, increases proportionally to v^2. Note that the number of collisions over one period is proportional to v. Thus, in order to estimate correctly the velocity variance, one has to average over the larger number of oscillation periods, the higher the particle velocity. The issue of how this can be done, however, has remained unresolved.

In order to estimate the variance in the first approximation, let us consider the velocity increment after two sequential collisions with the boundary. In this analysis, we assume that correlations among three and more increments are negligible. In the limit of a high velocity of a billiard particle, the correlator of sequential velocity increments can be estimated by the formula

$$\langle \Delta v_n \Delta v_{n+1}\rangle = u_0^2\frac{\pi^2}{4}\langle \cos^2\omega t_n(1 - O(\tau^2))\rangle = u_0^2\frac{\pi^2}{8} + O\left(\frac{u_0^2}{v^2}\right).$$

From this expression and Eq. (13), we derive

$$\sigma_r^2 = \frac{\langle (\Delta v_n + \Delta v_{n+1})^2\rangle}{2} \approx \left(\frac{4}{3} + \frac{\pi^2}{8}\right)u_0^2. \tag{24}$$

Figure 4 shows numerical and analytic estimates of the velocity increment variance in the stochastic (dashed line) and regular (solid lines) cases. In the case of stochastic oscillations, the numerical and analytic [Eq. (13)] estimates are identical, so the graph shows only numerical calculations of σ_s^2. Regular oscillations are characterized in this graph by the straight line defined by Eq. (24) and the broken line

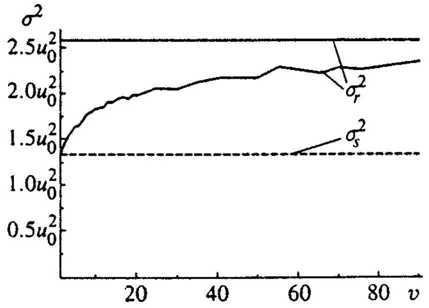

FIG. 4. Variance versus particle velocity in the Lorentz gas in the cases of stochastic (dashed line) and regular (solid lines) oscillations. The straight line shows the theoretical estimate of the variance in the regular case by Eq. (24). The calculations were performed at $u_0 = 0.01$, $a = 1$, and $R = 0.4$.

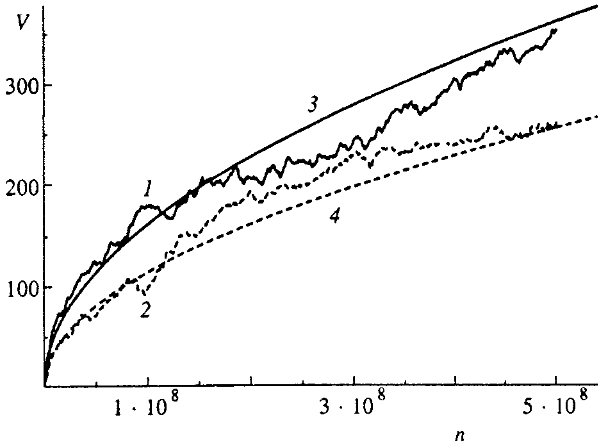

FIG. 5. Mean particle velocities as functions of the number of scattering events in the Lorentz gas (curves *1* and *2*) and their approximations by Eq. (18) (curves *3* and *4*). The dashed lines correspond to stochastic boundary oscillations, the solid lines correspond to regular oscillations. The averaging was performed over 100 process realizations with different velocity directions selected at random. The calculation were performed at $u_0 = 0.01$, $a = 1$, and $R = 0.4$.

calculated numerically. In order to take into account correlations between velocity increments, we calculated in the regular case the effective variance $\sigma_r^2 = \langle \Delta V^2 \rangle / N_{\max}$, where ΔV is the total velocity increment after $N_{\max}$ collisions. Given Eq. (23) describing the correlation function decay, we equated $N_{\max}$ to $10v^2/(\omega^2\sigma_l^2)$, which is a factor of ten higher than the characteristic correlation decoupling number. As is shown by the graphs, the variance σ_s^2 in the stochastic case is constant, whereas in the regular case (σ_r^2) it grows with the velocity. In addition, the variance in the regular case determined by Eq. (24) is slightly overestimated.

Thus, the numerical and analytic estimates given in this section indicate that particle acceleration should occur in chaotic billiards with periodically oscillating boundaries. We can most likely say that deterministic chaoticity is a sufficient condition for Fermi acceleration. Moreover, periodic oscillations of billiard boundaries lead to a higher particle acceleration.

3.4. Numerical results

This section presents numerically calculated particle velocity as a function of the number of scattering events and time in comparison with the analytic estimates given above. The calculations were performed by the Lorentz gas model with the following parameters: the amplitude of the scatterer boundary oscillation velocity $u_0 = 0.01$; the scatterer radius $R = 0.4$; the distance between their centers $a = 1$; the frequency of boundary oscillations $\omega = 1$; the initial velocity $v_0 = 1$. Thus, the mean free path calculated analytically for these parameters, $l = 0.6216815$. The numerical calculations of the mean free path [Eq. (6)] and its variance in this specific billiard configuration yield $l = 0.62163 \pm 0.00003$ and $\sigma_l^2 = 0.657 \pm 0.001$.

The difference in realizations was in the initial values of α and ϕ, which were selected at random. Two different cases were investigated: stochastic oscillations of scatterer boundaries with initial phases distributed uniformly and regular oscillations of boundaries. In both cases, the billiard ball (particle) dynamics was determined by the mapping derived in Secs. 2.1 and 2.2. The scatterer boundary oscillation velocity in the first case was defined by the formula $u_n = u_0 \cos\varphi_n$, where φ_n is a random parameter uniformly distributed over the interval $[0, 2\pi)$, and in the second case by the formula $u_n = u_0 \cos\omega t_n$, where t_n is the moment of collision between the particle and boundary. For each case 100 realizations of billiard dynamics were investigated. The averaged velocity as a function of the number of scattering events and time is plotted in Figs. 5 and 6, respectively. In both graphs, the solid lines plot the data for the regular case, and the dashed line corresponds to the case of random oscillations.

Figure 5 shows the averaged velocity of an ensemble of particles versus the number n of scattering events over the range of 5×10^8 iterations. It is clear that both curves are accurately approximated by the square-root function (18). In the case of stochastic oscillations, parameters M_s and σ_s were derived from Eq. (13), and in the regular case the limiting values M_r and σ_r were derived from numerical calculations described in the previous subsection.

The curves of the mean velocity versus time (Fig. 6) plot the data averaged over 100 realizations in the stochastic (dashed lines) and regular (solid lines) cases. The particle dynamics was simulated over a time interval of $[0, 3 \times 10^6]$ time units, and some trajectories of ''fast'' particles covered up to 3×10^9 iterations. The mean particle velocity was approximated using Eq. (20). The parameter M_s was calculated for stochastic oscillations by Eq. (13), and for regular oscillations as a limit of M_r obtained in the previous subsection. The curves show that the growth in the particle velocity is approximately linear, and the approximation of the average velocity by Eq. (20) is in reasonable agreement with computer simulations.

4. CONCLUSIONS

Billiards are fairly convenient models of a set of physical systems. For example, particle trajectories in billiards of specific configurations can be used in modeling many dy-

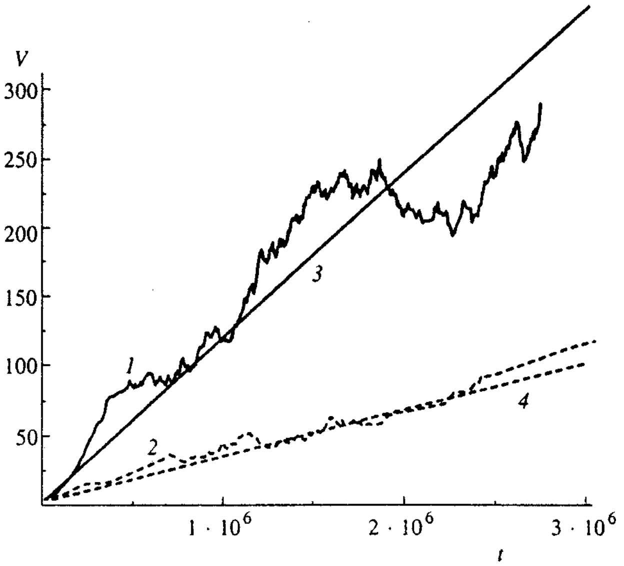

FIG. 6. The same data as in Fig. 5, but plotted against time. The approximation was performed by Eq. (20).

namical systems. Moreover, most approaches to the problems of mixing in many-body systems originate from billiard-like problems. A natural generalization of a billiard system is a billiard whose boundary is not fixed, but varies following a certain law. This is a relatively new field of research, which opens new prospects in studies of problems that have been known for a long time, but have been poorly investigated. For example, the problem of particle dynamics in a billiard whose boundary changes with time has a direct physical application as a model of nonequilibrium statistical mechanics. As follows from the existing literature, the dynamical properties of a billiard with perturbed boundaries are important: if its dynamics is chaotic, boundary perturbations can lead to an infinite growth in the particle velocity, i.e., such a billiard demonstrates Fermi acceleration.

In the present article, we have studied the problem of Fermi acceleration in dynamical systems generated by two-dimensional dispersing billiards with perturbed boundaries. A billiard with a boundary like that of the Lorentz gas oscillating in accordance with a certain law has been investigated. It is well known that the conventional Lorentz gas (i.e., that with an unperturbed boundary) has clearly demonstrated chaotic properties (mixing, decay of correlations, etc.). Perturbation of boundaries in such a billiard leads to the Fermi acceleration. This model has been studied in two versions, namely, those with stochastically and regularly oscillating scatterer boundaries. It has turned out that the acceleration is higher in the case of periodical boundary oscillations.

We can identify two basic acceleration mechanisms, which have been discovered in deriving the particle velocity distribution as a function of the number of scattering events in the case of stochastic oscillations (Sec. 3.3). The first is the mechanism deriving from the condition $\langle \Delta v\rangle > 0$ [Eq. (13)], which drives all particles to the side of higher velocities. The second is the dispersive (or fluctuational) mechanism controlled by two conditions: (a) $\langle \Delta v^2\rangle > 0$, therefore the peak in the velocity distribution spreads with time; (b) the absolute value of velocity cannot be negative, therefore the peak spread cannot be symmetrical, but its predominant direction is to the side of higher velocities, as a result, the simple normal distribution is replaced by distribution (16). Moreover, it follows from both analytic and numerical calculations that fluctuations and the mean increase in the particle velocity are larger in the case of regular scatterer boundary oscillations, which leads to a larger velocity growth. Thus, the mechanism due to correlations between sequential changes in the velocity has been suggested.

It is quite clear that the reasoning used in deriving the particle velocity as a function of the number of scattering events and time can be directly translated into another type of billiard in which a distribution of angle α (between the normal to the surface at the impact point and particle velocity) is known. Therefore, the technique developed in our work can be used in solving the problem of Fermi acceleration in the general case.

The presence of a chaotic condition in a system can change its statistical properties. A recent publication by Tsang and Ngai[10] considered a billiard in an area defined by a rectangle whose corners were replaced by quarter-circles of radius R (smoothed corners) and one side oscillated periodically. A particle travels within this area and collides with the boundaries. Each collision with the boundary is not perfectly elastic, and the particle loses a fraction of its energy proportional to a constant δ ($\delta \ll 1$). This model is similar to Ulam's model, but the presence of smoothed corners introduces random elements to the particle dynamics. Tsang and Ngai[10] investigated relaxation of a system to equilibrium. A similar investigation was performed earlier by Tsang and Lieberman[25] on the basis of Ulam's model. It was shown that the function $\Phi(t)=E(t)-E(\infty)$, which is the deviation of the mean energy from the equilibrium value, drops exponentially, $\Phi(t)\propto \exp(-t/\tau)$, which is quite natural of most physical systems. The investigation of this parameter in the billiard discussed in Ref. 10 revealed that its relaxation to equilibrium in this case is slower, $\Phi(t)\propto \exp[-(t/\tau)^{\beta}]$, where $\beta<1$ and drops with R. Given the results of this paper, we can understand the cause of the slower system relaxation. In fact, the random element in the system becomes more important at larger radii of circles at the corners, which leads to acceleration of particles. Therefore the system relaxation to its equilibrium, associated with the particle energy dissipation in the system, becomes slower. The approaches developed in the reported work create preconditions for determination of β, hence of the relaxation rate to equilibrium in a system where the chaotic dynamics is dominant.

Thus, on the basis of our investigations, we can put forth an important hypothesis: a random element in a billiard with a fixed boundary is a sufficient condition for the Fermi acceleration in the system when a boundary perturbation is introduced.

---

*⁾E-mail: loskutov@moldyn.phys.msu.su

[1] Ya. G. Sinay, Usp. Mat. Nauk **25**, 141 (1970).

[2] L. A. Bunimovich, Commun. Math. Phys. **65**, 295 (1979).

[3] L. A. Bunimovich and Ya. G. Sinai, Commun. Math. Phys. **78**, 479 (1981).

[4] L. A. Bunimovich, *Dynamical Systems*, Vol. 2 [in Russian], VINITI, Moscow (1985), p. 173.

[5] L. A. Bunimovich, Chaos **1**, 187 (1991).

[6] A. Tabachnikov, *Billiards*, France Math. Soc. Press, Lyon (1995).

[7] *Proc. of the Int. Conf. on Classical and Quantum Billiards*, J. Stat. Phys. **83** (1–2) (1996).

[8] J. Koiller, R. Markarian, S.Q. Kamphorst, and S. P. de Carvalho, Nonlinearity **8**, 983 (1995); J. Stat. Phys. **83**, 127 (1996).

[9] *New Trends in Hamiltonian Systems*, Word Scientific, Singapore (1996).

[10] K. J. Tsang and K. L. Ngai, Phys. Rev. E **56**, R17 (1997).

[11] L. G. Akinshin, K. A. Vasil'ev, A. Yu. Loskutov, and A. B. Ryabov, Physical Ideas of Russia [in Russian] **2–3**, 87 (1997).

[12] E. Fermi, Phys. Rev. **75**, 1169 (1949).

[13] G. M. Zaslavskiĭ and B. V. Chirikov, Dokl. Akad. Nauk SSSR **159**, 306 (1964) [Sov. Phys. Doklady **9**, 989 (1964)].

[14] S. M. Ulam, in *Proc. of the 4th Berkeley Symp. on Math. Stat. and Probability*, California Univ. Press (1961), Vol. 3, p. 315.

[15] A. Brahic, Astron. Astrophys. **12**, 98 (1971).

[16] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Springer-Verlag, New York (1992).

[17] G. M. Zaslavskiĭ, *Stochastic Irreversibility in Nonlinear Systems* [in Russian], Nauka, Moscow (1970).

[18] A. J. Lichtenberg, M. A. Lieberman, and R. H. Cohen, Physica D **1**, 291 (1980).

[19] L. D. Pustyl'nikov, Dokl. Akad. Nauk SSSR **292**, 549 (1987) [*sic*].

[20] L. D. Pustyl'nikov, Mat. Sb. **85**, 113 (1994).

[21] T. Krüger, L. D. Pustyl'nikov, and S. E. Troubetzkoy, Nonlinearity **8**, 397 (1995).

[22] P. R. Baldwin, J. Phys. A **24**, L941 (1991).

[23] N. Chernov, J. Stat. Phys. **88**, 1 (1997).

[24] P. L. Garrido, J. Stat. Phys. **88**, 807 (1997).

[25] K. J. Tsang and M. A. Lieberman, Physica D **11**, 147 (1984); Phys. Lett. A **103**, 175 (1984).

Translation provided by the Russian Editorial office

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Nonequilibrium Josephson effect in systems of tunnel superconducting junctions and in layered superconductors

D. A. Ryndyk*)

Institute of Microstructure Physics, Russian Academy of Sciences, GSP-105, 603600 Nizhniĭ Novgorod, Russia

(Submitted 25 May 1999)
Zh. Éksp. Teor. Fiz. **116**, 1798–1816 (November 1999)

This paper examines the time-dependent Josephson effect in systems of tunnel superconducting junctions and in layered superconductors (the intrinsic Josephson effect) with allowance for nonequilibrium superconductivity effects. Kinetic and quasihydrodynamic equations are derived that describe self-consistently the dynamics of Josephson phases and nonequilibrium quasiparticles. It is found that the state of nonequilibrium between the layers leads to an effective mechanism of the interaction between interlayer Josephson junctions, which can be used to synchronize the junctions. Current–voltage characteristics of chains of intrinsic junctions are obtained for different values of the parameters.
[S1063-7761(99)02211-8]

1. INTRODUCTION

It is well known (see the review articles in Refs. 1–3) that when subjected to a finite voltage a Josephson junction is a source of nonthermal quasiparticles, which relax in the ''banks.'' Hence near the junction there is always a region of nonequilibrium superconductivity. The excess of quasiparticles changes the energy gap. More than that, a difference arises in the populations of the electronlike and holelike branches of the electron excitation spectrum (the electron–hole imbalance), which changes the chemical potential of the superconducting condensate (the number of superconducting electrons changes to balance the excessive charge of the quasiparticles) and leads to penetration of the superconductor by the electric field. If the banks are fairly large, diffusion of the quasiparticles plays the main role (in dirty superconductors). The characteristic spatial scales are determined by the energy relaxation depth $l_\epsilon=\sqrt{D\tau_\epsilon}$ and the imbalance relaxation depth (the penetration depth of the electric field) $l_q=\sqrt{D\tau_q}$, with D the diffusion coefficient, τ_ϵ the inelastic relaxation time, and τ_q the imbalance relaxation time. But if the banks are smaller than l_q in the direction of diffusion, a spatially homogeneous nonequilibrium state forms that relaxes with characteristic times τ_ϵ and τ_q. In this case the system may be farther from equilibrium.

Quasiparticle interactions between junctions have been studied in systems of Mercereau–Notaris bridges[4,5] and in SNS junctions.[6,7] The nonequilibrium change in the energy gap in SIS systems was examined in Refs. 8–11 (an interesting particular case of this phenomenon is the stimulation, or enhancement, of superconductivity in a double tunnel junction with a nonequilibrium middle layer). Another factor that may be important is the violation of the Josephson relationship $d\varphi/dt=(2e/\hbar)V$ between the voltage across the junction and the phase difference at the junction in the nonequilibrium case (see below). In layered superconductors with Josephson interaction between the layers, nonequilibrium effects may be very large because the effective layer thickness is small (3–10 Å). In recent studies,[12–27] the intrinsic Josephson effect in high-T_c superconductors was observed directly. A number of researchers[28–32] remarked on the importance of allowing for nonequilibrium effects, especially the violation of the equilibrium Josephson relation, in interpreting the results of experiments in this field.

In Sec. 2 we will discuss the elementary theory of the nonequilibrium Josephson effect and make estimates for artificial tunnel structures and high-T_c superconductors. Section 3 is devoted to the derivation of kinetic and quasihydrodynamic equations of quasiparticle dynamics in systems of tunnel junctions. In Sec. 4 we will examine the dynamics of a chain of junctions and present the results of numerical calculations of current–voltage characteristics and of the high-frequency interaction of junctions.

2. ELEMENTARY THEORY

From the microscopic viewpoint, the main process in a nonequilibrium superconductor is the change in the quasiparticle distribution function, which is accompanied by changes in the macroscopic characteristics of the condensate: the energy gap Δ, the shift $\delta\mu$ of the chemical potential, and the invariant potential Φ related to this shift.

A change in the energy gap takes place when the part of the distribution function which is symmetric in energy, n_ϵ (n_ϵ describes the quasielectron energy distribution for $\epsilon>0$ and the quasihole distribution for $\epsilon<0$); in the simplest case it can be found from the Éliashberg self-consistency equation[33]

$$1=\lambda\int_\Delta^{\theta_D}\frac{1-n_\epsilon-n_{-\epsilon}}{(\epsilon^2-\Delta^2)^{1/2}}\,d\epsilon. \tag{1}$$

On the other hand, the shift of the chemical potential of the condensate in the superconductor is determined by the part of the distribution function which is symmetric in

TABLE I. Artificial structures ($T = 4.2$ K).

Structure	$N(0)$, 10^{34} $\text{erg}^{-1}\,\text{cm}^{-3}$	RS, $\Omega\,\text{cm}^2$	d_0, Å	ν, s^{-1}	τ_ϵ, s	$\nu\tau_\epsilon$
Nb–AlO$_x$–Al–AlO$_x$–Nb (Ref. 8)	0.81	$(3-7)\times10^{-7}$	100	$\geqslant 10^8$	$\sim 10^{-8}$	$\geqslant 1$
Nb/Al–AlO$_x$–Nb′Al–AlO$_x$–Nb (Ref. 9)	~ 1	$\sim 10^{-6}$	~ 100	$\sim 10^8$	$\sim 10^{-9}$	~ 0.1
Nb/Al/Nb′/Al–AlO$_x$–Nb″Al–AlO$_x$–Nb′/Al/Nb (Ref. 9)						
Nb/Al–AlO$_x$–Al′–AlO$_x$–Al/Nb (Ref. 10)	0.81	$(1-4)\times10^{-6}$	60	$\sim 10^8$	$\sim 10^{-8}$	~ 1
Nb/Al/Nb/Al–AlO$_x$–Al′–AlO$_x$–Al/Nb/Al/Nb (Ref. 10)						
Nb–AlO$_x$–Nb′–AlO$_x$–Nb (Ref. 11)	1.99	$\sim 10^{-6}$	100	$\geqslant 10^7$	5×10^{-10}	~ 0.01

energy.[34–36] In the simplest case of a pure superconductor in a spatially homogeneous quasistatic nonequilibrium state, the energy distribution of the paired electrons is given by the expression $v_k^2 = (1/2)(1 - \xi_k/\epsilon_k)$, where $\xi_k = \hbar^2 k^2/2m - \epsilon_F - \delta\mu$, $\epsilon_k = \sqrt{\Delta^2 + \xi_k^2}$ is the quasiparticle spectrum, and $\delta\mu$ is the shift of the chemical potential. Then for $\delta\mu \ll \epsilon_F$ the charge density is given by the formula

$$\rho = 2eN(0)\left[\delta\mu + \int_\Delta^\infty (n_\epsilon - n_{-\epsilon})\,d\epsilon\right], \tag{2}$$

where the first term on the right-hand side is the condensate charge and the second is the quasiparticle charge, and $N(0) = mp_F/2\pi^2\hbar^3$ is the density of states on the Fermi surface. Equation (2) is the second self-consistency equation. In the quasineutral case ($\rho \approx 0$) we have

$$\delta\mu = -\int_\Delta^\infty (n_\epsilon - n_{-\epsilon})\,d\epsilon. \tag{3}$$

This expression is a reflection of the features of the charge effects in superconductors: when an electron–hole imbalance develops, the quasineutrality is ensured due to the superconducting electrons.

The shift of the chemical potential can be expressed in terms of what i known as the invariant potential: $\delta\mu = e\Phi$, with

$$\Phi(t) = \phi + \frac{\hbar}{2e}\frac{\partial\theta}{\partial t}, \tag{4}$$

where ϕ is the electric potential, θ is the phase of the order parameter, and $\Phi = 0$ in an equilibrium state. For a pure superconductor, in the quasiclassical case this relationship follows directly from the shape of the quasiparticle spectrum.[35] Generally, it is the gauge-invariant potential Φ that enters into the dynamical equations. The expression for the charge becomes

$$\rho = -2e^2N(0)(\Phi - \Psi) = -\frac{1}{4\pi r_d^2}(\Phi - \Psi), \tag{5}$$

where $\Psi = (1/e)\int_\Delta^\infty (n_\epsilon - n_{-\epsilon})\,d\epsilon$ is the electron–hole imbalance potential,$^{1)}$ and r_d is the penetration depth of the electric field.

In a state with a shift of the chemical potential, the ordinary relationship $d\varphi_{ij}/dt = (2|e|/\hbar)V_{ij}$ the Josephson phase difference $\varphi_{ij} = [\theta_j - \theta_i]\text{sgn}e$ and the voltage $V_{ij} = \phi_i - \phi_j$ breaks down. Instead, using the definition of Φ, we obtain the Josephson nonequilibrium relationship

$$\frac{d\varphi_{ij}}{dt} = \frac{2|e|}{\hbar}V_{ij} + \frac{2|e|}{\hbar}(\Phi_j - \Phi_i). \tag{6}$$

Thus, in the general case, we must self-consistently describe the dynamics of the Josephson phases $\varphi_{ij}(t)$ and the distribution functions $n_\epsilon^{(i)}(t)$ with allowance for the conditions (1) and (5) and for the nonequilibrium expressions for the tunneling current between the layers.

The main parameters determining the degree of the nonlinearity of the effects of any type in tunnel structures and other systems with weak coupling are the products $\nu\tau_\epsilon$ and $\nu\tau_q$, where ν is what is known as the tunnel frequency, which for thin layers ($d_0 \ll l_\epsilon, l_q$) is given by the formula

$$\nu = \frac{1}{4e^2N(0)R_Nd_0}, \tag{7}$$

where R_N is the resistance of the junction in the normal state per unit surface area. For $\nu\tau_\epsilon > 1$ or $\nu\tau_q > 1$ strong distortions of the quasiparticle distribution function are possible.

The results of estimates recently made for artificial structures are listed in Table I. These data suggest the following. First, the highly nonequilibrium regime is realized in tunnel structures based on Al and Nb. Second, the parameter $\nu\tau_\epsilon$ can be much larger than unity, $\nu\tau_\epsilon \gg 1$. Indeed, $\nu \sim 10^8\,\text{s}^{-1}$ correspond to very low barrier transmissivity, $D \sim 10^{-7} - 10^{-6}$ ($\nu = v_F D/4d_0$), and the condition for junction weakness is met with a large margin ($\nu \ll \Delta \sim 10^{12} - 10^{13}\,\text{s}^{-1}$). Thus, an increase in value of the nonequilibrium parameter by a factor of 100 to 1000 with barrier transmissivity is quite possible (at least in principle).

It is difficult to make estimates for high-T_c superconductors because of the lack of a sufficient large body of data obtained in direct measurements of nonequilibrium effects and the fuzziness of the microscopic picture. For rough estimates we can use formula (7) and a similar formula that expresses ν in terms of the critical current at absolute zero in the BCS model:

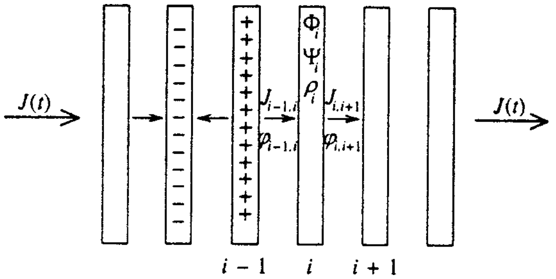

FIG. 1. A system of coupled Josephson layers.

$$\nu=\frac{J_c(0)}{2\pi e\Delta N(0)d_0}. \tag{8}$$

On the basis of Refs. 12–27 we assume that $J_c(0)\sim 500-20\,000\,\mathrm{A/cm^2}$, $V_c\sim 1-10\,\mathrm{mV}$, $R_N S\sim 10^{-5}-10^{-7}\,\Omega\,\mathrm{cm}^2$ (for structures with different degrees of anisotropy), $\Delta\sim 20\,\mathrm{meV}$, and $N(0)\sim 10^{32}-10^{33}$. We then arrive at the estimate $\nu\sim 10^8-10^{12}\,\mathrm{s}^{-1}$, depending on the degree of anisotropy and the way in which the estimate is made. When the inelastic relaxation time changes from $10^{-9}\,\mathrm{s}$ (at $T\sim 4.2\,\mathrm{K}$) to $10^{-12}\,\mathrm{s}$ (at $T\sim T_c$), the product $\nu\tau_\epsilon$ changes over a wide range from 10^{-4} in structures with the greatest anisotropy at temperatures close to critical (note, however, that $\nu\tau_q$ may be much larger in view of the factor T/Δ) to 10^3 in structures with weaker anisotropy at low temperatures, which suggests that under certain conditions a system with an intrinsic Josephson effect may be very far from equilibrium.

3. QUASIPARTICLE KINETICS IN SYSTEMS OF TUNNEL JUNCTIONS. THE QUASIHYDRODYNAMIC APPROXIMATION

In this paper we will examine the system of Josephson-coupled thin layers (Fig. 1) with a tunnel mechanism of current transfer between the layers. Thus, the state of the ith layer is determined by the interaction of that layer with the $(i-1)$st and $(i+1)$st layers. For any pair of layers this interaction is the same type, and below we will develop a microscopic theory of an arbitrary ith layer with allowance for the ''left,'' $(i-1)$st, layer. The effect of the ''right,'' $(i+1)$st, layer will be taken into account in the final expressions.

We will assume that the layer thickness d_0 is comparable to the penetration depth r_d of the electric field (as is the case in high-T_c superconductors) and is much smaller than the characteristic nonequilibrium relaxation depths, $d_0\ll l_\epsilon, l_q$. In this case the analysis of charge effects becomes much simpler, since we can average all quantities over the layer thickness.

3.1. The starting equations

The starting equations are the Éliashberg equations[33] for the two-time quasiclassical [integrated with respect to the energy variable $\xi=v_F(p-p_F)$] Green's functions $g_\epsilon(t)$, $f_\epsilon(t)$, $g_\epsilon^{A(R)}(t)$, and $f_\epsilon^{A(R)}(t)$ describing an arbitrary nonequilibrium state of the superconductor. In the most general form such equations were obtained via the Keldysh method[37] by Larkin and Ovchinnikov[38,39] (see also Refs. 2,3, and 40–44). The functions $g_\epsilon^{A(R)}$ and $f_\epsilon^{A(R)}$ describe the density of quasiparticle states in energy in the static case and are called spectral, while the ''Keldysh'' functions $g_\epsilon(t)$ and $f_\epsilon(t)$ describe the kinetics of the quasiparticles. Since we intend to go over to a kinetic (quasiclassical in time) equation, we introduce the Fourier representation in the time difference t_1-t_2 and retain the dependence on the time sum $t=(t_1+t_2)/2$. Moreover, in the ''dirty'' case ($\tau_p\Delta, \tau_p\omega\ll 1$ and $l_p\ll \xi_0, d_0$), we can average over the directions of the momentum on the Fermi surface. Thus, we obtain, for instance,2)

$$\begin{aligned} g_\epsilon(\mathbf{r},t)=\frac{i}{\pi}\int d\xi\int\frac{d\Omega_p}{4\pi}\int d(t_1-t_2)\int d(\mathbf{r}_1-\mathbf{r}_2)\\ \times G(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)\exp\{-i\epsilon(t_1-t_2)\\ +i\mathbf{p}\cdot(\mathbf{r}_1-\mathbf{r}_2)\}, \end{aligned} \tag{9}$$

The equations for the functions g_ϵ and $f_\epsilon(t)$ averaged over the layer thickness have the form (in a gauge in which the order parameter is real)

$$\frac{\partial g_\epsilon}{\partial t}=-ie\{\Phi g-g\Phi\}_\epsilon-i\{\Delta f^\dagger-f\Delta\}_\epsilon-I_1^{\mathrm{ph}}-I_1^{\mathrm{tun}}, \tag{10}$$

$$i2\epsilon f_\epsilon=ie\{\Phi f-f\Phi\}_\epsilon-i\{\Delta\overline{g}-g\Delta\}_\epsilon+I_2^{\mathrm{ph}}+I_2^{\mathrm{tun}}, \tag{11}$$

where I^{ph} and I^{tun} are, respectively the phonon collision integral and the tunneling source, $f^\dagger(t_1,t_2)=f^*(t_2,t_1)$ and $\overline{g}(t_1,t_2)=g(t_2,t_1)$ in the temporal representation, and the symbol $\{AB\}$ stands for the convolution in time,

$$\{AB\}(t_1,t_2)=\int A(t_1,t_3)B(t_3,t_2)\,dt_3, \tag{12}$$

which in the frequency representation has the form

$$\{AB\}_\epsilon(\omega)=\int A_{\epsilon+\omega/2-\omega_1/2}(\omega_1)B_{\epsilon-\omega_1/2}(\omega-\omega_1)\,\frac{d\omega_1}{2\pi}, \tag{13}$$

and in the quasiclassical case can be represented by a power series in ω, i.e.,

$$\begin{aligned} \{AB\}_\epsilon=A_\epsilon B_\epsilon+\frac{i}{2}\left(\frac{\partial A_\epsilon}{\partial\epsilon}\frac{\partial B_\epsilon}{\partial t}-\frac{\partial A_\epsilon}{\partial t}\frac{\partial B_\epsilon}{\partial\epsilon}\right)\\ -\frac{1}{8}\left(\frac{\partial^2A_\epsilon}{\partial\epsilon^2}\frac{\partial^2B_\epsilon}{\partial t^2}-2\frac{\partial^2A_\epsilon}{\partial\epsilon\,\partial t}\frac{\partial^2B_\epsilon}{\partial\epsilon\,\partial t}+\frac{\partial^2A_\epsilon}{\partial t^2}\frac{\partial^2B_\epsilon}{\partial\epsilon^2}\right)+\cdots. \end{aligned} \tag{14}$$

Here in (10) and (11) we must assume that formally $\Delta(t_1,t_2)=\Delta(t_1)\delta(t_1-t_2)$ and $\Phi(t_1,t_2)=\Phi(t_1)\delta(t_1-t_2)$. The following relationships hold:3)

$$g^*(t_1,t_2)=g(t_2,t_1),\quad f(t_1,t_2)=f(t_2,t_1),$$
$$f_\epsilon^\dagger=f_\epsilon^*,\quad f_\epsilon=f_{-\epsilon},\quad \overline{g}_\epsilon=g_{-\epsilon},\quad g_\epsilon^*=g_\epsilon,$$

which imply that $g_\epsilon(t)$ is a purely real function and $f_\epsilon(t)$ is an even function.

The tunneling source I^{tun} was obtained by Volkov[45] and in our notation has the form

$$\begin{aligned} I_1^{\mathrm{tun}}=\nu\{g^{(t)R}g-gg^{(t)A}+ff^{(t)\dagger A}-f^{(t)R}f^\dagger\\ +g^{(t)}g^A-g^Rg^{(t)}+f^Rf^{(t)\dagger}-f^{(t)}f^{\dagger A}\}, \end{aligned} \tag{15}$$

$$I_2^{\mathrm{tun}}=\nu\{g^{(t)R}f\overline{g}^{(t)A}+f^{(t)R}\overline{g}-gf^{(t)A}$$
$$+g^{(t)}f^{A}-f^{R}\overline{g}^{(t)}+f^{(t)}\overline{g}^{A}-g^{R}f^{(t)}\}. \tag{16}$$

Here the Green′s functions with the superscript (t) refer to the other (‘‘left″) superconductor. These functions contain an additional phase factor (in relation to the functions taken in a gauge in which the order parameter is real), which reflects the presence of the Josephson phase difference $\varphi(t)=[\theta'(t)-\theta(t)]\mathrm{sgn}e$ between the superconductors. This means that the following gauge transformation must be carried out:

$$\begin{aligned}
g^{(t)(R,A)}(t_1,t_2)&=\exp\left\{i[\varphi(t_1)-\varphi(t_2)]\mathrm{sgn}\frac{e}{2}\right\}\\
&\quad\times g'^{(R,A)}(t_1,t_2),\\
\overline{g}^{(t)(R,A)}(t_1,t_2)&=\exp\left\{-i[\varphi(t_1)-\varphi(t_2)]\mathrm{sgn}\frac{e}{2}\right\}\\
&\quad\times \overline{g'}^{(R,A)}(t_1,t_2),\\
f^{(t)(R,A)}(t_1,t_2)&=\exp\left\{i[\varphi(t_1)+\varphi(t_2)]\mathrm{sgn}\frac{e}{2}\right\}\\
&\quad\times f'^{(R,A)}(t_1,t_2),\\
f^{(t)\dagger(R,A)}(t_1,t_2)&=\exp\left\{-i[\varphi(t_1)+\varphi(t_2)]\mathrm{sgn}\frac{e}{2}\right\}\\
&\quad\times f'^{\dagger(R,A)}(t_1,t_2),
\end{aligned}\tag{17}$$

where the primed functions now satisfy the same gauge-invariant Eqs. (10) and (11) as the unprimed.

The self-consistency condition for the energy gap has the ordinary form

$$\Delta(t)=\lambda\int_{-\theta_D}^{\theta_D}\frac{d\epsilon}{4}\operatorname{Re}f_\epsilon(t), \tag{18}$$

and for the second self-consistency condition we must take the continuity equation with the charge

$$\rho(t)=-2eN(0)\left[e\Phi(t)+\int_{-\infty}^{\infty}\frac{d\epsilon}{4}g_\epsilon(t)\right]. \tag{19}$$

These conditions constitute a generalization of the expressions (1) and (5) to an arbitrary time-dependent case.

The resulting equations must be augmented with the expression for the current flowing between the layers. A convenient way of calculating the current is to integrate Eq. (10) with respect to energy. We will use this trick when we pass to the kinetic approximation.

The spectral functions are determined by equations of the form (10) and (11). We will assume that in the static case they are

$$g_\epsilon^{R(A)}=\frac{\epsilon\pm i\gamma}{\Delta}f_\epsilon^{R(A)}=\frac{\epsilon\pm i\gamma}{\sqrt{(\epsilon\pm i\gamma)^2-\Delta^2}}, \tag{20}$$

$$f_\epsilon^{\dagger R(A)}=f_\epsilon^{R(A)},\quad \overline{g}_\epsilon^{R(A)}=-g_\epsilon^{R(A)}, \tag{21}$$

$$g_\epsilon^{*R(A)}=-g_\epsilon^{A(R)},\quad f_\epsilon^{*R(A)}=-f_\epsilon^{A(R)},$$

with the sign of the square root determined by the condition $\mathrm{Im}\sqrt{\cdots}>0$. Below we will need combinations of the spectral functions, for which we introduce the special notation

$$\begin{aligned}
u_\epsilon&=\frac{1}{2}(g_\epsilon^R-g_\epsilon^A),\quad v_\epsilon=\frac{1}{2}(f_\epsilon^R-f_\epsilon^A),\\
w_\epsilon&=\frac{i}{2}(f_\epsilon^R+f_\epsilon^A),\quad y_\epsilon=\frac{i}{2}(g_\epsilon^R+g_\epsilon^A),\\
h_\epsilon&=\frac{1}{2}[1-(g_\epsilon^Rg_\epsilon^A-f_\epsilon^Rf_\epsilon^{\dagger A})],\quad b_\epsilon=\frac{i}{2}(g_\epsilon^Rf_\epsilon^A+\overline{g}_\epsilon^Af_\epsilon^R).
\end{aligned}\tag{22}$$

3.2. Kinetic approximation

At low frequencies $(\omega\ll\Delta)$, the system of the dynamical equations can be greatly simplified by passing to the kinetic equation for the energy distribution function for the quasiparticles. Such an equation for a system of tunnel junctions in the quasistatic case $(\omega\to 0)$ was derived by Bulyzhenkov and Ivlev,[46] Ivlev,[47] and Gulyan and Zharkov.[48] As shown by Larkin and Ovchinnikov[38,39] and Schmid and Shön,[49] in the time-dependent case the passage to the kinetic approximation can be achieved by introducing two real distribution functions f_1 and f_2, in terms of which the functions g_ϵ and f_ϵ are expressed so that the normalization condition

$$\left\{\begin{pmatrix} g^R & f^R\\ -f^{\dagger R} & \overline{g}^R\end{pmatrix}\begin{pmatrix} g & f\\ -f^{\dagger} & \overline{g}\end{pmatrix}\right\}$$
$$+\left\{\begin{pmatrix} g & f\\ -f^{\dagger} & \overline{g}\end{pmatrix}\begin{pmatrix} g^A & f^A\\ -f^{\dagger A} & \overline{g}^A\end{pmatrix}\right\}=0 \tag{23}$$

holds. As for the spectral functions $g_\epsilon^{R(A)}$ and $f_\epsilon^R(A)$, it is sufficient to express them in terms of $\Delta(t)$ via the formulas (20) of the static approximation.[39]

Shelankov[43] found that when the spin states are equally populated, one real distribution function of the general form $f_\epsilon(t)$ is sufficient, and the functions f_1 and f_2 have a definite symmetry. Using the Shelankov method,[43] we can express the Green's functions immediately in terms of the symmetric part $\beta_\epsilon(t)=(n_\epsilon+n_{-\epsilon}-1)\mathrm{sgn}\epsilon$ and the antisymmetric part $\alpha_\epsilon(t)=(n_\epsilon-n_{-\epsilon})\mathrm{sgn}\epsilon$ of the distribution function, as follows:

$$\begin{aligned}
g_\epsilon&=\{\beta_\epsilon g_\epsilon^A-g_\epsilon^R\beta_\epsilon\}-\alpha_\epsilon+\{g_\epsilon^R\alpha_\epsilon g_\epsilon^A\}-\{f_\epsilon^R\alpha_\epsilon f_\epsilon^{\dagger A}\},\\
f_\epsilon&=\{\beta_\epsilon f_\epsilon^A-f_\epsilon^R\beta_\epsilon\}+\{g_\epsilon^R\alpha_\epsilon f_\epsilon^A+f_\epsilon^R\alpha_\epsilon\overline{g}_\epsilon^A\}.
\end{aligned}\tag{24}$$

The Green's functions (24) automatically satisfy the normalization condition (23) for arbitrary α_ϵ and β_ϵ. Following this method of determining the distribution functions, we arrive at a system of dynamical kinetic equations. In the zeroth approximation in the frequency,

$$\begin{aligned}
g_\epsilon&=-(g_\epsilon^R-g_\epsilon^A)\beta_\epsilon-\alpha_\epsilon-(g_\epsilon^Rg_\epsilon^A-f_\epsilon^Rf_\epsilon^{\dagger A})\alpha_\epsilon\\
&=-2u_\epsilon\beta_\epsilon-2h_\epsilon\alpha_\epsilon,
\end{aligned}\tag{25}$$

$$
\begin{aligned}
f_\epsilon &= -(f^R_\epsilon - f^A_\epsilon)\beta_\epsilon + (g^R_\epsilon f^A_\epsilon + f^R_\epsilon \overline{g^A_\epsilon})\alpha_\epsilon \\
&= -2v_\epsilon\beta_\epsilon - 2ib_\epsilon\alpha_\epsilon,
\end{aligned}
$$

these expressions are sufficient for obtaining the kinetic equations to first order in frequency. Before we deal with the equations for the distributions functions, we must express exactly the term $\Delta(f^\dagger_\epsilon - f_\epsilon)$ (the first term in the frequency expansion of $\{\Delta f^\dagger - f\Delta\}_\epsilon$) in Eq. (10) via Eq. (11). As a result, Eq. (10) yields

$$
\begin{aligned}
\frac{\partial g_\epsilon}{\partial t} &= -ie\{\Phi g - g\Phi\}_\epsilon - i\{\Delta f^\dagger - f\Delta\}'_\epsilon + i\left[\frac{e\Delta}{2\epsilon}\{\Phi f - f\Phi\}_\epsilon\right. \\
&\quad \left. - \frac{\Delta}{2\epsilon}\{\Delta\overline{g} - g\Delta\}_\epsilon - \text{c.c.}\right] - \left[I_1^{\text{ph}} - \frac{\Delta}{2\epsilon}(I_2^{\text{ph}} + I_2^{*\text{ph}})\right] \\
&\quad - \left[I_1^{\text{tun}} - \frac{\Delta}{2\epsilon}(I_2^{\text{tun}} + I_2^{*\text{tun}})\right], \qquad (26)
\end{aligned}
$$

where $\{\Delta f^\dagger - f\Delta\}'_\epsilon = \{\Delta f^\dagger - f\Delta\}_\epsilon - \Delta(f^\dagger_\epsilon - f_\epsilon)$. If in this equation we use the expansion (14) and the expressions (25) and separate the even and odd parts, we arrive at a system of kinetic equations,

$$
\begin{aligned}
h_\epsilon\frac{d\alpha_\epsilon}{dt} &= -e\left[\frac{d(u_\epsilon\beta_\epsilon)}{d\epsilon} - \frac{\Delta}{\epsilon}\,\frac{d(v_\epsilon\beta_\epsilon)}{d\epsilon}\right]\frac{d\Phi}{dt} - \left[\alpha_\epsilon\frac{dh_\epsilon}{d\Delta}\right. \\
&\quad \left. + \frac{2\Delta}{\epsilon}\,\frac{d(h_\epsilon\alpha_\epsilon)}{d\epsilon}\right]\frac{d\Delta}{dt} + \nu Q_\alpha + I_\alpha, \qquad (27)
\end{aligned}
$$

$$
\begin{aligned}
u_\epsilon\frac{d\beta_\epsilon}{dt} &= -e\,\frac{d(h_\epsilon\alpha_\epsilon)}{d\epsilon}\,\frac{d\Phi}{dt} - \left[\beta_\epsilon\frac{du_\epsilon}{d\Delta} + \frac{d(v_\epsilon\beta_\epsilon)}{d\epsilon}\right]\frac{d\Delta}{dt} \\
&\quad + \nu Q_\beta + I_\beta, \qquad (28)
\end{aligned}
$$

where I_α and I_β are the phonon collision integrals, which are not written explicitly, and the tunneling sources Q_α and Q_β have been obtained in the low-frequency approximation and are given by the formulas (here $v = (\text{sgn}e/2)\,d\varphi/dt$)

$$
\begin{aligned}
Q_\alpha &= Q_{1\alpha} + Q_{2\alpha}\,\text{sgn}\,e\,\sin\varphi + Q_{3\alpha}\cos\varphi, \\
Q_{1\alpha} &= \left(u_\epsilon - \frac{\Delta}{\epsilon}v_\epsilon\right)[u'_{\epsilon-v}(\beta'_{\epsilon-v} - \beta_\epsilon) + u'_{\epsilon+v}(\beta_\epsilon - \beta'_{\epsilon+v})] \\
&\quad + \left(u_\epsilon - \frac{\Delta}{\epsilon}v_\epsilon\right)(h'_{\epsilon-v}\alpha'_{\epsilon-v} + h'_{\epsilon+v}\alpha'_{\epsilon+v}) - (u'_{\epsilon-v} \\
&\quad + u'_{\epsilon+v})h_\epsilon\alpha_\epsilon + \frac{\Delta}{\epsilon}(y'_{\epsilon-v} + y'_{\epsilon+v})b_\epsilon\alpha_\epsilon, \qquad (29) \\
Q_{2\alpha} &= -\left(v_\epsilon - \frac{\Delta}{\epsilon}u_\epsilon\right)(w'_{\epsilon+v} + w'_{\epsilon-v})\beta_\epsilon - \left(w_\epsilon - \frac{\Delta}{\epsilon}y_\epsilon\right) \\
&\quad \times(v'_{\epsilon+v}\beta'_{\epsilon+v} + v'_{\epsilon-v}\beta'_{\epsilon-v}) - \frac{\Delta}{\epsilon}(w'_{\epsilon+v} - w'_{\epsilon-v})h_\epsilon\alpha_\epsilon \\
&\quad - (v'_{\epsilon+v} - v'_{\epsilon-v})b_\epsilon\alpha_\epsilon - \left(v_\epsilon - \frac{\Delta}{\epsilon}u_\epsilon\right)(b'_{\epsilon+v}\alpha'_{\epsilon+v} \\
&\quad - b'_{\epsilon-v}\alpha'_{\epsilon-v}), \\
Q_{3\alpha} &= \left(v_\epsilon - \frac{\Delta}{\epsilon}u_\epsilon\right)[(v'_{\epsilon+v} - v'_{\epsilon-v})\beta_\epsilon + v'_{\epsilon-v}\beta'_{\epsilon-v} \\
&\quad - v'_{\epsilon+v}\beta'_{\epsilon+v}] + \frac{\Delta}{\epsilon}(v'_{\epsilon+v} + v'_{\epsilon-v})h_\epsilon\alpha_\epsilon - (w'_{\epsilon+v} \\
&\quad + w'_{\epsilon-v})b_\epsilon\alpha_\epsilon + \left(w_\epsilon - \frac{\Delta}{\epsilon}y_\epsilon\right)(b'_{\epsilon+v}\alpha'_{\epsilon+v} + b'_{\epsilon-v}\alpha'_{\epsilon-v}),
\end{aligned}
$$

$$
\begin{aligned}
Q_\beta &= Q_{1\beta} + Q_{2\beta}\,\text{sgn}\,e\,\sin\varphi + Q_{3\beta}\cos\varphi, \\
Q_{1\beta} &= u_\epsilon u'_{\epsilon-v}(\beta'_{\epsilon-v} - \beta_\epsilon) + u_\epsilon u'_{\epsilon+v}(\beta'_{\epsilon+v} - \beta_\epsilon) \\
&\quad + u_\epsilon(h'_{\epsilon-v}\alpha'_{\epsilon-v} - h'_{\epsilon+v}\alpha'_{\epsilon+v}) + (u'_{\epsilon+v} - u'_{\epsilon-v})h_\epsilon\alpha_\epsilon, \qquad (30) \\
Q_{2\beta} &= v_\epsilon(w'_{\epsilon-v} - w'_{\epsilon+v})\beta_\epsilon + w_\epsilon(v'_{\epsilon-v}\beta'_{\epsilon-v} - v'_{\epsilon+v}\beta'_{\epsilon+v}) \\
&\quad - (v'_{\epsilon+v} + v'_{\epsilon-v})b_\epsilon\alpha_\epsilon - v_\epsilon(b'_{\epsilon+v}\alpha'_{\epsilon+v} + b'_{\epsilon-v}\alpha'_{\epsilon-v}), \\
Q_{3\beta} &= v_\epsilon v'_{\epsilon+v}(\beta_\epsilon - \beta'_{\epsilon+v}) + v_\epsilon v'_{\epsilon-v}(\beta_\epsilon - \beta'_{\epsilon-v}) \\
&\quad - (w'_{\epsilon+v} - w'_{\epsilon-v})b_\epsilon\alpha_\epsilon + w_\epsilon(b'_{\epsilon+v}\alpha'_{\epsilon+v} - b'_{\epsilon-v}\alpha'_{\epsilon-v}).
\end{aligned}
$$

The self-consistency equations become

$$
\Delta = -\lambda\int_0^{\theta_D} v_\epsilon b_\epsilon\,d\epsilon, \qquad (31)
$$

$$
\begin{aligned}
\rho &= -2eN(0)\left[e\Phi(t) - \int_0^\infty h_\epsilon\alpha_\epsilon\,d\epsilon\right] \\
&= 2e^2N(0)[\Phi - \Psi] = -\frac{1}{4\pi r_d^2}(\Phi - \Psi), \qquad (32)
\end{aligned}
$$

with the electron–hole imbalance determined by the formula

$$
\Psi = \frac{1}{e}\int_0^\infty h_\epsilon\alpha_\epsilon\,d\epsilon. \qquad (33)
$$

Finally, for the current flowing between the layers we have

$$
\begin{aligned}
J &= J_1 + J_2\sin\varphi + J_3\cos\varphi, \\
J_1 &= 2eN(0)d_0\nu\int_{-\infty}^{\infty} d\epsilon\,[u_\epsilon u'_{\epsilon+v}(\beta_\epsilon - \beta'_{\epsilon+v}) \\
&\quad + u_\epsilon h'_{\epsilon+v}\alpha'_{\epsilon+v} - u'_{\epsilon+v}h_\epsilon\alpha_\epsilon], \qquad (34) \\
J_2 &= 2|e|N(0)d_0\nu\int_{-\infty}^{\infty} d\epsilon\,[-v_\epsilon w'_{\epsilon+v}\beta_\epsilon - w_\epsilon v'_{\epsilon+v}\beta'_{\epsilon+v} \\
&\quad - v'_{\epsilon+v}b_\epsilon\alpha_\epsilon - v_\epsilon b'_{\epsilon+v}\alpha'_{\epsilon+v}], \\
J_3 &= 2eN(0)d_0\nu\int_{-\infty}^{\infty} d\epsilon[v_\epsilon v'_{\epsilon+v}\beta_\epsilon - v_\epsilon v'_{\epsilon+v}\beta'_{\epsilon+v} \\
&\quad - w'_{\epsilon+v}b_\epsilon\alpha_\epsilon + w_\epsilon b'_{\epsilon+v}\alpha'_{\epsilon+v}].
\end{aligned}
$$

In the limit $\gamma/\Delta \ll 1$ (a good gap) and $\omega \to 0$ the kinetic equations derived in this section become the equations obtained in Refs. 46–48, and Eqs. (31) and (32) become (1) and (5).

3.3. Quasihydrodynamic approximation: the dynamics of the electron-hole imbalance

We examine the approximation in which the departure of the system from equilibrium is small and the kinetic equations can be linearized. Expanding in power series in α_η, $\beta_\epsilon^{(1)}$, v/Δ, and v/T and keeping only the leading terms, we obtain

$$h_\epsilon \frac{d\alpha_\epsilon}{dt} = -e\left[\frac{d(u_\epsilon \beta_\epsilon^{(0)})}{d\epsilon} - \frac{\Delta}{\epsilon}\frac{d(v_\epsilon \beta_\epsilon^{(0)})}{d\epsilon}\right]\frac{d\Phi}{dt} + \nu Q_\alpha^{(0)} + \nu Q_\alpha^{(1)} + I_\alpha^{(1)}, \tag{35}$$

$$u_\epsilon \frac{d\beta_\epsilon^{(1)}}{dt} = -\left[\frac{d(v_\epsilon \beta_\epsilon^{(0)})}{d\epsilon} + \beta_\epsilon^{(0)}\frac{du_\epsilon}{d\Delta}\right]\frac{d\Delta}{dt} + \nu Q_\beta^{(0)} + \nu Q_\beta^{(1)} + I_\beta^{(1)}, \tag{36}$$

$$Q_\alpha^{(0)} = \frac{u'_\epsilon(u_\epsilon - (\Delta/\epsilon)v_\epsilon)}{T\cosh^2(\epsilon/2T)}v + 2\left[\left(v_\epsilon - \frac{\Delta}{\epsilon}v_\epsilon\right)w'_\epsilon + \left(w_\epsilon - \frac{\Delta}{\epsilon}y_\epsilon\right)v'_\epsilon\right]\tanh\frac{\epsilon}{2T}\,\mathrm{sgn}\,e\sin\varphi + \frac{v'_\epsilon(v_\epsilon - (\Delta/\epsilon)u_\epsilon)}{T\cosh^2(\epsilon/2T)}v\cos\varphi, \tag{37}$$

$$Q_\alpha^{(1)} = 2\left(u_\epsilon - \frac{\Delta}{\epsilon}v_\epsilon\right)h'_\epsilon\alpha'_\epsilon - 2u'_\epsilon h_\epsilon\alpha_\epsilon + \frac{2\Delta}{\epsilon}y'_\epsilon b_\epsilon\alpha_\epsilon + \left[\frac{2\Delta}{\epsilon}v'_\epsilon h_\epsilon\alpha_\epsilon - 2w'_\epsilon b_\epsilon\alpha_\epsilon + 2\left(w_\epsilon - \frac{\Delta}{\epsilon}y_\epsilon\right)b'_\epsilon\alpha'_\epsilon\right]\cos\varphi, \tag{38}$$

$$Q_\beta^{(0)} = \left[2v_\epsilon\frac{dw'_\epsilon}{d\epsilon}\tanh\frac{\epsilon}{2T} + 2w_\epsilon\frac{d(v'_\epsilon\tanh(\epsilon/2T))}{d\epsilon}\right]\times v\,\mathrm{sgn}\,e\sin\varphi, \tag{39}$$

$$Q_\beta^{(1)} = 2u_\epsilon u'_\epsilon(\beta_\epsilon'^{(1)} - \beta_\epsilon^{(1)}) + [-2v'_\epsilon b_\epsilon\alpha_\epsilon - 2v_\epsilon b'_\epsilon\alpha'_\epsilon]\times\mathrm{sgn}\,e\sin\varphi + [2v_\epsilon v'_\epsilon(\beta_\epsilon^{(1)} - \beta_\epsilon'^{(1)})]\cos\varphi, \tag{40}$$

$$J_1 = \frac{1}{2eR_N}\int_{-\infty}^{\infty}d\epsilon\left[\frac{u_\epsilon u'_\epsilon}{2T\cosh^2(\epsilon/2T)}v + u_\epsilon h'_\epsilon\alpha'_\epsilon - u'_\epsilon h_\epsilon\alpha_\epsilon\right],$$

$$J_2 = \frac{1}{2|e|R_N}\int_{-\infty}^{\infty}d\epsilon\left[(v_\epsilon w'_\epsilon + w_\epsilon v'_\epsilon)\tanh\frac{\epsilon}{2T} - v'_\epsilon b_\epsilon\alpha_\epsilon - v_\epsilon b'_\epsilon\alpha'_\epsilon\right],$$

$$J_3 = \frac{1}{2eR_N}\int_{-\infty}^{\infty}d\epsilon\left[\frac{v_\epsilon v'_\epsilon}{2T\cosh^2(\epsilon/2T)}v + w_\epsilon b'_\epsilon\alpha'_\epsilon - w'_\epsilon b_\epsilon\alpha_\epsilon\right]. \tag{41}$$

Equation (35) for $\alpha_\epsilon(t)$ depends on $\beta_\epsilon^{(1)}$ only through $\Delta(t)$ and can be solved independently of (36). Moreover, we limit ourselves to the case where all variations in the energy gap in the equation for $\alpha_\epsilon(t)$ can be ignored and study the dynamics of the electron-hole imbalance.

The solution of the kinetic equation (35) in the general case is extremely cumbersome due to the dependence of the coefficients on ϵ. However, for $\Delta/\gamma \ll 1$ (gapless superconductivity), this dependence is unimportant, while for $\Delta/T \ll 1$ and $\gamma/T \ll 1$ ($T \simeq T_c$) the particles that play the main role are those with energies $\epsilon \sim T \gg \Delta, \gamma$, and the energy dependence is also unimportant; in such cases the tunneling source in the leading approximation is

$$Q_\alpha^{(0)} + Q_\alpha^{(1)} = \frac{v}{T\cosh^2(\epsilon/2T)} + 2(\alpha'_\epsilon - \alpha_\epsilon). \tag{42}$$

The form of the tunneling source and of the factor of $d\Phi/dt$ suggests that in the zeroth approximation in Δ/T and γ/T or Δ/γ, the expression

$$\alpha_\epsilon^{(i)}(t) = \frac{e\Psi_i(t)}{2T\cosh^2(\epsilon/2T)}, \tag{43}$$

is valid in all layers and automatically satisfies the definition (33) of Ψ (here $h_\epsilon \simeq 1$). In this case a closed equation for $\Psi(t)$ can be rigorously derived and a transition to the quasihydrodynamic description can be achieved. Integrating (35) and (41) with respect to ϵ, in the leading approximation we have

$$\frac{d\Psi_i}{dt} = \frac{d\Phi_i}{dt} + 2\nu\frac{\hbar}{2|e|}\left(\frac{d\varphi_{i-1,i}}{dt} - \frac{d\varphi_{i,i+1}}{dt}\right) + 2\nu(\Psi_{i-1} + \Psi_{i+1} - 2\Psi_i) - \tau_q^{-1}\Psi_i, \tag{44}$$

$$J_{i-1,i} = \frac{\hbar}{2|e|R_N}\frac{d\varphi_{i-1,i}}{dt} + \frac{\Psi_{i-1} - \Psi_i}{R_N} + J_c\sin\varphi_{i-1,i}. \tag{45}$$

In Eq. (44) we have added the second (''right'') source, introduced the layer indices, restored $\hbar$, and written the collision integral in the τ-approximation. For simplicity we have assumed that the parameters of all the layers and junctions are the same.

Combining Eqs. (44) and (45) and allowing for the expression (32) and the continuity equation, we arrive at the equation

$$\Psi_i = -2\tau_q \nu R_N J_c(\sin\varphi_{i-1,i} - \sin\varphi_{i,i+1}), \tag{46}$$

which corresponds to the imaginary part of the time-dependent Ginzburg–Landau equation. The term with $d\Psi/dt$ drops out of this equation because the factor of $d\Psi/dt$ is small in the parameters Δ/γ or Δ/T and γ/T. However, at high frequencies this term may, on the whole, prove to be of the same order as the other terms and must be retained. The final equation is

$$\tau_q\Gamma\frac{d\Psi_i}{dt} + \Psi_i = -2\tau_q\nu R_N J_c(\sin\varphi_{i-1,1} - \sin\varphi_{i,i+1}), \tag{47}$$

where

$$\Gamma=\int_0^\infty \frac{\Delta\nu_\epsilon \tanh(\epsilon/2T)}{\epsilon^2}\,d\epsilon. \tag{48}$$

Equations (47) and (45) contain no interference terms; these are known to be small in gapless superconductors. Allowing for them (for arbitrary values of the parameters) gives rise to more general equations, which are given in the Appendix.

4. DYNAMICS OF A CHAIN OF JUNCTIONS

A. 4.1. A system of equations for a chain of junctions

Using the equations of the dynamics of the electron–hole imbalance and the expressions for the current (all derived in Sec. 3), we can introduce a complete system of dynamical equations for a chain of junctions. Following Refs. 28,29,32, and 50–52, we use the continuity equation

$$d_0\frac{d\rho_i}{dt}=J_{i-1,i}-J_{i,i+1} \tag{49}$$

and the equation that relates the voltage (the electric field) across the layers to the layer charge,

$$V_{i,i+1}-V_{i-1,i}=\frac{4\pi d_0 d}{\epsilon_0}\rho_i, \tag{50}$$

where d is the distance between the layers, and ϵ_0 is the dielectric constant. These equations imply that

$$J_{i-1,i}+\frac{\epsilon_0}{4\pi d}\frac{dV_{i-1,i}}{dt}=J(t), \tag{51}$$

where $J(t)$ is the external current. Together with the Josephson nonequilibrium equation (6), the expressions for the charge [Eq. (32)] and current [Eq. (45)] [or (A2)], and the imbalance dynamical equation (47) [or (A1)], these equations constitute the complete system of dynamical equations for a chain of junctions.

Using Eqs. (45) and (47), in the dimensionless form we obtain

$$\beta\frac{d^2\varphi_{i,i+1}}{d\tau^2}+\frac{d\varphi_{i,i+1}}{d\tau}+\sin\varphi_{i,i+1}+\psi_i-\psi_{i+1} +\beta\left(\frac{d\mu_i}{d\tau}-\frac{d\mu_{i+1}}{d\tau}\right)=j(t), \tag{52}$$

$$\alpha\Gamma\frac{d\psi_i}{d\tau}+\psi_i=-\eta(\sin\varphi_{i-1,i}-\sin\varphi_{i,i+1}), \tag{53}$$

$$\mu_i+\zeta(2\mu_i-\mu_{i-1}-\mu_{i+1})=\psi_i+\zeta\left(\frac{d\varphi_{i-1,i}}{d\tau}-\frac{d\varphi_{i,i+1}}{d\tau}\right), \tag{54}$$

$$\sum_i\frac{d\varphi_{i-1,i}}{d\tau}=v(t), \tag{55}$$

where $j(t)$ is the external current in units of $J_c=(f_4T)/(eR_N)$, $v(t)$ is the external voltage in units of $V_c=R_NJ_c$, $\mu(t)=\Phi(t)/V_c$, $\psi(t)=\Psi(t)/V_c$, $\alpha=\tau_q\omega_c$, $\beta=\omega_c^2/\omega_p^2$, $\zeta=(\epsilon_0 r_d^2)/(d_0 d)$, $\omega_c=(2eR_NJ_c)/\hbar$, $\omega_p^2=(8\pi edJ_c)/(\hbar\epsilon_0 S)$, $\tau=\omega_c t$, and $\eta=2\nu\tau_q$. Note that there is a relationship between the coefficients, $\alpha\zeta=\beta\eta$, which follows from their definitions.

In the more general case, using Eqs. (A2) and (A1), we get

$$\beta\frac{d^2\varphi_{i,i+1}}{d\tau^2}+\frac{d\varphi_{i,i+1}}{d\tau}+\sin\varphi_{i,i+1}+\frac{c}{f_3}(\psi_i-\psi_{i+1}) -\frac{c_4}{f_3}(\psi_i+\psi_{i+1})\sin\varphi_{i,i+1}+\left[\frac{f_5}{f_3}\frac{d\varphi_{i,i+1}}{d\tau} +\frac{c_1}{f_3}(\psi_i-\psi_{i+1})\right]\cos\varphi_{i,i+1}+\beta\left(\frac{d\mu_i}{d\tau}-\frac{d\mu_{i+1}}{d\tau}\right)=j(t), \tag{56}$$

$$\begin{aligned}\alpha\frac{d\psi_i}{d\tau}&+\psi_i+\eta c(2\psi_i-\psi_{i-1}-\psi_{i+1})-\eta c_1[(\psi_{i-1}-\psi_i)\\ &\times\cos\varphi_{i-1,i}+(\psi_{i+1}-\psi_i)\cos\varphi_{i,i+1}]-\eta(c_2\cos\varphi_{i-1,i}\\ &+c_2\cos\varphi_{i,i+1}+2c_3)\psi_i+\eta(c_2+c_3\cos\varphi_{i-1,i})\psi_{i-1,i}\\ &+\eta(c_2+c_3\cos\varphi_{i,i+1})\psi_{i,i+1}=\eta f\left(\frac{d\varphi_{i-1,i}}{d\tau}-\frac{d\varphi_{i,i+1}}{d\tau}\right)\\ &+\alpha a\frac{d\mu_i}{d\tau}-\eta\frac{f_1f_3}{f_4}(\sin\varphi_{i-1,i}-\sin\varphi_{i,i+1})\\ &-\eta f_2\left(\frac{d\varphi_{i-1,i}}{d\tau}\cos\varphi_{i-1,i}-\frac{d\varphi_{i,i+1}}{d\tau}\cos\varphi_{i,i+1}\right),\end{aligned} \tag{57}$$

where now $V_c=R_NJ_c/f_3$ and $\omega_c=(2eR_NJ_c)/(f_3\hbar)$. These equations can be assumed to be phenomenological equations within a broad temperature interval.

4.2. Behavior of the current-voltage characteristics

Using the equations we have derived in the previous sections, we can analyze the dynamics of a chain of junctions for the values of the nonequilibrium parameter η. We limit ourselves to the case $\beta\gg 1$, in which the current-voltage characteristic of separate Josephson junction exhibits hysteresis. We discuss in detail the situation within the simple model specified by (52) and (53), when the imbalance dynamics is controlled by two parameters, η and $\alpha\Gamma$. The parameter ζ is assumed to be of order unity. Here are the main limiting cases.

In the limit $\eta\ll 1$, the current-voltage characteristics in Fig. 2a generally resemble those of a system of noninteracting junctions. There are separate branches corresponding to different numbers of junctions in the resistive state,[4)] and the size of the hysteresis is determined by the parameter β. However, there are qualitative differences, too.

First, the observed values of critical currents at which there is a transition to the subsequent curves are determined not by the initial spread of critical currents (which was fixed in the calculation to within 1%) but by the nonequilibrium effect of step pulling, which amounts to a situation in which in the course of the transition of one of the junctions into the resistive state there arises an electron-hole imbalance in the

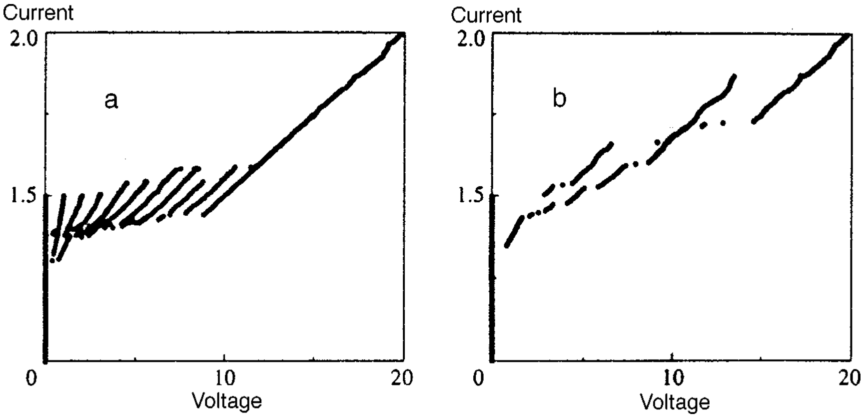

FIG. 2. Current–voltage characteristics of a chain of ten junctions with $\beta=10$ and $\zeta=1$ and small values of the nonequilibrium parameters: (a) $\eta=0.1$ and $\alpha\Gamma=0.01$, and (b) $\eta=0.5$ and $\alpha\Gamma=0.1$.

neighboring superconducting layers and an additional normal current, which develops because of the inhomogeneity of the imbalance, flows through the neighboring junctions. As a result, the supercurrent through these junctions decreases and a stronger external current is required to transform them into the resistive state. Thus, the spread of critical current proves to be of order ηJ_c, which is illustrated in Figs. 2a ($\eta=0.1$) and 2b ($\eta=0.5$).

Second, there is the ''excess current'' phenomenon. For instance, the curve corresponding to one junction being in the resistive state is an almost straight line (provided that the currents are large), as it is in the case of independent junctions, but the extrapolation of this straight line to $V=0$ yields $J_0\neq 0$. This well-known effect is the result of the nonequilibrium state, and for the last branch in Fig. 2a, corresponding to the case where all junctions are in the resistive state and the imbalance is zero, the excess current is also zero. Note that in their recent work on the intrinsic Josephson effect in high-T_c superconductors, Yan *et al.*[25] established the presence of excess current experimentally. As η increases, the step-entraining effect becomes more pronounced and the shape of the branches changes (Fig. 2b).

For moderate values of the nonequilibrium parameter η, i.e., $\eta\sim 1$ (Fig. 3), the shape of the branches changes substantially and the general form of the current-voltage characteristics depends on the value of the parameter $\alpha\Gamma$. A characteristic ''ripple'' appears, which is due to the dependence of the synchronization phases on the current (see below).

In the limit $\eta\gg 1$ the situation is determined by the second parameter, $\alpha\Gamma$. For $\alpha\Gamma\sim\eta$ or $\alpha\Gamma\gg\eta$ the giant-hysteresis regime is realized (Fig. 4a), in which $V\simeq NJ$ holds for $J>J_c$, where N is the number of junctions, irrespective of the number of junctions in the resistive state. The branches of the current-voltage characteristic corresponding to different states are very close to each other. But if $\alpha\Gamma\ll\eta$ holds in the dynamic state the imbalance oscillations are strong and the regime of chaotic oscillations is realized. The branches of the current-voltage characteristic in this regime disappear and the hysteresis effect is small (Fig. 4b).

4.3. High-frequency interaction of junctions

We will conclude this section by a discussion of the high-frequency interaction of junctions. This problem plays an important role in the interpretation of the microwave experiment in the intrinsic Josephson effect in high-T_c superconductors and from the practical viewpoint. Of special interest here is the regime of in-phase synchronization of a chain of junctions, in which the optimum conditions for generation and reception of microwaves are achieved. Our research has shown that in both weak ($\eta<1$) and strong ($\eta>1$) interaction there can be in-phase regimes, more complicated synchronous regimes with phase shifts (Figs. 5 and 6), inhomogeneous regimes with different control frequencies of the junctions, and chaotic regimes. When the interaction is weak and some of the junctions are in the resistive

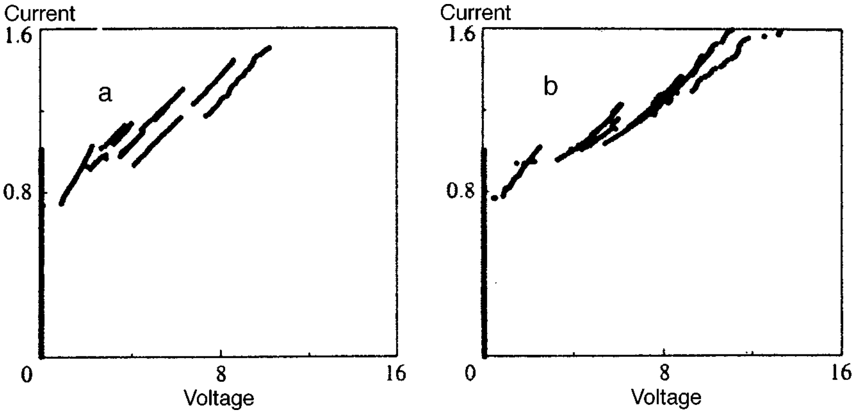

FIG. 3. Current–voltage characteristics of a chain of ten junctions with $\beta=10$ and $\zeta=1$ and intermediate values of the nonequilibrium parameters: (a) $\eta=1$ and $\alpha\Gamma=0.1$, and (b) $\eta=1$ and $\alpha\Gamma=1$.

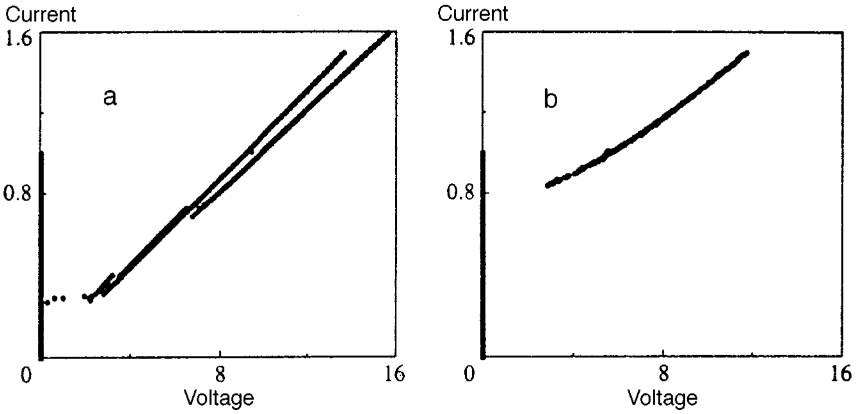

FIG. 4. Current–voltage characteristics of a chain of ten junctions with $\beta=10$ and $\zeta=1$ and large values of the nonequilibrium parameters: (a) $\eta=100$ and $\alpha\Gamma=100$, and (b) $\eta=100$ and $\alpha\Gamma=1$.

state, synchronization of a group of junctions is observed. Numerical calculations show that the type of synchronization depends on the parameters and the external current. For instance, for small values of η within the scope of the simplified system of equations (52)–(55), in-phase synchronization is observed (Fig. 5a), but allowance for the term with f_2, which enters into the more general equations (56) and (57), leads to a more complicated regime with phase shift (Fig. 5b). Interestingly, there are regimes where in-phase synchronization is observed only at some values of the current, while in the intervals between these values there is a shift in phase (Fig. 6). This explains the presence of ripples in the current-voltage characteristics (Fig. 3b).

5. CONCLUSION

In this paper we have systematically studied the theory of the nonequilibrium Josephson effect, from the microscopic equations for the Green′s functions to the macroscopic dynamical equations in which the electron-hole imbalance is taken into account in the quasihydrodynamic approximation. The calculated current-voltage characteristics demonstrate the nature of the effect of the electron-hole imbalance on the dynamics of one-dimensional chains of junctions. It is believed that the result will prove important in studies of submicrometer Josephson structures and the intrinsic Josephson effect in high-T_c superconductors. In a recent paper, Yurgens *et al.*[53] reported on their experimental studies of the current-voltage characteristics of a chain of intrinsic junctions in $Bi_2Sr_{1.5}La_{0.5}CuO_{6+\delta}$ (B-2201), which the superconducting ‘‘electrode″ is formed not by two or three layers of CuO_2, as it is in compounds of Bi and Tl studied earlier, but by only one layer. Because of this, probably, the nonequilibrium effects are much stronger and the current-voltage characteristics presented in Fig. 1b of Ref. 53 are similar to those depicted in Fig. 3a of the present work. A detailed study of the nonequilibrium intrinsic Josephson effect merits a separate publication.

The author is grateful to A. A. Andronov, V. V. Kurin, and A. S. Mel′nikov for their support and useful discussions. The work was partially sponsored by the Russian Fund for Fundamental Research (Projects 97-02-16928 and 99-02-16188), the Leading Scientific Schools Program (Project 96-15-96591), INTAS (Grant No. 96-0457), and the ICFPM Research Program.

APPENDIX

We can approximately derive the equations with interference terms by assuming that the distribution function $\alpha_\epsilon(t)$ is proportional to $\Psi(t)$, integrating (35) with respect to ϵ, and retaining all the terms:

$$\frac{d\Psi_i}{dt}=a\frac{d\Phi_i}{dt}+2\nu f\frac{\hbar}{2|e|}\left(\frac{d\varphi_{i-1,i}}{dt}-\frac{d\varphi_{i,i+1}}{dt}\right)-2\nu f_1\frac{T}{|e|}$$
$$\times(\sin\varphi_{i-1,i}-\sin\varphi_{i,i+1})-2\nu f_2\frac{\hbar}{2|e|}$$
$$\times\left(\frac{d\varphi_{i-1,i}}{dt}\cos\varphi_{i-1,i}-\frac{d\varphi_{i,i+1}}{dt}\cos\varphi_{i,i+1}\right)$$
$$+2\nu c(\Psi_{i-1}+\Psi_{i+1}-2\Psi_i)+2\nu c_1[(\Psi_{i-1}-\Psi_i)$$

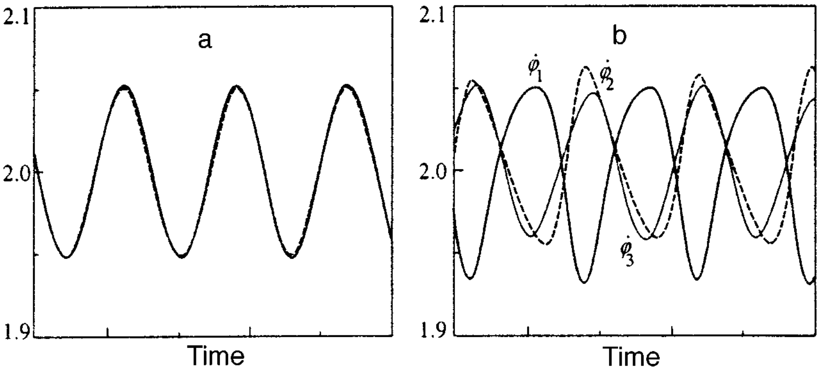

FIG. 5. Temporal dynamics of Josephson phases ($\dot\varphi_1$, $\dot\varphi_2$, and $\dot\varphi_3$ are depicted) with $\beta=10$, $\alpha\Gamma=0.1$, $\eta=0.5$, $\zeta=1$, and $j=2$: (a) without interference terms (the in-phase regime), and (b) with the ‘‘$f2$’’ interference term with $f_2=0.1$ (the regime with phase shift).

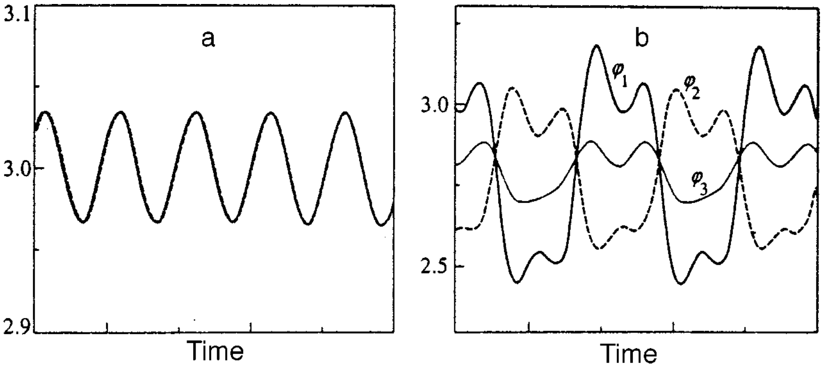

FIG. 6. Temporal dynamics of Josephson phases ($\dot{\varphi}_1$, $\dot{\varphi}_2$, and $\dot{\varphi}_3$ are depicted) with $\beta=10$, $\alpha\Gamma=1$, $\eta=1$, and $\zeta=1$: (a) at $j=3$ (the in-phaseregime), and (b) at $j=2.8$ (the regime with phase shift). The phase shift depends on the current.

$$\begin{aligned}&\times\cos\varphi_{i-1,i}+(\Psi_{i+1}-\Psi_i)\cos\varphi_{i,i+1}]\\ &+2\nu(c_2\cos\varphi_{i-1,i}+c_2\cos\varphi_{i,i+1}+2c_3)\Psi_i\\ &-2\nu(c_2+c_3\cos\varphi_{i-1,i})\Psi_{i-1,i}\\ &-2\nu(c_2+c_3\cos\varphi_{i,i+1})\Psi_{i,i+1}-\tau_q^{-1}\Psi_i. \qquad \text{(A1)}\end{aligned}$$

The expression for the current is

$$\begin{aligned}J_{i-1,i}=f_3\,&\frac{\hbar}{2|e|R_N}\frac{d\varphi_{i-1,i}}{dt}+c\,\frac{(\Psi_{i-1}-\Psi_i)}{R_N}\\ &+\frac{T}{|e|R_N}f_4\sin\varphi_{i-1,i}+\Bigg[f_5\,\frac{\hbar}{2|e|R_N}\frac{d\varphi_{i-1,i}}{dt}\\ &+c_1\frac{\Psi_{i-1}-\Psi_i}{R_N}\Bigg]\cos\varphi_{i-1,i}\,. \qquad \text{(A2)}\end{aligned}$$

Here the coefficients $a(T)$, $f(T)$, $f_1(T)$, $f_2(T)$, $f_3(T)$, $f_4(T)$, and $f_5(T)$ can be calculated exactly for all temperatures:

$$a(T)=\int_0^\infty\left[\frac{d(u_\epsilon\tanh(\epsilon/2T))}{d\epsilon}-\frac{\Delta}{\epsilon}\frac{d(v_\epsilon\tanh(\epsilon/2T))}{d\epsilon}\right]d\epsilon=1-\Gamma, \qquad \text{(A3)}$$

$$f(T)=\int_0^\infty\frac{u'_\epsilon(u_\epsilon-(\Delta/\epsilon)v_\epsilon)}{2T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A4)}$$

$$f_1(T)=-\frac{1}{T}\int_0^\infty\left[\left(v_\epsilon-\frac{\Delta}{\epsilon}u_\epsilon\right)w'_\epsilon+\left(w_\epsilon-\frac{\Delta}{\epsilon}y_\epsilon\right)v'_\epsilon\right]\tanh\frac{\epsilon}{2T}d\epsilon, \qquad \text{(A5)}$$

$$f_2(T)=-\int_0^\infty\frac{v'_\epsilon(v_\epsilon-(\Delta/\epsilon)u_\epsilon)}{2T\cosh^2E(\epsilon/2T)}d\epsilon, \qquad \text{(A6)}$$

$$f_3(T)=\int_0^\infty\frac{u_\epsilon u'_\epsilon}{2T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A7)}$$

$$f_4(T)=\frac{1}{T}\int_0^\infty(v_\epsilon w'_\epsilon+w_\epsilon v'_\epsilon)\tanh\frac{\epsilon}{2T}d\epsilon, \qquad \text{(A8)}$$

$$f_5(T)=\int_0^\infty\frac{v_\epsilon v'_\epsilon}{2T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A9)}$$

and the coefficients $c(T)$, $c_1(T)$, $c_2(T)$, and $c_3(T)$ depend on the shape of the distribution function and, if condition (43) is met, are given by the formulas

$$c(T)=\int_0^\infty\frac{u_\epsilon h_\epsilon}{2T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A10)}$$

$$c_1(T)=\int_0^\infty\frac{w_\epsilon b_\epsilon}{2T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A11)}$$

$$c_2(T)=\int_0^\infty\frac{\Delta v_\epsilon h_\epsilon}{2\epsilon T\cosh^2(\epsilon/2T)}d\epsilon, \qquad \text{(A12)}$$

$$c_3(T)=\int_0^\infty\frac{\Delta y_\epsilon b_\epsilon}{2T\cosh^2(\epsilon/2T)}d\epsilon. \qquad \text{(A13)}$$

These expressions make it possible to determine the corrections to Eqs. (47) and (45). From a more general viewpoint, Eqs. (A1) and (A2) can be thought of as phenomenological equations, since they contain all the main components of the imbalance source and the interlayer current.

*) E-mail: ryn@ipm.sci-nnov.ru

1) It is convenient to define the potential Ψ in this way in order to obtain the dynamical equations.

2) In this section we use a system of units in which $\hbar=1$.

3) The arguments of the functions in the temporal representation are always given explicitly, e.g., $g(t_1,t_2)$, while in the mixed representation (ϵ,t) the argument t may be dropped as, say, in g_ϵ.

4) A resistive state is one in which on the average the Josephson phase difference increases, $\langle d\varphi/dt\rangle\neq 0$.

---

[1] S. N. Artemenko and A. F. Volkov, Usp. Fiz. Nauk **128**, 3 (1979) [Sov. Phys. Usp. **22**, 295 (1979)].

[2] *Nonequilibrium Superconductivity*, D. N. Langenberg and A. I. Larkin (Eds.), North-Holland, Amsterdam (1986).

[3] A. M. Gulyan and G. F. Zharkov, *Superconductors in External Fields* [in Russian], Nauka, Moscow (1990).

[4] D. W. Palmer and J. E. Mercereau, Phys. Lett. A **61**, 135 (1977).

[5] S. N. Artemenko, A. F. Volkov, and A. V. Zaĭtsev, JETP Lett. **27**, 113 (1978).

[6] V. K. Kaplunenko, V. V. Ryazanov, and V. V. Shmidt, Zh. Éksp. Teor. Fiz. **89**, 1389 (1985) [Sov. Phys. JETP **62**, 804 (1985)].

[7] V. K. Kaplunenko, Fiz. Nizk. Temp. **12**, 382 (1986) [Sov. J. Low Temp. Phys. **12**, 217 (1986)].

[8] M. G. Blamier, E. C. G. Kirk, J. E. Evetts, and T. M. Klapwijk, Phys. Rev. Lett. **66**, 220 (1991).

[9] I. P. Nevirkovets, M. G. Blamire, and J. E. Evetts, Fiz. Nizk. Temp. **21**, 1258 (1995) [Low Temp. Phys. **21**, 963 (1995)].

[10] I. P. Nevirkovets, Phys. Rev. B **56**, 832 (1997).

[11] F. V. Komissinskiĭ and G. A. Ovsyannikov, JETP Lett. **61**, 600 (1995).

[12] R. Kleiner, F. Steinmeyer, G. Kinkel, and P. Müller, Phys. Rev. Lett. **68**, 2394 (1992); R. Kleiner and P. Müller, Phys. Rev. B **49**, 1327 (1994).

[13] Yu. I. Latyshev and A. F. Volkov, Physica C **182**, 47 (1991); Yu. I. Latyshev, J. E. Nevelskaya, and P. Monceau, Phys. Rev. Lett. **77**, 932 (1996).

[14] F. X. Régi, J. Schneck, J. F. Palmier, and H. Savary, J. Appl. Phys. **76**, 4426 (1994).

[15] J. Takeya and S. Akita, Physica C **235–240**, 3287 (1995).

[16] A. Yurgens, D. Winkler, Y. M. Zhang, N. Zavaritsky, and T. Claeson, Physica C **235–240**, 3269 (1995).

[17] Ji Ung Lee, J. E. Nordman, and G. Hohenwarter, Appl. Phys. Lett. **67**, 1471 (1995).

[18] M. Veith, T. Eick, W. Brodkorb, M. Manzel, H. Bruchlos, T. Köhler, H.-G. Schmidt, E. Steinbeiss, H.-J. Fuchs, K. Schlenga, G. Hechtfischer, and P. Müller, J. Appl. Phys. **80**, 3396 (1996).

[19] Schlenga, G. Hechtfischer, R. Kleiner, W. Walkenhorst, P. Müller, H. L. Johnson, M. Veith, W. Brodkorb, and E. Steinbeiss, Phys. Rev. Lett. **76**, 4943 (1996).

[20] A. Yurgens, D. Winkler, N. V. Zavaritsky, and T. Claeson, Phys. Rev. B **53**, R8887 (1996) [Phys. Rev. Lett. **79**, 5122 (1997)].

[21] K. Tanabe, Y. Hidaka, S. Karimoto, and M. Suzuki, Phys. Rev. B **53**, 9348 (1996).

[22] M. Rapp, A. Murk, R. Semerad, and W. Prusseit, Phys. Rev. Lett. **77**, 928 (1996).

[23] A. Odagawa, M. Sakai, H. Adachi, K. Setsune, T. Hirao, and K. Yoshida, J. Phys. Soc. Jpn. **36**, L21 (1997).

[24] M. Itoh, S. Karimoto, K. Namekawa, and M. Suzuki, Phys. Rev. B **55**, R12 001 (1997).

[25] S. I. Yan, L. Fang, M. S. Si, and J. Wang, J. Appl. Phys. **82**, 480 (1997).

[26] H. L. Jonson, G. Hechtfischer, G. Götz, R. Kleiner, and P. Müller, J. Appl. Phys. **82**, 756 (1997).

[27] K. Schlenga, R. Kleiner, G. Hechtfischer, M. Mößle, S. Schmitt, P. Müller, Ch. Helm, Ch. Preis, F. Forsthofer, J. Keller, H. L. Johnson, M. Veith, and E. Steinbeiß, Phys. Rev. B **57**, 14 518 (1998).

[28] T. Koyama and M. Tachiki, Phys. Rev. B **54**, 16 183 (1996).

[29] S. N. Artemenko and A. G. Kobelkov, Phys. Rev. Lett. **78**, 3551 (1997).

[30] D. A. Ryndyk, E-prints archive cond-mat/9511001 (1995).

[31] D. A. Ryndyk, JETP Lett. **65**, 791 (1997).

[32] D. A. Ryndyk, Phys. Rev. Lett. **80**, 3376 (1998).

[33] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **61**, 1254 (1971) [Sov. Phys. JETP **34**, 668 (1972)].

[34] M. Tinkham and J. Clarke, Phys. Rev. Lett. **28**, 1366 (1972); M. Tinkham, Phys. Rev. B **6**, 1747 (1972).

[35] A. G. Aronov and V. L. Gurevich, Fiz. Tverd. Tela (St. Petersburg) **16**, 2656 (1974) [Sov. Phys. Solid State **16**, 1722 (1974)]; A. G. Aronov, Yu. M. Gal'perin, V. L. Gurevich, and V. I. Kozub, Adv. Phys. **30**, 539 (1981).

[36] A. F. Volkov and A. V. Zaĭtsev, Zh. Éksp. Teor. Fiz. **69**, 2222 (1975) [Sov. Phys. JETP **42**, 1130 (1975)].

[37] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)].

[38] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **68**, 1915 (1975) [Sov. Phys. JETP **41**, 960 (1975)].

[39] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **73**, 299 (1977) [Sov. Phys. JETP **46**, 155 (1977)].

[40] L. P. Gor'kov and N. B. Kopnin, Zh. Éksp. Teor. Fiz. **64**, 356 (1973) [Sov. Phys. JETP **37**, 183 (1973)]; **65**, 396 (1973) [**38**, 195 (1974)].

[41] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **65**, 1704 (1973) [Sov. Phys. JETP **38**, 854 (1974)].

[42] C.-R. Hu, Phys. Rev. B **21**, 2775 (1980).

[43] A. L. Shelankov, Zh. Éksp. Teor. Fiz. **78**, 2359 (1980) [Sov. Phys. JETP **51**, 1186 (1980)].

[44] A. Schmid, in *Nonequilibrium Superconductivity, Phonons, and Kapitza Boundaries (Proc. NATO ASI)*, K. E. Gray (Ed.), Plenum Press, New York (1981), p. 423.

[45] A. F. Volkov, Zh. Éksp. Teor. Fiz. **68**, 756 (1975) [Sov. Phys. JETP **41**, 376 (1975)].

[46] I. É. Bulyzhenkov and B. I. Ivlev, Zh. Éksp. Teor. Fiz. **74**, 224 (1978) [Sov. Phys. JETP **47**, 115 (1978)].

[47] B. I. Ivlev, Zh. Éksp. Teor. Fiz. **75**, 1771 (1978) [Sov. Phys. JETP **48**, 893 (1978)].

[48] A. M. Gulyan and G. F. Zharkov, Zh. Éksp. Teor. Fiz. **89**, 156 (1985) [Sov. Phys. JETP **62**, 89 (1985)].

[49] A. Schmid and G. Shön, J. Low Temp. Phys. **20**, 207 (1975).

[50] L. N. Bulaevskii, M. Zamora, D. Baeriswyl, H. Beck, and J. R. Clem, Phys. Rev. B **50**, 12 831 (1994).

[51] L. N. Bulaevskii, D. Domínguez, M. P. Maley, A. R. Bishop, and B. I. Ivlev, Phys. Rev. B **53**, 14 601 (1996); L. N. Bulaevskii, D. Domínguez, M. P. Maley, A. R. Bishop, O. K. C. Tsui, and N. P. Ong, Phys. Rev. B **54**, 7521 (1996).

[52] A. S. Mel'nikov, Phys. Rev. Lett. **77**, 2786 (1996).

[53] A. Yurgens, D. Winkler, T. Claeson, T. Murayama, and Y. Ando, Phys. Rev. Lett. **82**, 3148 (1999).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Inhomogeneous ferromagnetic resonance modes in [Fe/Cr]$_n$ superlattices with a high biquadratic exchange constant

A. B. Drovosekov, O. V. Zhotikova, N. M. Kreines,*) and D. I. Kholin

P. L. Kapitsa Institute for Physical Problems, Russian Academy of Sciences, 117334 Moscow, Russia

V. F. Meshcheryakov

Moscow Institute of Radio Engineering, Electronics, and Automation, 117464 Moscow, Russia

M. A. Milyaev, L. N. Romashev, and V. V. Ustinov

Institute of Metal Physics, Ural Department of Russian Academy of Sciences, 620219 Ekaterinburg, Russia

(Submitted 1 July 1999)
Zh. Éksp. Teor. Fiz. **116**, 1817–1833 (November 1999)

In a set of [Fe/Cr]$_n$ superlattices, magnetization curves and spectra of ferromagnetic resonance under an in-plane magnetic field have been studied at room temperature. Along with the acoustic branch, several additional branches have been observed in resonance spectra. Resonance spectra have been calculated analytically for a structure with an infinite number of layers and numerically for finite numbers of layers in real samples using a model of biquadratic exchange taking account of the fourth-order magnetic anisotropy. A possibility of describing both static and resonance properties of the system in terms of this model has been demonstrated.

1. INTRODUCTION

Exchange interaction between thin ferromagnetic layers separated by a nonferromagnetic spacer has been studied for a long time.[1–4] Since the detection of noncollinear (90-degree) magnetic ordering in such structures,[5,6] the energy of interaction between two neighboring ferromagnetic layers separated by a thin spacer has been usually expressed in the form of two terms:

$$E=-J_1\frac{(\mathbf{M}_1\cdot\mathbf{M}_2)}{M_1M_2}-J_2\frac{(\mathbf{M}_1\cdot\mathbf{M}_2)^2}{(M_1M_2)^2}, \qquad (1)$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are the magnetizations of the interacting layers, J_1 and J_2 are phenomenological constants. Depending on the sign of J_1, the first term, named bilinear exchange, leads to either parallel or antiparallel alignment of neighboring magnetic layers. The second term, which is called biquadratic exchange, makes an angle of 90° between the magnetization vectors preferable if J_2 is negative. The competition between these two components of the ferromagnetic energy can lead, generally speaking, to an arbitrary angle between the magnetizations of two neighboring ferromagnetic layers.

The biquadratic exchange has been experimentally detected in a set of metallic multilayered systems. An appreciable amount of data obtained by the Kerr magnetometry, vibrating sample magnetometry, and SQUID magnetometry, measurements of magnetoresistance, ferromagnetic resonance (FMR), and Mandelstam–Brillouin light scattering can be interpreted on the basis of the biquadratic exchange model.[4] The discussion in the present paper is limited to the case of Fe/Cr multilayered systems, which have been studied most extensively and whose biquadratic exchange constant is fairly high.

Samples used in studies of exchange interaction between layers of such multilayer magnetic structures can be classified with two types. The first of them includes so-called ‘‘sandwiches,’’ which are composed of two iron layers separated by a chromium spacer. Most experiments have been conducted with such samples.[4–7] One advantage of the experiments with sandwiches is the simple interpretation of obtained data. On the other hand, the in-plane magnetic anisotropy in such samples is usually comparable to the effects of interlayer interaction, so the iron magnetizations are either collinear or oriented at 90° with respect to one another.

The second type of samples includes superlattices, i.e., periodic structures consisting of a large number (in reality, up to several dozens) of identical iron layers separated by chromium spacers. Since each iron layer interacts with its two neighbors at the same time, the effects of magnetic exchange in such structures are much stronger. For example, the magnetic field necessary for alignment of magnetic moments of all layers into a collinear structure can be higher than 10 kOe,[8–10] even though the field of the fourth-order magnetic anisotropy in iron is about 500 Oe. In this situation, even in the absence of magnetic field, the magnetization direction in each layer can deviate from the easy magnetic axis and be determined primarily by the minimum of the expression (1). As a result, a so-called ‘‘canted’’ magnetic structure with two magnetic sublattices with an ordering angle between them different from 0°, 90°, and 180° can be formed.[9–13] The large number of layers in superlattices, however, complicates the interpretation of measurements of both magnetization and ferromagnetic resonance.

The order of magnitude and oscillations in the bilinear exchange constant J_1 as a function of the nonferromagnetic spacer thickness are explained perfectly on the basis of the well-known RKKY theory.[2] At the same time, no comprehensive approach to the origin of the biquadratic exchange has yet been put forth. Some theories treat the chromium spacer as nonmagnetic. In this case, mechanisms both associated with the electronic structure of chromium itself and controlled by defects or roughness of the iron–chromium interface yield J_2 much lower than experimental values.[14–17] The model suggested by Słonczewski,[16] on the contrary, presumes a considerable exchange stiffness of the spacer material. This model has allowed the researchers to explain the noncollinear ordering in superlattices, but it leads to a formula for the coupling between iron layers different from Eq. (1). This model is supported by certain experimental data.[12] The work by Levchenko *et al.*[18] is also worth mentioning, which pointed out the inevitability of a peculiar domain structure in such systems, provided that sufficiently large atomically smooth areas of the iron–chromium interface are present.

Thus, the model of biquadratic exchange, which is extensively used at present in interpreting experimental data on Fe/Cr systems, has not been fully explained on the basis of existing microscopic theories. In this connection, a question arises about how accurate is the description of magnetic properties of multilayered magnetic structures by this model.

A set of publications, including recent ones (see Ref. 7 and references therein), have been dedicated to systematic studies of the biquadratic exchange in Fe/Cr/Fe sandwiches using several experimental techniques (magnetoresistance, Kerr effect, FMR, light scattering). But, as was noted above, the effects of interaction between layers in sandwiches are superposed on a strong in-plane anisotropy, which modifies the pattern of magnetic ordering described by Eq. (1) and gives rise to jumps and hysteretic loops on experimental curves.

In the present work, we studied FMR spectra and magnetization curves in a set of $[\mathrm{Fe/Cr}]_n$ superlattices, in most of which a canted magnetic ordering has been detected. The main goal of our study was to check out how adequate the model of biquadratic exchange is when the corrections to the measured effects due to the anisotropy are relatively small. In particular, we have studied the question of whether one can describe at the same time the static and resonance properties of a system using the same values of J_1 and J_2. Previously[9,10] we reported about observation of an optical and several additional resonant modes in FMR spectra corresponding to standing spin waves for two samples. Note that an FMR optical mode was previously detected only in sandwiches. A larger number of samples studied in the reported work, measurements, and inclusion of the fourth-order anisotropy in analytic and numerical calculations of resonance spectra, gave us an opportunity to compare the experimental data and calculations by the biquadratic exchange model in more detail.

2. SAMPLES AND EXPERIMENTAL TECHNIQUES

The $[\mathrm{Fe/Cr}]_n$ superlattices were fabricated at the *Katun-S* facility, which employs molecular beam epitaxy. The substrates were MgO single-crystal plates cut parallel to the (100) crystal plane. The buffer layer was a Cr film with a thickness of about 100 Å. The samples were grown at a substrate temperature of 210–220 °C. The sample quality and thicknesses of Fe and Cr layers were monitored using the small-angle X-ray scattering and X-ray spectroscopy of high coordinate resolution (the techniques are described in detail elsewhere[19]). The [001] crystal axis of the iron and chromium films was aligned with the normal to the sample plane.

We tested seven $[\mathrm{Fe/Cr}]_n$ samples. Below is the list of these samples with the thicknesses and total numbers of layers in the structures.

(1) $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{12}$
(2) $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(14\ \text{Å})]_{12}$
(3) $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{16}$
(4) $[\mathrm{Fe}(33\ \text{Å})/\mathrm{Cr}(7.7\ \text{Å})]_{16}$
(5) $[\mathrm{Fe}(48\ \text{Å})/\mathrm{Cr}(7.6\ \text{Å})]_{16}$
(6) $[\mathrm{Fe}(20\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{20}$
(7) $[\mathrm{Fe}(26\ \text{Å})/\mathrm{Cr}(9.2\ \text{Å})]_{24}$

We used two experimental techniques, namely, the ferromagnetic resonance and measurements of static magnetization curves. The experiments were conducted at room temperature. The magnetization curves were recorded by a vibrating-sample magnetometer under a magnetic field ranging from 0 to 17 kOe. The ferromagnetic resonance was studied in a frequency band of 9.5 to 37 GHz under magnetic fields of up to 18 kOe. In order to detect FMR in this frequency band, we used a set of six cylindrical microwave cavities in which the microwave modes H_{011}, H_{012}, and H_{013} were excited. A sample was placed at the bottom of the cavity so that the microwave magnetic field was aligned parallel to the sample plane. A dc magnetic field was also applied parallel to the sample plane and directed along the hard [110] or easy [100] axes of the iron magnetic anisotropy. By changing the sample position, we could satisfy either transverse or longitudinal pumping conditions.

3. EXPERIMENTAL RESULTS

In Fig. 1, symbols of different shapes show magnetostatic measurement data for three samples, namely $[\mathrm{Fe}(33\ \text{Å})/\mathrm{Cr}(7.7\ \text{Å})]_{16}$, $[\mathrm{Fe}(48\ \text{Å})/\mathrm{Cr}(7.6\ \text{Å})]_{16}$, and $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{16}$. The curves show the magnetization component M along the applied magnetic field as a function of the magnetic field strength H. The field is aligned with the hard magnetic axis in the sample plane (the crystal axis [110]). The solid and dashed lines plot calculations to be discussed below. Two of the three samples have considerable value of magnetization M_0 even at zero magnetic field, which indicates the presence of noncollinear magnetic ordering. The corresponding angle between magnetizations of neighboring iron layers, $\theta_0 = 2\cos^{-1}(M_0/M_S)$, in these samples is given in the graph (M_S is the saturation magnetization). In the third sample, the magnetic moment at zero magnetic field is also zero, but the nonlinear approach of all three curves to the saturation level is an indication of considerable biquadratic exchange in all the samples. In the series of seven multilayered films studied in our experiments, four had a nonvanishing zero-field magnetization.

In studying the ferromagnetic resonance, we detected, along with the homogeneous acoustic FMR mode, several

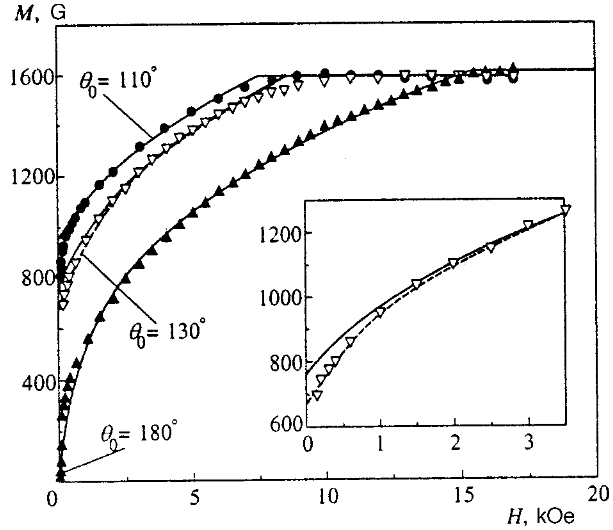

FIG. 1. Magnetization as a function of magnetic field for three samples. The symbols represent experimental data: (●) [Fe(33 Å)/Cr(7.7 Å)]$_{16}$; (▽) [Fe(48 Å)/Cr(7.6 Å)]$_{16}$; (▲) [Fe(20 Å)/Cr(10 Å)]$_{20}$. The solid lines show calculations by Eqs. (8) and (9), the dashed lines are numerical calculations. The inset shows $M(H)$ in the range of low magnetic fields for [Fe(48 Å)/Cr(7.6 Å)]$_{16}$ on an extended scale.

inhomogeneous modes[9,10] in all samples. Figure 2 shows a set of experimental curves recorded at different frequencies in the configuration of transverse resonance excitation in the [Fe(21 Å)/Cr(10 Å)]$_{12}$ sample under a field aligned with the hard axis of iron magnetization. In addition to the intense absorption line corresponding to the acoustic resonance mode (it is labeled by *1* in the graph), one can also see weaker and broader lines corresponding to excitation of inhomogeneous magnetization modes and marked by arrows (digits *2* and *3* on the lower curve). The narrow line marked by *4* is due to the paramagnetic substance, which was used to calibrate the Hall probe. The line width of the acoustic FMR mode was approximately 300 Oe. The line widths at higher magnetic fields (on the right of the acoustic mode) were larger (up to 1000 Oe), the corresponding resonances have not been observed in all samples, and they could be excited only in the longitudinal configuration. The resonances in weaker magnetic fields (on the left of the acoustic mode) were detected in all tested superlattices. Some of them could be detected only in the longitudinal configuration, and a fraction of these resonances could be observed in both longitudinal and transverse configurations. The intensities of lines seen in the transverse resonance excitation configuration were a factor of several tens lower under the longitudinal excitation. It is graphically illustrated by Fig. 2 in Ref. 9, which shows traces of FMR in one sample recorded with two different alignments of the dc and microwave magnetic fields. The frequencies of the FMR modes detected in our experiments for several superlattices are plotted by dots in Figs. 4, 5, and 8 as functions of magnetic field applied along both easy and hard magnetic axes. The anisotropy of absorption peak positions, i.e., the change in these positions due to the dc magnetic field rotation in the sample plane through an angle of 45°, was, as can be seen in the graphs, up to 1 kOe.

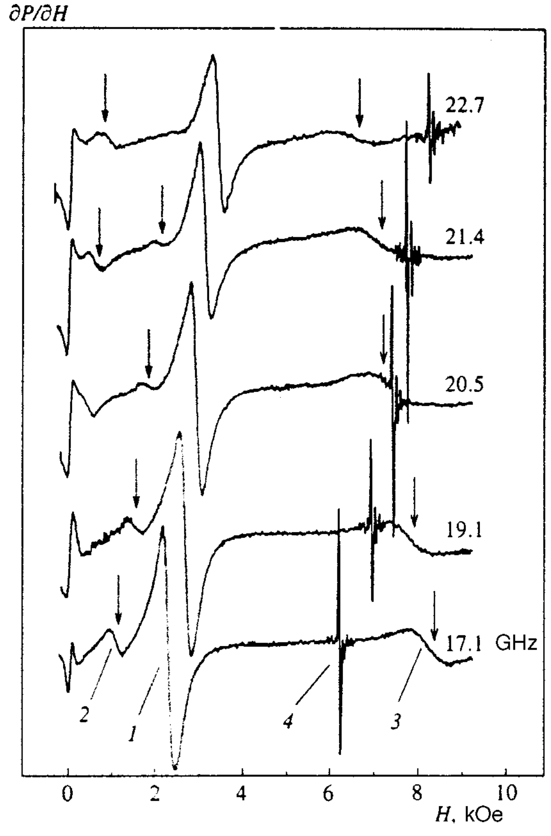

FIG. 2. Derivative of microwave absorption $\partial P/\partial H$ as a function of magnetic field in the configuration of longitudinal excitation at different frequencies for the [Fe(21 Å)/Cr(10 Å)]$_{12}$ sample.

4. ANALYTIC CALCULATION OF STATIC MAGNETIZATION AND FMR SPECTRUM OF A SUPERLATTICE

In order to interpret our experimental results, we calculated analytically the magnetization curves and FMR spectra of a superlattice on the basis of the biquadratic exchange model taking into account the fourth-order cubic anisotropy in the case of a magnetic field H parallel to the sample plane. We carried out a similar calculation, but without taking account of the cubic anisotropy, in our earlier work.[9] We assumed that the magnetization of each iron layer was homogeneous over its volume and equal to M_S, irrespective of the applied magnetic field, and the magnetization of each layer rotated as a whole. Under these assumptions, the magnetic energy of a multilayered structure per unit area of a film can be expressed as

$$E=-d\sum_{j=1}^{n}(\mathbf{H}\cdot\mathbf{M}_j)-\frac{J_1}{M_S^2}\sum_{j=1}^{n-1}(\mathbf{M}_j\cdot\mathbf{M}_{j+1})$$
$$-\frac{J_2}{M_S^4}\sum_{j=1}^{n-1}(\mathbf{M}_j\cdot\mathbf{M}_{j+1})^2+d\frac{N_{zz}}{2}\sum_{j=1}^{n}(\mathbf{M}_j\cdot\mathbf{z})^2$$
$$-d\frac{K_u}{M_S^2}\sum_{j=1}^{n}(\mathbf{M}_j\cdot\mathbf{z})^2-d\frac{K_1}{2M_S^4}\sum_{j=1}^{n}[(\mathbf{M}_j\cdot\mathbf{x})^4$$
$$+(\mathbf{M}_j\cdot\mathbf{y})^4+(\mathbf{M}_j\cdot\mathbf{z})^4], \qquad (2)$$

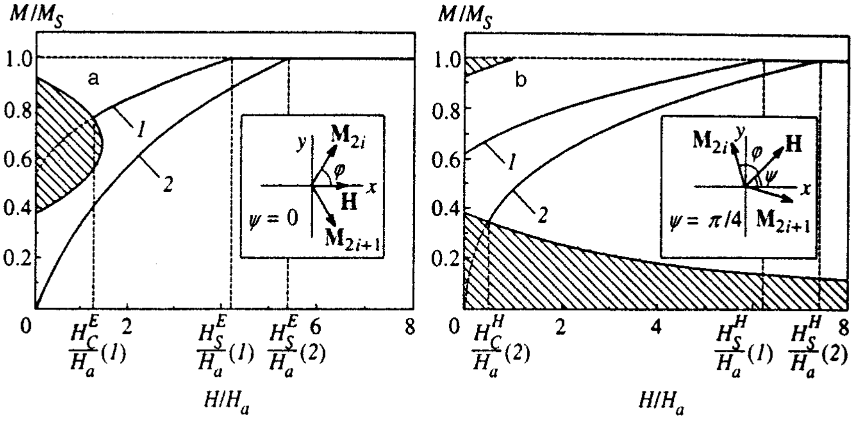

FIG. 3. Curves of $M(H)$ calculated by Eqs. (8) and (9) for two directions of applied magnetic field [(a) along the easy magnetization axis and (b) along the hard axis] for different values of exchange constants: (*1*) $J_1/(dM_SH_a)=-0.3$, $J_2/(dM_SH_a)=-0.5$; (*2*) $J_1/(dM_SH_a)=-1.0$, $J_2/(dM_SH_a)=-0.3$. The hatched areas are regions where the solutions are unstable. The insets schematically show the alignments of iron magnetization vectors with respect to the easy and hard magnetization axes for these solutions.

where $\mathbf{M}_j$ is the magnetization of the jth iron layer, d is the thickness of each iron layer, N_{zz} is the demagnetization factor (4π in this specific case), K_u characterizes the uniaxial anisotropy, and K_1 accounts for the fourth-order anisotropy with easy axes **x**, **y**, and **z**, where the **z**-axis is normal to the sample plane. In what follows, we introduce for convenience the surface anisotropy coefficient

$$K_{\mathrm{eff}}=N_{zz}-2K_u/M_S^2, \tag{3}$$

and also the effective anisotropy field of the fourth order:

$$H_a=2K_1/M_S. \tag{4}$$

For simplicity of our calculations, the following analysis will be performed in the limit of an infinite superlattice. Since the applied magnetic field is parallel to the sample plane and the demagnetization field in these samples is considerably higher than the anisotropy field, the magnetizations of all iron layers in equilibrium lies also in the film plane, so $\mathbf{M}_j\cdot\mathbf{z}=0$. In this case, the total energy can be expressed as

$$\begin{aligned}E=&-dHM_S\sum_j\cos\varphi_j-J_1\sum_j\cos(\varphi_j-\varphi_{j+1})\\&-J_2\sum_j\cos^2(\varphi_j-\varphi_{j+1})-\frac{1}{16}dH_aM_S\\&\times\sum_j\cos 4(\varphi_j+\psi),\end{aligned} \tag{5}$$

where ψ is the angle between the easy axis **x** and external field **H**, and φ_j is the angle between magnetic field **H** and magnetization **M** of the jth ferromagnetic layer. The equilibrium values of angles φ_j are derived, as is usual, from the energy minimum condition

$$\partial E/\partial\varphi_j=0. \tag{6}$$

We will consider two configurations of the highest symmetry, namely, with the magnetic field applied along the easy ($\psi=0$) or hard ($\psi=\pi/4$) axes of iron magnetization anisotropy (Fig. 3). One can easily prove that condition (6) leads to a two-sublattice ordering of ferromagnetic moments in a wide range of magnetic fields, and the sublattice magnetizations are aligned symmetrically with respect to the applied magnetic field. In this case, one can analytically calculate the magnetization curves and resonance spectrum of the structure. The angles φ_j can be expressed in this case as

$$\varphi_j=(-1)^j\varphi,\quad\text{where}\quad\varphi=\cos^{-1}(M/M_S). \tag{7}$$

Let us consider only the interesting case of $J_2<0$. Then, with due account of Eq. (7), condition (6) yields the following two solutions:

$$H=\left(-\frac{4J_1-8J_2}{dM_S^2}\pm\frac{H_a}{M_S}\right)M+\left(-\frac{16J_2}{dM_S^4}\mp 2\frac{H_a}{M_S^3}\right)M^3, \tag{8}$$

$$M=M_S. \tag{9}$$

Hereafter the upper signs correspond to the angle $\psi=0$, and the lower signs to $\psi=\pi/4$. The function $M(H)$, determined implicitly by Eq. (8), corresponds to the case when the angle between sublattice magnetizations is nonzero. Solution (9) describes the case of saturation, when the magnetizations of all iron layers are aligned with the applied magnetic field ($\varphi=0$). The regions of magnetic field where these solutions minimize the total system energy can be determined using the condition that the second differential must be positive definite:

$$\Delta^2E=\sum_{i,j}\frac{\partial^2E}{\partial\varphi_i\partial\varphi_j}\Delta\varphi_i\Delta\varphi_j>0\quad\text{for arbitrary}\quad\Delta\varphi_i\quad\text{and}\ \Delta\varphi_j. \tag{10}$$

One can prove that, for the two-sublattice configuration of magnetization described by Eq. (7), condition (10) is equivalent to the combination of two inequalities:

$$(H\cos\varphi\pm H_a\cos 4\varphi)+4\left(\frac{J_1}{dM_S}\cos 2\varphi+\frac{2J_2}{dM_S}\cos 4\varphi\right)>0, \tag{11}$$

$$H\cos\varphi\pm H_a\cos 4\varphi>0. \tag{12}$$

First let us consider limitations imposed by condition (11). For the unsaturated phase defined by Eq. (8), condition (11) reduces to

$$\frac{\partial H}{\partial M}=\left(-\frac{4J_1-8J_2}{dM_S^2}\pm\frac{H_a}{M_S}\right)+3\left(-\frac{16J_2}{dM_S^4}\mp 2\frac{H_a}{M_S^3}\right)M^2>0. \tag{13}$$

This is a formulation of a quite obvious fact that only branches of function $M(H)$ on which the magnetization increases with the magnetic field are stable. For a solution (9) corresponding to the saturated state, inequality (11) leads to the condition

$$H>H_S^{E,H}=-(4J_1+8J_2)/dM_S\mp H_a. \tag{14}$$

This condition means that the saturated state, in which the magnetic moments in all layers are aligned with the applied field, is stable when the external field is higher than the saturation field $H_S^{E,H}$, at which the curve defined by Eq. (8) crosses curve (9) (H_S^E is the saturation field along the easy axis, and H_S^H in the same along the hard axis).

Note that in the interesting case of a relatively high biquadratic exchange (to be exact, under the condition $-J_2>dH_aM_S/8$) condition (13) is always satisfied, and condition (11) reduces to the trivial statement that the alignment of all magnetic moments with the applied magnetic field is stable when the latter is higher than the saturation field $H_S^{E,H}$.

Now let us turn to condition (12). It is clear that it defines areas on the H/H_a–M/M_S plane in which a symmetrical configuration of two sublattices is unstable *a fortiori* (the hatched areas in Fig. 3). The crossing point of the magnetization curves calculated by Eqs. (8) and (9) with the boundaries of these areas determine the fields H_C^E and H_C^H for the easy and hard directions, respectively, below which the above formulas for the magnetization curves and FMR spectra do not apply (see below).

Note that, in the case of saturation fields much higher than the anisotropy field, which was realized in our experiments, condition (8) holds over a wide field range $H_C^{E,H}<H<H_S^{E,H}$, and at $H>H_S^{E,H}$ solution $M=M_S$ is stable.

In calculating the oscillation spectrum of the system under consideration, let us use the Landau–Lifshitz equations without the damping term:

$$\gamma^{-1}(\partial\mathbf{M}_j/\partial t)=-\mathbf{M}_j\times\mathbf{H}_j^{\mathrm{eff}}, \tag{15}$$

where H_j^{eff} is obtained by differentiating energy (2) with respect to M_j:

$$H_j^{\mathrm{eff}}=-(\partial E/\partial\mathbf{M}_j)d^{-1}. \tag{16}$$

In solving Eqs. (15) and (16), let us express the magnetization M_j as the sum of the static magnetization $M_j^{(0)}$ and the small term describing a wave propagating along the normal to the film plane:

$$\mathbf{M}_j=\mathbf{M}_j^{(0)}+\mathbf{m}_j\exp\{-i(\omega t-jq)\}. \tag{17}$$

Here parameter q is the product of the wave vector and the superlattice period. It can be treated as a phase difference between magnetization oscillations in neighboring iron layers. The case of $q=0$ corresponds to the acoustic mode of oscillations, and $q=\pi$ corresponds to the optical mode.

After substituting expression (17) in Eqs. (15) and (16) and linearizing the resulting equation with respect to $\mathbf{m}_j$, we can derive the frequencies $\omega_q(H)$ as functions of q and magnetic field H from the existence condition of a nontrivial solution:

$$\frac{\omega_q^2}{\gamma^2}=\left[(A+BM^2)\cos^2\frac{q}{2}+K_{\mathrm{eff}}+\frac{3H_a}{4M_S}\pm\frac{H_a}{4M_S^3}(M_S^2-4M^2)\right]\times\left[(A+BM^2)M^2\cos^2\frac{q}{2}+(A+3BM^2)(M_S^2-M^2)\sin^2\frac{q}{2}\pm\frac{H_a}{M_S^3}(M_S^2-M^2)(M_S^2-6M^2)\right]\quad\text{for }H_C^{E,H}<H<H_S^{E,H}, \tag{18}$$

$$\frac{\omega_q^2}{\gamma^2}=\left[H-H_S\sin^2\frac{q}{2}+K_{\mathrm{eff}}M_S+\frac{3H_a}{4}\pm H_a\left(\cos^2\frac{q}{2}-\frac{3}{4}\right)\right]\times\left(H-H_S\sin^2\frac{q}{2}\pm H_a\cos^2\frac{q}{2}\right)\quad\text{for }H>H_S^{E,H} \tag{19}$$

with the notations

$$A=-\frac{4J_1-8J_2}{dM_S^2},\quad B=-\frac{16J_2}{dM_S^4}. \tag{20}$$

At $H_a=0$, Eqs. (8) and (9) for the static magnetization and Eqs. (18) and (19) for the FMR frequency are identical to the corresponding expressions in Ref. 9, whereas at $H=0$ they coincide with the formulas in Ref. 20. Note once again that the resulting formulas apply to the fields higher than H_C, when we have a two-sublattice ordering of magnetic moments aligned symmetrically with respect to the external magnetic field.

5. DISCUSSION OF RESULTS

5.1. Comparison to analytic calculations

Our calculations were fitted to experimental data in the following way. Magnetization curves measured by a vibrating-sample magnetometer under a magnetic field aligned with the hard axis in the sample plane were approximated by Eqs. (8) and (9). The resulting constants J_1, J_2, and M_S were used in calculations of FMR spectra. The parameters H_a and K_{eff} were selected to obtain the best fits to the resonance acoustic branch measured along the easy and hard magnetization axes. This procedure was performed for all samples. Figure 1 shows, along with the experimental data, calculated magnetization curves for three samples (solid lines). The agreement between theoretical and experimental curves is quite satisfactory. The absence of a visible cusp on experimental curves of $M(H)$, which can be seen on

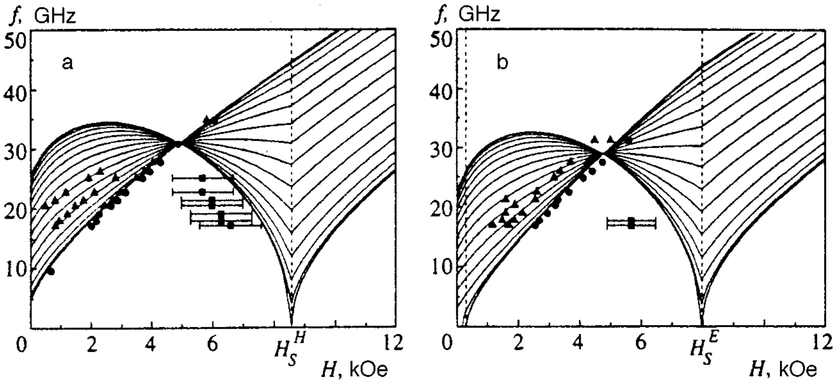

FIG. 4. FMR spectra of the $[\mathrm{Fe}(48\ \text{Å})/\mathrm{Cr}(7.6\ \text{Å})]_{16}$ sample polarized along the (a) easy and (b) hard magnetic axes. The points are experimental data, and solid lines are calculations by Eqs. (21) and (22) with the following parameters: $J_1=-0.58$ erg/cm^2; $J_2=-0.50$ erg/cm^2; $M_S=1590$ G; $K_{\mathrm{eff}}=12$; $H_a=300$ Oe.

the calculated curves at H_S, can be ascribed to variations in the iron and chromium layer thicknesses, which can blur this feature.

Calculations of the FMR spectrum of one sample featured in Fig. 1 ($[\mathrm{Fe}(48\ \text{Å})/\mathrm{Cr}(7.6\ \text{Å})]_{16}$) based on Eqs. (18) and (19) for two alignments of the applied magnetic field (with the easy and hard magnetization axes) are plotted in Fig. 4 by solid lines. As was stated above, in calculating these spectra we used the same parameters as for the magnetization curve in Fig. 1. The set of q's that we employed in plotting the set of curves in the graphs requires clarification. For a traveling wave in an infinite superlattice, q varies continuously between $-\pi$ and π, but in a finite structure magnetization modes are characterized by a set of discrete wave vectors. The graphs plot magnetization modes corresponding to q's ranging between 0 and π with a step of $\pi/(N-1)$, where N is the number of iron layers in the superlattice. In this case, the number of FMR modes is equal to the number of magnetic layers in the sample, which is identical to the number of inherent oscillation modes of the structure. As numerical calculations will show, there is little sense in defining q with good accuracy on the basis of model boundary conditions on the outer surfaces because the condition of the two-sublattice magnetic ordering in a finite structure is violated in a magnetic field.

In comparing the theoretical curves with experimental FMR spectra, we note above all that in all the samples the numbers of experimentally observed resonance branches were smaller than the numbers of iron layers in the structure. It is not surprising, however, that we failed to detect all possible magnetic modes because all the modes except the optical and acoustic ones can be excited in principle only owing to the finite number of iron layers or slight variations in their thicknesses, whereas a homogenous microwave field in both the longitudinal and transverse configurations is hardly an efficient tool for driving these modes.

The agreement between calculations and experimental data (points on the graphs) for the acoustic resonance mode is good for both orientations of the external magnetic field. For the optical mode (black squares) the difference between the experimental data for some samples (but not all of them) and theoretical predictions was larger than the absorption line width, shown in the graph by the horizontal bars (Fig. 4). The comparison between the experimental data and calculations for several intermediate modes higher than the acoustic one (triangles in the graph) is difficult because their corresponding q's are unknown. The criterion of agreement between measurements and calculations in this case should be the equality between the slopes of experimental curves of $f(H)$ and nearest theoretical curves. One can see in Fig. 4 that the experimental data reasonably conform to this criterion in the cases of magnetic field orientation along both easy and hard axes, although the difference was substantial for some of the samples. Thus, the suggested theory has demonstrated qualitative agreement with experimental data, although sometimes it cannot accurately describe resonant features of our samples.

5.2. Numerical calculations. Comparison with experiment

In order to fit the suggested model to real experimental data, we numerically calculated the spectrum of inherent magnetic modes of a superlattice incorporating a finite number of iron layers. The magnetic component of the system energy was expressed, as in analytic calculations, by Eq. (2), where the sums were performed over the finite set of layers in a real sample. Then this energy was minimized numerically by varying all angles φ_j between the magnetic moments of the layers and applied magnetic field. This approach is quite different from the two-sublattice model discussed above. The absence of one neighbor in the case of the highest and lowest iron layers leads to larger canting angles between the magnetizations of these layers and applied field. As a result, the two-sublattice ordering is replaced by a more complicated pattern.

Let us discuss numerical calculations in greater detail taking as an example the $[\mathrm{Fe}(48\ \text{Å})/\mathrm{Cr}(7.6\ \text{Å})]_{16}$ sample considered in the previous section. The magnetization curve of this sample calculated with due account of the real number of iron layers ($n=16$) is shown in Fig. 1 by the dashed line. Its deviation from the curve obtained in the limit of the infinite superlattice is insignificant, and it is notable only in the field range below 3 kOe. At the same time, the magnetization pattern in this case is radically different from that predicted by the two-sublattice model. Figure 5 shows the equilibrium angles φ_j between the magnetization vector of the jth iron layer and external magnetic field aligned with the hard magnetization axis. It is clear that these angles are notably different for the outer layers (in this specific case, the first and sixteenth). The angles φ_j for the inner layers also

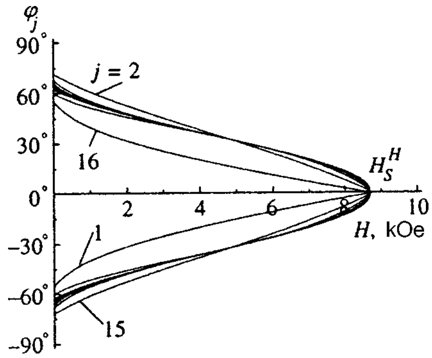

FIG. 5. Calculated equilibrium values of angles φ_j between magnetization vector in jth iron layer and magnetic field direction for the [Fe(48 Å)/Cr(7.6 Å)]$_{16}$ sample. The numbers near some of the curves indicate the layer number. The external magnetic field is applied along the hard magnetization axis.

vary considerably. Nonetheless, the alignment of all magnetic moments to the ferromagnetic phase takes place simultaneously, and the cusp on the magnetization curve at H_S persists. Under sufficiently strong fields, the calculations were not affected by either the field direction or the sample history, whereas at weaker magnetic fields (depending on the parameters J_1, J_2, and H_a) hysteresis loops were observed. Note, however, that all absorption lines recorded in our experiments had positions in the field range where the calculations were unambiguous, so the doubts concerning the sample history were irrelevant.

The curves in Fig. 6 plot numerical calculations of the FMR spectrum of the sample under consideration, [Fe(48 Å)/Cr(7.6 Å)]$_{16}$. The constants J_1, J_2, and M_S for these calculations were derived from the magnetization curve. Note above all the similarity of these curves to the analytic calculations plotted in Fig. 4. There are, however, notable differences between them. First, the dropping section of the curve of the optical mode (the lowest curve in the field range of 4 to 8 kOe) becomes doubly degenerate and separate from the higher resonance lines. Second, the intermediate resonance modes between the acoustic and optical FMR modes are also deformed in comparison with the case of an infinite superlattice, but now their positions are determined uniquely, without any independent variables like q. The analysis of their configurations shows that the total sample magnetization oscillates not only in the optical and acoustic modes, but also in all other FMR modes, which allows us to drive these intermediate modes by a uniform microwave field.

The agreement between the numerical calculations and experimental data for the sample under discussion is better because the lowest FMR mode obtained in this calculation and separated from the rest is closer to the ''optical'' resonance mode in the range of strong magnetic fields. Nonetheless, we have to admit that the calculated curves still notably deviate from the experimental points, especially in the range of higher magnetic fields. Moreover, although the calculations are not affected by any uncertainty associated with the variable q, difficulties may arise in ascribing resonant lines detected in experiments to specific modes obtained by numerical calculations.

The analysis of the results concerning the set of samples under investigation has led us to a conclusion that, using the same constants in the energy expression, it is sometimes impossible to obtain a satisfactory description of the magnetization curves and FMR spectra simultaneously. So the question arises of how good the description of experimental spectra may be if we vary the constants to obtain the best fits to the recorded spectra and how great the discrepancy between the measured and calculated magnetization curves can be in this case. Such a comparison for the [Fe(21 Å)/Cr(10 Å)]$_{12}$ sample, in which the discrepancy between the experimental spectrum and the spectrum calculated using the parameters derived from the magnetization curve was maximal, will be given below.

The full circles in Fig. 7 show measurements of this sample's magnetization, the solid line is a calculated curve which is best fitted to the magnetization measurements, and the dashed line is a calculated curve whose parameters were derived from the FMR spectrum (the function plotted by triangles will be discussed below). The spectra corresponding to these calculations of magnetization are shown, along with experimental data, in Fig. 8.

The sample [Fe(21 Å)/Cr(10 Å)]$_{12}$ is the only one in which we detected more than one resonance line in the region of stronger magnetic fields (on the right of the acoustic branch). It is clear that at parameters derived from the experimental magnetization curves (see the caption to Fig. 8) these dropping resonance branches are lower than the calculated curves (Fig. 8a1 and 8b1). The agreement between the experimental data and calculations, however, can be almost

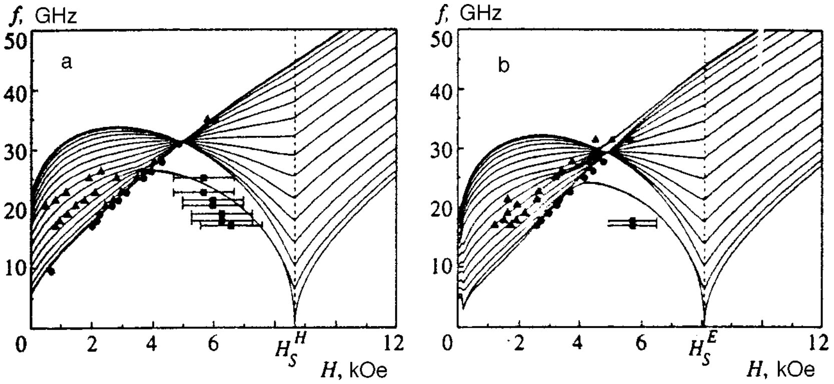

FIG. 6. FMR spectra in magnetic fields aligned with the (a) hard and (b) easy magnetization axes in the [Fe(48 Å)/Cr(7.6 Å)]$_{16}$ sample. The points are experimental data, the solid lines show calculations with the following parameters: $J_1 = -0.67$ erg/cm^2; $J_2 = -0.47$ erg/cm^2; $M_S = 1590$ G; $K_{\mathrm{eff}} = 12$; $H_a = 300$ Oe.

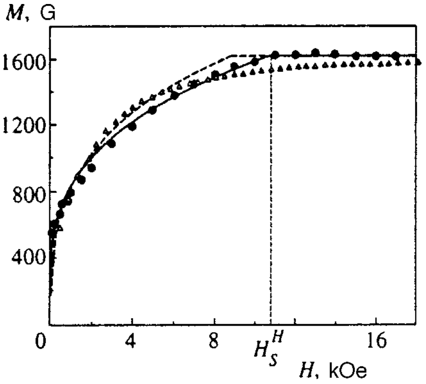

FIG. 7. Magnetization of the $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{12}$ sample versus magnetic field. The full circles plot experimental data, the solid line shows numerical calculations by the biquadratic exchange model with parameters derived from the magnetization curve, the dashed line shows similar calculations with parameters selected to obtain the best fit to the experimental spectrum (Fig. 8). The triangles represent the results of calculations based on Słonczewski's model.

ideal for all resonance modes observed in experiments at both field orientations if we vary parameters J_1 and J_2 to obtain the best fit (Fig. 8a2 and 8b2). The magnetization curve (Fig. 7) calculated at these values of parameters is within 10% from the experimental points, which, however, is larger than the experimental uncertainty. Note that the difference between J_1 an J_2 derived from the magnetization curves and independently from FMR spectra was within 10–20% in our samples.

At this point it is probably appropriate to compare the absolute values of constants J_1 and J_2 obtained in our experiments with the results by other authors. The bilinear exchange constants for our samples are around the peak of J_1 plotted versus the spacer thickness, and our data are in fairly good agreement with earlier publications.[2,4,7] This also applies to our measurements of J_2 (Refs. 2 and 4), although the spread of biquadratic exchange constants quoted in literature for the same thicknesses of the chromium spacer is notably larger.[7]

5.3. Comparison with Słonczewski's model

As was noted in Introduction, the model of ''proximity magnetism'' suggested by Słonczewski[16] to account for the noncollinear magnetic structure in superlattices of Cr and Mn layers yields an energy of interaction between neighboring iron layers different from the biquadratic exchange:

$$E=C_{+}(\theta)+C_{-}(\theta-\pi)^2. \tag{21}$$

Here θ is the angle between magnetization vectors in neighboring ferromagnetic layers, C_+ and C_- are phenomenological constants. The open triangles in Fig. 7 show the magnetization curve calculated for the $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{12}$ sample by Słonczewski's model. The constants C_+ and C_- were selected to obtain the best fit of calculations to experimental data. The graph clearly shows that the agreement with the curve predicted by Słonczewski's model is poorer than with the biquadratic exchange calculations.

Chirita *et al.*[21] calculated a spectrum of spin waves for a sandwich by the ''proximity magnetism'' model. The resulting frequency as a function of magnetic field for the optical mode is radically different from our measurements of $f(H)$ in the range of strong magnetic fields. Thus, it is obvious that, at least at room temperature, the measurements of our samples are much better described by the biquadratic exchange model than by Słonczewski's ''proximity magnetism'' model.

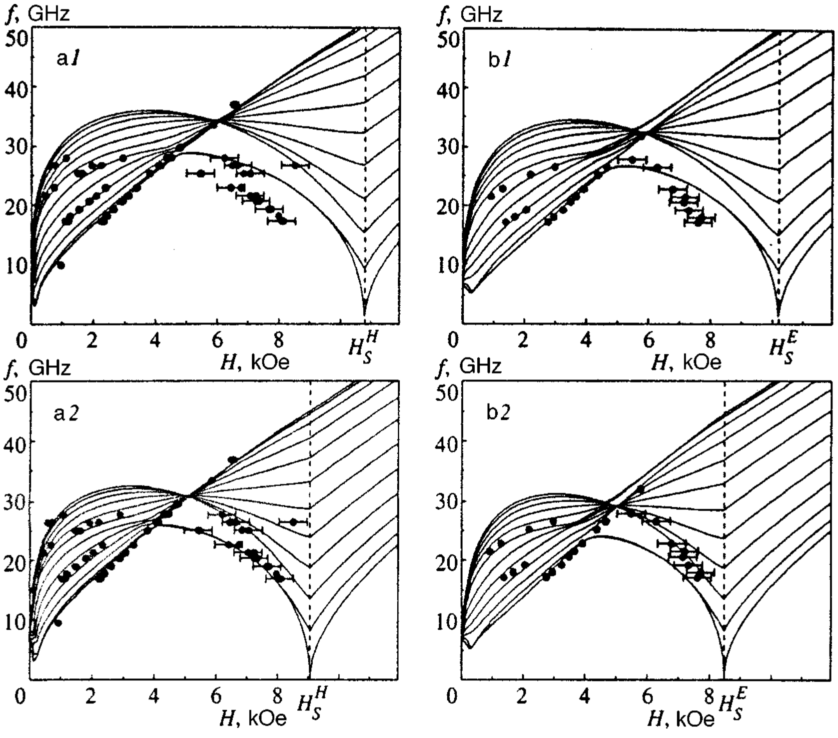

FIG. 8. FMR spectra in the field aligned with the (a*1*, a*2*) hard and (b*1*, b*2*) easy magnetization axis in the $[\mathrm{Fe}(21\ \text{Å})/\mathrm{Cr}(10\ \text{Å})]_{12}$ sample. The points plot experimental data, the solid lines are numerical calculations. The parameters for graphs a1 and b1 were derived from the magnetization curve ($J_1=-0.42$ erg/cm^2; $J_2=-0.24$ erg/cm^2; $M_S=1620$ G; $K_{\mathrm{eff}}=11$; $H_a=300$ Oe). The parameters for graphs a*2* and b*2* were selected to obtain the best fit to the experimental spectrum ($J_1=-0.38$ erg/cm^2; $J_2=-0.19$ erg/cm^2; $M_S=1620$ G; $K_{\mathrm{eff}}=11$; $H_a=300$ Oe).

6. CONCLUSIONS

Let us briefly summarize the reported investigation.

In a set of $[\mathrm{Fe/Cr}]_n$ superlattices with noncollinear magnetic ordering at room temperature, we measured magnetization curves and ferromagnetic resonance spectra under a magnetic field parallel to the sample surface. Along with the homogeneous mode (acoustic branch) in FMR spectra, we detected inhomogeneous resonance modes in our experiments.

On the basis of the biquadratic exchange model, we calculated magnetization curves and spin-wave spectra of an infinite superlattice with due account of the fourth-order magnetic anisotropy. It turned out that, in the configurations with the magnetic field aligned with the easy or hard magnetization axis, a two-sublattice ordering of iron magnetic moments symmetrical with respect to the applied magnetic field is stable over a wide range of magnetic fields. In this case, we obtained analytic expressions for the magnetization versus field and oscillation spectra in the system, which are in good agreement with experimental data.

Magnetization curves and oscillation spectra in superlattices incorporating finite numbers of iron layers, as in real samples, have been calculated numerically. These calculations have yielded theoretical curves which are in better agreement with experimental results.

In most cases, constants J_1 and J_2 derived from experimental magnetization curves also accurately describe resonant properties of superlattices. The difference between exchange constants derived from magnetization curves and independently from FMR spectra of the same sample is within 10–20%.

Magnetization curves calculated on the basis of the biquadratic exchange model are in better agreement with measurements of our samples than those based on Słonczewski's ''proximity magnetism'' model.[16]

In conclusion, we stress that the model used in the reported study is oversimplified. For example, it ignores the fact that the iron–chromium interface is not infinitely thin, but a region of an alloy with a thickness of several atomic layers.[22,23] The model treating iron magnetization in each layer as a single vector rotated as a whole is also very approximate.[2] Nonetheless, this model yields a fairly accurate description of the basic static and resonant properties of $[\mathrm{Fe/Cr}]_n$ magnetic superlattices and is in reasonable agreement with experimental data.

We express our sincere gratitude to N. G. Bebenin for useful discussions. The work was supported by the Russian Fund for Fundamental Research (Projects 98-02-16797 and 98-02-17517) and INTAS (Grant No. 96-0531).

*)E-mail: kreines@kapitza.ras.ru

---

[1] P. E. Wigen and Z. Zang, Brazilian J. of Physics **22**, 267 (1992).

[2] B. Heinrich and J. Cochran, Adv. Phys. **42**, 523 (1993).

[3] A. Fert, P. Grunberg, A. Barthelemy *et al.*, J. Magn. Magn. Mater. **140–144**, 1 (1995).

[4] S. O. Demokritov, J. Phys. D: Appl. Phys. **31**, 925 (1998).

[5] M. Ruhrig, R. Schafer, A. Hubert *et al.*, Phys. Status Solidi A **125**, 635 (1991).

[6] B. Heinrich, J. F. Cochran, M. Kovalevski *et al.*, Phys. Rev. B **44**, 9348 (1991).

[7] S. M. Rezende, C. Chesman, M. A. Lucena *et al.*, J. Appl. Phys. **84**, 958 (1998).

[8] V. V. Ustinov, L. N. Romashev, V. I. Minin *et al.*, Fiz. Met. Metalloved. **80**(2), 71 (1995).

[9] A. B. Drovosekov, N. M. Kreines, D. I. Kholin *et al.*, JETP Lett. **67**, 727 (1998).

[10] A. B. Drovosekov, D. I. Kholin, A. N. Kolmogorov *et al.*, J. Magn. Magn. Mater. **198–199**, 455 (1999).

[11] V. V. Ustinov, N. G. Bebenin, L. N. Romashev *et al.*, Phys. Rev. B **54**, 15958 (1996-II).

[12] A. Schreyer, J. F. Anker, Th. Zeidler *et al.*, Phys. Rev. B **52**, 16066 (1995).

[13] N. M. Kreines, A. N. Kolmogorov, and V. F. Mescheriakov, J. Magn. Magn. Mater. **177–181**, 1189 (1998).

[14] J. C. Słonczewski, Phys. Rev. Lett. **67**, 3172 (1991).

[15] J. C. Słonczewski, J. Appl. Phys. **73**, 5957 (1993).

[16] J. C. Słonczewski, J. Magn. Magn. Mater. **150**, 13 (1995).

[17] S. O. Demokritov, E. Tsymbal, P. Grunberg *et al.*, Phys. Rev. B **49**, 720 (1994).

[18] V. D. Levchenko, A. I. Morozov, A. S. Sigov, and Yu. S. Sigov, Zh. Éksp. Teor. Fiz. **114**, 1817 (1998) [JETP **87**, 985 (1998)].

[19] V. V. Ustinov, T. P. Krinitsina, V. V. Popov *et al.*, Fiz. Met. Metalloved. **84**(4), 161 (1997).

[20] N. G. Bebenin and V. V. Ustinov, Fiz. Met. Metalloved. **84**(2), 29 (1997).

[21] M. Chirita, G. Robins, R. L. Stamps *et al.*, Phys. Rev. B **38**, 869 (1998-II).

[22] B. Heinrich, J. F. Cochran, D. Venus *et al.*, J. Appl. Phys. **79**, 4518 (1996).

[23] V. V. Ustinov, V. A. Tsurin, L. N. Romashev, and V. V. Ovchinnikov, Pis'ma Zh. Tekh. Fiz. **25**, 88 (1999) [Tech. Phys. Lett. **25**, 459 (1999)].

Translation provided by the Russian Editorial office

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 89, NUMBER 5 NOVEMBER 1999

Electronic spectrum of a two-dimensional Fibonacci lattice

Yu. Kh. Vekilov,*) I. A. Gordeev, and É. I. Isaev

Moscow State Institute of Steels and Alloys, 117936 Moscow, Russia

(Submitted 25 June 1999)
Zh. Éksp. Teor. Fiz. **116**, 1834–1842 (November 1999)

The electronic spectrum and wave functions of a new quasicrystal structure—a two-dimensional Fibonacci lattice—are investigated in the tight-binding approximation using the method of the level statistics. This is a self-similar structure consisting of three elementary structural units. The ''central'' and ''nodal'' decoration of this structure are examined. It is shown that the electronic energy spectrum of a two-dimensional Fibonacci lattice contains a singular part, but in contrast to a one-dimensional Fibonacci lattice the spectrum does not contain a hierarchical gap structure. The measure of allowed states (Lebesgue measure) of the spectrum is different from zero, and for ''central'' decoration it is close to 1. The character of the localization of the wave functions is investigated, and it is found that the wave functions are ''critical.'' © *1999 American Institute of Physics.* [S1063-7761(99)02411-7]

1. INTRODUCTION

The electronic spectrum of a two-dimensional quasiperiodic structure has been studied for a Penrose lattice constructed from two structural units.[1] It has been shown that the density of states is strongly singular, but the measure of the allowed states of the spectrum (Lebesgue measure) is different from zero. At the same time, just as for a one-dimensional quasicrystal structure, the wave functions are critical, i.e., they are neither localized nor delocalized, but rather they decay according to a power law. Besides the Penrose lattice, it is of interest to investigate other two-dimensional structural quasicrystalline formations. In the present paper, the object of investigation is a two-dimensional quasiperiodic Fibonnaci sequence. The model is constructed as follows. Plotting a Fibonnaci sequence along the axes of a Cartesian coordinate system, we obtain a Fibonnaci lattice consisting of three structural units (Fig. 1). The areas of the three elementary structural units are in the ratio 1, τ, and $\tau^2=\tau+1$, where $\tau=(1+\sqrt{5})/2$ is the ''golden section.'' Just like a one-dimensional Fibonnaci lattice and a Penrose lattice, the two-dimensional Fibonnaci lattice is self-similar. Such a structure can be obtained in practice by, for example, lithography.

At present there does not exist a sufficiently general method for investigating the electronic properties of quasiperiodic structures. Since quasicrystals do not possess translational symmetry, the conventional methods for calculating the band structure of solids on the basis of Bloch's theorem are not directly applicable. For this reason, just as in Ref. 2, rational approximants of a two-dimensional Fibonnaci lattice were chosen as the object of investigation. The two-dimensional Fibonnaci lattice was treated as the structural limit of a sequence of rational approximants with increasing period.

2. ELECTRONIC SPECTRUM

The electronic properties of a two-dimensional Fibonnaci lattice were investigated in the tight-binding approximation by the method of the statistics of levels. To reduce the number of free parameters in the problem to a minimum, a Hamiltonian with constant transfer integrals for nearest neighbors was used. As the results for one-, two-, and three-dimensional quasicrystals show, this form of the Hamiltonian makes it possible to reproduce the characteristic features of a quasicrystal object and to examine qualitatively the effect of quasiperiodicity on the electronic structure of a quasicrystal with the corresponding dimension.[1–6] In this work, ''central'' and ''nodal'' decoration of the approximants by atoms of the same kind were examined: atoms with one s orbital per atom are located at the centers or at the nodes of the cells. For this case the Hamiltonian of the system can be written in the form

$$H=\sum_j |j\rangle\varepsilon_j\langle j|+\sum_{j,j\neq i}|i\rangle t_{ij}\langle j|.$$

When atoms of only one component are present in the system, the diagonal elements ε_j can be set equal to zero. Then the Schrödinger equation in the tight-binding approximation can be written in the form

$$\sum_j t_{ij}\psi_j=E\psi_i, \tag{1}$$

where the transfer integrals are proportional to r^{-2} (r is the distance between the atoms) only for nearest neighbors—for atoms separated by a distance not exceeding τ (taking account of the next neighbors does not introduce anything qualitatively new and complicates the calculation). To analyze the electronic properties of the object under study, the Hamiltonian matrix of Eq. (1) was diagonalized numerically for different approximants with periodic boundary conditions and the distribution of the energy levels was investigated.

The smoothness of the energy spectrum was investigated by the method of level statistics (LS method).[1,5,6] Here there are two key relations: The first one is the fraction of neighboring interlevel spacings $\Delta E\leqslant BN^{\beta}$

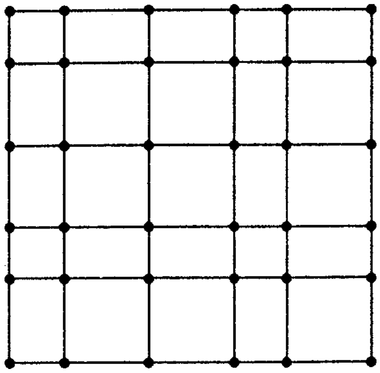

FIG. 1. Fragment of a two-dimensional Fibonacci lattice consisting of short S and long L segments, $L/S=\tau$, where $\tau=(1+\sqrt{5})/2$ is the ''golden section.''

$$D(\beta)=\frac{1}{N-1}\sum_{j=1}^{N-1}\theta\left(\beta-\log_N\left(\frac{\varepsilon_{j+1}-\varepsilon_j}{B}\right)\right), \tag{2}$$

where N is the number of atoms in the basis of the approximant, $B=\varepsilon_N-\varepsilon_1$ is the total width of the band, and θ is the Heaviside theta function. The second one is the fraction of the zone filled by interlevel spacings $\Delta E\leqslant BN^{\beta}$

$$F(\beta)=\frac{1}{B}\sum_{j=1}^{N-1}(\varepsilon_{j+1}-\varepsilon_j)\,\theta\left(\beta-\log_N\left(\frac{\varepsilon_{j+1}-\varepsilon_j}{B}\right)\right). \tag{3}$$

If $B=\text{const}$, then irrespective of the degree of smoothness of the spectrum, in the thermodynamic limit, these functions should satisfy the conditions[1,5,6]

$$D(\beta)=1 \quad \text{for } \beta>-1 \quad \text{and } F(\beta)=0 \quad \text{for } \beta<-1.$$

For crystal and amorphous systems (with smooth spectra) the curves $D(\beta)$ and $F(\beta)$ in the thermodynamic limit jump from 0 to 1 at $\beta=-1$.[1,5,6] For this reason, the energy spectrum is considered to be irregular or singular if in the thermodynamic limit the dependence of the interlevel spacings on the size of the system is different from the $1/N$ law.

The localization behavior of the wave functions of a two-dimensional Fibonnaci lattice was studied by the method of the statistics of $2p$-norms of the wave functions.[1,5] The $2p$-norm is defined as

$$\|\psi\|_{2p}\equiv\frac{\Sigma_n|\psi_n|^{2p}}{(\Sigma_n|\psi_n|^2)^p}, \tag{4}$$

where ψ_n are the amplitudes of the electronic wave function (expansion coefficients of the wave function in a tight-binding basis).

Treating the two-dimensional Fibonnaci lattice as the structural limit of a series of approximants with increasing period, the thermodynamic limit in the behavior of the curves describing the statistics of the distribution of $2p$-norms of the eigenvectors of the Hamiltonian of Eq. (1) can be found. Statistical analysis of the distribution of the $2p$-norms of the wave functions was performed by calculating $I_{2p}(\gamma)$,[1,5] describing the fraction of states for which the $2p$-norms satisfy $\|\psi\|_{2p}\leqslant N^{\gamma}$, i.e.,

$$I_{2p}(\gamma)\equiv\frac{1}{N}\sum_{n=1}^{N}\theta(\gamma-\log_N\|\psi^{(n)}\|_{2p}). \tag{5}$$

The classification of the wave functions was constructed using the normalization integral.[1] The wave functions are considered to be delocalized if

$$\int_{|\mathbf{r}|<R}|\psi(\mathbf{r})|^2 d\mathbf{r}\sim R^d,$$

where d is the dimension of the space, and localized if a normalization

$$\int_{\infty}|\psi(\mathbf{r})|^2 d\mathbf{r};$$

exists. The wave functions that cannot be normalized in an infinite system but are not delocalized are considered to be ''critical.''

3. RESULTS

Four approximants of a two-dimensional Fibonnaci lattice were investigated in this work: 5/3, 8/5, 13/8, and 21/13. For nodal decoration the unit cells of these approximants contain 64, 169, 441, and 1156 atoms, respectively, in the basis. For central decoration the unit cells of the approximants contain 49, 144, 400, and 1089 atoms, respectively, in the basis. Figure 2 shows plots of the integrated density of states, calculated for values of the energy levels at 21 **k**-points for the 8/5, 13/8, and 21/13 approximants of the Fibonnaci lattice.

It is evident in Fig. 2 that the curves of the integrated density of states converge quite rapidly and do not show the presence of a hierarchical gap structure, characteristic for a Cantor set of the spectrum of a one-dimensional quasicrystal. The quantities $D(\beta)$ and $F(\beta)$ were calculated for the 8/5, 13/8, and 21/13 approximants. The calculations show that the $D(\beta)$ curves for these three approximants almost converge to the thermodynamic limit in the region $\beta<-1$. This means that in the limit $N\to\infty$ all interlevel spacings are less than in conventional systems (for crystal and amorphous systems, in the thermodynamic limit the dependence of the magnitudes of these spacings on the size of the system is determined by the relation $\Delta E\sim 1/N^1$). A calculation of a dependence $F(\beta)$ for the approximants mentioned above in the region $\beta>-1$ shows that a finite fraction of the zone is filled by larger spacings than in conventional systems. Thus, the results obtained show that in the limit $N\to\infty$ the energy spectrum of the Fibonnaci lattice contains a singular part.

The density of states for the approximants 8/5, 13/8, and 21/13 was calculated by the method of triangles using a 21 **k**-point irreducible part of the Brillouin zone of the corresponding approximant. Curves of the electronic density of states are shown in Figs. 3 and 4. As follows from Fig. 4, for nodal decoration the form of the spectrum depends on the ''parity'' of the approximant, which attests to the existence of a topological feature of the two-dimensional Fibonacci lattice. The curves of the electronic density of states become less smooth and more ''peaked'' as the order of the approximant increases. This confirms the conclusion that the energy

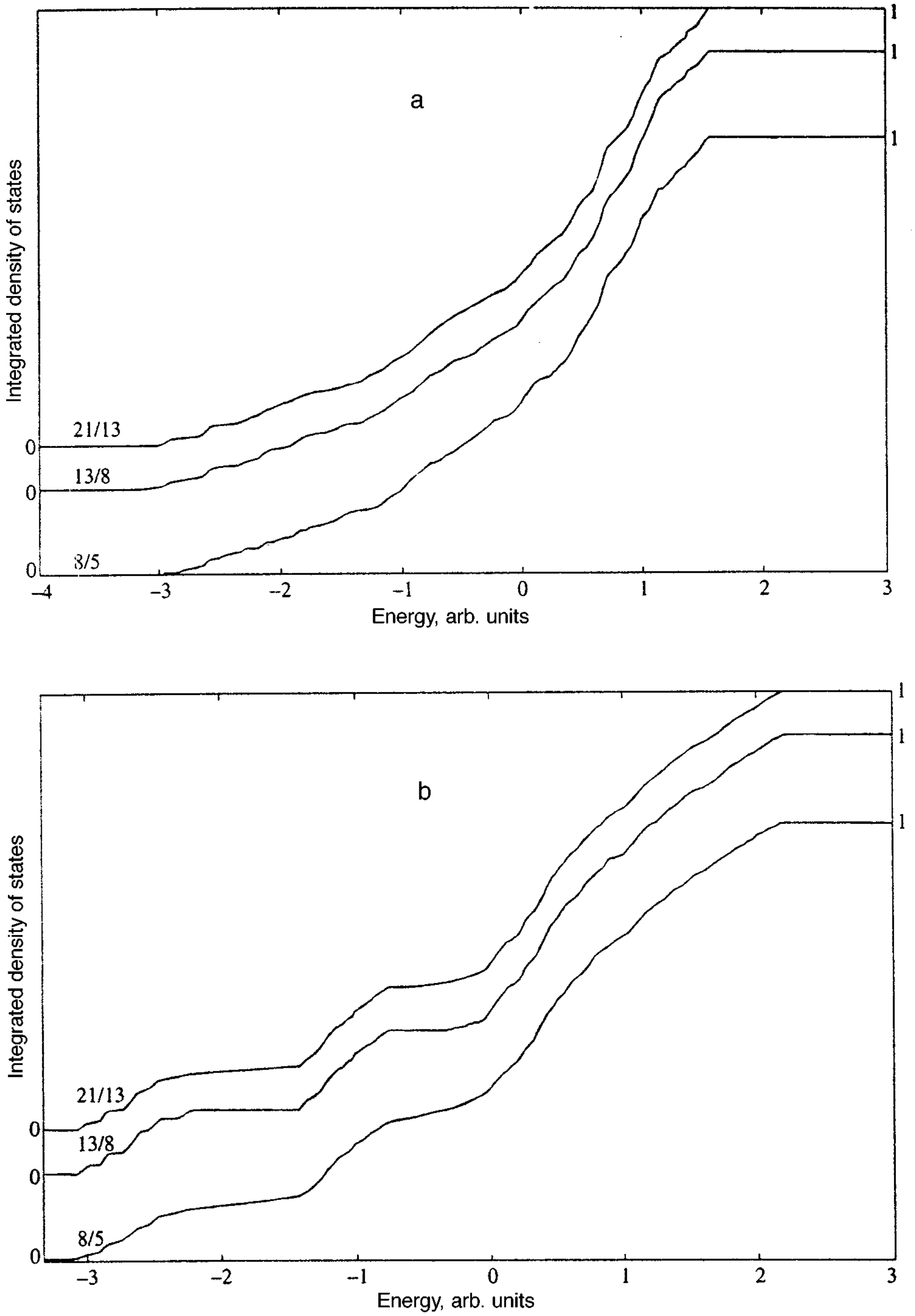

FIG. 2. Integral density of states for periodic approximants 8/5, 13/8, and 21/13 of a two-dimensional Fibonacci lattice with central decoration (a) and nodal decoration (b).

spectrum of the structure under study contains a singular part, since otherwise the electron density of states would converge to a smooth curve. The smoothness of the spectrum depends on the energy range: The energy spectrum is smoother at low energies and strong oscillations are present mainly at high energies. However, as the order of the approximant increases, the length of the smooth section of the energy spectrum decreases. For this reason, it can be assumed that in the thermodynamic limit (quasicrystal) strong oscillations of the density of states, which reflect the presence of a singular part in the spectrum, are present in the entire energy range of the electronic spectrum, attesting to the fact that the energy bands are flat, almost dispersion-free, with zero electron group velocity.

The calculation of the electronic density of states made it possible to determine the Lebesgue measure of the energy spectra of the corresponding approximants. The Lebesgue measure of the energy spectrum was calculated as the total length of the allowed sections of the spectrum. The energy spectra, orthonormalized to the Lebesgue measure $B=\varepsilon_N-\varepsilon_1$, of the approximants 5/3, 8/5, 13/8, and 21/13 were 1.0, 0.94, 0.98, and 1.0 for central decoration and 0.70, 1.0, 0.69, and 0.99 for nodal decoration. Therefore the measure of the

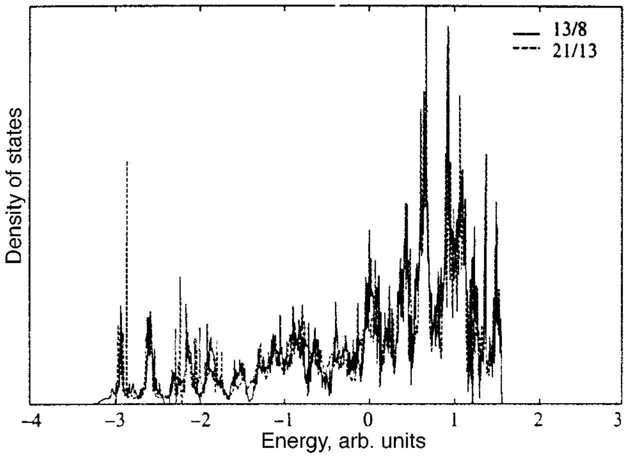

FIG. 3. Density of states for the periodic approximants 13/8 and 21/13 of a two-dimensional Fibonacci lattice with central decoration.

allowed sections of the spectrum is finite and close to 1 for central decoration.

For central decoration the dependence of the Lebesgue measure of the energy spectrum on the order of the approximant for a two-dimensional Fibonacci lattice is weak, in contrast to the one-dimensional quasicrystal, where the Lebesgue measure decreases as a power law as a function of the size of the system. The small difference in the Lebesgue measures of the energy spectra for the last two computed approximants shows that the energy spectrum of a two-dimensional Fibonacci lattice occupies a region of finite width on the energy scale. Moreover, values of the normalized Lebesgue measures of the energy spectra that are close to 1 indicate that there are no large gaps in the spectrum of a Fibonnaci lattice with central decoration.

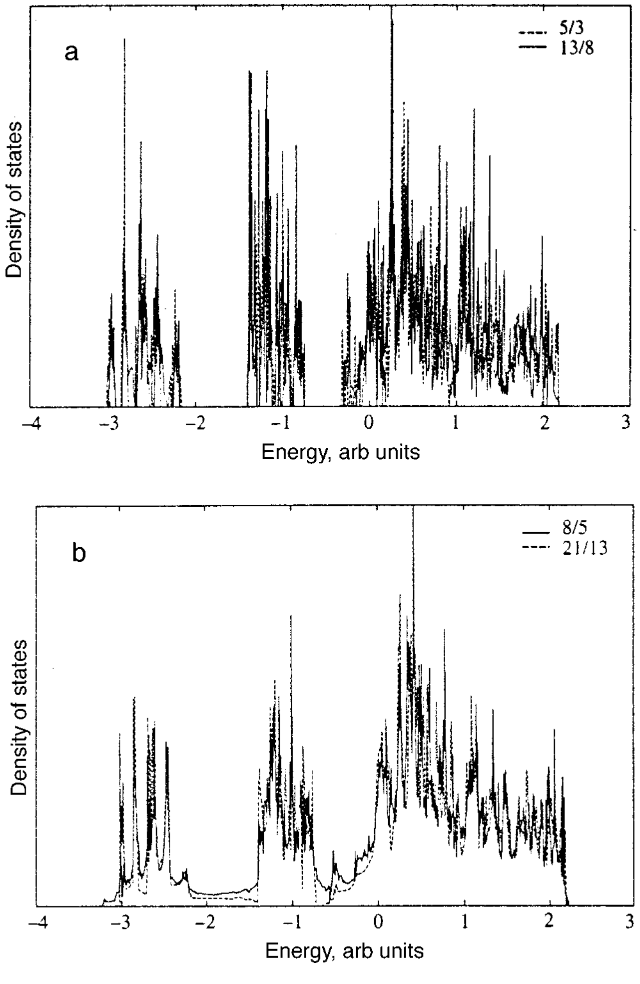

FIG. 4. Density of states for even periodic approximants 5/3 and 13/8 (a) and odd periodic approximants 8/5 and 21/13 (b) of a two-dimensional Fibonacci lattice with nodal decoration.

For nodal decoration a strong dependence on the parity of the approximant is observed. For odd approximants the Lebesgue measure, just as for central decoration, is close to 1. For even approximants the Lebesque measure is close to 0.7, which shows that a substantial gap is present in the spectrum. The results obtained are different from the Lebesque measures of the approximants of a Penrose lattice. The latter are approximately 0.62.[1] Accordingly, differences should be expected in the conducting properties of these two structures.

The quantities $I_8(\gamma)$ were calculated for the three approximants 8/5, 13/8, and 21/13 of the Fibonnaci lattice. The calculations show that in the thermodynamic limit the curves $I_8(\gamma)$ converge in a small region $\gamma \approx -2.2$ for central decoration and $\gamma \approx -2.3$ for nodal decoration. Moreover, as the numbers of the approximant increase, the curves $I_8(\gamma)$ approach a step function with the step located near $\gamma \approx -2.3$. This means that in the thermodynamic limit ($N \to \infty$) the $2p$-norms ($p=4$) of almost all wave functions of the Fibonacci lattice depend on the size of the system as $\|\psi\|_g \approx N^{\gamma}$, where $\gamma \approx -2.3$.

In the case of a delocalized state the $2p$-norm of the wave function, as follows from Eq. (4), depends on the size of the system as $\|\psi\|_{2p}^{\text{ext}} \approx N^{1-p}$, and the exponentially localized wave functions are characterized by the $2p$-norm $\|\psi\|_{2p}^{\text{exp loc}} \approx 1$. The dependence found for the $2p$-norms ($p=4$) of the wave functions of a Fibonacci lattice on the size of the system rules out delocalization and exponential localization of the states. Nonetheless, it is known that the dependence of the $2p$-norm of the wave function on the size of the system as $N^{\gamma(p,\alpha)}$ holds for wave functions whose squared moduli of the amplitude decrease as a power law[1,5] ($\alpha \geqslant 0$—localization exponent, where the cases $\alpha = 0$ and $\alpha \to \infty$ must be referred, respectively, to a delocalized state and to an exponential-localization state). For this reason, it was assumed that

$$|\psi|^2 \sim |\mathbf{r}|^{-2\alpha}. \tag{6}$$

The dependence $\gamma(p,\alpha)|_{p=4}$ with $-3 \leqslant \gamma \leqslant -1.7$ was calculated for sufficiently large quasiperiodic fragments on the basis of the computed $2p$-norm of the wave function (6). The results were obtained numerically for the 34/21 approximant of a Fibonacci lattice — before convergence to the thermodynamic limit. The calculations performed showed that $\gamma \approx -2.3$ corresponds to a localization exponent $\alpha \approx 0.48$ and $\alpha \approx 0.44$, respectively, for central and nodal decorations. According to the classification with respect to the normalization integral, wave functions exhibiting such behavior are critical, since the normalization of the functions with $|\psi|^2 \sim |\mathbf{r}|^{-2\alpha}$, $\alpha > 0$, in the two-dimensional case can be performed only for $\alpha > 1$.

In summary, for central decoration the energy spectrum does not contain large gaps, and the quasiperiodicity of the structure leads to strong oscillations in the entire energy range. Most wave functions of a two-dimensional Fibonacci lattice are critical and squared moduli of the amplitudes as a function of distance decrease according to a power law. The localization exponent α is 0.48 (central decoration) and 0.44 (nodal decoration), and it is the same in order of magnitude as the localization exponent $\alpha \approx 0.5$, obtained in Ref. 1 for a Penrose lattice. The energy spectrum of a Penrose lattice contains a singular part; the spectrum is smoother at low energies and strongly oscillatory at high energies. As the order of the approximant increases, the spectrum becomes more ''peaked.''[1] This agrees well with the results of the present work, which allows us to conclude that the properties of the electronic spectrum of various quasiperiodic two-dimensional structures, such as the Penrose lattice and the two-dimensional Fibonacci lattice, are general, although the conducting properties can be different.

4. CONCLUSIONS

The results of our investigation show that the electronic spectrum of a two-dimensional Fibonacci lattice possesses the same characteristic features as the spectrum of a Penrose lattice. The spectrum contains a singular part, but it is not self-similar. Most wave functions are ''critical,'' i.e., they are neither localized nor delocalized, but instead decay with increasing distance according to a power law. The measure of the allowed sections of the spectrum, in contrast to the one-dimensional structure, is finite. It is evident that the general properties of the spectra of different two-dimensional quasiperiodic structures should be similar. The spectrum of the structure studied differs from that of a Penrose lattice exist only with respect to the localization exponents of the wave functions and the measure of the allowed spacings. The wave functions of a two-dimensional Fibonacci lattice are less localized, and the meaure of the allowed spacings with central decoration is close to 1. The latter indicates that it could be possible to observe Ohm's law experimentally.

In conclusion, we thank D. V. Olenev for his attention to this work and for valuable remarks. This work was supported by the Swedish Royal Academy of Sciences.

*)E-mail: vekilov@trf.misa.ac.ru

---

[1] H. Tsunetsugu, T. Fujiwara, K. Ueda, and T. Tokihiro, Phys. Rev. B **43**, 8879 (1991).

[2] D. V. Olenev, É. I. Osaev, and Yu. Kh. Vekilov, Zh. Éksp. Teor. Fiz. **113**, 1009 (1998) [JETP **86**, 550 (1998)].

[3] P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, JETP Lett. **41**, 145 (1985).

[4] M. Kohmoto, B. Sutherland, and C. Tang, Phys. Rev. B **35**, 1020 (1987).

[5] H. Tsunetsugu, T. Fujiwara, K. Ueda, and T. Tokihiro, J. Phys. Soc. Jpn. **55**, 1420 (1986).

[6] L. Yamamoto and T. Fujiwara, Phys. Rev. B **51**, 8841 (1995).

Translated by M. E. Alferieff