Polarization of spin-1/2 particles in an axisymmetric magnetic field

A. Ya. Silenko*)

Nuclear Problems Research Institute, Belorussian State University, 220080 Minsk, Belarus
(Submitted 11 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 1153–1161 (October 1998)

The present paper shows that the nature of the polarization of charged spin-1/2 particles moving in a uniform magnetic field changes dramatically in a relatively weak transverse axisymmetric magnetic field. The direction along which the spin projection is quantized has a fixed orientation with respect to the axes of a cylindrical coordinate system and can form a substantial angle with the direction of the uniform magnetic field. The presence of spin quantization is proved both by the fact that the commutator of the Hamiltonian operator and the projection of the polarization operator in the direction of quantization is zero and by analyzing the Bargmann–Michel–Telegdi equation for this given case. Finally, the possibilities of detecting this effect and utilizing it are discussed. © *1998 American Institute of Physics.* [S1063-7761(98)00110-3]

1. INTRODUCTION

It is known that charged particles moving in a uniform magnetic field in a plane perpendicular the field's vector $\mathbf{H}^{(1)}$ have a fixed (quantized) value of the spin projection in the direction of the field. In this paper we will show that for spin-1/2 particles the nature of polarization changes dramatically in the presence of even a relatively weak transverse axisymmetric magnetic field whose field vector $\mathbf{H}^{(2)}$ is perpendicular to $\mathbf{H}^{(1)}$. Such a field is generated by a straight current collinear with $\mathbf{H}^{(1)}$ or by the current flowing in a toroidal solenoid whose axis is collinear with $\mathbf{H}^{(1)}$. The problem in this case is axisymmetric, and the symmetry axis is parallel to $\mathbf{H}^{(1)}$ and coincides with the direction of the straight conductor or the axis of the toroidal solenoid. As a result, the total magnetic field is nonuniform. Notwithstanding this fact, however, if a particle is moving in a plane perpendicular to $\mathbf{H}^{(1)}$ along a circular path whose center is on the symmetry axis, the projection of the particle's spin is quantized. In the given case the direction along which the projection of spin has a fixed value does not coincide with a single Cartesian axis but retains a constant orientation with respect to the axes of the cylindrical system of coordinates in which the moving particle is at rest. For electrons, positrons, and muons this direction may form a significant angle with $\mathbf{H}^{(1)}$ even if $|\mathbf{H}^{(2)}| \ll |\mathbf{H}^{(1)}|$.

In this paper we will use the weak-field approximation, which amounts to a condition that is usually met:

$$|\mathbf{H}^{(1)}+\mathbf{H}^{(2)}| \ll H_0, \quad H_0 = m^2c^3/|e|\hbar = 4.41\times 10^{13}\ \text{Oe}.$$

Throughout the work we use the relativistic system of units $\hbar = c = 1$.

2. PARTICLE POLARIZATION IN QUANTUM THEORY

For the sake of convenience we will adopt a cylindrical system of coordinates. The z axis of this system coincides with the symmetry axis of the problem, and the origin coincides with the center of the circular path of the particle (Fig. 1). In this geometry particles never move along the z axis and the vector $\mathbf{H}^{(2)}$ of the additional magnetic field is always directed along the tangent to the particle path.

Even if we use quantum mechanics to examine particle polarization, particle motion can be described semiclassically, which means we can speak of a particle path (see Refs. 1–3). The accuracy of such a description can easily be determined via the Heisenberg uncertainty principle. Classical theory makes it possible to find particle momentum with a relative error

$$\frac{|\delta p|}{p} \sim \frac{1}{rp} = \frac{1}{l},$$

where l is the orbital angular momentum of the particle, and r is the characteristic size of the region in which a particle moves (the radius of the circular orbit in our case). The radius of the circular orbit is given by the formula

$$r = -\frac{p_\phi}{eH_z}, \tag{1}$$

where p_ϕ is the projection of the particle momentum. Hence the relative error acquired in the process of going over to the semiclassical description is of order

$$\frac{|\delta p|}{p} \sim \frac{|e|H_z}{p^2} = \frac{H_z}{H_0(\gamma^2-1)}, \tag{2}$$

where γ is the Lorentz factor. In most cases of practical interest, this ratio is very small ($H_z/H_0 \approx 5\times 10^{-10}$ for $H_z = 2\times 10^4$ Oe). Equation (2) also yields the error acquired when one ignores the commutators of the operators of dynamical variables, since the absolute value of the commutator of the operators of position $\mathbf{r}$ and momentum $\mathbf{p} \equiv i\nabla$ is 3 ($|[\mathbf{r},\mathbf{p}]=3|$) and the ratio $|[\mathbf{r},\mathbf{p}]|/rp$ is of order $H_z/[H_0(\gamma^2-1)]$, which means that the ordering of the given operators in quantum mechanical expressions can be

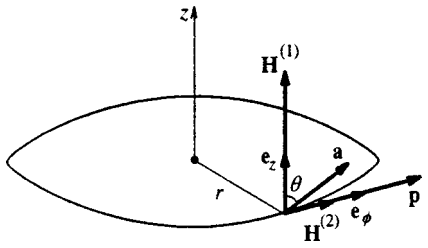

FIG. 1. Particle polarization in an axisymmetric magnetic field: r is the radius of the orbit, $\mathbf{e}_\phi$ and $\mathbf{e}_z$ are the unit vectors of the cylindrical coordinate system, $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$ are the vectors of the uniform and transverse nonuniform magnetic fields, $\mathbf{p}$ is the particle momentum, $\mathbf{a}$ is the vector that specifies the direction of the spin's orientation, and θ is the angle between $\mathbf{a}$ and the z axis.

assumed arbitrary. Note that particle polarization has a very small effect on particle motion in a magnetic field.[2]

We will study particle polarization without imposing restrictions on the particle energy and will use the Foldy–Wouthuysen (FW) representation, whose merit is the simplicity in describing polarization effects. The polarization operator in this representation reduces to the matrix $\mathbf{\Pi}=\beta\mathbf{\Sigma}$ (Refs. 3 and 4), where

$$\mathbf{\Pi}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & -\boldsymbol{\sigma}\end{pmatrix},\quad \beta=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \mathbf{\Sigma}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix},$$

with σ the Pauli matrix, and 0 and ± 1 the respective 2-by-2 matrices. If to describe the state of a particle we use only the upper spinor, which is always possible in the FW representation, the polarization operator is proportional to the spin operator.

Spin motion is essentially quantum because the operators Σ_i and Σ_j do not commute if $i\neq j$.

The projection of the polarization vector in a certain direction is conserved if the corresponding projection of the polarization operator commutes with the Hamiltonian operator. The Hamiltonian operator in the FW representation for relativistic spin-$\frac{1}{2}$ particles with an anomalous magnetic moment (AMM) that are in motion in an electromagnetic field was found in Ref. 5–7. In Refs. 5 and 7 this operator was obtained for relativistic particles (in Ref. 5 without allowing for derivatives of the external field strengths, and in Ref. 7 with allowance for such derivatives). A Hamiltonian that allows for relativistic corrections, including corrections that take into account the derivatives of field strengths, was derived in Ref. 6. The results of Refs. 5–7 agree with one another.

The expression found in Refs. 5 and 7 for the Hamiltonian of relativistic particles moving in a magnetic field without allowance for terms with derivatives of field strengths has the form

$$\mathscr{H}=\beta\epsilon-\left(\frac{\mu_0 m}{\epsilon}+\mu'\right)\mathbf{\Pi}\cdot\mathbf{H}+\frac{\mu'}{\epsilon(\epsilon+m)}(\mathbf{\Pi}\cdot\mathbf{p})(\mathbf{H}\cdot\mathbf{p}), \tag{3}$$

where $\epsilon=\sqrt{\boldsymbol{\pi}^2+m^2}$ is the kinetic energy operator, $\boldsymbol{\pi}=\mathbf{p}-e\mathbf{A}$ and $\mathbf{p}$ are the kinetic momentum and momentum operators, with $\mathbf{A}$ the vector potential of the field, and $\mu_0=e/2m$ and $\mu'=\mu-\mu_0$ are the Dirac and anomalous magnetic moments, with μ the total magnetic moment. Going over to an arbitrary ordering of the operators in (3) is equivalent to ignoring the commutators of these operators. For an axisymmetric magnetic field

$$|[p_i,\mathbf{H}]|=\left|\frac{\partial\mathbf{H}}{\partial x_i}\right|\sim\frac{|\mathbf{H}^{(2)}|}{r}. \tag{4}$$

Combining this with (1), we get

$$\frac{|[p_i,\mathbf{H}]|}{|p_i\mathbf{H}|}\sim\frac{|e\mathbf{H}^{(2)}|}{p^2}=\frac{|\mathbf{H}^{(2)}|}{H_0(\gamma^2-1)}.$$

The ratio $|[p_i,\mathbf{H}]|/|p_i\mathbf{H}|$ specifies the relative error introduced by a permutation of noncommuting operators in (3), and the fact that this ratio is small makes it possible to use an arbitrary ordering of the operators.

The corrections to the Hamiltonian containing terms linear in the field with field strength derivatives are given by the formula[7]

$$\mathscr{H}'=\beta\frac{\mu'}{4\epsilon(\epsilon+m)}\{2(\text{curl }\mathbf{H})\cdot\mathbf{p}+\mathbf{\Sigma}\cdot(\text{curl curl }\mathbf{H})\}.$$

At $\epsilon\approx m$ this expression becomes the formula derived in Ref. 6.

Since in our case curl $\mathbf{H}=0$, we have $\mathscr{H}'=0$, which means that the Hamiltonian operator contains no corrections corresponding to the nonuniformity of the field.

We write the Hamiltonian (1) as a sum of two terms:

$$\mathscr{H}=\mathscr{H}_1+\mathscr{H}_2,\quad \mathscr{H}_1=\beta\epsilon-\frac{\mu_0 m}{\epsilon}\mathbf{\Pi}\cdot\mathbf{H}, \tag{5}$$

$$\mathscr{H}_2=-\mu'\mathbf{\Pi}\cdot\mathbf{b},\quad \mathbf{b}=\mathbf{H}-\frac{(\mathbf{H}\cdot\mathbf{p})\mathbf{p}}{\epsilon(\epsilon+m)},$$

where $\mathscr{H}_1$ is the Hamiltonian of a Dirac particle. Next we calculate the commutators of $\mathscr{H}$ and the operators $\Pi_\phi=-\Pi_x\sin\phi+\Pi_y\cos\phi$ and Π_z. For $\sin\phi$ and $\cos\phi$ we must take into account only the commutators with the operator $\beta\epsilon$. The following commutation relations hold (here $\{\cdots,\cdots\}_+$ stands for an anticommutator):

$$[\boldsymbol{\pi}^2,\Pi_\phi]=i\beta\left\{\pi_\phi,\frac{\Sigma_\rho}{\rho}\right\}_+, \tag{6}$$

$$[\Pi_z,\Pi_\phi]=-2i\Sigma_\rho, \tag{7}$$

where $\Sigma_\rho=\Sigma_x\cos\phi+\Sigma_y\sin\phi$.

The operators π_ϕ and ρ are defined on the class of functions that are the eigen-wave-functions of $\mathscr{H}$. Hence, with allowance for the semiclassical nature of the motion and the possibility of ignoring the noncommutativity of dynamical variables, we can introduce an approximation in which the operators π_ϕ and ρ are replaced by the values of the particle momentum p_ϕ and the radius r of the orbit, respectively. Then

$$[\boldsymbol{\pi}^2,\Pi_\phi]\approx 2i\beta\frac{p_\phi}{r}\Sigma_\rho.$$

If we allow for the expression for r, the above equation reduces to

$$[\boldsymbol{\pi}^2,\Pi_\phi]\approx -2i\beta eH_z\Sigma_\rho.$$

Following a method similar to the one used in Ref. 2, we arrive at the relationship

$$[\boldsymbol{\pi}^2,\Pi_\phi]=[\epsilon^2,\Pi_\phi]=\{\epsilon,[\epsilon,\Pi_\phi]\}_+\approx 2\epsilon[\epsilon,\Pi_\phi],$$

so that

$$[\epsilon,\Pi_\phi]\approx\frac{1}{2\epsilon}[\boldsymbol{\pi}^2,\Pi_\phi]\approx -i\beta\frac{eH_z}{\epsilon}\Sigma_\rho.$$

Since $\mathbf{\Pi}\cdot\mathbf{H}=\Pi_\phi H_\phi+\Pi_z H_z$, we use formula (4) and get

$$[\mathscr{H}_1,\Pi_\phi]=0. \tag{8}$$

Since the vectors $\mathbf{p}$ and $\mathbf{e}_\phi$ are collinear, the fact that the commutator of $\mathscr{H}_1$ and Π_ϕ is zero agrees with the fact that the projection of the polarization operator in the direction of the momentum of Dirac particles is conserved (the latter has been proved for the Dirac representation by Sokolov *et al.*[8]).

Note that H_ϕ is independent of ϕ but may depend on ρ and z. However, the commutator $[\boldsymbol{\pi},H_\phi]$ is small compared to $\boldsymbol{\pi}H_\phi$.

Combining Eqs. (5), (7), and (8), we arrive at the following relationships:

$$[\mathscr{H},\Pi_\phi]=[\mathscr{H}_2,\Pi_\phi]\approx -2i\mu'\mathbf{\Sigma}\cdot(\mathbf{b}\times\mathbf{e}_\phi)$$
$$=2i\mu'H_z\Sigma_\rho, \tag{9}$$

$$[\mathscr{H},\Pi_z]=-2i\frac{\mu_0 m}{\epsilon}H_\phi\Sigma_\rho-2i\mu'\mathbf{\Sigma}\cdot(\mathbf{b}\times\mathbf{e}_z)$$
$$=-2i\frac{\mu}{\gamma}H_\phi\Sigma_\rho.$$

We now introduce a vector $\mathbf{a}$ in such a way that the projection of the polarization operator on this vector, Π_a, assumes a quantized value. Hence the commutator of $\mathscr{H}$ and Π_a is zero:

$$[\mathscr{H},\Pi_a]=0. \tag{10}$$

Equations (9) imply that $\mathbf{a}$ can be written as follows:

$$\mathbf{a}=a_\phi\mathbf{e}_\phi+a_z\mathbf{e}_z,$$

where a_ϕ and a_z are scalar operators. Since

$$\Pi_a=\frac{\mathbf{\Pi}\cdot\mathbf{a}}{|\mathbf{a}|}=\frac{\Pi_\phi a_\phi+\Pi_z a_z}{|\mathbf{a}|}$$

and since we ignore the fact that the operators of dynamical variables do not commute, the condition (10) yields

$$\mu'H_z a_\phi=\frac{\mu}{\gamma}H_\phi a_z.$$

Using the standard notation

$$g=\frac{\mu}{s}\frac{2m}{e}=\frac{4\mu m}{e},$$

we can transform the above relationship to

$$\frac{a_\phi}{a_z}=\frac{g}{(g-2)\gamma}\frac{H_\phi}{H_z}. \tag{11}$$

The angle θ between the direction of $\mathbf{a}$ and the z axis (see Fig. 1) is specified by the formula

$$\tan\theta=\frac{g}{|g-2|\gamma}\left|\frac{H_\phi}{H_z}\right|. \tag{12}$$

Since the operator Π_a commutes with the Hamiltonian, in stationary states it has definite values λ. Since

$$\Pi_a\Pi_a=(\mathbf{\Pi}\cdot\mathbf{a})(\mathbf{\Pi}\cdot\mathbf{a})/\mathbf{a}^2=1,\quad \Pi_a\Pi_a\psi=\lambda^2\psi,$$

we have $\lambda=\pm 1$. The projection of the spin operator (in the FW representation, the spin operator is proportional to the polarization operator) in the direction of $\mathbf{a}$ takes quantized values $\pm 1/2$. Since the vector $\mathbf{a}$ has constant projections on the axes of the cylindrical coordinate system, it rotates with an angular velocity

$$\boldsymbol{\omega}=-\frac{e}{\epsilon}\mathbf{H}^{(1)},$$

which is equal to the angular velocity of particle rotation in the magnetic field. In stationary states, the average values of the projections of the polarization and spin operators on directions perpendicular to $\mathbf{a}$ are equal to zero.

3. RELATION TO THE BARGMANN–MICHEL–TELEGDI EQUATION

Let us now establish how the above results are related to the Bargmann–Michel–Telegdi (BMT) equation,[9] which describes spin motion in an electromagnetic field. A characteristic feature of this motion is that the polarization vector ξ varies, and in our case this variation is given by the equation

$$\frac{d\boldsymbol{\xi}}{dt}=\left(\frac{d\boldsymbol{\xi}}{dt}\right)_{\rm BMT}$$
$$=\frac{e}{2m}\left(g-2+\frac{2}{\gamma}\right)[\boldsymbol{\xi}\times\mathbf{H}]$$
$$+\frac{e}{2m}(g-2)\frac{\gamma}{\gamma+1}(\mathbf{v}\cdot\mathbf{H})([\mathbf{v}\times\boldsymbol{\xi}]). \tag{13}$$

This equation can be derived not only semiclassically[1] but also by methods of quantum theory.[4] It can be shown that in our case the corrections to Eq. (13) obtained by Good[10] and Nyborg[11] have a negligible effect on spin motion. These corrections amount to adding to the right-hand side of Eq. (13) the term[1)]

$$\left(\frac{d\boldsymbol{\xi}}{dt}\right)_{\rm GN}=\frac{\mu\gamma}{m(\gamma+1)}$$
$$\times[\boldsymbol{\xi}\times(\mathbf{v}\times\nabla)]\left[\boldsymbol{\xi}\cdot\mathbf{H}-\frac{\gamma}{\gamma+1}(\boldsymbol{\xi}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{H})\right]. \tag{14}$$

Using Eqs. (1) and (4), we find that the ratio

$$\left|\left(\frac{d\boldsymbol{\xi}}{dt}\right)_{\mathrm{GN}}\right| \Big/ \left|\left(\frac{d\boldsymbol{\xi}}{dt}\right)_{\mathrm{BMT}}\right|$$

does not exceed, in order of magnitude,

$$\frac{\gamma v|\mathbf{H}^{(2)}|}{rmH}=\frac{p|\mathbf{H}^{(2)}|}{rm^2H}\sim\frac{|\mathbf{H}^{(2)}|}{H_0}\ll 1.$$

Thus, when we examine spin motion in an axisymmetric magnetic field, the corrections to the BMT equation given by (14) can be ignored.

From Sec. 2 it follows that the projection of the polarization vector in the direction of **a** remains constant ($d\xi_a/dt=0$). Equation (13) can be written as follows:

$$\frac{d\boldsymbol{\xi}}{dt}=\boldsymbol{\Omega}\times\boldsymbol{\xi}, \tag{15}$$

$$\boldsymbol{\Omega}=-\frac{e}{2m}\left(g-2+\frac{2}{\gamma}\right)\mathbf{H}+\frac{e}{2m}(g-2)\frac{\gamma}{\gamma+1}(\mathbf{v}\cdot\mathbf{H})\mathbf{v}.$$

The unit vector $\mathbf{a}_0=\mathbf{a}/|\mathbf{a}|$ rotates with an angular velocity ω:

$$\frac{d\mathbf{a}_a}{dt}=\boldsymbol{\omega}\times\mathbf{a}_0=-\frac{e}{\epsilon}\mathbf{H}^{(1)}\times\mathbf{a}_0. \tag{16}$$

Note that the equation of particle motion, which yields ω, is valid both in classical theory and in quantum theory.[2,13] Since

$$\frac{d\xi_a}{dt}=\frac{d}{dt}(\boldsymbol{\xi}\cdot\mathbf{a}_0)=(\boldsymbol{\Omega}\times\boldsymbol{\xi})\cdot\mathbf{a}_0+\boldsymbol{\xi}\times(\boldsymbol{\omega}\times\mathbf{a}_0)$$

$$=[(\boldsymbol{\Omega}-\boldsymbol{\omega})\times\boldsymbol{\xi}]\cdot\mathbf{a}_0=0,$$

the vectors $\mathbf{a}_0$ and **a** are collinear with the vector $\mathbf{o}=\boldsymbol{\Omega}-\boldsymbol{\omega}$, which is the angular velocity of spin precession in the reference frame in which the moving particle is at rest. Equations (15) and (16) imply that

$$\mathbf{o}=-\frac{e}{2m}\left[\frac{g}{\gamma}H_\phi\mathbf{e}_\phi+(g-2)H_z\mathbf{e}_z\right]. \tag{17}$$

This formula describes spin motion in a cylindrical coordinate system. Since $\mathbf{a}_0=\mathbf{o}/|\mathbf{o}|$, formula (11) fully agrees with (17). Thus, the results obtained via a rigorous quantum mechanical method coincide with those obtained via the BMT equation, which also holds in quantum theory.

4. DISCUSSION AND CONCLUSIONS

At present it is believed that quantization of spin projection is possible only in a uniform field. The present paper shows that the spin projection of particles moving in an axisymmetric magnetic field in a plane perpendicular to the symmetry axis also has a fixed (quantized) value. Here particle polarization exhibits a number of characteristic features. What sets this case apart from that of a uniform field is that the direction along which the spin projection is quantized has a fixed orientation with respect to the axis of a cylindrical coordinate system in which the particle moving along a circular orbit is at rest, rather to a Cartesian coordinate system. Note that this direction forms a fixed angle with the particle momentum vector. The fact that there is spin quantization can be proved either by calculating the commutator of the Hamiltonian operator and the projection of the polarization operator in the direction of quantization (the direction of **a**) and finding that this commutator is zero, or by the fact that the corresponding projection of the polarization vector is conserved (which follows from the BMT equation).

In our case the corrections due to the nonuniformity of the magnetic field are small and have no effect on particle polarization.

Of special interest is the fact that for electrons, positrons, and muons the change in the nature of quantization of spin in the presence of a transverse magnetic field is extremely strong. For these particles the ratio of the total magnetic moment to the anomalous is

$$\frac{\mu}{\mu'}=\frac{g}{g-2}\approx\frac{2\pi}{\alpha}\sim 10^3.$$

Since already at $H_\phi\sim 10^{-3}H_z$ the particles are polarized at a substantial angle to the z axis, and at $H_\phi=10^{-2}H_z$, $\gamma=1.2$, for electrons we have $\theta=82°$, i.e., the direction in which the electrons are polarized is almost perpendicular to the direction of the uniform field. Due to the huge value of the effect, its observation in experiments becomes much easier.

The change in the nature of polarization affects the evolution of the particle polarization vector. By measuring the temporal variation of this vector we can determine the direction along which the spin projection is quantized, since **a** is collinear with the vector **o** of the angular velocity of spin precession in the reference frame in which the particle is at rest. The orientation of the polarization vector of the particle beam remains unchanged only when it is collinear with **a**.

Another method used in studying the effect of a transverse axisymmetric magnetic field on particle polarization in stationary states is related to radiative polarization of particles in a magnetic field (the Sokolov–Ternov effect[14]). In a uniform magnetic field the degree of polarization of particles in a beam is as high as 92.4 %. Naturally, radiative polarization also occurs when there is a transverse magnetic field. In stationary states, the particles are polarized either parallel or antiparallel to **a**. The probability of a transition with spin flip depends on the value of the spin projection in the direction of **a**, which leads to partial polarization of the beam. The direction of polarization collinear with **a** forms an angle with the z axis, and by measuring this angle we can determine the extent to which the theory agrees with the experiment.

Note that the characteristic features of radiative polarization in the presence of a transverse magnetic field can find practical applications in forming beams of electrons and positrons partially polarized at a given angle to the direction of particle momentum. This angle, equal to $\pi/2-\theta$, depends on the strength of the transverse field and can be calculated by Eq. (12).

Of course, a transverse magnetic field makes the motion of particles along circular paths somewhat unstable. However, since the components of the particle velocity along the ρ and z axes are small, this factor has only a small effect on the particle polarization, although actually the particles move not along a circle but along the surface of a very thin and narrow ring (its shape may differ from that of a toroid).

Another possibility of utilizing a transverse axisymmetric magnetic field is related to the geonium quasiatom. The configuration of fields in the geonium quasiatom (a uniform magnetic field and a quadrupole electric field), which specifies a Penning trap, makes it possible to attain a record-breaking accuracy in measuring the electron magnetic moment. Measuring polarization in the presence of a transverse magnetic field leads to a shift in the energy levels of geonium and makes it possible to select a configuration of the levels that enhances the accuracy of measuring the magnetic moments of the particles inside the Penning trap. This problem requires a separate investigation.

[1)]The corrections to the BMT equation yielded by (14) (which was derived by classical methods) are not identical to the correction derived in quantum theory.[12]

*)E-mail: silenko@inp.minsk.by

[1] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Nauka, Moscow (1989), p. 179 [Pergamon Press, Oxford (1991)].
[2] B. N. Baĭer, V. M. Katkov, and V. S. Fadin, *Emission of Radiation by Relativistic Electrons*, Atomizdat, Moscow (1973), p. 65.
[3] I. M. Ternov, V. R. Khalilov, and V. N. Rodionov, *Interaction of Charged Particles with an External Electromagnetic Field*, Moscow Univ. Press, Moscow (1982), p. 43.
[4] D. M. Fradkin and R. H. Good, Rev. Mod. Phys. **33**, 343 (1961).
[5] S. R. deGroot and L. G. Suttorp, *Foundations of Electrodynamics*, North-Holland, Amsterdam (1972).
[6] A. I. L'vov, *Preprint P. N. Lebedev Physics Institute No. 344*, Moscow (1987), p. 33.
[7] A. Ya. Silenko, Teoret. Mat. Fiz. **105**, 46 (1995) [Theor. Math. Phys. **105**, 1224 (1995)].
[8] A. A. Sokolov, I. M. Ternov, V. G. Bagrov, and R. A. Rzaev, in *Synchrotron Radiation*, Nauka, Moscow (1966), p. 72.
[9] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).
[10] R. H. Good, Phys. Rev. **125**, 2112 (1962).
[11] P. Nyborg, Nuovo Cimento **31**, 1209 (1964).
[12] A. Ya. Silenko, Poverkhnost' No. 2, 111 (1998) [Surface Investig. **13**, 251 (1998)].
[13] A. Messiah, *Quantum Mechanics*, North-Holland, Amsterdam (1962).
[14] A. A. Sokolov and I. M. Ternov, Dokl. Akad. Nauk SSSR **153**, 1052 (1963) [Sov. Phys. Dokl. **8**, 1203 (1963)].

Translated by Eugene Yankovsky

Concentration of lithium vapor in a high-temperature sapphire capillary based on light-induced drift

O. A. Vostrikov, K. A. Nasyrov, S. P. Pod'yachev, and A. M. Shalagin*)

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia
(Submitted 14 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1162–1170 (October 1998)

The accumulation of atomic lithium vapor in a sapphire capillary based on light-induced drift is experimentally investigated. To suppress lithium adsorption on the walls of the capillary and prolong the life of the atomic state, the capillary was heated to high temperatures. The phenomenon of an ''optical piston'' was observed, indicating a high degree of concentration of atomic lithium vapor. The experimental results are consistent with theoretical ideas, and they suggest that heated cells hold some promise for efficient isotope separation and the detection of atomic micro-impurities. © *1998 American Institute of Physics.* [S1063-7761(98)00210-8]

1. INTRODUCTION

Light-induced drift was first predicted in 1979 (Ref. 1). In that same year it was experimentally observed in atomic gases,[2] and soon thereafter in molecular gases.[3] Since then, light-induced drift has been vigorously investigated, both experimentally and theoretically (a review of the literature can be found, for example, in Refs. 4–6). Recently, interest has grown in the question of separating and accumulating isotopes of various chemical elements in their atomic states with the help of light-induced drift. In this regard, it is worth noting Refs. 7–9, which report separation of Na isotopes and accumulation of a radioactive Na isotope in a heated capillary, and also Ref. 10, which examines the separation of Li isotopes under conditions in which the walls are far from the sample itself.

The major impediment to the accumulation of one or another chemical element with the help of light-induced drift is adsorption on the walls of the cell or capillary, where accumulation properly takes place. To eliminate or at least reduce the influence of adsorption, it is necessary either to specially coat the walls of the cell[11–13] or heat the cell.[7–9] The second alternative is preferable due to its universal suitability for many chemical elements.

The present paper investigates the accumulation of atomic Li vapor with the help of light-induced drift in a heated sapphire capillary.

2. EXPERIMENT

For the purpose of this study, we assembled an experimental setup (see Fig. 1) similar to the one used in Ref. 9. Radiation from a dye laser *2* (an Ar^+ laser *1* was used as the pump) was focused by a collecting lens *3* ($F=1$ m) into the working cell *4* and directed into a sapphire capillary *5*, which was closed at one end. The length of the capillary was 10 cm, and its inner diameter was 1.5 mm. The capillary was placed inside a tantalum coil *6* with a pitch of 1 mm. Leads were connected to the coil through windows in the cell. A voltage from the power supply *7* was applied to the coil through the leads, which made it possible to heat the capillary. Typical temperatures reached 1300–1600 K. The temperature was monitored by a remote pyrometer.

The necessary lithium vapor concentration was produced with the help of an electrically heated lithium container *10*. A diaphragm *11* separated the region of the cell where the lithium vapor was generated from the region housing the heated sapphire capillary.

The working cell *4* was connected to a vacuum station for inlet and outpumping of the buffer gas. The vacuum was 10^{-5} Torr or better. We used argon as the buffer gas (pressure 20 Torr).

DCM dye was used as the working medium in the dye laser. The laser frequency was tuned with the help of an intra-cavity Fabry–Perot interferometer. The laser linewidth was ≈ 1.5 GHz. The power of the dye laser ranged from 50 to 150 mW.

To record the accumulation of lithium vapor we used a camera *9* mounted on a tripod 60 cm from the working cell. To separate the fluorescence radiation from the background of the incandescent coil, we used an interference filter *8*.

The experiment consisted in the following. The laser frequency was tuned to the D_2 line of ^{7}Li (note that for such tuning under the conditions of our experiment only the given isotope is excited). The offset from exact resonance was chosen to be such that the ^{7}Li isotope migrated in the direction of laser beam propagation. Figure 2 plots the dependence of the fluorescence intensity on the frequency of the exciting radiation. The ^{6}Li line is enhanced because the spectrum was recorded above the saturation intensity. The vertical line shows the experimental laser frequency. The offset was chosen to maximize light-induced drift.

At the initial instant, just after turning the laser on, only the characteristic glow of the sapphire capillary and coil were observed (Fig. 3a) and fluorescence of the lithium vapor was absent. After some time a luminous region appeared at the closed end of the capillary (Fig. 3b). With the passage

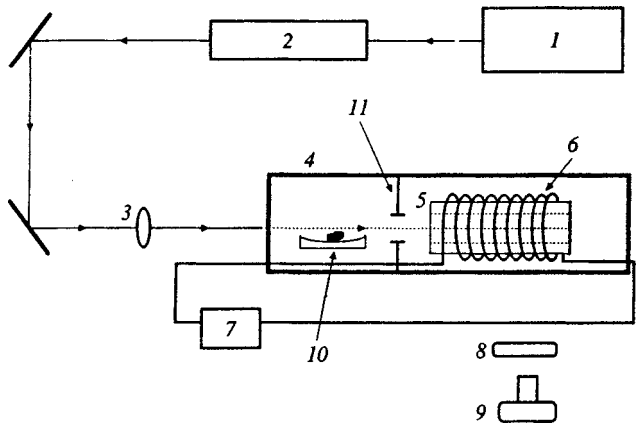

FIG. 1. Experimental setup: *1* — Ar^+ laser (pump laser), *2* — dye laser, *3* — focusing lens, *4* — cell, *5* — sapphire capillary, *6* — tantalum wire, heated by a current from the power supply *7*, *8* — interference filter, *9* — camera, *10* — lithium vapor source, *11* — diaphragm.

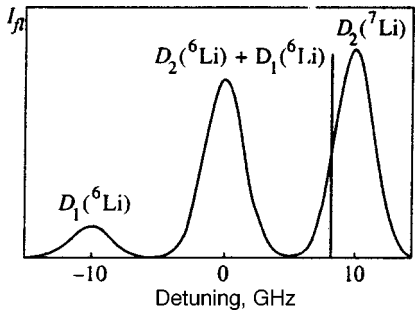

FIG. 2. Lithium fluorescence spectrum. Vertical line indicates position of the laser frequency in the experiment on concentrating lithium vapor.

of time, the size and brightness of this region increased (Figs. 3c and 3d, respectively). With the passage of time, the brightest part of the luminous region clearly shifts toward the open end of the capillary. After some time, the luminous region ''breaks away'' from the closed end of the capillary (Fig. 3e). The luminous region then moves toward the open end of the capillary (Fig. 3f).

The typical time τ for the process to develop from start to finish (Figs. 3a and 3f) depended on the temperature of the sapphire capillary and the vapor concentration at the entrance to the capillary. The latter was chosen to be such that the lithium diffusing into the capillary did not produce any appreciable visible luminescence (Fig. 3a). As a result, at $T \approx 1500$ K the time τ was several minutes.

These observations can be explained by invoking ideas about the phenomenon of light-induced drift in the following way. Lithium atoms initially move in the direction of the wave vector toward the closed end of the capillary as a result of light-induced drift. The greater the number of lithium atoms that accumulate there, the more the light absorption region migrates in the opposite direction: radiation is essentially completely absorbed over a distance $l_{\text{eff}} \approx 2$ cm along the capillary. Thus emerges the well-known ''light piston'' effect (see Refs. 14–17).

3. THEORY

We consider a two-component gaseous medium, in which one component can interact resonantly with radiation, and the second component serves as a buffer gas. Under conditions in which the concentration of the absorbing component gas is much less than that of the buffer gas, it is possible to neglect collisions of atoms of the first component with one another. We also assume that a two-level model describes the internal states of the atoms of the first component; the model takes only two quantum states of the atom into account (the ground state and the first excited state), with resonant absorption taking place between them. We denote the lower and upper states by n and m, respectively.

The equations for the velocity distribution functions ρ_m and ρ_n for the atoms in the indicated states have the form[4]

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} + \Gamma_m\right)\rho_m = Np(\mathbf{v}) + S_m,$$

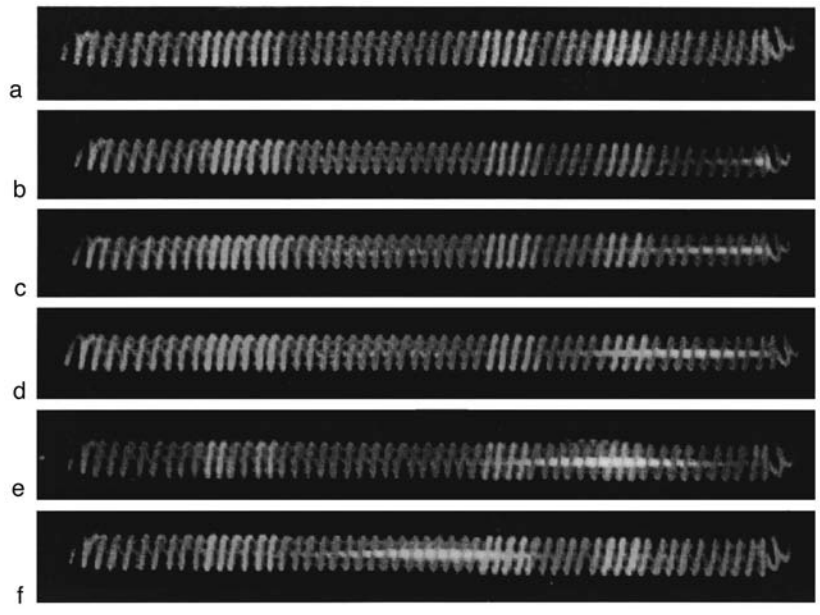

FIG. 3. Time evolution of accumulation of lithium vapor in a sapphire capillary and formation of an ''optical piston.''

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\rho_n-\Gamma_m\rho_m=-Np(\mathbf{v})+S_n, \tag{1}$$

$$Np(\mathbf{v})=\frac{2|G|^2\Gamma}{\Gamma^2+(\Omega-\mathbf{k}\cdot\mathbf{v})^2}(\rho_n-\rho_m).$$

Here we have introduced the following notation: $\Omega=\omega_L-\omega_{mn}$ is the frequency offset of the laser radiation ω_L from the center of the absorption line at the transition frequency ω_{mn}; Γ_m is the spontaneous decay rate of the atom from the upper to the lower state; Γ is the uniform half-width of the absorption line for the transition $n-m$; $G=Ed_{mn}/\hbar$ is the Rabi frequency; N is the concentration of resonant atoms; and S_i is the collision integral of the atoms in the state i. The first term on each right-hand side in Eqs. (1) describes the excitation of an atom by monochromatic radiation, where $p(\mathbf{v})$ is a probability of absorption of radiation per unit time by an atom with velocity $\mathbf{v}$. To describe collisions of the atoms with the buffer gas, we adopt the model of strong elastic collisions, in which the collision integral has the form

$$S_i(\mathbf{v})=-\nu_i\rho(\mathbf{v})+\nu_iW(\mathbf{v})N_i, \tag{2}$$

where ν_i is the collision frequency for an atom in state i, $W(\mathbf{v})$ is the Maxwellian velocity distribution, and

$$N_i=\int \rho_i(\mathbf{v})d\mathbf{v} \tag{3}$$

is the population of the ith state. The sense of this model is that after its first collision, an atom finds itself in an equilibrium velocity distribution.

Next we bring in the fluxes $\mathbf{j}_{m,n}$ of the atoms in states m and n and the total flux $\mathbf{j}$ of absorbing atoms

$$\mathbf{j}_i=\int \mathbf{v}\rho_i(\mathbf{v})\,d\mathbf{v}, \quad \mathbf{j}=\mathbf{j}_m+\mathbf{j}_n. \tag{4}$$

To analyze the experimental situation in the present case, it is sufficient to consider the steady-state solution of system (1). On the basis of (1) we compose equations for the fluxes $\mathbf{j}_m$ and $\mathbf{j}$. To this end, we multiply each of the equations by $\mathbf{v}$ and integrate over $\mathbf{v}$. As a result, we arrive at the following equations (see also Ref. 4):

$$\mathbf{j}_m=\frac{N}{\Gamma_m+\nu_m}\int \mathbf{v}\,p(\mathbf{v})\,d\mathbf{v}, \tag{5}$$

$$\mathbf{j}=\mathbf{j}_{dr}-D\,\nabla N, \quad \mathbf{j}_{dr}=\frac{\nu_n-\nu_m}{\nu_n}\mathbf{j}_m, \tag{6}$$

$$D=v_T^2/2\nu_n, \quad v_T=\sqrt{2k_BT/M}.$$

Here $\mathbf{j}_{dr}$ is the atom flux due to light-induced drift.

In deriving these equations, we have assumed that the radiation-perturbed velocity distribution function does not deviate too much from its equilibrium value. Hence it follows, in particular, that the spatial inhomogeneity scale of the concentration of absorbing particles substantially exceeds the mean free path. We can therefore neglect the spatial derivative term in the equation for $\mathbf{j}_m$.

Under the stated conditions, we can also start out with the steady-state, spatially homogeneous system of equations (1) in the calculation of $p(\mathbf{v})$. Their solution in the strong collision model does not entail any difficulties.[4,18] As a result we obtain

$$p(\mathbf{v})=N\frac{Y(\mathbf{v})\,W(\mathbf{v})}{\tau_1+\tau_2\langle Y\rangle}, \tag{7}$$

$$Y(\mathbf{v})=\frac{\Gamma^2\chi}{\Gamma^2(1+\chi)+(\Omega-\mathbf{k}\cdot\mathbf{v})^2}, \quad \langle Y\rangle=\int Y(\mathbf{v})\,W(\mathbf{v})\,d\mathbf{v},$$

$$\chi=\frac{2\tau_1|G|^2}{\Gamma}, \quad \tau_1=\frac{\nu_n+\nu_m}{\nu_n(\Gamma_m+\nu_m)}, \quad \tau_2=\frac{2}{\Gamma_m}-\tau_1.$$

We call the quantity χ the saturation parameter.

Substituting Eq. (7) into Eq. (5), and then substituting Eq. (5) into Eq. (6), we obtain

$$\mathbf{j}_{dr}=Nv_T\frac{\mathbf{k}}{k}\frac{\nu_n-\nu_m}{\nu_n}\frac{\tau_1\langle Y_1\rangle}{\tau_1+\tau_2\langle Y\rangle}, \tag{8}$$

$$\langle Y_1\rangle=\int \frac{\mathbf{k}\cdot\mathbf{v}}{kv_T}Y(\mathbf{v})\,W(\mathbf{v})\,d\mathbf{v}.$$

Equation (8) holds for any ratio of the homogeneous and inhomogeneous linewidths. For the important case in which Doppler broadening dominates ($\Gamma\sqrt{1+\chi}\ll\Omega_D$) and the offset Ω is bounded by the Doppler linewidth $\Omega_D=kv_T$, which corresponds to our experimental conditions, Eq. (8) admits of the following simplification:

$$\langle Y_1\rangle=\frac{\Omega}{\Omega_D}\langle Y\rangle, \tag{9}$$

$$\langle Y\rangle=\frac{\sqrt{\pi}\,\Gamma}{\Omega_D}\frac{\chi}{\sqrt{1+\chi}}\exp[-(\Omega/\Omega_D)^2].$$

The condition $\Omega_D\gg\Gamma\sqrt{1+\chi}$ implies that $\langle Y\rangle\ll1$, and instead of (8) we may use the relation

$$\mathbf{j}_{dr}=\frac{\mathbf{k}}{k}\frac{\nu_n-\nu_m}{\nu_n+\nu_m}\frac{\Omega}{k}N\langle Y\rangle. \tag{10}$$

We also introduce into consideration the light-induced drift speed $u_{dr}=j_{dr}/N$, whereupon

$$u_{dr}=\frac{\nu_n-\nu_m}{\nu_n+\nu_m}\frac{\Omega}{k}\langle Y\rangle. \tag{11}$$

We now apply these results to the process of accumulation assisted by light-induced drift in a long, thin capillary closed at one end. Let the radiation enter through the open end of the capillary and propagate along its axis (the z axis). Under steady-state conditions, we should set $\mathbf{j}=0$ in Eq. (6). This leads to the form

$$D\frac{\partial}{\partial z}N=u_{dr}N. \tag{12}$$

Accumulation of particles near the closed end of the capillary can produce an optically dense medium, in which the radiative intensity varies substantially over the length of the capillary. As a consequence, the quantity u_{dr} becomes a function of the spatial coordinate z. Under the given conditions, it is necessary simultaneously to take into account

variation of both the particle concentration and the radiative intensity, i.e., together with Eq. (12) it is necessary to consider the equation describing the radiative intensity I of the propagating beam:

$$\frac{d}{dz}I=-P, \quad P=\hbar\omega_L N\int p(\mathbf{v})\,d\mathbf{v}. \tag{13}$$

Here P is the radiative energy absorbed per unit time per unit volume. In the approximation of large Doppler broadening

$$P=\hbar\omega_L N\frac{\langle Y\rangle}{\tau_1}. \tag{14}$$

Taking Eqs. (14) and (9) into account, Eqs. (12) and (13) have the integral

$$N-N_0=\frac{\tau_1\Omega}{k}\,\frac{\nu_n-\nu_m}{\nu_n+\nu_m}\,\frac{I_0-I}{D\hbar\omega_L}. \tag{15}$$

Here N_0 and I_0 are the concentration of resonant atoms and the radiative intensity before the light enters the capillary. In a buffer gas of argon, $(\nu_n-\nu_m)/\nu_n$ is known[19] to be approximately -0.2 for lithium. Consequently, for the regime of accumulation and storage of resonant atoms, it is necessary to choose the ''redshifted'' offset of the radiation frequency, $\Omega<0$.

It is clear from (15) that for prescribed values of the offset Ω and initial radiative intensity I_0 there exists a maximum possible concentration of accumulated atoms:

$$N_{\max}=\frac{\tau_1\Omega}{k}\,\frac{\nu_n-\nu_m}{\nu_n+\nu_m}\,\frac{I_0}{D\hbar\omega_L}. \tag{16}$$

Note that in the present approximation the maximum possible concentration of accumulated atoms is directly proportional to the radiative intensity.

Utilizing relation (15), it is possible to obtain the radiative intensity in closed form, which is most conveniently written as an equation for the saturation parameter:

$$\frac{d}{dz}\chi=-\frac{1}{l}\,\frac{(\chi_0-\chi)\,\chi}{\sqrt{1+\chi}}. \tag{17}$$

Here l is some length that characterizes the scale on which absorption takes place:

$$l=\frac{D(\nu_n+\nu_m)\,\Omega_D k\exp[(\Omega/\Omega_D)^2]}{\sqrt{\pi}\,\Gamma\,\Omega(\nu_n-\nu_m)}. \tag{18}$$

As a function of the offset Ω, this length reaches its minimum at $\Omega=-\Omega_D/\sqrt{2}$:

$$l_{\min}=\frac{\sqrt{2}\,D(\nu_n+\nu_m)\,k\sqrt{e}}{\sqrt{\pi}\,\Gamma\,|\nu_n-\nu_m|}. \tag{19}$$

In Fig. 4, we have plotted the dependence of the radiative intensity and the fluorescence intensity on the spatial coordinate z with scale length l; these were obtained by solving Eq. (17) numerically with $\chi_0=100$. As can be seen from this solution, the size of the fluorescence region is determined by the scale length l.

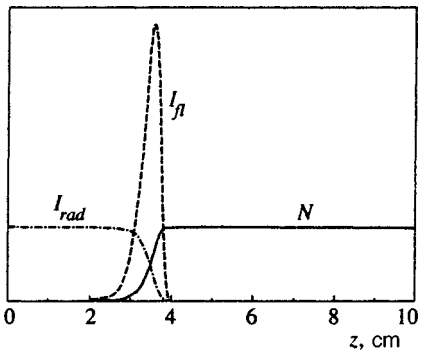

FIG. 4. Distribution of the radiative intensity ($I_{\rm rad}$), fluorescence intensity ($I_{\rm fl}$), and lithium vapor concentration (N) inside the capillary. Numerical solution of Eq. (17) for $\chi_0=100$.

Using these results, we can proceed to a consideration of the quasisteady situation, which is realized for $N_0\ll N_{\max}$. In this case it may be assumed that the spatial distribution of accumulated atoms corresponds at any given moment to steady-state conditions ($N_0\to 0$), except that the region occupied by the accumulated atoms with concentration $N_{\max}$ slowly expands toward the open end of the capillary at speed S, with

$$S=u_{dr}N_0/N_{\max}. \tag{20}$$

4. DISCUSSION

The present experiment was carried out under the following conditions: the pressure of the argon buffer gas was $p=20$ Torr, the laser power was 100 mW, corresponding to an intensity $I=10$ W/cm^2 at the entrance to the capillary. The diffusion coefficient of lithium in argon at the indicated pressure and with $T=1500$ K is taken to be $D\approx 400$ cm^2/s (Ref. 19). It is possible to estimate ν_n from the value of D. The homogeneous half-width of the line under these conditions is $\Gamma\approx 3\times 10^8$ s^{-1}, and the Doppler half-width $\Omega_D=2\times 10^{10}$ s^{-1}. For the indicated parameters we find $\chi_0\simeq 100$. Thus, we are working under conditions of large Doppler line broadening with field broadening taken into account. The maximum possible concentration of accumulated atoms, calculated with Eq. (16) for $\Omega=-\Omega_D/\sqrt{2}$, is $N_{\max}\simeq 1.7\times 10^{13}$ cm^{-3}. An estimate of $l_{\min}$ based on Eq. (19) yields $l_{\min}\simeq 2$ cm. We therefore estimate $u_{dr}\simeq 3\times 10^3$ cm/s for the maximum possible light-induced drift rate.

It can be seen that the theoretical estimate $l_{\min}\simeq 2$ cm and the value estimated from the experimental data (in the photograph) $l_{\rm eff}\simeq 2$ cm coincide.

The onset of the ''optical piston'' regime suggests that a high concentration of resonant atoms is produced beyond the fluorescence region at the closed end of the capillary. If we suppose that the concentration of resonant atoms just outside the capillary entrance is $N_0=10^{10}$ cm^{-3} (most likely it is significantly less), and if we take $N_{\max}$ to be our estimate of the concentration of resonant atoms at the closed end of the capillary, then the degree of concentration of the resonant atoms achieved using light-induced drift is at least 10^3. Such high degrees of concentration are extremely promising for the separation and concentration of weakly represented iso-

topes from a natural mixture of any chemical element, and also for concentration of micro-impurities of atomic gases.

Furthermore, in the experiment we noted that after turning off the lithium vapor source, the lifetime of the optical piston, or of the accumulated atoms in the capillary, is some tens of minutes. This suggests that chemical loss of lithium atoms is largely suppressed in a heated capillary. It also suggests the feasibility of using high-temperature cells for the aforementioned purposes.

At present, our laser stabilization and detection systems do not permit more detailed measurements. In particular, we lack experimental information on the quantities N_0 and $N_{\max}$ and on residual adsorption on the capillary walls. We intend to improve the overall system so that it will be possible to perform such measurements, and also accumulate the rare isotope of lithium to the same level of concentration and reach a coefficient of separation of at least 10^3, and a degree of concentration greater than 10^3. The results of the present experiment suggest that these goals are attainable.

This work was carried out with the support of the Russian Fund for Fundamental Research (Grant No. 96-02-19556) and the State Scientific–Technical Program ''Laser Physics'' (Grant No. 7.41).

In this work we used the ''AMETIST'' cw dye laser built by the Laser Systems Laboratory of Novosibirsk State University (`http://www.cnit.nsu.ru/nwww/lls/english/dlt.htm`).

*)E-mail: shalagin@iae.nsk.su

[1] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[2] V. D. Antsygin, S. N. Atutov, F. Kh. Gel'mukhanov, G. G. Telegin, and A. M. Shalagin, JETP Lett. **30**, 282 (1979).
[3] V. N. Panfilov, V. P. Strunin, P. L. Chapovskiĭ, and A. M. Shalagin, JETP Lett. **33**, 48 (1981).
[4] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Nonlinear Spectroscopy*, North-Holland, Amsterdam (1991).
[5] L. J. F. Hermans, Int. Rev. Phys. Chem. **11**, 289 (1992).
[6] E. R. Eliel, Adv. At., Mol., Opt. Phys. **30**, 199 (1993).
[7] C. Hradechny, J. Slovak, T. Tethal, A. M. Shalagin, and I. M. Yermolaev, Appl. Radiat. Isot. **43**, 1259 (1992).
[8] C. Hradechny, T. Tethal, I. M. Yermolaev, S. G. Zemlyanoi, and P. Zuzaan, Appl. Radiat. Isot. **45**, 257 (1994).
[9] F. Buřič, C. Hradečný, J. Slovák, T. Těthal, and I. M. Yermolayev, Phys. Rev. A **54**, 3250 (1996).
[10] S. N. Atutov, P. V. Kolinko, and A. M. Shalagin, Laser Phys. **3**, 855 (1993).
[11] S. N. Atutov, Phys. Lett. A **119**, 121 (1986).
[12] S. N. Atutov, I. M. Ermolaev, and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **92**, 1215 (1987) [Sov. Phys. JETP **65**, 679 (1987)].
[13] S. N. Atutov, P. V. Kolinko, and A. M. Shalagin, Opt. Commun. **107**, 218 (1994).
[14] F. Kh. Gel'mukhanov and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **78**, 1674 (1980) [Sov. Phys. JETP **51**, 839 (1980)].
[15] H. G. C. Werij, J. P. Woerdman, J. J. M. Beenakker, and I. Kuščer, Phys. Rev. Lett. **52**, 2237 (1984).
[16] H. G. C. Werij, J. E. M. Haverkort, and J. P. Woerdman, Phys. Rev. A **33**, 3270 (1986).
[17] W. A. Hamel, A. D. Streater, and J. P. Woerdman, Opt. Commun. **63**, 43 (1987).
[18] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).
[19] S. N. Atutov, B. V. Bondarev, S. M. Kobtzev, P. V. Kolinko, S. P. Podyachev, and A. M. Shalagin, Opt. Commun. **115**, 276 (1995).

Translated by Paul F. Schippnick

Theory and interpretation of the quantum teleportation effect

D. N. Klyshko*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 23 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1171–1187 (October 1998)

Recently published optical experiments which investigate the effect of ''quantum teleportation'' are analyzed. It is shown that the occurrence of teleportation (copying) of the polarization of one photon onto another requires that an optical shutter be added to the experimental setup, opening automatically only upon the occurrence of certain favorable events. An instructive model is proposed, along with a classical analog of the effect, and various treatments are discussed. It is emphasized that the concepts of ''reduction of the wave function'' and ''quantum nonlocality'' are not necessary for a quantitative description of the effect: the standard formalism proves to be sufficient. © *1998 American Institute of Physics.* [S1063-7761(98)00310-2]

1. INTRODUCTION

The unexpected possibility of irreversibly copying a quantum state from one individual system onto another system isomorphic with the first was recently described by Bennett *et al.*[1] The idea received further development in Refs. 2–8. In contrast to the reversible exchange of states between two-level atoms and a cavity field (transfer of a qubit of information) observed in Ref. 9, here part of the information is irreversibly converted into a classical form. In essence, Ref. 1 proposed a method of preparing an individual quantum system in a prescribed state ψ with the important feature that the information about which state in particular has been prepared is of a quantum nature, i.e., it is written down in the form of the state of another system and remains unknown. The first optical experiments along these lines were described recently.[7,8] The interpretation of the effect adopted in Refs. 1–8, as well as its name, is based on the widely held notion that the instantaneous reduction (collapse) of the wave function as a result of a measurement leads to quantum nonlocality.

Figure 1 presents a simplified experimental setup.[7] Three quasi-monochromatic stationary beams of light *A*, *B*, and *C* are incident upon the optical system, where the photons (black dots) arrive simultaneously in groups of three. The *A* photons are completely polarized. The *B* and *C* photons are depolarized; however, there is a unique correlation between their polarizations. A 50% nonpolarizing semitransparent mirror (beamsplitter) *BS* mixes the *A* and *B* photons. As a result of the initial correlation between the *B* and *C* photons and the action of the beamsplitter, all three photons are correlated. Under these conditions, information $\mathbf{e}_A$ about the polarization of an *A* photon turns out to be encoded in the probabilities of three-way coincidences. This effect is investigated using two polarization converters T_A and T_C, a polarizing prism P_C in beam *C*, three photodetectors D_Z, and a scheme of three-way coincidences *CC*. The experiment[7] entails observing the dependence of the number of three-way coincidences N_3 (during a certain fixed time interval) on the parameters T_A and T_C and on the time delay τ in one of the three channels (as τ increases, the degree of the polarization interference, as always, decreases). Thus, the scheme is in essence a kind of polarized intensity interferometer operating in photon-counting mode.

The three-photon interference observed in the scheme and depicted in the figure should possess a remarkable property: the number of coincidences N_3 depends on the converters T_A and T_C in the same way as if they were set up in sequence in one beam or as if the polarization of photon *A* at the output of T_A were transferred to photon *C* at the entrance to T_C: $\mathbf{e}_A \rightarrow \mathbf{e}_C$. In other words, the phase and degree of the polarization observed in the three-way coincidences are determined by the product of the Jones matrices $\mathbf{T}_C\mathbf{T}_A$. If T_C performs the inverse transformation to T_A ($\mathbf{T}_C\mathbf{T}_A=1$), then N_3 will depend on neither T_A nor T_C, and the degree of the interference will be zero.

The detectors D_{Cx} and D_{Cy}, and the converter T_C, can be considered to be a device for measuring the polarization of the *C* photons, but unconditional readings in D_{Cj} will not detect any polarization because the *C* photons are not polarized. However, conditional readings in D_{Cj}, i.e., those taking place simultaneously with readings in D_A and D_B, exhibit complete polarization. For example, from the dependence of N_3 on T_C it is possible to measure the polarization vector $\mathbf{e}_A=(\alpha_x,\alpha_y)$ of the *A* photons (or, equivalently, the Stokes vector $\langle \mathbf{S}_A \rangle$). However, 75% of the time both photons *A* and *B* fall on the same detector, D_A or D_B. Such ''unfavorable'' events do not exhibit a polarization copying effect.

Photons *B* and *C* are prepared in an entangled EPR state giving complete correlation in the polarization. The mean frequencies of the fields in the *A* and *B* beams should be the same (so that they can interfere in the beamsplitter *BS*), but the frequency of the field in the *C* beam can be arbitrary. All three photons should also be correlated in their time of arrival at the beamsplitter with an accuracy determined by the coherence time and timing resolution of the coincidence scheme.[5] To this end, Ref. 7 used foursomes of photons emitted during type-II parametric scattering (Ref. 10) in the second order of the pump intensity (one of the four photons

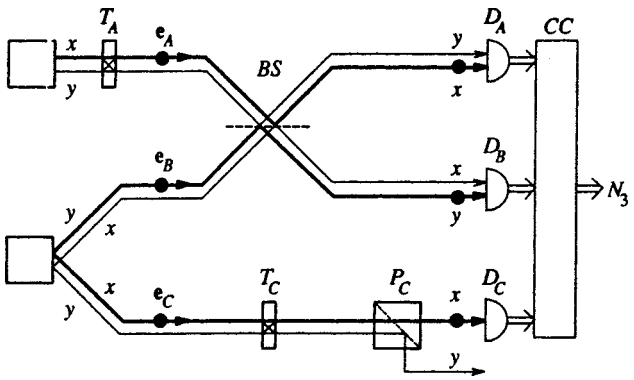

FIG. 1. Simplified layout of the experiment in Ref. 7 and an instructive model for copying polarization vectors $\mathbf{e}_A \to \mathbf{e}_C$. Only one photon (circles) is present at the entrance to each of the beams A, B, and C. Photon A has arbitrary polarization $\mathbf{e}_A$. The basis vectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are chosen such that $\mathbf{e}_x = \mathbf{e}_A$. Simultaneous detection of the photons by detectors D_A and D_B implies that photons A and B did not interfere in the beamsplitter; consequently, they have orthogonal polarizations, $\mathbf{e}_A \perp \mathbf{e}_B$ (otherwise they would have fallen onto the same detector). Photons B and C were prepared in states with orthogonal polarizations, $\mathbf{e}_B \perp \mathbf{e}_C$; therefore $\mathbf{e}_C = \mathbf{e}_A$. T_A and T_C are polarization converters, P_C is a polarizing prism, D_Z are detectors, CC is a triple coincidence counter, N_3 is the number of triple coincidences during a certain time interval, and x and y are polarization indices.

was superfluous); this setup utilized pumping of a nonlinear crystal by short pulses and additional filtering of the spectrum of the scattering field to increase the coherence time.

The present paper contains a formal analysis of the scheme depicted in Fig. 1, and draws attention to the fact that to demonstrate the teleportation effect (instead of an interference effect) it is necessary to replace the scheme of three-way coincidences by a scheme of two-way coincidences and add an optical shutter (modulator) to beam C. The shutter should control the scheme of two-way coincidences at the detectors D_A and D_B and let through C photons only upon the occurrence of a certain subset of ''favorable'' events in these detectors—simultaneous readings in D_A and D_B. Thus, teleportation, i.e., copying the polarization of the A photons, is only feasible using a nonunitary transformation of the field C.

An elementary instructive model of the polarization copying effect is also proposed, along with the corresponding classical analog. It is also emphasized that for a quantitative calculation of the teleportation effect the concept of reduction of the wave function is superfluous, as it probably also is for a quantitative description of all the known phenomena of quantum physics observed to date (possible exceptions are discussed in Ref. 11). Recall in this regard that a critical stance vis-à-vis a literal understanding of reduction as a ''real'' process has been taken more than once; see Refs. 12–16.

Section 2 examines an instructive model and a classical analog of the effect. Section 3 calculates the operation of the scheme depicted in Fig. 1, and some variations of it in the Heisenberg picture. Section 4 briefly repeats these calculations in the Schrödinger picture and analyzes some methodological issues inherent in the transition from the one picture to the other in problems of quantum optics. Section 5 compares two possible approaches to the description and interpretation of the effect—the generally accepted approach[1–8] and the approach proposed in the present paper.

2. ELEMENTARY MODEL OF THE EFFECT AND CLASSICAL ANALOGY

The polarization copying effect $\mathbf{e}_A \to \mathbf{e}_C$ has an instructive (but nonrigorous) explanation (Fig. 1) that suggests a very similar classical experiment. This explanation is a direct consequence of two well-known effects—anticorrelation of photons with identical polarization at the output of a beamsplitter and correlation of photons with orthogonal polarization of EPR–Bohm type. Indeed, the detection of two photons in the output beams A' and B' (i.e., the lack of anticorrelation between the photons) implies that the initial photons in the A and B beams did not interfere in the beamsplitter and consequently had orthogonal polarization $\mathbf{e}_A \perp \mathbf{e}_B$. From here on, the polarization of the B and C photons can be taken to be orthogonal: $\mathbf{e}_B \perp \mathbf{e}_C$. From $\mathbf{e}_A \perp \mathbf{e}_B$ and $\mathbf{e}_B \perp \mathbf{e}_C$ it follows that $\mathbf{e}_A = \mathbf{e}_C$.

Let us dwell on this conclusion in more detail. Let the field in the B and C beams be described by the state[1]

$$|\psi\rangle_{BC} = (|B_x, C_y\rangle - |B_y, C_x\rangle)/\sqrt{2}. \tag{2.1}$$

Here $|B_x, C_y\rangle \equiv b_x^+ c_y^+ |0\rangle$, b_x^+ and c_y^+ are photon creation operators in the x mode in beam B and in the y mode in beam C, and x and y are indices of two arbitrary orthogonal types of polarization ($\mathbf{e}_j^* \mathbf{e}_k = \delta_{jk}$).

One peculiarity of this state is its invariant form under any transformation of the polarization basis. Indeed, let

$$|B_x\rangle = t^*|B_1\rangle - r|B_2\rangle, \quad |C_x\rangle = t^*|C_1\rangle - r|C_2\rangle,$$
$$|B_y\rangle = r^*|B_1\rangle + t|B_2\rangle, \quad |C_y\rangle = r^*|C_1\rangle + t|C_2\rangle, \tag{2.2}$$

where 1 and 2 are indices of the new basis and the transformation coefficients t and r satisfy $|t|^2 + |r|^2 = 1$. Substituting Eq. (2.2) into Eq. (2.1) gives

$$|\psi\rangle_{BC} = (|B_1, C_2\rangle - |B_2, C_1\rangle)/\sqrt{2}.$$

It is easy to convince oneself that in the state (2.1) the three components of the Stokes vectors $\mathbf{S}_B$ and $\mathbf{S}_C$ have opposite signs (regardless of basis):

$$\langle \psi | S_{Bn} S_{Cn} | \psi \rangle = -1. \tag{2.3}$$

Here

$$S_{B1} \equiv b_x^+ b_x - b_y^+ b_y, \quad S_{B2} \equiv b_x^+ b_y + b_x b_y^+,$$
$$S_{B3} \equiv (b_x^+ b_y - b_x b_y^+)/i, \tag{2.4}$$

and similarly for $\mathbf{S}_{Cn}$. At the same time, $\langle S_{Bn} \rangle = \langle S_{Cn} \rangle = 0$, and the B and C photons are completely depolarized.

For the sake of clarity, we assume that all three photons A, B, and C have arbitrary prescribed polarizations $\mathbf{e}_Z$ and corresponding Stokes vectors $\mathbf{S}_Z$, which vary randomly from trial to trial (we stipulate at once that this assumption is incompatible with quantum theory; see Sec. 5). In this case, property (2.3) can be interpreted to mean that the Stokes vectors point in opposite directions, $\mathbf{S}_C = -\mathbf{S}_B$, i.e., that the polarization vectors are orthogonal, $\mathbf{e}_B \perp \mathbf{e}_C$. We choose a basis in which $\mathbf{e}_x = \mathbf{e}_A$, and assume, in line with (2.1), that

there are only two equiprobable possibilities: either $\mathbf{e}_B=\mathbf{e}_x=\mathbf{e}_A$ and $\mathbf{e}_C=\mathbf{e}_y$, or $\mathbf{e}_B=\mathbf{e}_y$ and $\mathbf{e}_C=\mathbf{e}_x=\mathbf{e}_A$. But the first case should exhibit anticorrelation of the readings of detectors D_A and D_B: the two photons can only be found together—either both are found at D_A, or both are found at D_B; see (3.8) below. This is the photon anticorrelation effect (or two-photon interference effect), first observed in Ref. 17. It has a simple classical analog, the anticorrelation of intensity fluctuations at the output of the beamsplitter as a result of energy conservation[18]. Consequently, when taking readings with both detectors D_A and D_B the latter case occurs, i.e., $\mathbf{e}_C=\mathbf{e}_x=\mathbf{e}_A$ (heavy lines in Fig. 1).

Let us consider the analogous classical method of copying the polarization of one light beam onto another without measuring it. Let there be three ideal lasers A, B, and C, emitting polarized beams of light, where the intensities I_0 and frequencies ω_0 of beams A and B are identical. Polarization converters are placed in beams B and C, both controlled by a common random number generator such that the two beams are always orthogonal: $\mathbf{e}_B(t)\perp\mathbf{e}_C(t)$. As a result, the points mapping the polarization onto the Poincaré sphere cover it uniformly—any polarization of beams B and C is equiprobable, but their Stokes vectors are always oppositely directed (i.e., the polarizations are completely correlated). We have thus obtained an analog of property (2.3) of the state (2.1) (here averaging using the wave function $|\psi\rangle$ has been replaced by classical averaging over time or over the ensemble).

Next, beams A and B are mixed in the beamsplitter, and the light intensities $I'_A(t)$ and $I'_B(t)$ in the output beams A' and B' are monitored by two analog detectors. Because of fluctuations of the polarization vector $\mathbf{e}_B(t)$ the intensities $I'_A(t)$ and $I'_B(t)$ also fluctuate, but always with opposite phase, so their total intensity is preserved, $I'_A(t)+I'_B(t)=2I_0$. At those times when $I_A(t)$ is equal to $I_B(t)$ (with some prescribed error $\Delta I/I_0$), the beams entering the beamsplitter do not interfere, and consequently their polarizations are orthogonal, $\mathbf{e}_A\perp\mathbf{e}_B$. But in this case $\mathbf{e}_B\perp\mathbf{e}_C$ as well, so that $\mathbf{e}_A=\mathbf{e}_C$. At these times, the shutter that otherwise blocks beam C opens automatically. As a result, we obtain light pulses (with random intervals and duration) at frequency ω_C and polarization $\mathbf{e}_A$. The important difference between this model and the quantum model is the bounded copying accuracy, which is inversely proportional to ΔI and the relative opening time of the shutter (i.e., its duty cycle). In the quantum domain, copying in the ideal case is perfect.

3. HEISENBERG PICTURE

In using the Heisenberg picture, it is assumed that the elements of the optical channel alter the wave function of the initial field: $\psi\to\psi'=U\psi$. Here U is the evolution operator of the field, describing the action of the beamsplitters, the polarization converters, etc. (without sources and detectors); the primes denote quantities belonging to the output modes of the channel. The superfluousness of the reduction concept for the calculation becomes evident when we make use of the equivalent Heisenberg picture, in which the operation of the scheme is described by the variation of the field operators:

$$a_k\to a'_k=U^+a_kU.$$

This transformation can also be represented in the form

$$a'_k=\sum_m T^*_{km}a_m$$

(see Refs. 18 and 19). Here T is a phenomenological transformation matrix (spectral Green's function), which enables one to find linear relations between the output μ'_n and input μ_n moments of the same order n; symbolically

$$\mu_n\to\mu'_n=\mathbf{T}^n\mu_n.$$

As a result, the observed output moments μ'_n or the associated probabilities p' can be defined in terms of $\mathbf{T}$ and the initial (input) state vector $|\psi\rangle$ of the incident optical field. The relation between the classical matrix $\mathbf{T}$ and the evolution operator U will be spelled out below; see Eq. (4.6). This formalism is also applicable in the presence of dissipation, whereupon the matrix $\mathbf{T}$ is nonunitary, $T_{km}\neq T^*_{mk}$ (Refs. 18 and 19).

In radio engineering terms, the optical system is a multiport network. The matrix $\mathbf{T}$ coincides with the analogous matrix in classical optics, so the descriptions of the linear transformation of the statistics of the field by the optical system in the quantum theory and in the classical theory are identical, differences being manifested only in the relative values of the input moments. Thus, the quantum details, if they exist at all, are already embedded in the light source, and the optical system together with the detectors can be considered a measuring instrument for studying the statistics of the source. However, this classification loses meaning in feedforward and feedback optoelectronic circuits that use a light-modulator. For example, the layout in Fig. 1, when one gates beam C under the control of detectors D_A and D_B, is transformed from a measurement system to a preparation system.

We assume the components of our scheme, including the detectors, to be ideal. Actual experiments can be described using so-called one-photon wave packets, i.e., quasistationary states

$$|\psi(t)\rangle_1=\int d\mathbf{k}\,f(\mathbf{k})\exp(-i\omega_{\mathbf{k}}t)a^+_{\mathbf{k}}|0\rangle,$$

but for our purposes it is sufficient to restrict the discussion to the single-mode approximation. Granted, we then lose the possibility of tracking the sequence of events in time, but the multimode description[5] only ensures that the requirements of the special theory of relativity are met and enables us to specify the simultaneity requirements for the emission of the three photons (the wave packets at the beamsplitter should overlap). Note that it is precisely the generation of three-photon light (in contrast to two-photon light, used in Ref. 8) that is the main impediment to implementing an experiment of the type described in Ref. 7.

The detectors—photon counters—are situated in the output modes behind the beamsplitter BS and polarization converter T_C. They can be used to measure the probabilities $p'(n_1,n_2,\ldots)=p'(\{n_k\})$ of finding specific numbers of

photons $\{n_k\}$ in the output modes. We denote the corresponding Fock states for the output modes by the symbol $|\dots)$:

$$|n_1,n_2,\dots)\equiv|n_1)\otimes|n_2)\dots \equiv(n_1!n_2!\dots)^{-1/2}(a'^{+}_1)^{n_1}(a'^{+}_2)^{n_2}\dots|0). \tag{3.1}$$

We assume the vacuum vectors of the two bases to be identical: $|0\rangle\equiv|0)$. The matrix $\mathbf{T}$ governs the relationship between the two bases; for example, ${}_m\langle 1|1)_k=T_{km}$. This follows from

$$|1)_k\equiv a'^{+}_k|0)=\sum_m T_{km}a^{+}_m|0\rangle=\sum_m T_{km}|1\rangle_m. \tag{3.2}$$

According to the Born postulate, the probability $p'=|q'|^2$ is defined by the inner product q' of the corresponding Fock bra vector $(n_1,n_2,\dots|$ and the initial ket vector of the field $|\psi\rangle$:

$$q'(n_1,n_2,\dots)=(n_1,n_2,\dots|\psi\rangle =(n_1!n_2!\dots)^{-1/2}\langle 0|a'^{n_1}_1a'^{n_1}_2\dots|\psi\rangle. \tag{3.3}$$

In what follows, the primes on p and q, which serve to remind us that the detectors are situated in the output modes, will be omitted. All operators a'_k in (3.3) commute, and the order in which they are written is irrelevant.

We assume that to facilitate a complete analysis, polarizing prisms P_A and P_B have been inserted in beams A' and B' behind the beamsplitter (see Fig. 1), and accordingly in front of the corresponding detectors. We denote and order the six considered modes as follows: A_x, A_y, B_x, B_y, C_x, C_y; for example, $p(100110)\equiv p(A_x,B_y,C_x)$. Here x and y are the indices of two arbitrary orthogonal polarizations; it is possible, in particular, to choose $\mathbf{e}_x\equiv\mathbf{e}_A$.

In the case of one-photon states the probabilities are identical to the corresponding moments, for example

$$p(A_x,B_y,C_x)=\langle\psi|N'_{Ax}N'_{By}N'_{Cx}|\psi\rangle=|\langle 0|a'_xb'_yc'_x|\psi\rangle|^2.$$

Here a'_i, b'_j, c'_k are the photon annihilation operators in the corresponding output modes, and $N'_j\equiv a'^{+}_j a'_j$. Indeed, the substitution

$$a^{+}a\to a^{+}Ia=a^{+}(|0\rangle\langle 0|+|1\rangle\langle 1|+|2\rangle\langle 2|+\dots)\,a$$

yields

$$\langle N\rangle=\langle 1|a^{+}a|1\rangle=\langle 1|a^{+}|0\rangle\langle 0|a|1\rangle|=|\langle 0|a|1\rangle|^2.$$

The $\mathbf{T}$ matrix of our optical system (see Fig. 1) (neglecting T_A) is defined by the unitary transformation

$$a'_j=(a_j+b_j)/\sqrt{2},\quad c'_x=t^*_Cc_x+r^*_Cc_y,$$
$$b'_j=(-a_j+b_j)/\sqrt{2},\quad c'_y=-r_Cc_x+t_Cc_y, \tag{3.4}$$

where t_C and r_C are elements of the (classical) Jones matrix $\mathbf{T}_C$ of the converter T_C ($|t_C|^2+|r_C|^2=1$). Substituting (3.4) into (3.3) yields the probability amplitudes of all elementary events observed in the experiment $(n_1,\dots,n_6)$, which satisfy $\Sigma n_i=3$.

For example, the probability amplitude for detecting one photon in the output modes A_x, B_y, C_x is

$$q(A_x,B_y,C_x)=\langle 0|(a_x+b_x)(-a_y+b_y)(t^*_Cc_x+r^*_Cc_y)|\psi\rangle/2. \tag{3.5}$$

The effect under discussion takes place when only two of the eight matrix elements here are nonzero. Let one photon be present in each beam at the system input. Then

$$\langle 0|a_xa_yc_j|\psi\rangle=\langle 0|b_xb_yc_j|\psi\rangle=0.$$

Also let

$$\langle 0|(a_xb_yc_y)|\psi\rangle=\langle 0|a_yb_xc_x|\psi\rangle=0.$$

Then Eq. (3.5) takes the form

$$q(A_x,B_y,C_x)=\langle 0|(t^*_Ca_xb_yc_x-r^*_Ca_yb_xc_y|\psi)\rangle/2. \tag{3.6}$$

According to this expression, the converter T_C affects the polarization of the photons in beams A and C in an identical way.

On the other hand, the probability amplitude for detecting two photons in the output mode $|A_x)$ and one photon in the output mode $|C_x)$ under the same conditions is

$$\begin{aligned}q(2A_x,C_x)&=(2A_x,C_x|\psi\rangle=\langle 0|(a'_x)^2c'_x|\psi\rangle/\sqrt{2}\\&=\langle 0|(a_x+b_x)^2(t^*_Cc_x+r^*_Cc_y)|\psi\rangle/2\sqrt{2}\\&=\langle 0|a_xb_x(t^*_Cc_x+r^*_Cc_y)|\psi\rangle/\sqrt{2}\\&=\langle 0|r^*_Ca_xb_xc_y|\psi\rangle/\sqrt{2}.\end{aligned} \tag{3.7}$$

These events do not depend on t_C, i.e., they do not exhibit the required effect, and therefore must be eliminated using a coincidence or shutter scheme.

Note that neither of the coincidences $(A_x,B_x,*)$ and $(A_y,B_y,*)$ takes place; the operator

$$a'_jb'_j=(a_j+b_j)(-a_j+b_j)=b_j^2-a_j^2, \tag{3.8}$$

acting on the one-photon states $|A_j,B_j\rangle$, yields zero. This is again a manifestation of the photon anticorrelation effect.[17,18]

Let us now specify the input state. Let[1]

$$|\psi\rangle=|\psi\rangle_A|\psi\rangle_{BC}=[\alpha_x|A_x\rangle+\alpha_y|A_y\rangle]\times[|B_x,C_y\rangle-|B_y,C_x\rangle]/\sqrt{2}. \tag{3.9}$$

This state possesses the required properties:

$$b_xc_y|\psi\rangle_{BC}=-b_yc_x|\psi\rangle_{BC}=|0\rangle_{BC},$$
$$b_xc_x|\psi\rangle_{BC}=b_yc_y|\psi\rangle_{BC}=0.$$

Hence, using Eq. (3.5) and analogous expressions, we find

$$\begin{aligned}&q(A_y,B_x,C_x)=-q(A_x,B_y,C_x)=(t^*_C\alpha_x+r^*_C\alpha_y)/\sqrt{8},\\&q(A_x,A_y,C_x)=-q(B_x,B_y,C_x)=(-t^*_C\alpha_x+r^*_C\alpha_y)/\sqrt{8},\\&q(2A_x,C_x)=-q(2B_x,C_x)=r^*_C\alpha_x/2,\\&q(2B_y,C_x)=-q(2A_y,C_x)=t^*_C\alpha_y/2.\end{aligned} \tag{3.10}$$

Probability amplitudes of the form $q(*,*,C_y)$, according to (3.4), can be found by making the substitutions $t^*_C\to -r_C$, $r^*_C\to t_C$, where

$$p(*,*,C_x)+p(*,*,C_y)=1.$$

From (3.10) we obtain probabilities of all 16 observable events:

$$\begin{aligned}
p(A_x,B_y,C_x)&=p(A_y,B_x,C_x)\\
&=[|t_C\alpha_x|^2+|r_C\alpha_y|^2+2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)]/8,\\
p(A_x,B_y,C_y)&=p(A_y,B_x,C_y)\\
&=[|r_C\alpha_x|^2+|t_C\alpha_y|^2-2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)]/8,
\end{aligned}\tag{3.11}$$

$$\begin{aligned}
p(A_x,A_y,C_x)&=p(B_x,B_y,C_x)\\
&=[|t_C\alpha_x|^2+|r_C\alpha_y|^2-2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)]/8,\\
p(A_x,A_y,C_y)&=p(B_x,B_y,C_y)\\
&=[|r_C\alpha_x|^2+|t_C\alpha_y|^2+2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)]/8,
\end{aligned}\tag{3.12}$$

$$\begin{aligned}
p(2A_x,C_x)&=p(2B_x,C_x)=|r_C\alpha_x|^2/4,\\
p(2A_x,C_y)&=p(2B_x,C_y)=|t_C\alpha_x|^2/4,\\
p(2A_y,C_x)&=p(2B_y,C_x)=|t_C\alpha_y|^2/4,\\
p(2A_y,C_y)&=p(2B_y,C_y)=|r_C\alpha_y|^2/4.
\end{aligned}\tag{3.13}$$

The sum of all elementary probabilities of the form $p(*,*,C_x)$ defines the marginal probability of detecting a C_x photon (in the absence of modulators):

$$\begin{aligned}
p(C_x)=\langle N'_{Cx}\rangle&=2[p(A_x,B_y,C_x)+p(A_x,A_y,C_x)\\
&+p(2A_y,C_x)+p(2B_x,C_x)]=(|t_C\alpha_x|^2\\
&+|r_C\alpha_y|^2+|r_C\alpha_x|^2+|t_C\alpha_y|^2)/2=1/2.
\end{aligned}\tag{3.14}$$

Analogously, $p(c_y)=1/2$, i.e., the C photons, as expected, remain completely depolarized: the converters T_A and T_C have no effect on the unconditional readings in detectors D_{Cj} of beam C.

However, conditional readings in D_{Cj} detect a definite polarization. According to (3.11), the converter T_C acts on the four events $p(A_x,B_y,*)$ and $p(A_y,B_x,*)$ (occurring in 1/4 of all trials) in exactly the same way as if it were located in beam A at the entrance to the system after T_A. The joint effect of T_A and T_C on these events is described by the product of the Jones matrices $\mathbf{T}_C\mathbf{T}_A$. This also furnishes an operational definition for the effect observed in Ref. 7. By repeating the procedure repeatedly for different T_C, it is possible to measure the polarization $\mathbf{e}_A$ of the A photons imposed by their source and the converter T_A.

It is possible to make the dependence of all four events $p(A_x,A_y,*)$ and $p(B_x,B_y,*)$ on T_A and T_C the same. To this end, when these events occur it is necessary according to (3.12) to include an additional controllable converter T'_C with Jones matrix σ_z before T_C (which is equivalent to changing the sign of α_y).[1] Such a device ''corrects'' the polarization of the C photons in (3.12) by increasing the fraction of favorable events from 1/4 to 1/2.

At the same time, the eight events represented by (3.13) with two photons incident on one detector (which occurs in 50% of the trials) do not exhibit the polarization copying effect. The influence of such events can be eliminated by using an optical shutter that opens automatically when ''favorable'' events like (3.11) occur (and (3.12) in the presence of T'_C). The shutter is also needed to block beam C in the absence of readings in both detectors D_A and D_B due to their nonideality or due to losses in the channel. Finally, a corresponding delay should be introduced before the shutter in beam C.

The shutter in essence replaces the scheme of three-way coincidences. All of the photons that have passed through the shutter and the converter T'_C will have polarization identical to that of the A photons, and copying takes place: $\mathbf{e}_A\rightarrow\mathbf{e}_C$. In this case, the device will in fact prepare single photons in beam C having known (but random) creation times and unknown polarization copying the polarization of the A photons. The polarization analyzer for the C photons, consisting of the prism P_C and detectors D_{Cj}, is now autonomous and can be position d anywhere along beam C.

Let us find the Stokes parameters $\langle S_{Cn}\rangle$ and the degree of polarization of beam C using the controllable converter T'_C, but without the shutter for eliminating events (3.13). Taking the change of sign of α_y in (3.12) into account, we have

$$p(C_x)=4p(A_x,B_y,C_x)+2p(2B_x,C_x)+2p(2A_y,C_x).\tag{3.15}$$

Now, Eqs. (3.11)–(3.13) yield

$$\begin{aligned}
p(C_x)&=\frac{1}{2}[1+2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)],\\
p(C_y)&=\frac{1}{2}[1-2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)],
\end{aligned}\tag{3.16}$$

instead of (3.14). To determine the Stokes parameters $\langle S_{C1}\rangle$, $\langle S_{C2}\rangle$, and $\langle S_{C3}\rangle$, it is necessary to measure the difference

$$p(C_x)-p(C_y)=2\mathrm{Re}\,(t_Cr_C^*\alpha_x^*\alpha_y)\tag{3.17}$$

for

1) $t_C=1$, $r_C=0$; 2) $t_C=r_C=1/\sqrt{2}$;

3) $t_C=1/\sqrt{2}$, $r_C=i/\sqrt{2}$.

Hence,

$$\begin{aligned}
\langle S_{C1}\rangle&=0,\\
\langle S_{C2}\rangle&=\mathrm{Re}\,(\alpha_x^*\alpha_y)=\frac{1}{2}\langle S_{A2}\rangle,\\
\langle S_{C3}\rangle&=\mathrm{Im}\,(\alpha_x^*\alpha_y)=\frac{1}{2}\langle S_{A3}\rangle.
\end{aligned}\tag{3.18}$$

Noting that

$$\langle S_{C0}\rangle=p(C_x)+p(C_y)=1,$$

we find the degree of polarization of beam C:

$$\begin{aligned}
P_C&\equiv(\langle S_{C1}\rangle^2+\langle S_{C2}\rangle^2+\langle S_{C3}\rangle^2)^{1/2}/\langle S_{C0}\rangle\\
&=|\alpha_x^*\alpha_y|=\frac{1}{2}\sin\theta_A,
\end{aligned}\tag{3.19}$$

where $\theta_A \equiv 2\arctan|\alpha_y/\alpha_x|$ is the polar angle of the point mapping the state of the A photon onto the Poincaré sphere. For example, if the A photons are linearly polarized, $\theta_A = \pi/2$ and $P_C = 1/2$, and for circular polarization, $\theta_A = 0$ or π and $P_C = 0$. Thus, even in the case of ideal circuit elements and detectors, a controllable unitary converter T'_C, giving the transformation $\mathbf{e}'_C = \sigma_z \mathbf{e}_C$ for the two-way coincidences (A_x, A_y) and (B_x, B_y), is insufficient for exact copying of the polarization of the A photon; a shutter is still needed to absorb the C photons upon the occurrence of any one of the events $(2A_x)$, $(2A_y)$, $(2B_x)$, and $(2B_y)$.

Another variant of mixing and analysis of the A and B beams is possible, in which the polarizing prisms are located not behind the beamsplitter BS but in front of it. In this case, it is possible to use two independent beamsplitters to mix the beams $A_x - B_x$ and $A_y - B_y$ or $A_x - B_y$ and $A_y - B_x$, and two detectors at the outputs of the beamsplitters.[2] However, as can easily be seen by repeating the above calculation, in this case events in which two photons are incident on one detector again make it necessary to use a shutter (see Eqs. (3.7)), and limit the maximum efficiency of the scheme to 1/2.

We now consider a simplified version of the experimental setup in Fig. 1, with no polarizing prisms P_A or P_B, and only one detector each in the A and B output beams (this scheme is used in the experiments in Ref. 7). In this case it is not necessary to use the correcting unitary transformation $\mathbf{T}'_C$. According to (3.11)–(3.13), the probability of detecting one photon in beam A and one in beam B regardless of their polarizations, as well as a third photon in mode C_x, is

$$p(A,B,C_x) = 2p(A_x,B_y,C_x) = [\,|t_C\alpha_x|^2 + |r_C\alpha_y|^2 + 2\mathrm{Re}\,(t_C r_C^* \alpha_x^* \alpha_y)\,]/4. \tag{3.20}$$

These events manifest the copying effect, in contrast to events in which the two photons wind up in the same output beam—either A or B—with probability

$$p(2A,C_x) = p(2B,C_x) = 2[p(2A_x,C_x) + p(2A_y,C_x) + p(A_x,A_y,C_x)] = [1 + |r_C\alpha_x|^2 + |t_C\alpha_y|^2 - 2\mathrm{Re}\,(t_C r_C^* \alpha_x^* \alpha_y)]/4. \tag{3.21}$$

To eliminate these events a shutter is again necessary. The maximum fraction of favorable events is now

$$p(A,B,C_x) + p(A,B,C_y) = 1/4.$$

Thus, the above scheme with additional analysis of the polarization of beams A and B and use of the modulator T'_C enables one to increase the maximum efficiency of copying from 1/4 to 1/2.

In the case of ideal photon counters, an even simpler scheme is possible, in which there is only one detector in the output of the beamsplitter. Two photons from beams A and B after the beamsplitter can both wind up in either the A output channel or the B output channel, or one can wind up in the A output channel and one in the B output channel (see Fig. 1). The detection of just one photon in beam A means that the latter event occurred. Its probability is given in Eq. (3.20):

$$p(A,C_j) = p(A,B,C_j).$$

4. SCHRÖDINGER PICTURE

To go to the Schrödinger picture it is necessary to represent the initial state (3.9) in the basis of output modes (3.1). In general, an arbitrary initial state of the field can be represented in the form $|\psi\rangle = F(\mathbf{a}^+)|0\rangle$, where $F(\mathbf{x})$ is some analytic function and the operators $\mathbf{a} \equiv (a_1, a_2, \ldots)$ pertain to the input modes of the channel. We define the operators at the channel output using the unitary transformation matrix

$$\mathbf{a}'^+ = \mathbf{T}\mathbf{a}^+ \quad (\mathbf{T}\mathbf{T}^+ = I).$$

Substituting the inverse transformation

$$\mathbf{a}^+ = \mathbf{T}^{-1}\mathbf{a}'^+ = \mathbf{T}^+\mathbf{a}'^+$$

into the function $F(\mathbf{a}^+)$, we find the state of the field in the basis of Fock states of the output modes defined in (3.1):

$$|\psi\rangle = F(\mathbf{a}^+)|0\rangle = F(\mathbf{T}^{-1}\mathbf{a}'^+)|0\rangle. \tag{4.1}$$

In the case under consideration the form of the function $F(\mathbf{a}^+)$ can be found from Eq. (3.9):

$$|\psi\rangle = 2^{-1/2}(\alpha_x a_x^+ + \alpha_y a_y^+)(b_x^+ c_y^+ - b_y^+ c_x^+)|0\rangle. \tag{4.2}$$

Inverting (3.4) yields

$$a_j^+ = (a'^+_j - b'^+_j)/\sqrt{2}, \quad c_x^+ = t_C^* c'^+_x - r_C c'^+_y,$$
$$b_j^+ = (a'^+_j + b'^+_j)/\sqrt{2}, \quad c_y^+ = r_C^* c'^+_x + t_C c'^+_y \tag{4.3}$$

$(j = x, y)$. Substituting these expressions into Eq. (4.2), we find the state of the field $|\psi\rangle$ at the system output in the form of a superposition of Fock states of the output modes $|n_j)$. For simplicity, let $T_C = 1$. Then

$$|\psi\rangle = \{[\,|A_y,B_x) - |A_x,B_y)][\,\alpha_x|C_x) + \alpha_y|C_y)] + [|B_x,B_y) - |A_x,A_y)][\,\alpha_x|C_x) - \alpha_y|C_y)\,] + \sqrt{2}[\,|2B_y) - |2A_y)\,]\alpha_y|C_x) + \sqrt{2}[\,|2A_x) - |2B_x)\,]\alpha_x|C_y)\}/\sqrt{8}. \tag{4.4}$$

Here $|2A_x) = (a'^+_x)^2|0\rangle/\sqrt{2}$ is the state with two photons in the output mode A_x.

The experiment measures the probabilities $p(n_1, n_2, \ldots) = p\{n_k\}$ of detecting certain sets of photon numbers $\{n_k\}$ in the output modes. According to the Born postulate (3.3), the probability amplitude $q(n_1, n_2, \ldots)$ of these events is equal to the product of the bra vector $(n_1, n_2, \ldots|$ and the ket vector (4.4). For example,

$$q(A_y, B_x, C_x) = (A_y, B_x, C_x|\psi\rangle = \alpha_x/\sqrt{8}.$$

To consider the dependence of $q(A_y, B_x, C_j)$ on T_A and T_C, it is convenient to first find the projection of $|\psi\rangle$ onto the subspace describing the detection of photons only in the modes A_y and B_x:

$$(A_y, B_x|\psi\rangle = [\,\alpha_x|C_x) + \alpha_y|C_y)]/\sqrt{8} \equiv |\psi\rangle_{C\,eff}/\sqrt{8}. \tag{4.5}$$

This quantity is an unnormalized vector in the space C. Here we have also introduced the normalized effective vector $|\psi\rangle_{C\mathrm{eff}}$ for the field C, which describes the effect of T_A and

T_C on the probability of the subset of events $(A_y, B_x) \equiv (0110^{**})$. The vector $|\psi\rangle_{C\text{eff}}$ has the same form as the initial state vector $|\psi\rangle_A$ for beam A, so the converters T_A and T_C affect $p(A_y, B_x, C_x)$ in the same way.

At the same time, according to Eq. (4.4), the events (A_x, A_y) and (B_x, B_y) yield the effective vector $\alpha_x|C_x\rangle - \alpha_y|C_y\rangle$ with the wrong sign α_y (which can be corrected by the transformation $T_C' = \sigma_z$, Ref. 1), and the events $(2A_j)$ and $(2B_j)$ generally do not exhibit the copying effect.

Above we assumed that (3.9) and (4.4) are one and the same vector $|\psi\rangle$ represented in different bases $|n\rangle$ and $|n)$. This is the so-called passive viewpoint regarding the transformation of the vector space. From the active point of view there is only one basis, $|n) \equiv |n\rangle$, and as the field propagates along the channel (or, equivalently, with the passage of time) its state vector varies under the action of the evolution operator, $|\psi\rangle \to |\psi\rangle' = U|\psi\rangle$ (here $|\psi\rangle'$ coincides with the right-hand side of (4.4) upon replacing $|n)$ by $|n\rangle$). We now express the effect of the evolution operator U on $|\psi\rangle$ in terms of the transformation matrix $\mathbf{T}$. To this end, we make the following substitution in $|\psi\rangle = F(\mathbf{a}^+)|0\rangle$:

$$|0\rangle \to U^+U|0\rangle = U^+|0\rangle$$

(assuming that there are no external forces producing photons from the vacuum). Now

$$|\psi\rangle' = UF(\mathbf{a}^+)U^+|0\rangle = F(U\mathbf{a}^+U^+)|0\rangle.$$

The operator $U\mathbf{a}^+U^+$ can be transformed in the following way, using the relation $\mathbf{T}\mathbf{a}^+ = U^+\mathbf{a}^+U$:

$$U\mathbf{a}^+U^+ = U[\mathbf{T}^{-1}(U^+\mathbf{a}^+U)]U^+ = \mathbf{T}^{-1}\mathbf{a}^+.$$

As a result,

$$|\psi\rangle' = U|\psi\rangle = F(\mathbf{T}^{-1}\mathbf{a}^+)|0\rangle. \tag{4.6}$$

This expression coincides with (4.1) upon making the substitution $\mathbf{a}^+ \to \mathbf{a}'^+$. The mean value of an arbitrary operator $G(\mathbf{a})$ observed at the output is given in the Heisenberg and Schrödinger pictures by

$$\langle G(\mathbf{a})\rangle' = {}'\langle\psi|G(\mathbf{a})|\psi\rangle' = \langle\psi|G(\mathbf{a}')|\psi\rangle. \tag{4.7}$$

5. DISCUSSION

Two basic approaches to the explanation and interpretation of the effect under discussion are possible. We call them ''metaphysical'' and ''minimalist.''

1. It is customarily assumed[1,2,5,7] that during the course of the experiment, it is possible, using an observation on photons A and B, to measure the Bell operator, which has the four eigenvectors

$$|\Psi^{(\pm)}\rangle_{AB} = 2^{-1/2}[|A_x, B_y\rangle \pm |A_y, B_x\rangle],$$
$$|\Phi^{(\pm)}\rangle_{AB} = 2^{-1/2}[|A_x, B_x\rangle \pm |A_y, B_y\rangle]. \tag{5.1}$$

The inverse transformation enables one to express the Fock states with one photon in each mode in terms of the Bell operators:

$$|A_x, B_y\rangle = [|\Psi^{(+)}\rangle_{AB} + |\Psi^{(-)}\rangle_{AB}]/\sqrt{2},$$
$$|A_x, B_x\rangle = [|\Phi^{(+)}\rangle_{AB} + |\Phi^{(-)}\rangle_{AB}]/\sqrt{2},$$
$$|A_y, B_x\rangle = [|\Psi^{(+)}\rangle_{AB} - |\Psi^{(-)}\rangle_{AB}]/\sqrt{2},$$
$$|A_y, B_y\rangle = [|\Phi^{(+)}\rangle_{AB} - |\Phi^{(-)}\rangle_{AB}]/\sqrt{2}. \tag{5.2}$$

Substituting these expressions into the initial state vector (3.9), which can be represented in the form $|\psi\rangle = |\psi\rangle_A|\Psi^{(-)}\rangle_{BC}$, we obtain an expansion in the Bell basis for the A and B beams (Ref. 1):

$$|\psi\rangle = \frac{1}{2}\{|\Psi^{(-)}\rangle_{AB}[-\alpha_x|C_x\rangle - \alpha_y|C_y\rangle] + |\Psi^{(+)}\rangle_{AB} \times [-\alpha_x|C_x\rangle + \alpha_y|C_y\rangle] + |\Phi^{(-)}\rangle_{AB}[\alpha_y|C_x\rangle + \alpha_x|C_y\rangle] + |\Phi^{(+)}\rangle_{AB}[-\alpha_y|C_x\rangle + \alpha_x|C_y\rangle]\}. \tag{5.3}$$

According to the reduction hypothesis, as a result of such a measurement of the Bell operator the initial three-photon state $|\psi\rangle$ is projected onto one of the four vectors $|\Psi^{(\pm)}\rangle_{AB}$, $|\Phi^{(\pm)}\rangle_{AB}$. According to Eq. (5.3), $|\psi\rangle$ thereupon collapses to one of the states (compare equations (4.4) and (4.5))

$$|\psi\rangle_{C1} = -\alpha_x|C_x\rangle - \alpha_y|C_y\rangle, \quad |\psi\rangle_{C2} = -\alpha_x|C_x\rangle + \alpha_y|C_y\rangle,$$
$$|\psi\rangle_{C3} = \alpha_y|C_x\rangle + \alpha_x|C_y\rangle, \quad |\psi\rangle_{C4} = -\alpha_y|C_x\rangle + \alpha_x|C_y\rangle, \tag{5.4}$$

where the number n labeling the state $|\psi\rangle_{Cn}$ is known. Thus, the act of measuring the Bell operator brings about the instantaneous collapse of the initial three-photon state to one of the one-photon states: $|\psi\rangle_{ABC} \to |\psi\rangle_{Cn}$, whereupon the polarization vector of the C photon $\mathbf{e}_{Cn}$ acquires the components $(\pm\alpha_x, \pm\alpha_y)$. The field or detectors in beam C ''find out'' about this as the result of superluminal interactions embodying ''quantum nonlocality.''

Classical information about the number $n = 1-4$ labeling the state $|\psi\rangle_{Cn}$ can be used to automatically ''correct'' the polarization of the C photons using the polarization converter T_C', whose Jones matrix $\mathbf{T}_n$ is equal to I, σ_z, σ_x, or $-i\sigma_y$. In this case

$$\mathbf{e}_{Cn} \to \mathbf{T}_n\mathbf{e}_{Cn} = \mathbf{e}_{C1} = \mathbf{e}_A.$$

As a result, information about the polarization of the A photon $\mathbf{e}_A$ is instantaneously transferred to the C photon, as if it were not located some distance away.[7]

In this approach, the Born postulate (3.3) is artificially divided into two successive steps (compare equation (4.5)) and the mathematical procedure of projecting onto some subspace is put in correspondence with a ''real'' event—instantaneous reduction. The possibility of both photons A and B falling into one mode and the necessity of using a shutter are not mentioned in the cited works.

2. The second approach to the description of the effect is based on the standard calculation (see Ref. 18) of the effects of an actual optical system (of beamsplitters and polarization converters) on the operators (Sec. 3) or the wave function (Sec. 4) of the field, and calculation of actually measured quantities, i.e., probabilities (see Ref. 18). According to the Born postulate (3.3) and the theory of photodetection, the probabilities of all events observed with photon counters D_A and D_B are determined by the projections of $|\psi\rangle$ onto the Fock states of the output modes $|A_j, B_k)$, $|2A_j)$, $|2B_k)$.

These states, represented by means of the transformation matrix in the basis of input modes $|\ldots\rangle$ (see (3.2)), differ from the Bell states (5.1) on account of additional components, so that the Bell states do not correspond to real observations (at least, not in the scheme of the experiment in Ref. 7).

The set of elementary probabilities found above in the one-photon approximation (3.11)–(3.13) for all possible combinations of readings of the photon counters describe the principal features of the observed phenomena. This approach also facilitates an instructive description of the effect in terms of a prescribed polarization and photon correlation (see Sec. 2).

According to the ''minimalist'' interpretation, the effect is considered to be a manifestation of quantum correlation between the three light beams, which is not in need of ''explanation'' using the mysterious, *ad hoc* concept of ''quantum nonlocality,'' which implies a peculiar interaction between remote devices akin to telepathy. Recall that analogous controllable correlations are also possible in classical models, with the difference between these correlations and quantum correlations of EPR–Bohm type being quite subtle.[20]

More consistent is the ''non-Kolmogorov'' concept of quantum mechanics,[21] which captures in a unified way a general property of quantum probabilistic models,namely the lack of elementary joint probabilities for noncommuting observables in the presence of marginal probabilities.[20,21] For example, it is possible to measure or calculate three probabilities $p_m(s_m)$ for the Stokes operators S_m of a plane wave $(m=1,2,3)$. In the one-photon state,

$$s_m=\pm 1, \quad p_m(\pm 1)=(1\pm\langle S_m\rangle)/2.$$

However, one can neither measure nor calculate the joint distribution $p(s_1,s_2,s_3)\geqslant 0$, which according to the Kolmogorov axiom of additivity defines the marginal distributions

$$p_1(s_1)=\sum_{s_2,s_3} p(s_1,s_2,s_3).$$

Consequently, quantum models describing experiments with single photons (or particles with spin 1/2) are non-Kolmogorov, and assigning a photon a set of definite, *a priori* properties $\{s_k\}$ in a single trial (i.e., the polarization vector **e**) does not make sense within the traditions of ''na'ive realism.''

A similar conclusion about the non-Kolmogorov behavior of two or more photons (or other systems) belonging to several different beams and described, in particular, by EPR states of the type (2.1) follows from a consistent approach to the resolution of a number of quantum paradoxes and their ''minimalist'' resolution.[21] Such an approach seems to be a reasonable alternative to an appeal to nonlocality, as it encompasses several regularities in one fell swoop: the lack of joint distributions for noncommuting operators (to which physicists have for a long time acquiesced under the rubric of ''dualism''), violation of the Bell inequalities, and other manifestations of ''quantum nonlocality'' (recall that the simplest proof of the Bell inequalities is based on the existence of a joint distribution for four observables—like the distribution of two Stokes parameters for two photons[20,21]). The lack of nonnegative joint distributions also follows directly from the properties of sets of quantum moments for certain states (regardless of the criteria of the sort represented by violations of the Bell inequality), i.e., the quantum moment problem does not always have a solution.[20,21]

When using a coincidence scheme (instead of a shutter; see Fig. 1), the consistency of detecting events in time in three detectors is irrelevant. The detectors in the three beams are equivalent, and separating the appearance of counts in them into two stages—first, the appearance of two counts in D_A and D_B bringing about reduction of the wave function, which then, on account of quantum nonlocality, influences the behavior of the third detector—is justified by neither a consistent theory nor experiment. We are equally justified in assuming that reduction takes place first in the detectors D_C (in the absence of a shutter they can be located closer to the three-photon light source than the detectors D_A and D_B). If D_A is much closer to the source than D_B, then one can ask at what instant reduction takes place: upon readout of D_B or D_A? It may well be that a chain of successive reductions takes place:

$$|\psi\rangle_{ABC}\rightarrow|\psi\rangle_{BC}\rightarrow|\psi\rangle_C.$$

Clearly these questions are of rhetorical significance, since the events are separated by spacelike intervals and concepts of it earlier and it later do not apply.

From a ''minimalist'' standpoint, reduction of the wave function is an extraneous hypothesis: as was shown above, the observable effects are completely described by the standard formalism in the Heisenberg and Schrödinger pictures by the Born postulate; see Eqs. (4.7). Accordingly, one is free to choose the carrier of information about the polarization vector $\mathbf{e}_A$ from one end of the optical system to the other to be the wave function of the field $(\psi\rightarrow\psi')$ or the field operators $(a_k\rightarrow a_k')$, just as in the case of a single polarized beam. Of course, not all observable events violate special relativity (if only because the quantum Green's functions for the free field have the same form as the classical).

The equivalence of the three detectors breaks down if modulators are used, that is, a shutter and/or an additional converter T_C'. The shutter controls pulses from the detectors D_A and D_B, and in essence replaces the triple-coincidence scheme by blocking photons in the output beam C in ''unfavorable'' cases. In this case, the device in fact serves to prepare single photons with the (unknown) polarization $\mathbf{e}_C$, which copy the polarization of the A photons $\mathbf{e}_A$. Classical signals controlling the shutter, which now and then lets photons through, take part in the copying process $\mathbf{e}_A\rightarrow\mathbf{e}_C$. Note that the action of the shutter can be described by a nonunitary transformation of the field (in contrast to phase plates and beamsplitters), so that the field C behind the shutter must be described by a mixed state.[18,19] Information is ''teleported'' in at most 50% of the trials. Due to nonideality of the detectors and other system elements, their real fraction will in fact be much less.[7]

We may point to a similar but more primitive device, which prepares single polarized photons with known creation times (but without the copying effect) using a two-photon

source and a shutter;[22–24] in this case the detection of one photon in beam A causes the shutter along the path of its twin photon in beam B to open.

6. CONCLUSION

The aforementioned necessity of adding a shutter to experiments demonstrating quantum teleportation is a purely technical problem, easily solved by present-day technology. At the same time, from a conceptual point of view, only the absence or presence of the shutter differentiates three-photon interference observed using a triple-coincidence scheme in Ref. 7 from teleportation, by which we mean the production of photons that mimic the polarization of the incident photons. The unitary transformations $T_C'=\sigma_n$ (see Ref. 1) in the optical case are not of fundamental significance—they only enable us to double the efficiency with which copies are produced.

In a quantitative description of experiments like those described in Refs. 7 and 8, the concepts of instantaneous reduction of the wave function and quantum nonlocality are superfluous; they are merely a traditional auxiliary means for interpreting quantum theory and obtaining an instructive, graphic picture of what ''really'' happens. Of course, the choice between the ''metaphysical'' and ''minimalist'' interpretations described in Sec. 5 is a matter of taste (which of course distinguishes interpretation from theory). Favoring our approach are an instructive correlation model and the analogous classical effect (see Sec. 2), the deduced need for a shutter (see Sec. 3), and Occam's Razor. Whatever one's choice, it would be well to remain mindful of the existence of alternative points of view.

At present, all experimental data known to the author are accurately described by the standard algorithms of quantum theory and Born's postulate. An actual need for the reduction postulate arises only in an attempt to provide a quantitative description of a narrow class of ''time-of-flight'' correlation experiments, in which repeated observations of a single system are made in succession with high temporal resolution, where because of interactions the observed operators do not commute.[11]

The present work was carried out with the financial support of the Russian Fund for Fundamental Research (in the State Committee on Science and Technology subject area ''Fundamental Metrology,'' Grant No. 96-02-16334-a) and the Russian Foundation for the Support of Leading Scientific Schools (School on ''Quantum Phenomena in Nonlinear Optics,'' Grant No. 96-15-96673).

*)E-mail: dnk@klyshko.phys.msu.su

[1] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[2] H. Weinfurter, Europhys. Lett. **25**, 559 (1994).
[3] L. Davidovich, N. Zagury, M. Brune, J. M. Raimond, and S. Haroche, Phys. Rev. A **50**, R895 (1994).
[4] J. I. Cirac and A. S. Parkins, Phys. Rev. A **50**, R4441 (1994).
[5] S. L. Braunstein and A. Mann, Phys. Rev. A **51**, R1727 (1995); **53**, 630(E) (1996).
[6] S. L. Braunstein and H. J. Kimble, Phys. Rev. Lett. **80**, 869 (1998).
[7] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Nature (London) **390**, 575 (1997).
[8] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, Phys. Rev. Lett. **80**, 1121 (1998).
[9] X. Maître, E. Hagley, G. Nogues, C. Wunderlich, P. Goy, M. Brune, J. M. Raimond, and S. Haroche, Phys. Rev. Lett. **79**, 769 (1997).
[10] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, Phys. Rev. Lett. **75**, 4337 (1995).
[11] D. N. Klyshko, Phys. Lett. A **123**, 179 (1998); Laser Phys. **7**(6) (1997).
[12] H. Margenau, Ann. Phys. (N.Y.) **23**, 469 (1963).
[13] D. Home and M. A. B. Whitaker, Phys. Lett. A **128**, 1 (1988).
[14] L. E. Ballentine, Int. J. Theor. Phys. **27**, 211 (1988).
[15] *Quantum Mechanics without Reduction*, M. Sini and J. Levy-Leblond (eds.), Hilger, Bristol (1990).
[16] M. Namiki and S. Pascazio, Phys. Rev. A **44**, 39 (1993).
[17] C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).
[18] D. N. Klyshko, Usp. Fiz. Nauk **164**, 1187 (1994).
[19] D. N. Klyshko, Phys. Lett. A **137**, 334 (1989).
[20] N. V. Evdokimov, D. N. Klyshko, V. P. Komolov, and V. A. Yarochkin, Usp. Fiz. Nauk **166**, 91 (1996).
[21] D. N. Klyshko, Phys. Lett. A **218**, 119 (1996); Laser Phys. **6**, 1056 (1996).
[22] B. Ya. Zel'dovich and D. N. Klyshko, JETP Lett. **9**, 40 (1969).
[23] D. N. Klyshko, Kvant. Elektron. **4**, 1056 (1977) [Sov. J. Quantum Electron. **7**, 591 (1977)].
[24] C. K. Hong and L. Mandel, Phys. Rev. Lett. **56**, 58 (1986).

Translated by Paul F. Schippnick

Effective permittivity of a discrete random medium with anisotropic inclusions

O. V. Bagatskaya,*) N. P. Zhuk,*) and S. N. Shul'ga*)

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted 20 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 1188–1201 (October 1998)

We use the method of substitution of field variables in the bilocal approximation to find the effective permittivity of a two-phase composite random medium in the form of an ensemble of small, arbitrarily anisotropic spherical inclusions distributed inside an isotropic matrix. To illustrate the results we calculate the damping of the plane waves of the mean field in such a medium. © *1998 American Institute of Physics.* [S1063-7761(98)00410-7]

1. INTRODUCTION

The stochastic method of describing material media, which is aimed at explaining the space–time structure of a real medium and the chaotic variation from point to point or with the passage of time of the properties of the medium, is widely used in the theory of propagation of electromagnetic waves in the turbulent ionosphere,[1] in the microwave probing of the Earth's solid surface,[2,3] in the radio-wave monitoring of polymeric composite materials,[4,5] in analyzing the electromagnetic properties of polycrystalline materials,[6,7] and in the theory of artificial media.[8,9] In all these applications, electromagnetic anisotropy is a characteristic feature of the medium in which the waves propagate.

As is known,[10–13] the propagation of a statistically mean field through a random medium with properties undergoing chaotic spatial variations is described by equations characteristic of a deterministic medium with spatial dispersion. The material parameters of this deterministic, or effective, medium characterize the properties of the random medium in relation to the mean field and are called the effective parameters of the latter medium.

A method for calculating the effective permittivity of a medium with permittivity fluctuations was proposed by Lifshits *et al.*[6] Note that this method was later ''rediscovered'' by Bourret[14] and Keller and Karal,[15] although earlier it had been used in Refs. 6, 16, and 17. Bearing in mind the direction of our investigation, which takes into account the effect of the electromagnetic anisotropy of the medium on the propagation and scattering of electromagnetic waves, we note the various papers[18–22] in which the method of Lifshits *et al.*[6] is used to find and study the effective permittivity of uniaxial[18–20] and gyrotropic[21,22] random media, whose permittivity tensors in a certain system of coordinates have a form that is typical of a uniaxial crystal or a magnetically active plasma and are characterized, respectively, by three (diagonal) or five (three diagonal and two off-diagonal) nonvanishing components.

The method adopted in this paper for calculating the effective permittivity may be called the ''ordinary'' theory of multiple scattering for an electromagnetic field, ordinary in the sense that it is similar to the theory of multiple scattering for a scalar (acoustic) field in a random medium.[10–13] In this method, the effective permittivity is represented by a series expansion in powers of a certain parameter, which for an acoustic field proves to be small even for strong fluctuations of the properties of the medium, provided that these fluctuations are small-scale.[10–13] Because of the strong singularity (of the Dirac delta-function type) of the electric Green's function at the point occupied by the source,[23,24] the characteristic parameter in this theory for an electromagnetic field proves to be proportional to the intensity of the fluctuations.[25,26] To complete the picture, we note that the bilocal approximation proper was introduced by Lifshits and Rozentsveĭg,[27,28] who used it to find the effective parameters of a microheterogeneous elastic medium.

To calculate the effective permittivity of media with strong fluctuations of their properties, one usually uses the method of substitution of field variables (or renormalization), proposed by Finkel'berg[29,30] and developed in Refs. 31–43. It amounts to isolating the singular component (of the Dirac delta-function type) in the electric matrix Green's function and introducing equations for a new field variable for which the Green's function coincides with the regular part of the electric matrix Green's function. Note that the singular and regular parts of the electric Green's function can be interpreted (and this is a convenient feature of the Green's function) as generalized functions generated by, respectively, the nondecreasing and vanishing-at-infinity parts of the spectral matrix Green's function.[31,39,44] As in the above approach, the adopted method does not provide simple and at the same time exact expressions for the effective permittivity, i.e., the effective permittivity is still represented by a perturbation-theory expansion, but at least in our case the characteristic parameter remains small even for strong fluctuations of the material properties, provided that the spatial scale of such fluctuations is small.[31,32,35]

In view of our interest in random media with electromagnetic anisotropy, we note that the method of substitution of field variables has been used in calculations of the effective permittivity of an electrically isotropic random medium with anisometric perturbations and fixed[32,35] or random[38] orientation of the statistical-symmetry axes, of a gyrotropic medium of the type of a magnetically active plasma[31] or a rotating medium,[37] and of an arbitrarily anisotropic random medium with fixed or randomly directed statistical-symmetry

axes.[39] The results of Zhuck[39] can only be applied to the special model of a continuous medium considered in that paper, while the other, canonical, model of a discrete random medium was ignored in Ref. 39. We would like to note at this point that there exists a well-known technique[45,46] that makes it possible to incorporate a discrete random medium into the class of models that can be analyzed by the method of substitution of field variables. From this the goal of the present investigation follows naturally, i.e., to calculate the effective permittivity of a microheterogeneous composite that consists of small anisotropic particles distributed within an isotropic medium and to study the dissipative properties of such a composite.

Below we use the results of Refs. 29–39 to describe a method for calculating the effective permittivity of a statistically homogeneous continuous random medium with an arbitrary anisotropy of the electrical and statistical properties and strong but small-scale (in comparison to the wavelength) fluctuations. The method is then applied to a model of a discrete random medium obtained by uniformly distributing an ensemble of chaotically oriented small spherical inclusions of an arbitrarily anisotropic homogeneous insulator in an isotropic matrix. As a result we find the effective permittivity of the two-phase composite mentioned earlier. The formulas can be interpreted as generalizations of similar formulas in Ref. 45, which deals with electrically isotropic inclusions. Using the formula for the effective permittivity and the method of contour integration, for a random composite without dissipative losses we calculate the real and imaginary parts of this quantity and the ''diffraction'' correction to the propagation constant for the mean field. The imaginary part of this correction describes the damping of the mean field due to scattering by the random inclusions.

2. THE METHOD OF CHANGING THE FIELD VARIABLES FOR AN ANISOTROPIC RANDOM MEDIUM

2.1. Statement of the effective permittivity problem

Consider the equations

$$\operatorname{curl}\operatorname{curl}\mathbf{E}_r - k_0^2\hat{\varepsilon}^{(r)}\mathbf{E}_r = \frac{4\pi i k_0}{c}\mathbf{J}, \tag{1}$$

$$\mathbf{H}_r = \frac{1}{ik_0}\operatorname{curl}\mathbf{E}_r \tag{2}$$

for a random electromagnetic field $\mathbf{E}_r$ and $\mathbf{H}_r$ generated by extraneous electric sources $\mathbf{J}$ in an infinite dielectric medium with permittivity $\hat{\varepsilon}^{(r)}$. We assume that the permittivity tensor $\hat{\varepsilon}^{(r)}$ of the medium has all of its nine components $\varepsilon_{mn}^{(r)}$, which are random functions of the radius vector $\mathbf{x}=(x_1,x_2,x_3)$.

The effective permittivity operator $\hat{\varepsilon}^{(e)}$ is defined by the identity[25,26]

$$\langle\hat{\varepsilon}^{(r)}\mathbf{E}_r(\mathbf{x})\rangle \equiv \hat{\varepsilon}^{(e)}\langle\mathbf{E}_r(\mathbf{x})\rangle. \tag{3}$$

Having this operator, we can easily show, on the basis of Eqs. (1) and (2), that the mean-field excitation is described by the equations

$$\operatorname{curl}\operatorname{curl}\langle\mathbf{E}_r\rangle - k_0^2\hat{\varepsilon}^{(e)}\langle\mathbf{E}_r\rangle = \frac{4\pi i k_0}{c}\mathbf{J}, \tag{4}$$

$$\langle\mathbf{H}_r\rangle = \frac{1}{ik_0}\operatorname{curl}\langle\mathbf{E}_r\rangle, \tag{5}$$

which are a characteristic feature of a spatially dispersive medium with nonlocal permittivity $\hat{\varepsilon}^{(e)}$.

In a statistically homogeneous medium, extraneous sources in the form of a spatial harmonic,

$$\mathbf{J}(\mathbf{x}) = \mathbf{J}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{6}$$

generate a mean field of the same form:[39]

$$\langle\mathbf{E}_r(\mathbf{x})\rangle = \mathbf{E}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \langle\mathbf{H}_r(\mathbf{x})\rangle = \mathbf{H}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{7}$$

where $\mathbf{k}$ is an arbitrarily specified three-dimensional wave vector, and $\mathbf{J}(\mathbf{k})$, $\mathbf{E}(\mathbf{k})$, and $\mathbf{H}(\mathbf{k})$ are the vector amplitudes of the sources and the electromagnetic field, respectively.

Applying the operator $\hat{\varepsilon}^{(e)}$ to the mean-electric-field vector, we get

$$\hat{\varepsilon}^{(e)}\langle\mathbf{E}_r(\mathbf{x})\rangle = e^{i\mathbf{k}\cdot\mathbf{x}}\hat{\varepsilon}^{(e)}(\omega,\mathbf{k})\mathbf{E}(\mathbf{k}), \tag{8}$$

where $\hat{\varepsilon}^{(e)}(\omega,\mathbf{k})$ is the effective permittivity tensor in the spectral region; the dependence on ω and $\mathbf{k}$ points to the frequency and spatial dispersions of the effective medium. The characteristic scale Δk of variation of this tensor as a function of the spectral parameter $\mathbf{k}$ is of order $1/L$, where L is the correlation interval for the perturbations of the medium. Combining this fact with (8), we find that in the long-wavelength mode, where $kL \ll 1$ (or in other words, for small-scale perturbations), the properties of the random medium in relation to the mean field in the spatial-harmonic form (7) or in the form of a linear combination of such harmonics, are described by the effective permittivity tensor $\hat{\varepsilon}^{(e)}$ (Refs. 25, 26, and 39):

$$\hat{\varepsilon}^{(e)}(\omega) = \lim_{k\to 0}\hat{\varepsilon}^{(e)}(\omega,\mathbf{k}). \tag{9}$$

Setting up the effective permittivity tensor of a medium with small-scale perturbations is a very important problem of the theory of multiple scattering of electromagnetic waves in a random medium.

2.2. Renormalized scattering equation

In the first stage of constructing the effective permittivity tensor we follow the well-known pattern developed in Refs. 35, 36, and 39, which amounts to transforming the stochastic differential equation (1) into an integral equation for a new field variable $\mathbf{F}$. To this end we introduce an anisotropic reference medium, or a comparison medium, which fills the entire space and has a constant deterministic permittivity tensor $\hat{\varepsilon}$. (Below this tensor is defined as the solution of Eq. (23).) Let $\mathbf{E}_b(\mathbf{x})$ be the electric vector of the field generated by the extraneous sources $\mathbf{J}(\mathbf{x})$ in the reference medium, and let $\hat{G}(\mathbf{x}-\mathbf{x}')$ be the electric matrix Green's function of this medium. In our case we can assume that the Green's function is a fixed inverse Fourier transform:

$$\hat{G}(\mathbf{x}-\mathbf{x}')=(2\pi)^{-3}\int d^3k\,\exp[i\mathbf{k}(\mathbf{x}-\mathbf{x}')]\,\hat{G}(\mathbf{k}), \quad (10)$$

where $\hat{G}(\mathbf{k})$ is the spectral Green's function, which can be found by solving the equation

$$[\mathbf{k}[\mathbf{k}\hat{G}(\mathbf{k})]]+k_0^2\hat{\varepsilon}\hat{G}(\mathbf{k})=-\hat{I} \quad (11)$$

with $\hat{I}$ the identity matrix. Note that the singularities (poles) of the integrand in (10), which coincide with the zeros of $\Delta(\mathbf{k})$ of Eq. (16) below, are assumed to lie in the complex plane because of dissipative losses (possibly negligible losses) in the reference medium. The same interpretation, if not stated otherwise, can be applied to other integrals (e.g., (30)) whose integrands have nonintegrable singularities.

Using the Green's function $\hat{G}(\mathbf{x}-\mathbf{x}')$, we can employ a standard procedure for replacing the differential equation (1) by the integral equation

$$\mathbf{E}_r(\mathbf{x})=\mathbf{E}_b(\mathbf{x})+k_0^2\int d^3x'\,\hat{G}(\mathbf{x}-\mathbf{x}')\times[\hat{\varepsilon}^{(r)}(\mathbf{x}')-\hat{\varepsilon}]\mathbf{E}_r(\mathbf{x}'). \quad (12)$$

The solution of Eq. (11) in coordinate-free form[47] is

$$\hat{G}(\mathbf{k})=-\frac{\mathbf{k}\otimes\mathbf{k}}{k_0^2\zeta(\mathbf{k})}+\hat{G}^{(1)}(\mathbf{k}). \quad (13)$$

Here and in what follows the Cartesian product signified by $\otimes$ is a tensor of the simplest form: for a vector $\mathbf{k}=(k_1,k_2,k_3)$,

$$\mathbf{k}\otimes\mathbf{k}=\begin{bmatrix}k_1k_1 & k_1k_2 & k_1k_3\\ k_2k_1 & k_2k_2 & k_2k_3\\ k_3k_1 & k_3k_2 & k_3k_3\end{bmatrix},$$

$$\hat{G}^{(1)}(\mathbf{k})=\frac{\hat{D}(\mathbf{k})}{\zeta(\mathbf{k})\Delta(\mathbf{k})}, \quad (14)$$

$$\hat{D}(\mathbf{k})=(\zeta\hat{I}-\mathbf{k}\otimes\mathbf{k}\hat{\varepsilon})(\zeta\hat{I}-\hat{\varepsilon}\mathbf{k}\otimes\mathbf{k})+k_0^2(\mathbf{k}\otimes\mathbf{k}\det\hat{\varepsilon}-\zeta\,\mathrm{adj}\,\hat{\varepsilon}), \quad (15)$$

$$\Delta(\mathbf{k})=k^2\zeta+k_0^2(\mathbf{k}\hat{\varepsilon}\hat{\varepsilon}\mathbf{k}-\zeta\,\mathrm{Tr}\,\hat{\varepsilon})+k_0^4\det\hat{\varepsilon}, \quad (16)$$

$$\zeta(\mathbf{k})=\mathbf{k}\hat{\varepsilon}\mathbf{k}. \quad (17)$$

In these expressions, $\det\hat{\varepsilon}$ and $\mathrm{adj}\hat{\varepsilon}$ denote the determinants of the matrix $\hat{\varepsilon}$ and the adjoint of $\hat{\varepsilon}$ (Ref. 47).

Now we introduce a (constant) renormalization matrix $\hat{S}$, defined in Eq. (31) below, and specify the spectral function $\hat{G}^{(2)}(\mathbf{k})$ according to the following expression:

$$\hat{G}^{(2)}(\mathbf{k})=\hat{G}(\mathbf{k})+\frac{\hat{S}}{k_0^2}=-\frac{\mathbf{k}\otimes\mathbf{k}}{k_0^2\zeta(\mathbf{k})}+\frac{\hat{S}}{k_0^2}+\hat{G}^{(1)}(\mathbf{k}), \quad (18)$$

while the corresponding spatial Fourier transform is defined as

$$\hat{G}^{(2)}(\mathbf{x}-\mathbf{x}')=\hat{G}(\mathbf{x}-\mathbf{x}')+\frac{\hat{S}}{k_0^2}\,\delta(\mathbf{x}-\mathbf{x}'), \quad (19)$$

where $\delta(\mathbf{x}-\mathbf{x}')$ is the three-dimensional Dirac delta function. Plugging (19) into (12) and performing simple algebraic transformations, we obtain

$$\mathbf{F}(\mathbf{x})=\mathbf{E}_b(\mathbf{x})+k_0^2\int d^3x'\,\hat{G}^{(2)}(\mathbf{x}-\mathbf{x}')\hat{\xi}(\mathbf{x}')\mathbf{F}(\mathbf{x}'), \quad (20)$$

where

$$\mathbf{F}=[\hat{I}+\hat{S}(\hat{\varepsilon}^{(r)}-\hat{\varepsilon})]\mathbf{E}_r, \quad (21)$$

$$\hat{\xi}=(\hat{\varepsilon}^{(r)}-\hat{\varepsilon})[\hat{I}+\hat{S}(\hat{\varepsilon}^{(r)}-\hat{\varepsilon})]^{-1}. \quad (22)$$

Equation (20) is the sought integral equation for the new field variable $\mathbf{F}$ with a random perturbation $\hat{\xi}$. At $\hat{S}=0$ this equation obviously becomes the initial equation (12).

To ensure that the random perturbation $\hat{\xi}$ is small at least on the average, we require that $\langle\hat{\xi}\rangle=0$. In view of (22), this requirement can be transformed into[39]

$$\langle(\hat{\varepsilon}^{(r)}-\hat{\varepsilon})[\hat{I}+\hat{S}(\hat{\varepsilon}^{(r)}-\hat{\varepsilon})]^{-1}\rangle=0. \quad (23)$$

From Eq. (34) below it follows that $\hat{S}\equiv\hat{S}(\hat{\varepsilon})$. This means that (23) is actually an equation for determining the permittivity tensor of the reference medium, $\hat{\varepsilon}$.

2.3. Effective perturbation operator and effective permittivity

Iterating and averaging Eq. (20) in the bilocal approximation,[6,14,15] we arrive at the equation

$$\langle\mathbf{F}(\mathbf{x})\rangle=\mathbf{E}_b(\mathbf{x})+k_0^2\int\int d^3x'\,d^3x''\,\hat{G}^{(2)}(\mathbf{x}-\mathbf{x}')\hat{\xi}^{(e)}\times(\mathbf{x}'-\mathbf{x}'')\langle\mathbf{F}(\mathbf{x}'')\rangle, \quad (24)$$

where

$$\hat{\xi}^{(e)}(\mathbf{x}'-\mathbf{x}'')=k_0^2\langle\hat{\xi}(\mathbf{x}')\hat{G}^{(2)}(\mathbf{x}'-\mathbf{x}'')\hat{\xi}(\mathbf{x}'')\rangle. \quad (25)$$

Comparing (24) with the averaged version of Eq. (20), we find that

$$\langle\hat{\xi}(\mathbf{x})\mathbf{F}(\mathbf{x})\rangle\equiv\hat{\xi}^{(e)}\langle\mathbf{F}(\mathbf{x})\rangle, \quad (26)$$

where $\hat{\xi}^{(e)}$ is an integral operator acting on $\mathbf{x}$ and having a kernel $\hat{\xi}^{(e)}(\mathbf{x}-\mathbf{x}')$. Combining (8) with Eqs. (21) and (22), we get

$$\langle\mathbf{F}\rangle=[\hat{I}+\hat{S}(\hat{\varepsilon}^{(e)}-\hat{\varepsilon})]\langle\mathbf{E}_r\rangle, \quad \langle\hat{\xi}\mathbf{F}\rangle=(\hat{\varepsilon}^{(e)}-\hat{\varepsilon})\langle\mathbf{E}_r\rangle. \quad (27)$$

If we now plug these expressions into Eq. (26), we obtain a relationship that links the operators $\hat{\varepsilon}^{(e)}$ and $\hat{\xi}^{(e)}$. In the spectral domain this relationship becomes

$$\hat{\varepsilon}^{(e)}(\boldsymbol{\omega},\mathbf{k})-\hat{\varepsilon}=\hat{\xi}^{(e)}(\boldsymbol{\omega},\mathbf{k})+\hat{\xi}^{(e)}(\boldsymbol{\omega},\mathbf{k})\hat{S}[\hat{\varepsilon}^{(e)}(\boldsymbol{\omega},\mathbf{k})-\hat{\varepsilon}], \quad (28)$$

where $\hat{\xi}^{(e)}(\boldsymbol{\omega},\mathbf{k})$ is the Fourier transform of $\hat{\xi}^{(e)}(\mathbf{x}-\mathbf{x}')$. In the long-wavelength approximation (as $k\to0$) we have

$$\hat{\varepsilon}^{(e)}(\boldsymbol{\omega})-\hat{\varepsilon}=\hat{\xi}^{(e)}(\boldsymbol{\omega})+\hat{\xi}^{(e)}(\boldsymbol{\omega})\hat{S}[\hat{\varepsilon}^{(e)}(\boldsymbol{\omega})-\hat{\varepsilon}]. \quad (29)$$

The elements of the matrix $\hat{\xi}^{(e)}(\boldsymbol{\omega})=\lim_{k\to0}\hat{\xi}^{(e)}(\boldsymbol{\omega},\mathbf{k})$ can be found from (18) and (25) via the formula

$$\xi_{mn}^{(e)}(\omega)=\int d^3k\, B_{mpqn}(\mathbf{k})\left[-\frac{k_p k_q}{\zeta(\mathbf{k})}+S_{pq}+k_0^2 G_{pq}^{(1)}(\mathbf{k})\right], \quad (30)$$

which contains the spectral functions $B_{mpqn}(\mathbf{k})$ of the random perturbation $\hat{\xi}(\mathbf{x})$:

$$B_{mpqn}(\mathbf{k})=(2\pi)^{-3}\int d^3x \exp[-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')] \times C_{mpqn}(\mathbf{x}-\mathbf{x}'). \quad (31)$$

Here

$$C_{mpqn}(\mathbf{x}-\mathbf{x}')\equiv\langle\xi_{mp}(\mathbf{x})\xi_{qn}(\mathbf{x}')\rangle \quad (32)$$

are the correlation functions of the random perturbations ξ_{mn}. For our further discussions it is convenient to denote the value of the correlation function (32) at $\mathbf{x}=\mathbf{x}'$ by

$$\Gamma_{mpqn}=\int d^3k\, B_{mpqn}(\mathbf{k}). \quad (33)$$

Now that we have $\hat{\xi}^{(e)}(\omega)$ we can define the effective permittivity tensor $\hat{\varepsilon}^{(e)}(\omega)$ as the solution of Eq. (29). To ensure that the expressions we derive are valid in the case of strong fluctuations, it is advisable to select the matrix $\hat{S}$ in such a way that

$$\Gamma_{mpqn}S_{pq}=\int d^3k\, B_{mpqn}(\mathbf{k})\frac{k_p k_q}{\zeta(\mathbf{k})}. \quad (34)$$

Actually this is a system of nine equations for determining the nine unknown quantities S_{pq} as functions of ε_{mn}. To obtain an equation for the unknown quantities ε_{mn}, we plug $S_{pq}=S_{pq}(\hat{\varepsilon})$ into (23). According to (22) and (31)–(33), the quantities Γ_{mpqn} and B_{mpqn} are related to the unknown quantities S_{pq} through ξ_{mn}. Hence Eqs. (34) are generally nonlinear in S_{pq}. Combining Eqs. (34) with (30) and (14), we arrive at the final expression for $\xi_{mn}^{(e)}$:

$$\xi_{mn}^{(e)}(\omega)=k_0^2\int d^3k\, B_{mpqn}(\mathbf{k})\frac{D_{pq}(\mathbf{k})}{\zeta(\mathbf{k})\Delta(\mathbf{k})}. \quad (35)$$

If we assume that the $\xi_{mn}^{(e)}(\omega)$ are small (the condition needed for this assumption to be true is analyzed below), we arrive at an approximate solution of Eq. (29) for $\varepsilon_{mn}^{(e)}(\omega)$, the components of the effective permittivity tensor found by perturbation techniques:

$$\hat{\varepsilon}^{(e)}(\omega)=\hat{\varepsilon}+\hat{\xi}^{(e)}(\omega). \quad (36)$$

A combination of this formula and (35) solves the problem of the effective permittivity of a random anisotropic medium.

When the frequency dispersion of the random medium can be ignored, the last term on the right-hand side of Eq. (36), in accordance with (35), tends to zero with frequency as ω^2, while the first term, the permittivity of the reference medium, remains unchanged. Hence the permittivity of the reference medium, $\hat{\varepsilon}$, can be interpreted as the value of the effective permittivity of the random medium in the static limit, while the additional term $\hat{\xi}^{(e)}(\omega)$ can be interpreted as the ''diffraction'' contribution to this quantity due to the scattering of the field by random perturbations.[25,35]

2.4. Strong perturbations of the properties of the medium

The limits of applicability of the above results can easily be found.[35] First, however, we must establish the parameter used in the perturbation expansion in the calculation of the effective permittivity and make this parameter small. If we do this, we can be sure that truncating the expansion at the first two terms (the right-hand side of Eq. (36)) is justified. To find the perturbation-theory parameter, we need only estimate the ratio of any two consecutive term in the expansion, say, the terms on the right-hand side of Eq. (36). In order to arrive at such an estimate it is convenient to introduce a positive determinate quantity σ_ε, equal to the characteristic value of permittivity fluctuations. Then we can assume that approximately $B_{mpqn}(\mathbf{k})$ is equal to $\sigma_\varepsilon^2 L^2$ for $k<1/L$ and $B_{mpqn}(\mathbf{k})\approx 0$ for $k>1/L$, where L is the characteristic spatial scale of the fluctuations. According to what has been said earlier, the domain of integration in (35) is limited by the sphere $k<1/L$, so that the characteristic value of the integration variable k is $1/L$. For small-scale fluctuations ($k_0 L\ll 1$) this value is much larger than k_0, which we can ignore in the integrand in (36) in comparison to the large value $k\sim 1/L$. The resulting integral is equal, in order of magnitude, to $\sigma_\varepsilon^2(k_0L)^2$. Since on the scale of the parameters σ_ε and k_0L the order of the first term on the right-hand side of Eq. (36) is unity, the ratio of the second term on the right-hand side of Eq. (36) to the first is approximately $\sigma_\varepsilon^2(k_0L)^2$. Hence the right-hand side of Eq. (36) is a section of the series expansion in powers of the parameter we have found here, and the smallness of this parameter,

$$\sigma_\varepsilon^2(k_0L)^2\ll 1, \quad (37)$$

justifies the truncation of this perturbation series after a finite number of terms; namely, after the first two terms, as in (36). A remarkable feature of the condition (37) is that it is an indication that (35) and (36) can be used in the case of strong fluctuations of the properties ($\sigma_\varepsilon\gg 1$), provided that the fluctuations are small-scale ($k_0L\ll\sigma_\varepsilon^{-1}\ll 1$). This positive feature of the method of substitution of field variables is well-known.[25,26,35]

In conclusion of this section we note that for the special case of statistically isometric perturbations Eq. (34) provides an explicit way for finding the renormalization matrix S_{pq} (see Refs. 25, 26, and 39). Clearly, if we allow for the fact that the correlation functions of isometric perturbations (Eq. (32)) depend only on the distance[12]

$$R=|\mathbf{x}-\mathbf{x}'| \quad (38)$$

between the points $\mathbf{x}$ and $\mathbf{x}'$ and the spectral functions (31) depend only on the absolute value k of the spectral parameter $\mathbf{k}$, then, plugging into (34) the expression for Γ_{mpqn} from (33) and going over to integration with respect to the spherical coordinates k, θ, and φ, we can write Eq. (34) as

$$\int_0^\infty dk\, k^2 B_{mpqn}(\mathbf{k})\left[4\pi S_{pq}-\int d\Omega\,\frac{n_p n_q}{\mathbf{n}\hat{\varepsilon}\mathbf{n}}\right]=0, \quad (39)$$

where $d\Omega=\sin\theta\, d\varphi\, d\theta$ is the solid-angle element, and $n_p=k_p/k$ is the pth component of the unit vector along $\mathbf{k}$. The solution of Eq. (39) for S_{pq} can be found explicitly:

$$S_{pq}=\frac{1}{4\pi}\int d\Omega\,\frac{n_p n_q}{\mathbf{n}\hat{\varepsilon}\mathbf{n}}. \tag{40}$$

We see that $S_{pq}=S_{qp}$, i.e., in our case the renormalization matrix is symmetric: $\hat{S}^{\mathrm{T}}=\hat{S}$, where the superscript ''T'' indicates transposition.

3. EFFECTIVE PERMITTIVITY OF A TWO-PHASE COMPOSITE

3.1. Description of model

We will now use the tools developed in Sec. 2 to analyze a discrete random medium obtained by immersing a statistical ensemble of identical homogeneous anisotropic inclusions in a homogeneous isotropic enclosing medium, the matrix. We denote the permittivity of the medium by $\varepsilon^{(1)}$ and the permittivity tensor of an inclusion in the reference frame comoving with the inclusion, by $\hat{\varepsilon}^{(c)}$, with the elements of the latter tensor being $\varepsilon_{jk}^{(c)}$.

Let $\theta^{(r)}(\mathbf{x})$ be the indicator, or characteristic, function (see Ref. 46, p. 79) of the ensemble of the random inclusions; the function is unity if $\mathbf{x}$ belongs to an inclusion and zero for all other $\mathbf{x}$. The inclusions are thought to be distributed in space on the average uniformly and isotropically, so that their concentration (relative volume occupied by the inclusions),

$$\langle\theta^{(r)}(\mathbf{x})\rangle=v_2, \tag{41}$$

is independent of the point $\mathbf{x}$, while the two-point indicator function

$$\langle\theta^{(r)}(\mathbf{x})\,\theta^{(r)}(\mathbf{x}')\rangle=p(R) \tag{42}$$

depends only on R. Strictly speaking, the latter assumption is true only for spherical inclusions,[45,46] so that below we assume that the inclusions are spherical.

The statistical topology of the composite is determined by the dimensionless two-point correlation function (see Ref. 46, p. 74)

$$\varphi(R)=\frac{p(R)-v_2^2}{v_2(1-v_2)}, \tag{43}$$

whose Fourier transform $\tilde{\varphi}(k)$ is defined in the same way as in (31). Note that $\tilde{\varphi}(k)$ defined as the spectral density of the correlation function $[\theta^{(r)}(\mathbf{x})-v_2]/\sqrt{v_2(1-v_2)}$ of a real random process assumes, for real k, only real nonnegative values.[10]

The components of the permittivity tensor $\hat{\varepsilon}^{(r)}(\mathbf{x})$ of the random medium in the laboratory (absolute) reference frame x_1,x_2,x_3 are given by the expressions

$$\varepsilon_{lm}^{(r)}(\mathbf{x})=\varepsilon^{(1)}\delta_{lm} \tag{44}$$

if $\mathbf{x}$ is in the enclosing medium, and

$$\varepsilon_{lm}^{(r)}(\mathbf{x})=\alpha_{lj}^{(r)}\alpha_{mk}^{(r)}\varepsilon_{jk}^{(c)} \tag{45}$$

if $\mathbf{x}$ is within an inclusion. Here δ_{lm} is the Kronecker delta, and α_{lm} is the cosine of the angle between the lth axis of the laboratory reference frame and the mth axis of the comoving reference frame of a separate inclusion (a random quantity). Although all inclusions are thought of as being identical, they differ in position and orientation. We assume that all possible orientations of the inclusions are equiprobable and that the orientation of a separate inclusion does not depend on the inclusion's position or on the orientation and position of any other inclusion.

For this model of a discrete random medium, it is only logical to select the reference medium as being electrically neutral with a scalar permittivity ε (the dielectric constant). The same conclusion can be drawn from Eqs. (23) and (40). The latter equation suggests that the renormalization matrix $\hat{S}$ is proportional to the identity matrix:

$$\hat{S}=\frac{1}{3\varepsilon}\hat{I}. \tag{46}$$

If we allow for this fact, the random perturbation matrix (22) within the isotropic matrix-medium is also proportional to the identity matrix:

$$\hat{\xi}(\mathbf{x})=\xi^{(1)}\hat{I}, \tag{47}$$

where

$$\xi^{(1)}=3\varepsilon\,\frac{\varepsilon^{(1)}-\varepsilon}{2\varepsilon+\varepsilon^{(1)}}. \tag{48}$$

But if point $\mathbf{x}$ is within an inclusion, to calculate the components of the random perturbation matrix (22) in the laboratory reference frame it is convenient to represent these components in terms of the components of the same matrix in the reference frame comoving with the inclusion:

$$\xi_{lm}^{(r)}(\mathbf{x})=\alpha_{lj}^{(r)}\alpha_{mk}^{(r)}\xi_{jk}^{(c)}. \tag{49}$$

The elements $\xi_{jk}^{(c)}$, which form the matrix $\hat{\xi}^{(c)}$, are given by an expression that follows from (22) written in the reference frame comoving with the particular inclusion:

$$\hat{\xi}^{(c)}=3\varepsilon(\hat{\varepsilon}^{(c)}-\varepsilon\hat{I})(2\varepsilon\hat{I}+\hat{\varepsilon}^{(c)})^{-1}. \tag{50}$$

We see that the matrix $\hat{\xi}^{(c)}$ and its elements $\xi_{jk}^{(c)}$ are deterministic quantities.

Averaging $\hat{\xi}(\mathbf{x})$ over the ensemble with allowance for (47) and (49) yields

$$\langle\xi_{lm}^{(r)}(\mathbf{x})\rangle=\delta_{lm}[v_1\xi^{(1)}+v_2\xi^{(2)}], \tag{51}$$

where

$$\xi^{(2)}=\mathrm{Tr}\;\hat{\xi}^{(c)}=\xi_{pp}^{(c)}, \tag{52}$$

$v_1=1-v_2$ is the relative volume occupied by the enclosing medium, and over repeated indices there is summation. Combining this with (23), we arrive at a nonlinear equation for the dielectric constant ε of the reference medium:

$$v_1\xi^{(1)}+v_2\xi^{(2)}=0. \tag{53}$$

As for the perturbation correlation functions, for the adopted model we can write (see Ref. 46, Chap. 4) these functions in terms of the dimensionless two-point correlation function (43):

$$C_{mpqn}(\mathbf{x}-\mathbf{x}')=v_1v_2\varphi(R)D_{mpqn}, \tag{54}$$

$$D_{mpqn}=\xi^{(1)\,2}\delta_{mp}\delta_{qn}+\langle\alpha_{mj}\alpha_{pk}\alpha_{qs}\alpha_{nt}\rangle_{\mathrm{or}}\xi^{(c)}_{jkst}-\delta_{mp}\xi^{(1)} \times\langle\alpha_{qs}\alpha_{nt}\rangle_{\mathrm{or}}\xi^{(c)}_{st}-\delta_{qn}\xi^{(1)}\langle\alpha_{mj}\alpha_{pk}\rangle_{\mathrm{or}}\xi^{(c)}_{jk}. \tag{55}$$

Here the subscript ''or'' indicates averaging over the orientations of a scatterer (the averaging is done, say, by employing the proper probability densities for the Euler angles determining the orientation of the axes of the comoving reference frame). Note that here is the place where the assumption that the inclusions are homogeneous is needed. More precisely, the dependence of the correlation functions (54) only on the difference of spatial arguments, which is a characteristic feature of statistically homogeneous media[10,12] and which is needed if we want the method of substitution of field variable to work, is present only for spatially homogeneous inclusions.

Note that the above theory, on which our reasoning is based, presupposes that the spatial wave is small: $k_0L\ll 1$. In view of (54), L coincides with the correlation length of the dimensionless two-point correlation function (43), which in turn is larger than the diameter of a single inclusion or equal to that diameter. Thus, for our reasoning and the results given by formulas (58) and (61) to be valid, the spherical inclusions must be small on the scale of the wavelength.

The physics of the problem implies that, in view of the uniform distribution of the random orientations of the scatterers, the effective medium must be isotropic and that formula (45) actually has a much simpler structure corresponding to this assumption, i.e.,

$$\xi^{(e)}_{mn}=\delta_{mn}\,\delta\varepsilon, \tag{56}$$

where

$$\delta\varepsilon=\frac{1}{3}\,\xi^{(e)}_{mm}. \tag{57}$$

Calculating $\xi^{(e)}_{mm}$ by (35) and writing the expressions (15) and (16) explicitly for the case of an isotropic reference medium, we get

$$\delta\varepsilon=\frac{8\pi}{3}v_1v_2Dk_0^2\int_0^\infty\tilde{\varphi}(k)\frac{k^2\,dk}{k^2-k_0^2\varepsilon}, \tag{58}$$

$$D=\xi^{(1)\,2}+\frac{1}{3}\xi^{(c)}_{np}\xi^{(c)}_{pn}-\frac{2}{3}\xi^{(1)}\xi^{(2)}. \tag{59}$$

According to (36), the effective permittivity of the discrete medium being discussed can be written as follows:

$$\hat{\varepsilon}^{(e)}(\omega)=\varepsilon^{(e)}(\omega)\hat{I}, \tag{60}$$

$$\varepsilon^{(e)}(\omega)=\varepsilon+\delta\varepsilon. \tag{61}$$

The effective medium proves to be electrically isotropic and is characterized by the scalar permittivity $\varepsilon^{(e)}(\omega)$. For the special case of an isotropic enclosing medium and isotropic inclusions, Eq. (53) and the expression (58) become the well-known results of Ref. 45.

3.2. Mean-field damping

Let us analyze the effective permittivity (61) for a situation in which there are no dissipative losses in the random composite. In this case the permittivity of the enclosing medium, $\varepsilon^{(1)}$, is real: $\varepsilon^{(1)}=(\varepsilon^{(1)})^*$, and the permittivity tensor $\varepsilon^{(c)}$ of the anisotropic inclusion in the comoving reference frame is Hermitian (but not necessarily real): $\varepsilon^{(c)}=(\varepsilon^{(c)})^{T*}$ (here and in what follow the asterisk stands for complex conjugation). Under these conditions, as can easily be verified, $D=D^*$, and the quantity D of (59) proves to be real.

It is natural to expect, then, that in the absence of dissipation in the random medium, the quasistatic value of the effective permittivity also describes a dissipation-free medium. Indeed, in the absence of losses in the random medium, the only reason for the mean field to get weaker is the transfer of energy of the medium to the scattered component of the field due to diffraction on inclusions. However, in the quasistatic limit, the contribution of the diffraction phenomena to the effective permittivity disappears (as noted earlier), and the quasistatic value of the effective permittivity (which ε is) must describe a medium without losses. Hence in the case at hand the permittivity ε of the reference medium is a real quantity. (A indirect indication of this is the fact that, as Eqs. (48) and (52) imply, if ε obeys Eq. (53), the complex-valued quantity ε^* obeys the same equation. Here, however, we postulate a stronger statement; namely, we say that these two roots, ε and ε^*, coincide.)

Note that, strictly speaking, the final expression (58) in Sec. 3.2 refers to the dissipative random medium and a reference medium, while in the absence of dissipation the result must be obtained from (58) by passage to a limit in which the losses of the reference medium tend to zero. The integral on the right-hand side of Eq. (58) has poles of the first order at the points $k=k_0\sqrt{\varepsilon}$ and $k=-k_0\sqrt{\varepsilon}$, where the first pole ''lands'' on the positive half of the real axis from above as the losses get smaller ($\mathrm{Im}\varepsilon\to+0$; everywhere in this paper we use the branch of the square root of a complex number Z for which $0\leqslant\arg\sqrt{Z}<\pi$). Using the half-residue lemma to calculate the limiting value of the integral in (58), we have the following expressions in the absence of dissipation in the random and reference media:

$$\delta\varepsilon=\delta\varepsilon'+i\,\delta\varepsilon'', \tag{62}$$

$$\delta\varepsilon'=\frac{8\pi}{3}v_1v_2Dk_0^2\int_0^\infty\tilde{\varphi}(k)\frac{k^2\,dk}{k^2-k_0^2\varepsilon} \approx\frac{8\pi}{3}v_1v_2Dk_0^2\int_0^\infty\tilde{\varphi}(k)\,dk, \tag{63}$$

$$\delta\varepsilon''=\frac{(2\pi)^2}{3}\sqrt{\varepsilon}\,v_1v_2\tilde{\varphi}(0)k_0^3D, \tag{64}$$

where $\delta\varepsilon'$ and $\delta\varepsilon''$ are real values, and the integral in (63) is calculated as a principal-value integral at point $k=k_0\sqrt{\varepsilon}$.

Thus, the ''diffraction'' correction $\delta\varepsilon$ to the quasistatic value of the effective permittivity and hence the permittivity proper are complex-valued. The real part of the ''diffrac-

tion'' correction is given by the integral over the wave numbers and therefore is formed by scattering into the continuous wave spectrum. The imaginary part of the effective permittivity, coinciding with $\delta\varepsilon''$, is given by the half-residue of the integrand in (58) at the pole $k=k_0\sqrt{\varepsilon}$ corresponding to the propagation constant of the plane wave in the reference medium.

Let us now find the propagation constant $k^{(e)}$ of the plane wave of the mean field in the effective medium with a permittivity $\varepsilon^{(e)}(\omega)$. The quantity we are interested in is specified by the expressions

$$k^{(e)}=k_0\sqrt{\varepsilon^{(e)}(\omega)}=k_0\sqrt{\varepsilon}\,(1+\gamma^{(e)}), \tag{65}$$

$$\gamma^{(e)}\approx \delta\varepsilon/2\varepsilon\,, \tag{66}$$

$$\mathrm{Re}\ \gamma^{(e)}=\frac{4\pi}{3\varepsilon}v_1v_2Dk_0^2\int_0^\infty\tilde{\varphi}(k)\,dk, \tag{67}$$

$$\mathrm{Im}\ \gamma^{(e)}=\frac{2\pi^2}{3\sqrt{\varepsilon}}v_1v_2\tilde{\varphi}(0)k_0^3D. \tag{68}$$

Obviously, the real part of the dimensionless quantity $\gamma^{(e)}$ determines the shift in the phase velocity of the wave with respect to the phase velocity in the reference medium due to multiple scattering by random particles. The imaginary part of $\gamma^{(e)}$ describes the emergence of damping of the mean field due to scattering by the same particles. Equation (65) clearly shows that the imaginary part of the effective propagation constant depends on frequency as ω^4, which is typical of field damping as a result of Rayleigh scattering[48] and agrees with the adopted model of small inclusions.

*)E-mail: staff@univer.kharkov.ua

[1] T. Oguchi, Proc. IEEE **71**, 1029 (1983).
[2] Yu. M. Mel'nik (ed.), *Radar Methods in Earth Studies*, Sovetskoe Radio, Moscow (1980).
[3] A. K. Fung, *Microwave Remote Sensing*, Artech House, London (1994).
[4] T. L. Krohn and L. M. Medgyesi-Mitschang, IEEE Trans. Antennas Propag. **AP-37**, 219 (1989).
[5] *The Mechanics of Composites and Construction Elements*, Vol. 1: *Mechanics of Composites*, Naukova Dumka, Kiev (1982).
[6] I. M. Lifshits, M. I. Kaganov, and V. M. Tsukernik, *Annals of the Kharkov State University* (Proceedings of the Physics Section of the Physics–Mathematics Department), Vol. 2 (1950), p. 41.
[7] I. M. Kaganova and M. I. Kaganov, Waves Random Media **3**, 117 (1993).
[8] *Dielectric Properties of Heterogeneous Materials*, A. Priou (ed.), Elsevier, New York (1992).
[9] R. G. Barrera, J. Giraldo, and W. Luis Mochán, Phys. Rev. B **47**, 8528 (1993).
[10] V. I. Tatarskiĭ, *Wave Propagation in a Turbulent Atmosphere*, Dover, New York (1968).
[11] Yu. A. Ryzhov, V. V. Tamoĭkin, and V. I. Tatarskiĭ, Zh. Éksp. Teor. Fiz. **48**, 656 (1965) [Sov. Phys. JETP **21**, 433 (1965)].
[12] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskiĭ, *Principles of Statistical Radiophysics*, Vol. 3: *Elements of Random Fields*, Springer-Verlag, Berlin (1989).
[13] V. I. Tatarskiĭ and M. E. Gertsenshteĭn, Zh. Éksp. Teor. Fiz. **44**, 676 (1963) [Sov. Phys. JETP **17**, 458 (1963)].
[14] R. C. Bourret, Nuovo Cimento **26**, 1 (1962).
[15] J. B. Keller and F. C. Karal, J. Math. Phys. **7**, 661 (1966).
[16] É. A. Kaner, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **2**, 827 (1959).
[17] F. G. Bass, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **2**, 1015 (1959).
[18] D. Dence and J. E. Spence, IEEE Trans. Antennas Propag. **AP-19**, 302 (1971).
[19] E. Dence and J. E. Spence, IEEE Trans. Antennas Propag. **AP-20**, 110 (1972).
[20] E. N. Ermakova and V. V. Tamoĭkin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 1417 (1975).
[21] Yu. A. Ryzhov and V. V. Tamoĭkin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **7**, 605 (1964).
[22] Yu. A. Ryzhov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **21**, 316 (1978).
[23] F. V. Bunkin, Zh. Éksp. Teor. Fiz. **32**, 338 (1957) [Sov. Phys. JETP **5**, 277 (1957)].
[24] A. D. Yaghjian, Proc. IEEE **68**, 248 (1980).
[25] Yu. A. Ryzhov and V. V. Tamoĭkin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **13**, 356 (1970).
[26] J. A. Kong, *Electromagnetic Wave Theory*, Wiley-Interscience, New York (1990).
[27] I. M. Lifshits and L. N. Rozentsveĭg, Zh. Éksp. Teor. Fiz. **16**, 969 (1946).
[28] I. M. Lifshits and L. N. Rozentsveĭg, Zh. Éksp. Teor. Fiz. **21**, 1184 (1951).
[29] V. M. Finkel'berg, Zh. Tekh. Fiz. **34**, 509 (1964) [Sov. Phys. Tech. Phys. **9**, 396 (1964)].
[30] V. M. Finkel'berg, Zh. Éksp. Teor. Fiz. **46**, 725 (1964) [Sov. Phys. JETP **19**, 494 (1964)].
[31] Yu. A. Ryzhov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **9**, 39 (1966).
[32] Yu. A. Ryzhov and V. V. Tamoĭkin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **9**, 205 (1966).
[33] V. M. Finkel'berg, Zh. Éksp. Teor. Fiz. **53**, 401 (1967) [Sov. Phys. JETP **26**, 268 (1968)].
[34] B. S. Abramovich, Zh. Éksp. Teor. Fiz. **77**, 87 (1979) [Sov. Phys. JETP **50**, 44 (1979)].
[35] L. Tsang and J. A. Kong, Radio Sci. **16**, 303 (1981).
[36] A. Stogryn, IEEE Trans. Antennas Propag. **AP-31**, 985 (1983).
[37] Z. D. Genchev, Waves Random Media **2**, 99 (1992).
[38] S. V. Nghiem, R. Kwok, J. A. Kong, and R. T. Shin, Radio Sci. **28**, 687 (1993).
[39] N. P. Zhuck, Phys. Rev. B **21**, 15 636 (1994).
[40] N. P. Zhuck and A. S. Omar, in *Proc. '95 Int. Symp. on Electromagnetic Theory*, St. Petersburg, Russia (1995), p. 631.
[41] B. Michel and A. Lakhtakia, Phys. Rev. E **51**, 5701 (1995).
[42] N. P. Zhuck and A. S. Omar, IEEE Trans. Antennas Propag. **AP-44**, 1142 (1996).
[43] B. Michel and A. Lakhtakia, J. Phys. D **29**, 1431 (1996).
[44] A. Stogryn, Radio Sci. **18**, 1283 (1983).
[45] H. H. Lin, M. E. Veysoglu, S. H. Yueh *et al.*, J. Electromagn. Waves Appl. **8**, 801 (1994).
[46] T. D. Shermergor, *Theory of Elasticity of Microheterogeneous Media*, Nauka, Moscow (1977).
[47] F. I. Fedorov, *Optics of Anisotropic Media*, Academy of Sciences of the Belorussian SSR, Minsk (1958).
[48] K. S. Shifrin, *Light Scattering in a Turbid Medium*, Gostekhmeteoizdat, Moscow–Leningrad (1951).

Translated by Eugene Yankovsky

Formation of two-dimensional current sheets under high initial pressure conditions

S. Yu. Bogdanov,*) V. B. Burilina, and A. G. Frank

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted 29 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 1202–1214 (October 1998)

The possibilities of current-sheet formation in two-dimensional magnetic fields with a null line as well as the characteristic features of the plasma dynamics under high initial pressure conditions (helium, $P_0 \approx 300$ mtorr) are investigated for the first time. It is shown that current-sheet formation and efficient compression of the plasma into a sheet require that the magnetic field gradient be sufficiently large. A brightly emitting compact region with electron density $N_e \sim 9 \times 10^{16}$ cm^{-3}, an order of magnitude higher than the gas atom density, was observed to form at the center of the layer. © *1998 American Institute of Physics.* [S1063-7761(98)00510-1]

1. INTRODUCTION

Magnetic reconnection is a fundamental physical process which is the basis for various phenomena in plasma physics and astrophysics.[1–6] Flares on the sun and other stars, substorms in the magnetospheres of the Earth and planets, disruptive instability in toroidal magnetic confinement systems (tokamaks) that result in a loss of hot plasma, as well as many time-dependent phenomena in plasma focus and *Z*- and Θ-pinch high-current discharges are associated with magnetic reconnection. In all these phenomena, as a rule, the stored magnetic field energy is converted rapidly and efficiently into thermal and kinetic plasma energy and into accelerated particle fluxes and radiation of various wavelengths.

Magnetic-field line reconnection can occur in regions of space where oppositely (or differently) directed field lines come close to one another. These regions are characterized by high electric current density and small scales, so that dissipative processes are very influential here even under the conditions of high plasma conductivity. Here one of the main properties of a highly conducting plasma can break down—the freezing-in of the magnetic field in matter. Magnetic-field line reconnection can change the topology of the magnetic field. For this reason, even if the reconnection process itself occurs in relatively small regions of space, the changes occurring in the topology of the magnetic field as a result of reconnection can radically alter the behavior and dynamics of the entire system as a whole.

Regions with high electric current density, which separate oppositely (or differently) directed magnetic fields and which accumulate excess magnetic energy, ordinarily take the form of quasi-one-dimensional current sheets.[1,6–8] In other words, the distribution of the electric current density in any cross section perpendicular to the direction of the current is characterized by two considerably different scales. The largest scale (the width of the sheet) determines the amount of stored magnetic energy, while the smallest scale (the thicknes of the sheet) determines the characteristic dissipation time.[1] Magnetic reconnection processes, in which a substantial portion of the excess magnetic energy is rapidly transformed into plasma energy and accelerated-particle energy, can occur precisely within current sheets. In this connection, the study of the conditions and possibilities of the formation of spatially localized current sheets in highly conducting magnetized plasma is of fundamental importance.

From the standpoint of the localization of current sheets the so-called singular magnetic field lines,[1,6,9] along which a nonzero electric field can exist in the plasma coordinate system, merit serious attention. It is natural to expect that electric current in plasma can concentrate near singular lines as a result of the formation of current sheets.

The best known example of a singular line is the null line of a two-dimensional (2D) magnetic field (Fig. 1). The spatial dependence of the field can be represented near this line ($x=0$, $y=0$) in the form

$$\mathbf{B} = -h\{y;x;0\}, \quad |\mathbf{B}| = h|\mathbf{r}|, \tag{1}$$

where h is the radial gradient of the magnetic field.

The dynamics of plasma in magnetic fields containing null lines has been actively investigated for many years, both theoretically[1–9] and experimentally.[10–12] As a result, it has been established that a flat current sheet, which accumulates the magnetic energy released in flare-type processes accompanying explosive breakdown of the sheet,[1,10,12] can indeed form near a null line.

At the same time, efficient compression of plasma into a sheet,[13] where the electron density is 10 or more times higher than in the initial plasma,[14] has been observed experimentally. Typically, the gas-kinetic pressure of plasma concentrated in a sheet is ordinarily balanced by the magnetic pressure outside the sheet, i.e., the parameter

$$\beta = 8\pi N_e k(T_e + T_i/\overline{Z_i})/B^2 \tag{2}$$

reaches its maximum value $\overline{\beta} \simeq 1$ (here N_e is the electron density, T_e and T_i are the electron and ion temperatures, $\overline{Z_i}$ is the effective ion charge). The formation of quasi-one-dimensional sheets containing a dense current-carrying

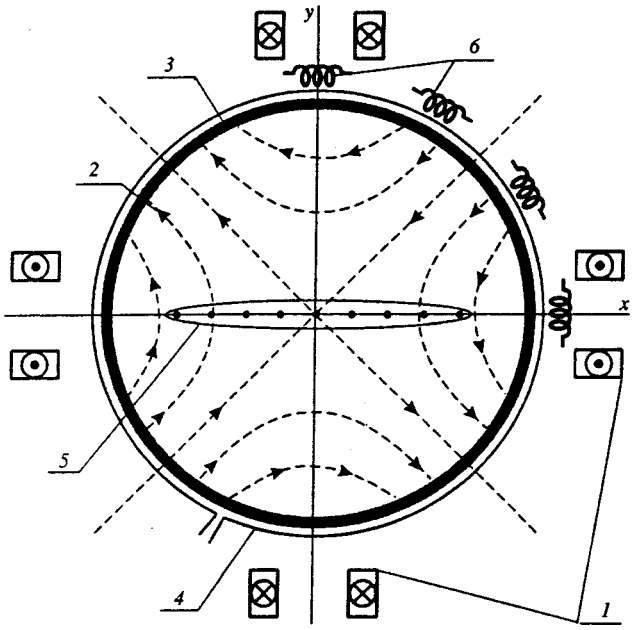

FIG. 1. Diagram (section) of the experimental setup for studying the formation and evolution of current sheets in two-dimensional magnetic fields with a null line: *1* — Conductors with currents (straight quadrupole), *2* — field lines of the two-dimensional vacuum magnetic field with a null line, *3* — vacuum chamber, *4* — θ-discharge loop, *5* — current sheet, *6* — magnetic probes.

plasma is also of great interest for various applications: simulation of astrophysical phenomena, such as solar flares (see Refs. 1 and 10), under laboratory conditions; development and construction of fundamentally new systems for controlled thermonuclear fusion;[15] development of new sources of visible- and ultraviolet-range radiation; generation of directed supersonic plasma flows; and so on.

Theoretical treatments devoted to the problem of magnetic reconnection have usually employed the strong magnetic field approximation, in which the gas-kinetic pressure of the initial plasma is assumed to be negligibly low.[16,7,1] Similar conditions were also chosen for most experimental investigations, so that the plasma dynamics at the sheet formation stage was determined predominantly by $\mathbf{j}\times\mathbf{B}$ forces.[10–14] It was observed that current sheets in which dense plasma is concentrated are formed in both linear[17] and nonlinear[18] regimes, which differed primarily by the ratio of the initial magnetic field and the field of the electric current in the plasma (see also Ref. 12). In all cases the maximum electron density in a sheet $N_e^{\max}=(0.8-2)\times10^{16}\,\mathrm{cm}^{-3}$ was much higher than the density of both the initial plasma and the plasma surrounding the sheet.

In the present paper we report the results of experiments in which the possibilities of producing current sheets in two-dimensional magnetic fields with a null line (1) under conditions of comparatively high initial gas pressure (helium, $P_0=300$ mtorr) were investigated for the first time.

It could be anticipated that the increase by a factor of 5–20 in the initial pressure compared with previous work[10,12,17,18] and the corresponding higher initial plasma density can have several effects. In the first place, hydrodynamic processes play a relatively larger role in relation to magnetohydrodynamic processes. Indeed, the size of the region near the null line of the magnetic field, where the sound velocity v_s is greater than the Alfvén velocity v_A, is

$$r_s\simeq\sqrt{4\pi\gamma NkT}/h \tag{3}$$

and increases with the initial plasma density. In other words, under these conditions the strong magnetic field approximation, which holds for $r\gg r_s$, can be satisfied only at an appreciable distance from the null line.

The second obvious effect is that the sheet formation process slows down. This is due to an increase in the characteristic Alfvén time

$$t_A=\sqrt{4\pi N_i M_i}/h. \tag{4}$$

Finally, the relative role of dissipative processes can increase as a result of the increase in t_A (4). This can be described quantitatively, for example, as a decrease in the magnetic Reynolds number

$$\mathrm{Re}_m=t_\sigma/t_A, \tag{5}$$

where

$$t_\sigma\simeq4\pi\sigma L^2/c^2 \tag{6}$$

is the ohmic dissipation time in a volume of characteristic size L.

These considerations all indicate that the question of the formation of a sheet of dense plasma under conditions of high initial pressure is very nontrivial and apparently requires a special choice of initial and boundary conditions.[19]

At the same time, if the plasma compression efficiency is assumed to remain the same as before,[10,12,14,17,18] then electron density in the sheet should increase considerably, reaching values $N_e\sim10^{17}\ \mathrm{cm}^{-3}$. Such a plasma would have several advantages from the standpoint of both plasma diagnostics and various applications, especially since the position of the sheet is fixed and does not change with time, which distinguishes such a sheet from, for example, rapidly moving plasma shells in Z- and Θ-pinch systems.

If it is assumed that $\overline{\beta}\simeq1$ holds for the sheet as before, then the expected increase in plasma density should decrease T_i and T_e—the electron and ion temperatures (for close values of the magnetic field gradient h and plasma current I_p)—and equalize them. The longitudinal electron thermal conductivity, which was the main channel for energy losses of a sheet in previous experiments,[20] should play a lesser role under conditions such that N_e increases and T_e decreases, while other loss channels, for example, losses due to radiation, can become dominant.

2. EXPERIMENTAL SETUP AND DIAGNOSTICS METHODS

The possibilities of current-sheet formation under high initial helium pressure conditions were studied in the standard experimental arrangement for such investigations (see Refs. 10 and 11). The quasistationary two-dimensional quadrupole magnetic field (1) possessed a gradient h equal to 280 or 570 G/cm and the null line lay on the axis of a cylindrical quartz vacuum chamber 18 cm in diameter (Fig. 1). The initial plasma was produced using a Θ discharge with strong

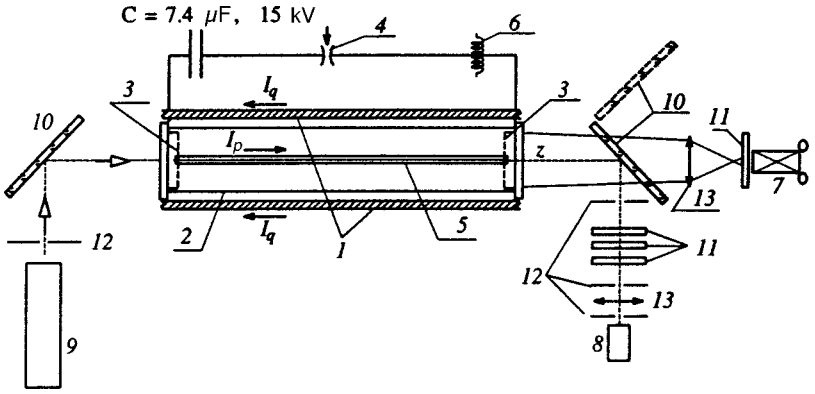

FIG. 2. Diagram of experimental setup (side view): *1* — Conductors of the quadrupole magnetic field, *2* — vacuum chamber, *3* — grid electrodes, *4* — controllable discharger, *5* — current sheet, *6* — Rogowski loop, *7* — image tube, *8* — photomultiplier (FEU), *9* — He–Ne laser, *10* — mirrors, *11* — light filters, *12* — diaphragms, *13* — achromatic objective lenses.

preionization; helium at pressure $P_0=300$ mtorr served as the working gas; the initial plasma density was close to the atomic density of the neutral gas $N_e^0 \leqslant N_a^0 \simeq 10^{16}$ cm^{-3}. Electric current was excited in the plasma by a pulsed voltage applied between two grid electrodes which were introduced into the chamber at both ends (Fig. 2) and were separated by a distance of 60 cm. The half-period of the plasma current was $T/2=5$ μs, and the maximum current was 60 kA (see Fig. 4a).

Magnetic probes positioned on the outside of the vacuum chamber (Fig. 1) were used to determine the configuration of the electric current in the plasma as well as to determine the thickness of the current sheet and its variations in time (see Fig. 4b and Ref. 21).

Two-dimensional images of the plasma in different spectral emission lines at successive times were obtained with an image tube (gain $\simeq$400, exposure time $\simeq$80 ns) in combination with narrow-band spectral interference filters with $\Delta\lambda_{1/2} \simeq 1-1.2$ nm (Fig. 3; see also Ref. 22).

The temporal variations of the intensities of the spectral lines He I 587.6 nm and He II 468.6 nm and the continuum emitted from a region 0.4 cm in diameter near the axis of the vacuum chamber were obtained using appropriate interference filters and a FÉU-79 photomultiplier (Fig. 4c). The spectral composition of the characteristic plasma radiation which passed through the filters was additionally analyzed with a MDR-3 monochromator (see Ref. 22 for a more detailed discussion).

In the present work the refraction of $\lambda=632.8$ nm laser radiation by a flat sheet of dense plasma produced in a two-dimensional magnetic field with a null line was detected for the first time. A He–Ne laser beam was directed along the null line, i.e., along the length of the plasma sheet in the direction of the *z* axis (Fig. 2). A system of diaphragms made it possible to record only the laser radiation that was deflected by small angles relative to the initial direction. In a number of cases the intensity of the laser radiation was substantially diminished after passing through the plasma sheet (Fig. 4d). This made it possible to determine the electron density gradient in the *xy* plane and to estimate the maximum value of the density.

3. PRINCIPAL RESULTS

The possibility of a flat plasma sheet forming under high initial helium pressure conditions was clearly demonstrated using plasma images obtained in different spectral lines (see Fig. 3 as well as Ref. 22). Immediately after the onset of excitation of an electric current in the plasma ($t \approx 0.2-0.4$ μs), the distribution of the radiation intensity in the *xy* plane, perpendicular to the current direction, acquired the shape of a sheet with a local intensity minimum at the center of the sheet. At subsequent times ($t \geqslant 1$ μs) the width of the sheet increased, reaching values $2\Delta x \approx 12$ cm, when the sheet thickness $2\Delta y < 1$ cm, while the radiation intensity was virtually uniform over the width, for example, in the He I line (Fig. 3a). Hence it follows that in a two-dimensional magnetic field with gradient $h=570$ G/cm a plasma sheet was formed in 1–1.5 μs, i.e., essentially during the characteristic Alfvén time t_A (4). For smaller gradients the formation time of the plasma sheet was much longer, so that during the entire first half-period of the current the plasma emission varied along the *x* axis and possessed a local minimum at the center of the sheet.

We note also that the sheet thickness $2\Delta y$ was 2–3 times smaller than $2r_s$ (see Eq. (3)), i.e., the diameter of the region near the null line where the sound velocity exceeded the Alfvén velocity.

In contrast to plasma images obtained in the He I line, the images in the He II 468.6 nm line (Fig. 3b) showed a substantial variation along the surface of the sheet and the formation of a bright compact object at the center of the sheet—the core ($t=0.6, 2.5$ μs). We emphasize that most images presented in Fig. 3b were obtained with much lower exposures than the images in Figs. 3a and 3c; this was manifested specifically in the fact that the emission of the near-wall continuum, as a rule, was absent in Fig. 3b, in contrast to Figs. 3a and 3c. The images in the He II line also acquired the shape of a flat sheet as exposure increased, while the central region was overexposed.

A flat plasma sheet was observed clearly at $t=1.2$, 1.5, and 2.1 μs in the spectral lines of N II (Fig. 3c). The appearance of a dark band in the central plane of the sheet ($t=2.1, 3.0$ μs) attested to a higher electron temperature in this region and ''burnout'' of the N II lines, similarly to the observations in Refs. 23 and 24.

Comparing the images presented in Figs. 3a, 3b, and 3c and obtained at closely spaced intervals in time, it can be concluded that the plasma radiation in different spectral lines originated in different spatial regions separated from one another. Thus, the radiation in the He II line corresponded to

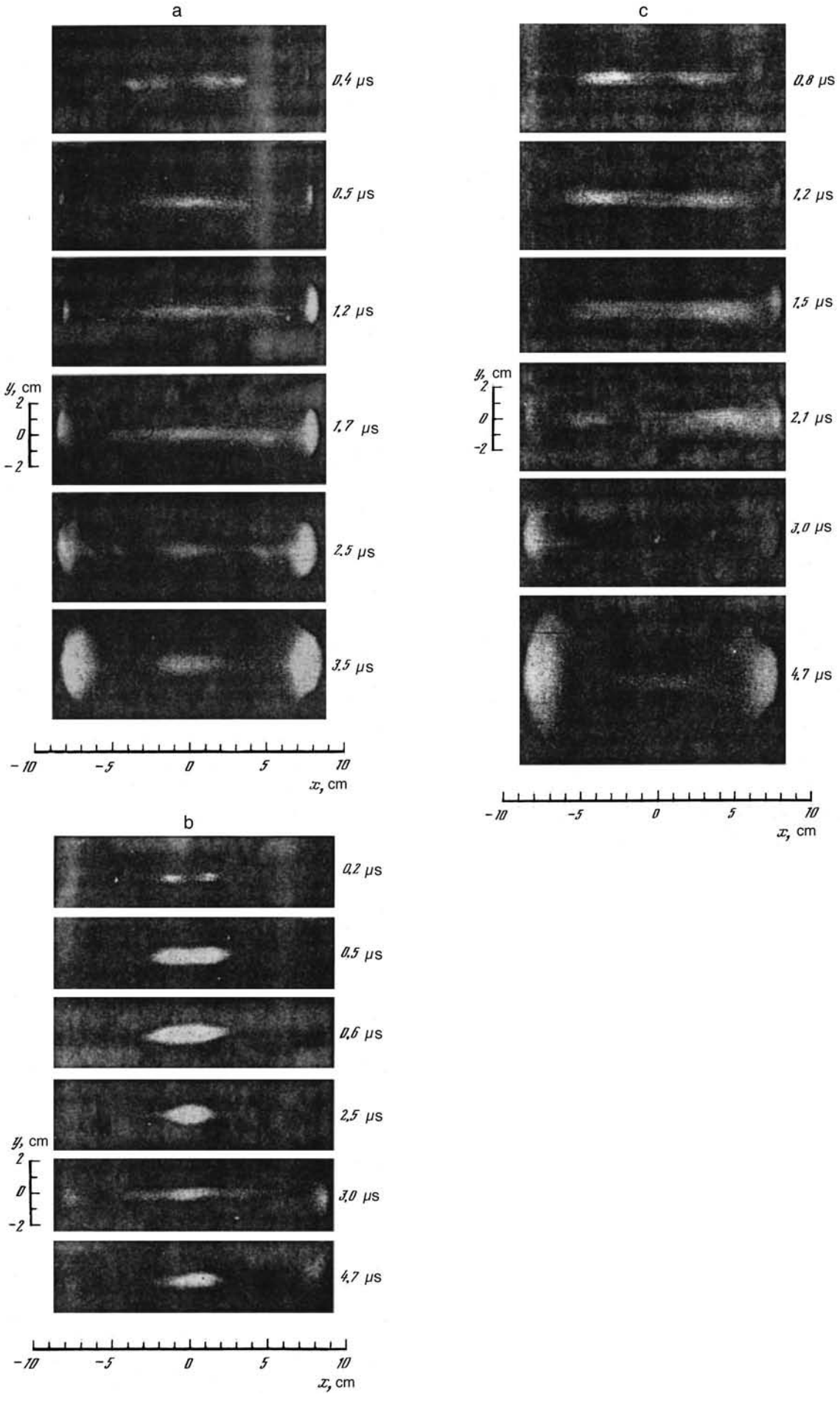

FIG. 3. Plasma images obtained at successive times in different spectral lines: a — He I (587.6 nm), b — He II (468.6 nm), c — N II (567.6–568.6 nm).

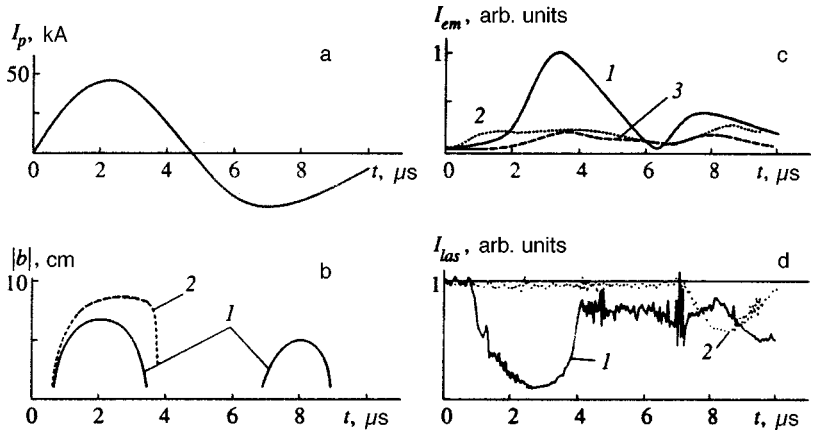

FIG. 4. Time dependence of: a — total electric current I_p in the plasma; b — half-width $|b|$ of the current sheet calculated from the results of magnetic measurements for different values of the magnetic field gradient $h=570$ (*1*) and 280 (*2*) G/cm; c — intensities of plasma radiation in the spectral lines for He II (468.6 nm) —*1* and He I (492.2 nm) — *2*; continuum intensity, $\lambda \simeq 480.6$ nm, $\Delta\lambda_{1/2}=1.5$ nm — *3*; d — intensity of He–Ne laser ($\lambda=632.8$ nm) after passage of the laser beam along the plasma sheet for different values of the magnetic field gradient: $h=570$ (*1*), 280 (*2*) G/cm.

the central plane of the sheet ($y\simeq 0$), while the He I and N II lines were emitted predominantly from regions displaced in the y direction on both sides of the central plane.

As follows from the magnetic measurements, the electric current distribution in the xy plane, perpendicular to the current direction, also acquired the shape of a sheet whose width was much greater than its thickness. Figure 4b shows the time dependence of the half-width $|b|$ of the current sheet for two gradients of the initial magnetic field, $h=570$ and 280 G/cm. One can see that $|b|$ increased as h decreased, since at $h=280$ G/cm the lateral ends of the sheet came right up against the walls of the vacuum chamber, i.e., $|b|\simeq R_c$. For $h=570$ G/cm the width of the current sheet $2|b|\leqslant 12$ cm $<2R_c$. In this case the sheet typically was formed during both the first and second half-periods of the current in the plasma, and the orientation of the sheet in the xy plane changed by 90° in accordance with the change in the direction of the current. The half-width $|b|$ of the current sheet, obtained on the basis of magnetic measurements corresponded, to within the accuracy of the measurements, to the half-width of a neutral current sheet[7]

$$b=\sqrt{4I_p/ch}\,. \tag{7}$$

This fact makes it possible to employ other relations which are valid for neutral current sheets.[7,10,12] Specifically, it is possible to estimate the magnetic field at the surface of the sheet:

$$B_x(x\simeq 0)\simeq 3.4 \text{ kG}.$$

Since, as noted above, $\overline{\beta}\simeq 1$ typically holds for current sheets (see Eq. (2)), the energy content of the plasma can be estimated hence as

$$N_e(T_e+T_i/\overline{Z_i})\simeq 2.8\times 10^{17} \text{ eV/cm}^3.$$

Thus, it can be concluded from magnetic measurements and by analyzing plasma images obtained in different spectral lines that the formation of a current sheet and efficient compression of plasma into a sheet can also occur under high initial gas pressure conditions, provided only that the gradient h of the initial two-dimensional magnetic field with a null line is large enough. Apparently, it is important to satisfy the condition $t_A<T/4$, where T is the period of the current in the plasma. Under otherwise identical conditions, increasing the gradient h substantially increases the intensity of the radiation from the plasma concentrated in the sheet. This indicates more efficient plasma compression.

Of great interest is a characteristic feature of the plasma sheet observed in the form of a bright compact core with the dimensions $2\delta x\simeq 3.3$ cm and $2\delta y\simeq 1.1$ cm in the central region (see Fig. 3b, $t=2.5$ μs). As noted above in the discussion of plasma images in different spectral lines, the intensity of the core radiation in the He II line was much higher than the intensity of radiation from neighboring sections of the sheet. The time dependence of the intensities of the spectral lines of He I and He II and the continuum intensity, which came from the region near the axis, is presented in Fig. 4c. The signal strengths reflected the intensity ratios qualitatively; the radiation in the He II 468.6 nm line was 5 to 6 times more intense than the radiation in the He I 587.6 nm line and in the continuum. The large increase in the He II line intensity in the interval from 2 to 3.4 μs is interesting; the intensity of the continuum increased at the same time, while the radiation in the He I line remained unchanged in the interval $\sim 1.0-5.5$ μs. It could be inferred that the observed increase in the intensity of the He II spectral line was due predominantly to an increase in the electron density N_e in the central region of the current sheet within the bright core. To check this hypothesis an experiment probing the plasma sheet with 632.8 nm laser radiation was performed.

Direct measurements of the spatial distributions of the electron density, performed previously by holographic interferometry,[14,17,18] showed that for plasma sheets which developed in two-dimensional magnetic fields at comparatively low initial pressure, $P\leqslant 50$ mtorr, sharp density gradients in a direction perpendicular to the surface of the sheet are typical and reach values

$$\frac{\partial N_e}{\partial y}\simeq 5\times 10^{16} \text{ cm}^{-4}.$$

The plasma radiation distributions in the xy plane, which are presented in Fig. 3, indicated the existence of quite sharp gradients of the luminosity of the sheet and also, apparently, of the electron density under high initial pressure conditions. It was natural to assume the plasma sheet to be uniform in the direction of the z axis, as was observed in previous experiments.[10,12]

If a narrow light beam is directed along the z axis, then as a result of passing through the plasma sheet the beam can be deflected from the initial direction. In visible radiation the plasma is a transparent phase nonuniformity, so that the deflection angle $\delta\psi$ is determined by the gradient of the refractive index in a direction perpendicular to the propagation direction of the beam:[25,26]

$$\delta\psi=\frac{1}{n}\int_{z_1}^{z_2}\nabla_\perp n dz. \tag{8}$$

Here n is the refractive index, and z_1 and z_2 are the boundaries of the plasma along the propagation path of the beam. The refractive index of plasma for visible-range radiation which depends mainly on the electron density N_e (with degree of plasma ionization exceeding $\sim$5%)

$$n=\sqrt{1-\frac{\omega_{pe}^2}{\omega^2}}\simeq 1-\frac{1}{2}\frac{\omega_{pe}^2}{\omega^2}=1-1.8\times 10^{-22}N_e\simeq 1, \tag{9}$$

where ω_{pe} is the plasma frequency and ω is the angular frequency of the probe radiation. The red line of the He–Ne laser satisfies $\omega\simeq 3\times 10^{15}$ s^{-1} and the refractive index is essentially 1 for any reasonable values of N_e. Correspondingly, the deflection angle of the laser beam from the initial direction is very small, so that the relation (8) can be written as

$$\delta\psi=L\nabla_\perp n=-1.08\times 10^{-20}\nabla_\perp N_e. \tag{10}$$

Here $L=z_2-z_1=60$ cm is the length of the interelectrode gap where the plasma sheet was formed.

In practice the experiment ismounted as follows: The He–Ne laser beam was directed strictly along the null line of the magnetic field, and after passing through the plasma-containing region the beam was passed through a special diaphragm, which cut off radiation with deflection angles relative to the system axis greater than $\varphi\simeq 1.8\times 10^{-3}$ rad. Next, the laser radiation was directed through a system of diaphragms and filters and was detected with a FÉU-79 photomultiplier (Fig. 2). Let the signal level in the absence of plasma be 1. In some cases the laser radiation was observed to be considerably attenuated as a result of the formation of the plasma sheet, as one can see from Fig. 4d. If the initial gradient was equal to 280 G/cm, then an approximately 30% intensity decrease was seen only $\approx 7-8$ μs after the onset of the current in the plasma, during the second half-period. However, for a large magnetic-field gradient, for example, $h=570$ G/cm, the attentuation of the laser radiation became much stronger, reaching $\approx$90%, and was even observed during the first half-period of the current, at $t\simeq 1.5-4$ μs. We note that the maxima of the plasma radiation in the He II line and in the continuum (Fig. 4c) were correlated with the maximum attenuation of the laser radiation which has passed through the plasma sheet (Fig. 4d).

A very important point is that the decrease in the intensity of the laser radiation was observed only with the use of a special diaphragm which did not pass radiation with angles of deflection relative to the system axis greater than the angle φ. In the absence of such a diaphragm, the intensity of the laser radiation which passed through the plasma sheet did not change. Hence it can be concluded that the attenuation of the signal shown in Fig. 4d was due predominantly to the deflections of the light beam away from the initial direction by angles $\delta\psi\geqslant 1.8\times 10^{-3}$ rad, i.e., refraction of the laser radiation by the plasma sheet. According to Eq. (10), such angles corresponded to an electron density gradient $\partial N_e/\partial y\geqslant 1.7\times 10^{17}$ cm^{-4}. This means that when a plasma sheet formed under high initial pressure conditions the electron density gradient was at least three times greater than the maximum gradient obtained previously.[14,17]

As noted above, the greatest attenuation of the laser radiation was observed in the regimes and at times when a region of maximum brightness of the He II line radiation with transverse size $2\Delta y\simeq 1.1$ cm was clearly observed at the center of the plasma sheet. Assuming that

$$\frac{\partial N_e}{\partial y}\approx\frac{N_e^{\max}}{\Delta y},$$

the maximum electron density can be estimated as $N_e^{\max}\approx 0.9\times 10^{17}$ cm^{-3}. This value was 5 to 10 times greater than the maximum densities N_e recorded for lower initial pressure,[12,14,17,18] and it was almost an order of magnitude greater than the initial density of the neutral gas $N_a^0\simeq 10^{16}$ cm^{-3}. We emphasize that a tenfold increase of the electron density compared with the initial value (as noted above, $N_e^0\leqslant N_a^0$) was observed in the central region of the sheet within the bright core. In other sections of the sheet the density was apparently lower. Note that the value obtained for $N_e^{\max}$ is only approximate. However, this estimate was substantiated using specially performed spectral measurements (see Ref. 27).

Therefore it can be concluded from the character of the temporal variations of the intensity of the plasma radiation in the spectral line He II 468.6 nm (Fig. 4c) and the attenuation of the He–Ne-laser radiation which has passed through the plasma sheet (Fig. 4d) that the electron density in the central region of the sheet gradually increased and reached $\approx 0.9\times 10^{17}$ cm^{-3}.

The plasma temperature was estimated first according to the average conductivity and second on the basis of the pressure balance, i.e., from the relation $\overline{\beta}\simeq 1$. When averaged over the spatial region where the electric current was concentrated and over a time interval equal to a half-period of the current, The plasma conductivity was $\overline{\sigma}\leqslant 10^{14}$ s^{-1}, which for Spitzer conductivity corresponded to $T_e\simeq 4$ eV. Pressure balance gives

$$T_e+T_i/\overline{Z}_i\geqslant 3 \text{ eV},$$

i.e., it can be assumed that the electron temperature in any case did not exceed 5 eV, and the effective ionization was $Z_i\simeq 1$, as indicated, specifically, from the fact that the emission in the He II line essentially did not burn out (see Figs. 3b and 4c). For $N_e>5\times 10^{16}$ cm^{-3} $T_e\leqslant 5$ eV the frequency of Coulomb collisions $\nu_e\geqslant 2\times 10^{11}$ s^{-1}, and the electron and ion mean free path lengths $l_e\simeq l_i\simeq 10^{-3}$ cm, i.e., much less than the half-thickness Δy of the plasma sheet (see Fig. 3). The equalization time of the electron and ion temperatures

under these conditions is very short, $\tau_E \leqslant 2\times 10^{-8}$ s, i.e., the temperatures should be essentially the same, $T_i \simeq T_e$.

The characteristic time due to the thermal conductivity of the plasma in the absence of (or along) the magnetic field can be estimated as[28]

$$t_{\kappa_e''} \sim L^2 N_e / \kappa_e'', \tag{11}$$

where κ_e'' is the thermal conductivity

$$\kappa_e'' \simeq 3.16 \frac{N_e k T_e}{m \nu_e}. \tag{12}$$

Coulomb collisions sayisfy $\nu_e \sim N_e T_e^{-3/2}$, so that when N_e increases and at the same time T_e decreases, the quantity $t_{\kappa_e''}$ increases considerably:

$$t_{\kappa_e''} \sim L^2 N_e T_e^{-5/2}. \tag{13}$$

Assume $L \simeq 1$ cm, which corresponds to the thickness (less than the transverse size) of the plasma sheet as well as the characteristic size of the core at the center of the sheet. Then, for the above plasma parameters we have $t_{\kappa_e''} \sim 10$ μs, which is greater than, for example, the half-period of the plasma current, i.e., the plasma temperature could have been substantially different in different sections of the sheet separated in space by distances greater than ~ 1 cm. It is natural to infer that this effect is responsible for the formation of a bright core in the central region of the sheet. It is also obvious from the estimates presented that energy losses due to heat conduction could not balance the energy released in the current sheet by ohmic dissipation. In all probability, the main channel for energy losses of the sheet formed under high initial pressure conditions were radiation losses.

4. CONCLUSIONS

In the present work we investigated the possibility in principal of forming flat current sheets in two-dimensional magnetic fields with a null line and studied them under high initial gas pressure conditions (helium, $P_0 = 300$ mtorr). The characteristic properties of the high-pressure regime are that, first, the plasma is not magnetized in a comparatively wide spatial region; second, the gas-kinetic pressure plays a larger role than electrodynamic forces, which in the previous experiments completely determined the plasma dynamics; and, finally, increasing the initial pressure slows down the sheet formation process.

Magnetic measurements and analysis of the plasma images obtained in different spectral lines show that the formation of a current sheet and efficient compression of plasma into a sheet can also occur under high initial pressure conditions, provided that the gradient of the initial two-dimensional magnetic field with a null line is large enough. It was established that the sheet formation time is of the order of the characteristic Alfvén time. It was shown that increasing the magnetic-field gradient increases considerably the intensity of the radiation from the plasma concentrated in the sheet, i.e., it leads to more efficient plasma compression. These results support the idea that the character of current-sheet formation in plasma in nonuniform magnetic fields is universal (see also Ref. 29).

The refraction of laser radiation with wavelength $\lambda = 632.8$ nm as a result of passage along a flat sheet of dense plasma was detected for the first time. This made it possible to determine the gradient of the electron density in the sheet as $\partial N_e / \partial y \geqslant 1.7\times 10^{17}$ cm^{-4} and to estimate the maximum density as $N_e^{\max} \approx 9\times 10^{16}$ cm^{-3}. The maximum electron density at the center of the plasma sheet exceeded the initial electron density and also the gas density by approximately an order of magnitude, attesting to the effectiveness of plasma compression under these conditions.

Of great interest is a characteristic feature of the plasma sheet observed in the form of a bright compact core, whose He II line radiation was much more intense than the radiation from neighboring sections of the sheet.

The increase in the electron density was accompanied by a decrease in the plasma temperature, which resulted in a large decrease in the electron heat conduction along the surface of the sheet. The formation of a bright compact core at the center of the sheet is apparently linked to the decrease in heat conduction, while losses to radiation can be assumed to be the main channel for energy losses of the current sheet formed under high initial pressure conditions.

We are deeply grateful to G. M. Batanov, N. P. Kiriĭ, and H.-J. Kunze for fruitful discussions and V. S. Markov for assisting in the experiment.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-18546a) and INTAS (Grant No. INTAS-96-456).

*)E-mail: bogdanov@fpl.gpi.ru

[1] S. I. Syrovatskii, Annu. Rev. Astron. Astrophys. **19**, 163 (1981).
[2] E. Priest, *Solar Magnetohydrodynamics*, Reidel, Dordrecht, 1982 [Russian translation, Mir, Moscow, 1985].
[3] A. A. Galeev in *Basic Plasma Physics* North–Holland, Amsterdam, 1984; Énergoatomizdat, Moscow, 1984, Vol. 2, p. 331.
[4] B. B. Kadomtsev, Usp. Fiz. Nauk **151**, 3 (1987); Rep. Prog. Phys. **50**, 115 (1987).
[5] K. Schindler, M. Hesse, and J. Birn, J. Geophys. Res. **93**, 5547 (1988).
[6] D. Biskamp, *Nonlinear Magnetohydrodynamics*, Cambridge Univ. Press, (1993).
[7] S. I. Sirovatskiĭ, Zh. Éksp. Teor. Fiz. **60**, 1727 (1971) [Sov. Phys. JETP **33**, 933 (1971)].
[8] E. Parker, *Cosmical Magnetic Fields: Their Origin and Their Activity*, Clarendon Press, Oxford, 1979 [Russian translation, Mir, Moscow, 1982].
[9] E. R. Priest and T. G. Forbes, Sol. Phys. **119**, 211 (1989).
[10] A. G. Frank in *Problems of Plasma Physics and Plasma Electronics* [in Russian], edited by L. M. Kovrizhnykh, Trudy FIAN **160**, 93, Nauka, Moscow, 1985.
[11] R. L. Stenzel and W. Gekelman, Physica D **12**, 133 (1984).
[12] S. Yu. Bogdanov, N. P. Kiriĭ, and A. G. Frank in *Magnetic Reconnection in Two-Dimensional and Three-Dimensional Configurations* [in Russian], Trudy IOFAN, **51**, 3, Fizmatgiz, Moscow, 1996.
[13] S. Yu. Bogdanov, N. P. Tokarevskaya, A. G. Frank, and A. Z. Khodzhaev, Fiz. Plazmy **1**, 133 (1975) [Sov. J. Plasma Phys. **1**, 71 (1975)].
[14] G. V. Dreĭden, N. P. Kiriĭ, V. S. Markov *et al.*, Fiz. Plazmy **3**, 45 (1977) [Sov. J. Plasma Phys. **3**, 26 (1977)].
[15] A. I. Morozov and A. G. Frank, Fiz. Plazmy **20**, 982 (1994) [Plasma Phys. Rep. **20**, 879 (1994)].
[16] S. I. Sirovatskiĭ, Astron. Zh. **43**, 340 (1966) [Sov. Astron. **10**, 270 (1966)].
[17] S. Yu. Bogdanov, G. V. Dreĭden, N. P. Kiriĭ *et al.*, Fiz. Plazmy **18**, 1269 (1992) [Sov. J. Plasma Phys. **18**, 654 (1992)].

[18] S. Yu. Bogdanov, G. V. Dreĭden, N. P. Kiriĭ *et al.*, Fiz. plazmy **18**, 1283 (1992) [Sov. J. Plasma Phys. **18**, 661 (1992)].
[19] A. G. Frank, S. Yu. Bogdanov, V. B. Burilina, and N. P. Kyrie, Bull. Phys. Amer. Soc. **41**, 1371 (1996).
[20] N. P. Kiriĭ, V. S. Markov, and A. G. Frank, JETP Lett. **56**, 82 (1992).
[21] V. B. Burilina, V. S. Markov, and A. G. Frank, Fiz. Plazmy **21**, 36 (1995) [Plasma Phys. Rep. **21**, 33 (1995)].
[22] S. Yu. Bogdanov, Yu. F. Bondar', V. B. Burilina *et al.*, Zh. Tekh. Fiz. **64**, 30 (1994) [Tech. Phys. **39**, 877 (1994)].
[23] S. Yu. Bogdanov, V. S. Markov, A. I. Morozov, and A. G. Frank, Pis'ma Zh. Tekh. Fiz. **21**, 5 (1995) [Tech. Phys. Lett. **21**, 995 (1995)].
[24] S. Yu. Bogdanov, V. B. Burilina, N. P. Kiriĭ *et al.*, Fiz. Plazmy **24**, 467 (1998) [Plasma Phys. Rep. **24**, 427 (1998)].
[25] L. A. Vasil'ev, *Schlieren Methods*, Israel Program for Scientific Translations, Jerusalem, 1971; Nauka, Moscow, 1968, p. 324.
[26] Yu. I. Ostrovskiĭ, M. M. Butusov, and G. V. Ostrovskaya, *Interferometry by Holography*, Springer–Verlag, New York, 1980; Nauka, Moscow, 1977, p. 339.
[27] Sht. Bücher, N. P. Kiriĭ, H.-J. Kunze, and A. G. Frank, Fiz. Plazmy **25**(3) (1999) [Plasma Phys. Rep. **25**(3) (1999)], in press.
[28] S. I. Braginskiĭ in *Reviews of Plasma Physics*, Vol. 1, Consultants Bureau, New York, 1965, p. 205; Gosatomizdat, Moscow, 1963, p. 183.
[29] A. G. Frank, S. Yu. Bogdanov, V. B. Burilina, and N. P. Kyrie in *Proc. Int. Conf. on Plasma Physics*, Nagoya, Japan 1996 (1997), v. 1, p. 506.

Translated by M. E. Alferieff

Dynamics of thin exploded-wire plasma with a cold dense core

G. V. Ivanenkov,*) A. R. Mingaleev, S. A. Pikuz, V. M. Romanova, and T. A. Shelkovenko

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

W. Stepniewski

Institute of Plasma Physics and Laser Microfusion Warsaw, 49 Hery 23, Poland

D. A. Hammer

Laboratory of Plasma Studies, 369 Upsom Hall, Cornell University Ithaca, New York 14853, USA
(Submitted 26 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1215–1229 (October 1998)

Ideas are put forward regarding the possibility of a cold dense core, surrounded by a plasma corona, forming near the axis at the initial stage of a nanosecond electric explosion of metal wires, and the influence of such a radial structure on the plasma compression dynamics is discussed. Experimental evidence supporting these suppositions is presented. It includes both indirect confirmations, based on optical and x-ray diagnostics data, and direct observations of the core by new means of x-ray probing employing an X pinch as a source of radiation. © *1998 American Institute of Physics.* [S1063-7761(98)00610-6]

1. INTRODUCTION

The development of powerful high-voltage generators such as PBFA Z, SATURN, PROTO II, GAMBLE II, BLACKJACK-5, ANGARA-5, and GIT-4, now makes it possible to produce up to 10 MA and higher current pulses, introducing multiterawatt power into the diode in $\simeq 100$ ns. This has opened up wide prospects for using such generators in diverse experiments with hot radiating dense Z-pinch plasma. The standard load for these generators consists of different configurations of thin metal wires—single- and multiwire liners and X pinches. In such generators, after the electric explosion is completed and after the subsequent rapid radial expansion, accompanied by a strongly pronounced skin effect, the current-carrying plasma is compressed by the self-magnetic field. The hot spots formed as a result of the sausage instability contain dense, hot plasma with high parameters ($n_e > 10^{22}$ cm^{-3}, $T_e \simeq 1$ keV, Z up to 51). Powerful bursts of radiation open up the possibility of using these objects as sources of hard ultraviolet and soft x radiation in multicharged ion spectroscopy, in x-ray optics and lithography investigations, and as media with a population inversion in the short-wavelength region. There are also prospects for using them in controlled thermonuclear fusion and for laboratory simulation of the action of powerful x-ray pulses accompanying an atmospheric nuclear explosion. The problems arising in the path toward the realization of these goals are often of a general-physical significance for investigations of dense plasma and condensed phases of matter as well as for the development of new diagnostics methods using x-ray optics and spectroscopy.

2. FORMULATION OF THE PROBLEM

Dense exploded-wire plasmas harbor many mysteries despite efforts to investigate them. For example, direct observation of the initial stage of an electrical explosion still encounters difficulties requiring apparatus with very high dynamic range. However, this is the stage where the transition of the matter from a metallic into a plasma state occurs and the conditions are established for subsequent development of a dense plasma column. There is absolutely no information about the distribution of the parameters or even the nature of the physical state of the matter formed in the process. At the same time it was noted in Ref. 1 that the explosion of thick, 1 cm long, aluminum wires with diameter >50 μm could have been incomplete, and a dense cold core of low-conductivity matter was observed directly in the x-ray images of the plasma. Actually, here the process only just began (surface explosion, expansion of coronal plasma, skin effect), and there was not enough time for it to evolve into the pinch stage: Instead of hot spots and thermal x-ray emission lines, only a surface plasma emission reminiscent of skin-layer emission was observed in the continuous spectrum.

Even more complicated is the case[2] in which an exploding copper wire 20 μm in diameter and 4 cm long exhibited both hot spots and extended emitting structures, whose thermal emission included characteristic soft x-ray lines. The regions with the two types of emission of could be clearly distinguished over quite a long distance, and in Ref. 2 it was suggested that in this situation a core surrounded by a corona is present. However, this structure was not observed directly, and the question of its universality remained unclear for a long time. Recently,[3,4] evidence of the existence of a radial structure appeared in the form of sharp increase in the plasma density near the axis and in teh immediate vicinity of the hot spots for typical wire dimensions, diameter 20 μm and length 1 cm. This effect becomes stronger as the wire material becomes heavier, and it is manifested as the appearance at the initial location of the wires of dense and cold

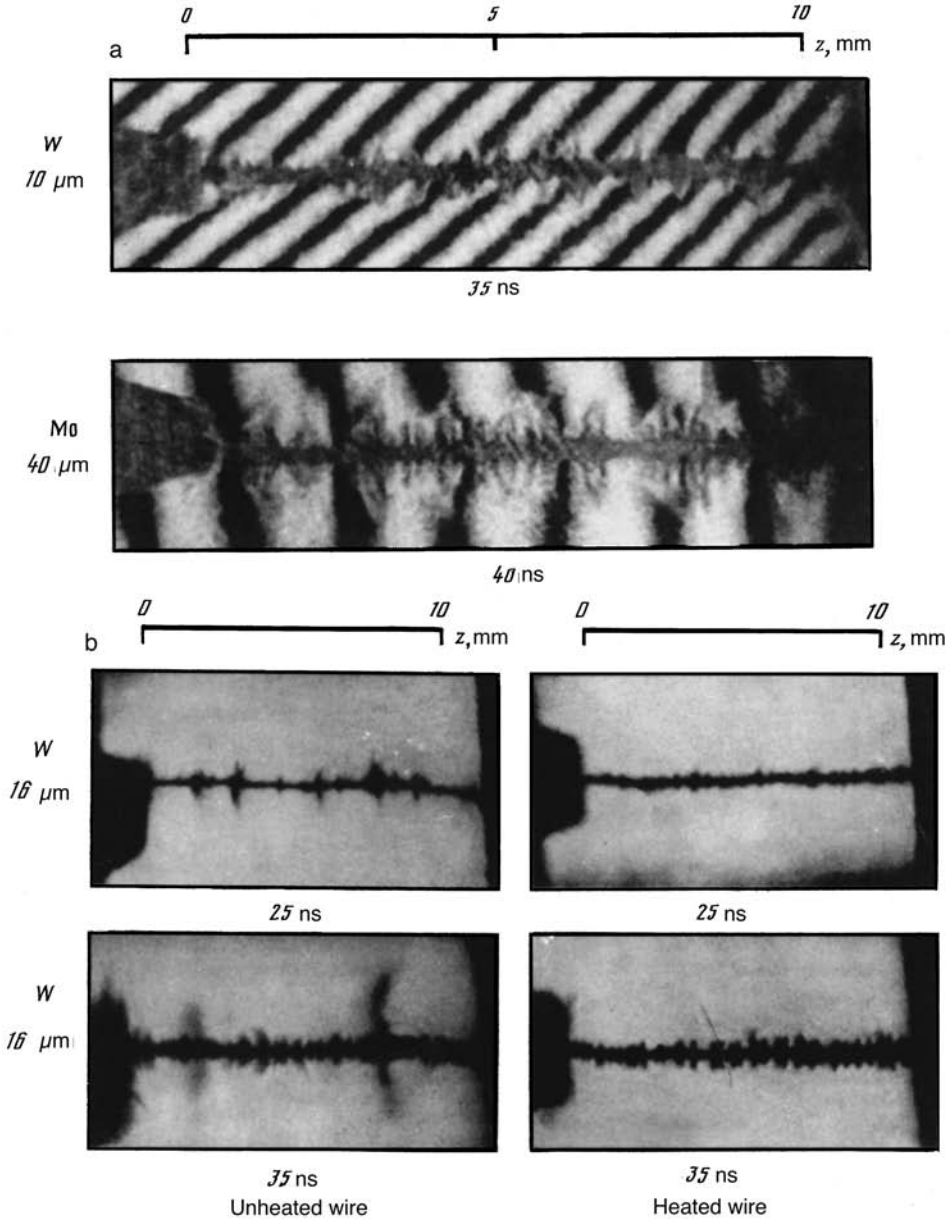

FIG. 1. Examples of laser interferograms (a) and schlieren images (b) of the plasma from exploded metal wires.

''bodies'' having diameters five to six times greater than the initial diameter and surrounded by hotter plasma.

For the present it is difficult to assert that such near-axis formations are the same in experiments performed using different facilities. It can only be concluded that there exists a more complicated distribution of the matter of an exploded wire than the distribution ordinarily used for simulation, including numerical modeling. Analysis of the stationary states of emitting Z pinches[5] leads to the same conclusion. Such states also include heterogeneous equilibria—a cold dense central region, from which heat is removed mainly by radiation, and a hot corona surrounding it with conductive heat transfer in the direction of the axis predominating. These solutions, which exist only for quite heavy substances, were obtained under the assumption, which does not strongly limit them, that the dissipation parameters are power-law functions of temperature. More importantly, the paths leading to the formation of such equilibria are still unclear, as is also, incidentally, the possibility of using them to analyze the strongly time-dependent states in experiments with wires.

It is simplest to assume that the formation of a sharp radial nonuniformity of matter is related to the completion of the initial stage of the explosion, when by virtue of the skin effect in the expanding and rapidly ionizing vapor current flows out of the central part, which has not had enough time to vaporize. Estimates and qualitative analysis, made in Ref. 6 for the wires of diameter 10–50 μm ordinarily used in experiments, lead to this conclusion. They attest to the appearance of a high-resistance core of dense matter, having a temperature of the order of tenths of the Fermi energy, in the rapidly current-heated metallic load. The subsequent scenario differs from the standard one through the interaction of the core with the shock wave resulting from the compression of the coronal plasma. This effect and its influence on the ultimate compression are some of the subjects of the present paper.

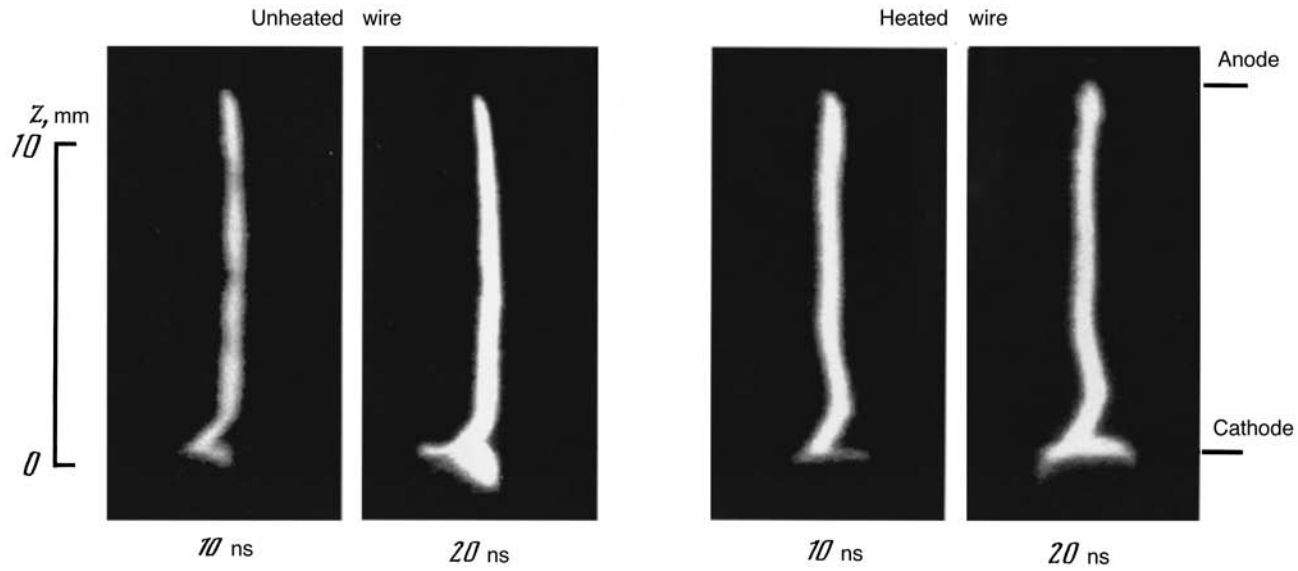

FIG. 2. Optical self-emission of exploded wires (W, 16 μm) at different times during the discharge.

3. EXPERIMENTAL RESULTS

Most data on wire explosion dynamics are currently obtained from optical measurements. Deep layers of the plasma column are inaccessible to optical measurements because of the long wavelength of the laser radiation: The actual density limit reached by such probing is determined by the nonuniformity of the plasma and the aperture angles of the optical scheme employed and is only $(1-2)\times 10^{20}$ cm^{-3}, which is much lower than the critical density $\simeq(1-2)\times 10^{21}$ cm^{-3}. For this reason, in the present work we employed, together with the optical method, x-ray probing, which is applicable for plasma with density up to 10^{23} cm^{-3} with essentially any values of the gradient.

The experiments were performed on different devices: the ''Don'' (170 kA, 200 kV, 80 ns) and BIN (280 kA, 400 kV, 150 ns) high-voltage generators at the Lebedev Institute and XP (480 kA, 200 kV, 100 ns) at Cornell University. Diagnostics allowing the plasma to be studied in both the optical and x-ray ranges were used. In what follows, only the new x-ray probe method used in experiments on the XP setup are described. The other diagnostic systems have been described in detail in previous works.[1,7]

3.1. Optical measurements

We point out indirect evidence of the existence of a core. An unstable plasma shell, whose shape is the same as that of the initial wire and which acquires with time the typical structure of a sausage instability, can be seen in the interferograms and schlieren photographs presented in Fig. 1. The overall character of the process is essentially independent of the material and the initial diameter of the wire, but sometimes ejecta of low-density plasma, in some cases reaching distances of several millimeters from the axis, are observed. At the same time, the form of the plasma column depends strongly on the state of the wire surface: When it is cleaned immediately prior to a shot by heating with a constant current, the plasma ''coat'' becomes more dense and stable (specifically, there are no transverse ejecta) and the resulting structure of the sausage instability becomes more regular (Fig. 1b). The main result obtained by comparing these images is that the variations of the column shape in both cases are superposed on massive and immobile base of the same type.

Another interesting fact which we shall present is that in the optical range of the characteristic radiation a thick luminous column with $\simeq 0.5$ mm is observed throughout virtually the entire time of current growth ($\simeq 70$ ns for the BIN machine) (Fig. 2), after which the column rapidly disintegerates. Since the corona, whose ions are stripped of outer electrons, is optically transparent, it can be inferred that the source of emission is a central fairly cold object.

3.2. X-ray measurements

A characteristic feature of the x-ray pinhole photographs of exploded-wire plasma is the presence of bright regions—so-called ''hot spots,'' which are arranged along the discharge axis near the initial position of the wires. The picture becomes much more regular when the wire surfaces are cleaned. For $\hbar\omega > 0.8-1.3$ keV a luminous cylindrical structure with characteristic diameter $\simeq 150-300$ μm, several times greater than the initial size of the wire (Fig. 3), is observed in the image obtained with quite good spatial resolution $\simeq 50$ μm. The overall character of the emission is essentially unrelated with the wire material (Fig. 3a), but the diameter of the cylindrical formation depends on the initial conductivity of the wire (Fig. 3b): It is much larger for metals with good conductivity (Al, Cu, Au) than for metals with low conductivity (Ni, Pd, W).

Moreover, there is the fact that wires with a low initial conductivity, under otherwise the same conditions, emit more intense x-ray radiation in the hard region ($\lambda < 5$ Å). Besides the large-scale structure of the hot spots with a characteristic dimension 0.5–2 mm along the axis observed in pinhole images with spatial resolution 50–100 μm by many

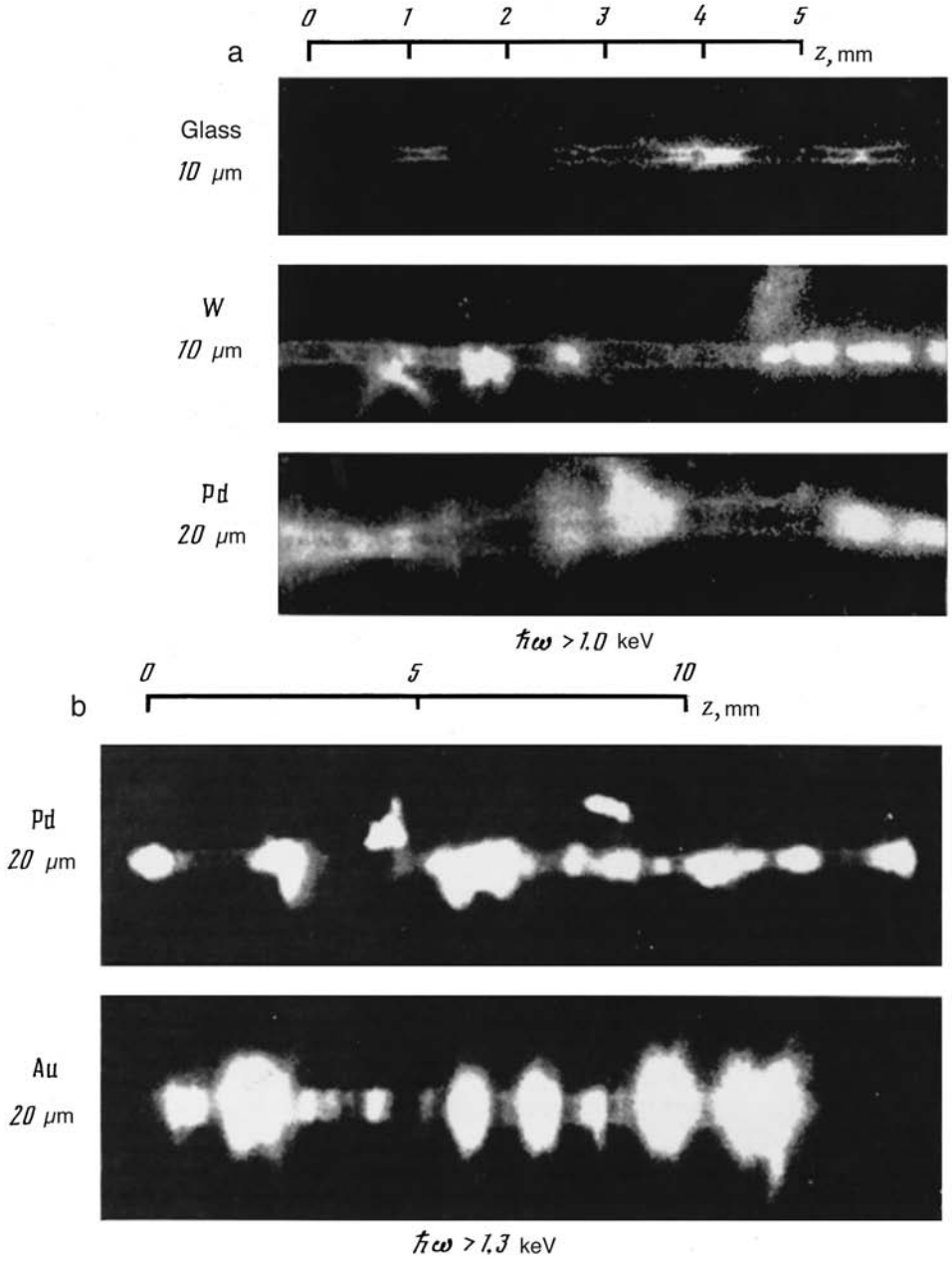

FIG. 3. Time-integrated x-ray pinhole image of exploded wires.

investigators, a small-scale substructure was observed in the present work. It follows from the data obtained with a pinhole camera with resolution 4–30 μm that in many cases the hot spots have a fine structure. Specifically, they consist of groups of two or three ''subdots,'' smaller than the pinhole diameter, which are separated by 20–100 μm along the wire axis (Fig. 4). A similar result was also obtained using a transmission diffraction grating produced in the opening of a 20 μm in diameter pinhole (Fig. 5).

3.3. X-ray probing

The dense regions of exploded wires were investigated by the x-ray schlieren photography. An X pinch located in the diode gap parallel to the load investigated (Fig. 6) served as the source of radiation. The diode anode was separated by a slit into two parts, which were mounted on return-current half-cylinders, which made it possible to measure separately the currents flowing through the source and the load. Mo and Pd were chosen as the wire material for the X pinch, since preliminary investigations showed that for $Z_n \simeq 40-45$ the region of the X-pinch emission has minimum size and adequate brightness in the wavelength range 2–5 Å having an acceptable level of hard radiation with $\lambda < 2$ Å. For comparison, we note that in an X pinch from heavy elements ($Z_n \simeq 70$) hard radiation is generated by an electron beam, arising in the process of compression of the plasma,[8] and is emitted from a region of size >1 mm, which is unacceptable for obtaining an image with high spatial resolution. In our experiments the spatial resolution was at a level of several microns with an exposure duration of 1–3 ns. The plasma images with a magnification of 7–8 were recorded on Kodak RAR 2497 or Kodak GWL photographic films, screened from the visible light by beryllium or titanium filters.

Figure 7 displays an image of the explosion of a 40 μm

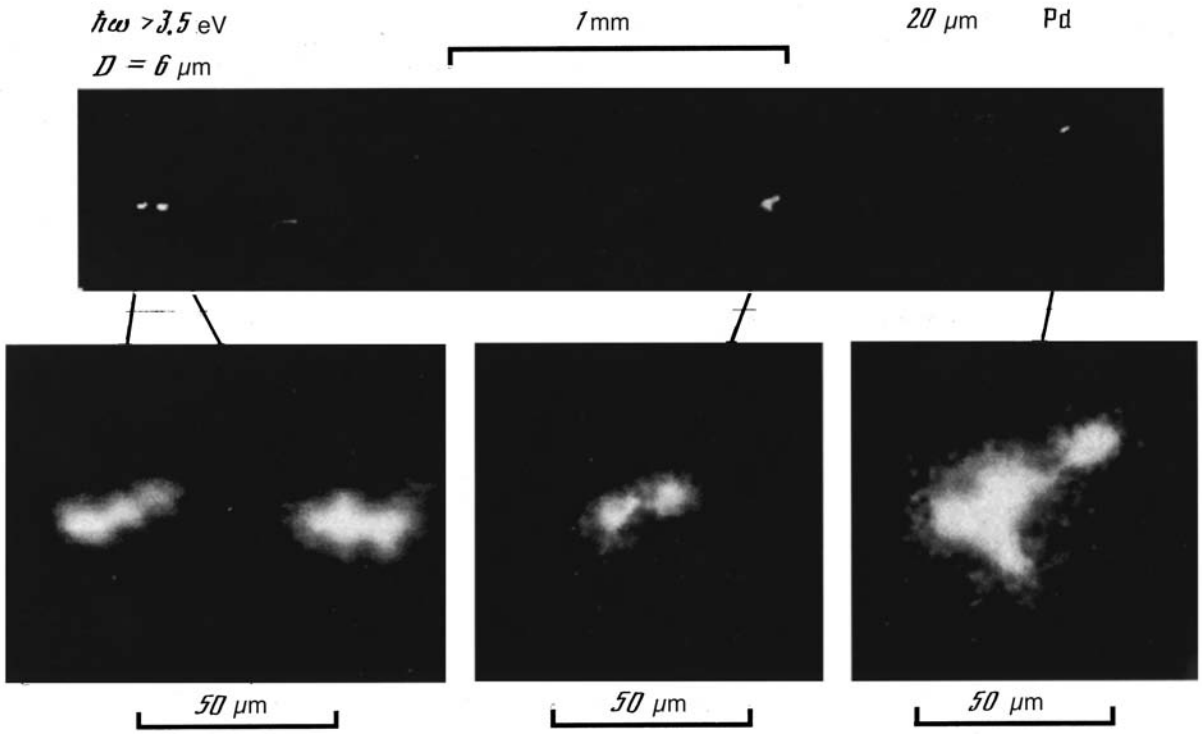

FIG. 4. Fine structure of a hot spot obtained with a high-resolution pinhole camera.

in diameter tungsten wire 50 ns after current onset. A dense core, completely screening the source radiation, can be seen. The diameter of the core is three to seven times greater than the initial diameter. The variations of the diameter of the core with periodicity 0.7–2 mm as well as the sharpness of the boundary of the core are interesting. The smearing of the edge of the image was les than 7–10 μm. Disturbances of the more rarefied and unstable corona with characteristic size 0.3–0.5 mm, corresponding to nonuniformities of the plasma column in the optical images in Fig. 1, were also observed. The uniformity with which material was distributed in the core along the axis depended on the purity of the wire surface, indicating that the origin of the core is related with the process of the initial electrical explosion of the conductor.

Quantitative analysis of such an image is a difficult problem, requiring careful calibration of the photographic materials and precise measurement of the spectral composition of the probing radiation in each experiment. Such an analysis was not performed in the present work, but simple estimates of the radiation absorption in the corona, taking account the transmission of the filters and the sensitivity of the photographic film, gave a corona density $n_i \simeq (4-20) \times 10^{19}$ cm^{-3} near the core and $\simeq (1-4) \times 10^{18}$ cm^{-3} near its visible boundary. These values, in principle, agree with the results of optical measurements for distances $\simeq 0.5$ mm from the wire axis, but they show that the core is essentially inaccessible to optical probing. Assuming that the core contains most of the initial material, we obtain an approximately 100-fold drop in density at the core–corona boundary.

The image shown in Fig. 7 of the explosion corresponds to the stage before the formation of the hot spots observed in this experiment $\simeq 70$ ns after the start of the current pulse. It is interesting to make the generate the probe radiation at the time when the hot spots form. This was done by varying the linear mass of the load by varying its diameter or using several closely spaced wires. Figure 8 shows an image of an explosion of a pair of close, 13 μm in diameter tungsten wires separated by 20–40 μm. X-radiation from such a load was observed both before and after the probe pulse, but the core is noticeable only in part of the image. Such a complicated form indicates that the core evolves during the discharge process. Evidently the core has its own internal structure. Figure 9 shows an image of a Ti ($Z_n = 22$) wire 40 ns after the start of the current pulse. Because of the weak absorption of radiation in titanium the corona is not visible here, but the internal structure of the core, which is different for different sections of the wire, does reveal itself. To investigate this structure the spatial resolution must be improved to a level of order 1–2 μm. Here we state only the fact that it exists. The choice of a titanium filter played an important role in obtaining this image: Near the *K* absorption edge Ti possesses a transmission window 2.5–4.0 Å–precisely in the range of the minimal dimensions of the emitting region of the X pinch.

The results presented show that x-ray schlieren photography yields much more data on the deep layers of dense plasma than the standard optical laser probing. In our experiments we also used a spherically curved crystal spectrograph (mica, 10 cm radius of curvature, 100 μm beryllium filter), making it possible to obtain soft x-ray spectra in the wavelength range of the order of several angstroms. The shape and form of the spectra, just as the plasma parameters found from them, fell into the typical range for such measurements (see, for example, Ref. 7).

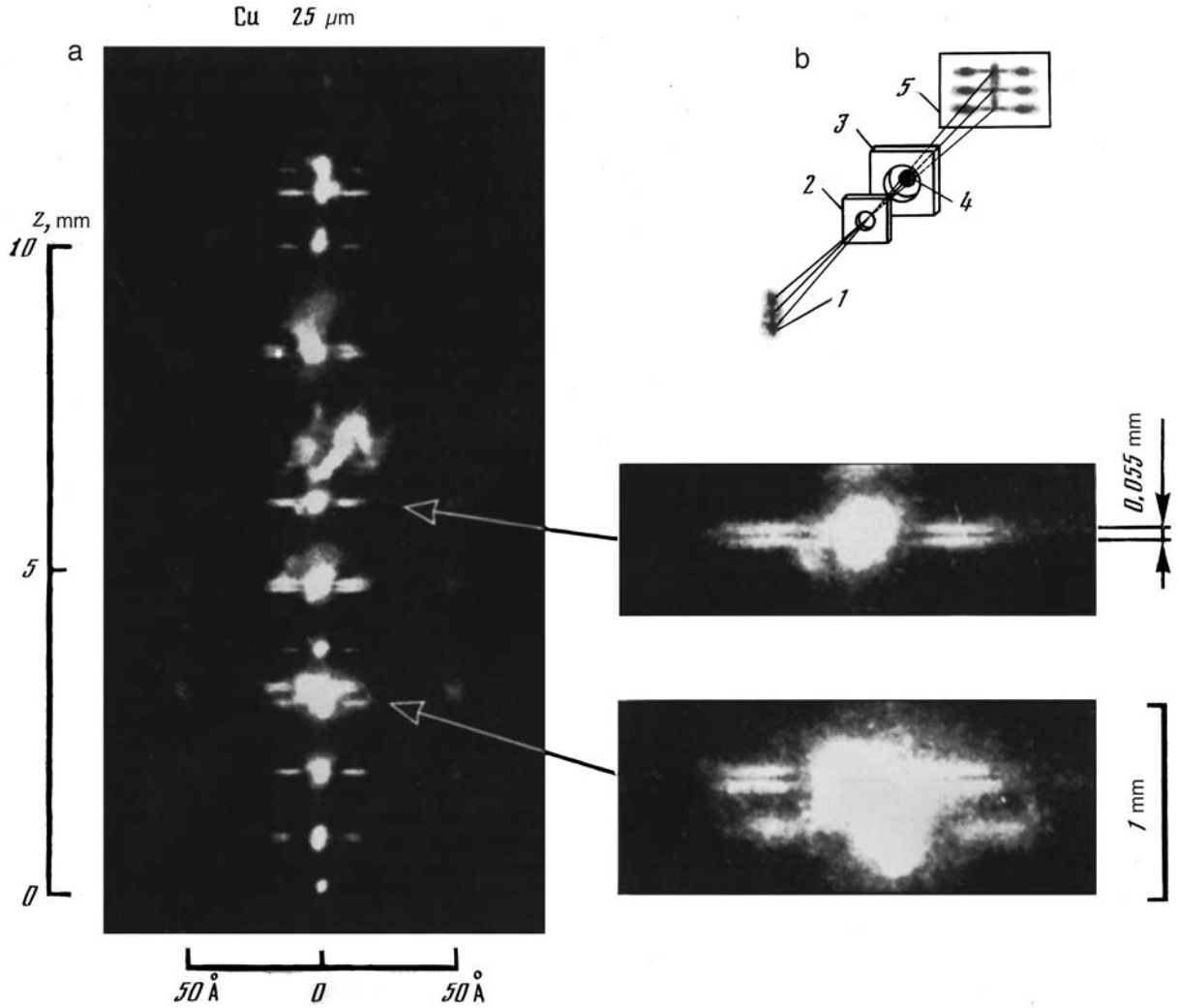

FIG. 5. Observation of the explosion of a copper wire using a transmission diffraction grating. The numbers in the diagram (b) show: *1* — plasma object, *2* — screen with a 50 μm in diameter diaphragm, *3* — substrate, *4* — tungsten diaphragm and grating, *5* — image.

4. QUALITATIVE ANALYSIS AND HYPOTHESES

The formation of a core can be related to the properties of the electrical explosion of the wire. Estimates[6] show that for fast ($\simeq$10 ns) application of voltage $>$100 kV the vaporization front can penetrate to the axis only if the diameter is very short, <5 μm. On the other hand, the skin effect is important only of the wire diameter is quite large (50 or more μm). For this reason, the nonvaporized central part of the metal is superheated above the boiling temperature, and the phase trajectory reaches the critical point via volume ebullition.

The somewhat idealized $P-V$ phase diagram, presented in Fig. 10 (V is specific volume, C is the critical point), illustrates these processes. In this diagram the curve ABC describes the path of ohmic heating of the matter on the wire axis, crossing the binodal at the point B. On account of the monotonic nature of the heating, each point B' on this curve corresponds to a definite time, and $B'B''D'E'$ describes the corresponding change in the thermodynamic parameters in the radial direction from the axis to the surface, where vaporization $B''D'$ occurs, and farther—into the vapor region

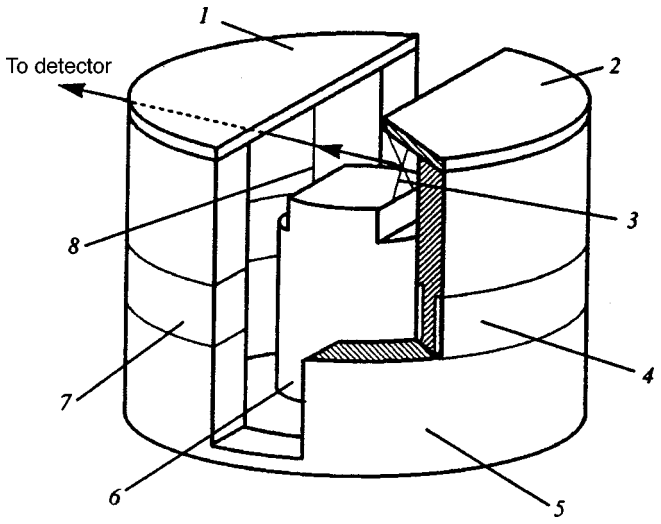

FIG. 6. Diagram of x-ray schlieren photography using an X pinch: *1*, *2* — anodes, *3* — source (X pinch), *4*, *7* — Rogowski loop, *5* — half-cylindrical return-current conductor, *6* — cathode, *8* — load.

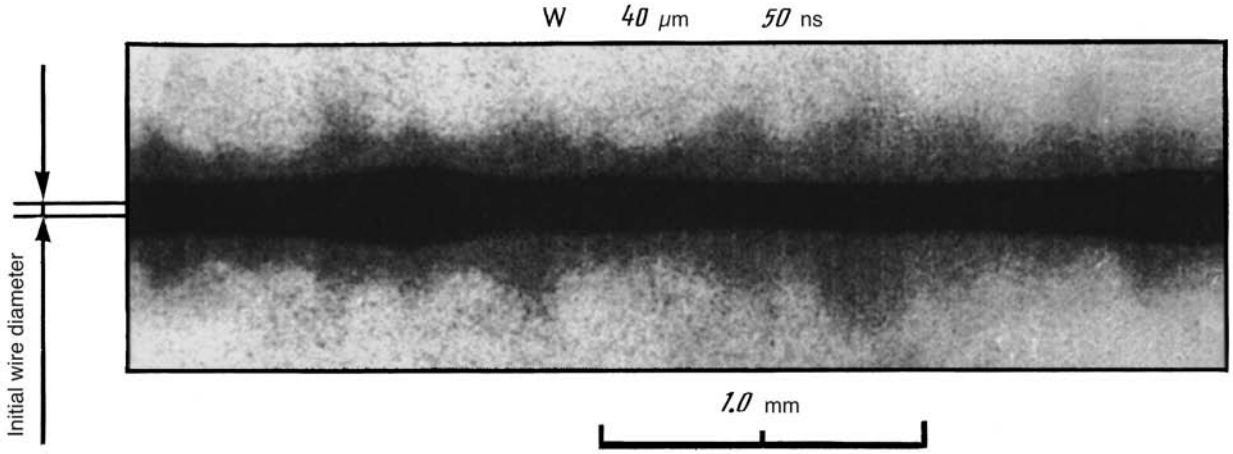

FIG. 7. X-ray schlieren image of the structure of the core-corona of an exploded tungsten wire.

expansion. When the point B lying below the critical point is reached, surface vaporization stops and the isotherm BDE separates the region of expanding vapor from the center, which continues to heat up.

Naturally, a separate diagram is associated with each point on the axis of the system. In the critical range of the parameters the conductivity of the matter at the center is very low (the resistance to the current reaches several $k\Omega$/cm), and the vapor, which ionizes in the course of breakdown, starts to overtake the current. This, as well as the confinement of the current to the skin layer during the subsequent expansion of the plasma formed, stop further heating of the center, and the center remains relatively cold right up to the arrival of the shock wave arising during compression. Ultimately, the distribution of material over the cross section becomes hetereogeneous—a central core surrounded by a plasma corona.

If the optical self-emission is related to the core (Fig. 2), then the core must have formed no later than during the first 10 ns after current onset and existed throughout the entire phase of current growth. The well understood behavior of a shock wave in plasma will serve as the starting point for understanding what follows.[9] Since the parameter $(me/m_i)^{1/2}$ is small, the structure of the shock wave is determined by slow e–i energy exchange processes and ionization on the one hand and the high electron mobility on the other. As a result, on the shock front the electron temperature, which undergoes a jump $\simeq m_e D^2$ (D is the velocity of the front), lags behind the ion temperature ($\Delta T_i \simeq m_i D^2$).

Subsequent relaxation requires $(m_i/m_e)^{1/2}$ times more time than momentum and energy transfer to ions in a viscous shock. This, as well as multiple repetition of this effect in the process of multiple ionization, determine the structure of the front. Heating of the plasma in front of the viscous shock by hot electrons and photons penetrating in front of it must also be added here. Under our conditions the dimensions of the front and zone of electronic and radiation heating reach tens of μm against the background of a plasma corona with a radius of $\simeq 100$ μm. As it approaches the core, this wide front starts to interact with the sharp drop in the density of the matter in its path, while its precursor—the wave of electronic and radiative heat conduction — slowly heats the mas-

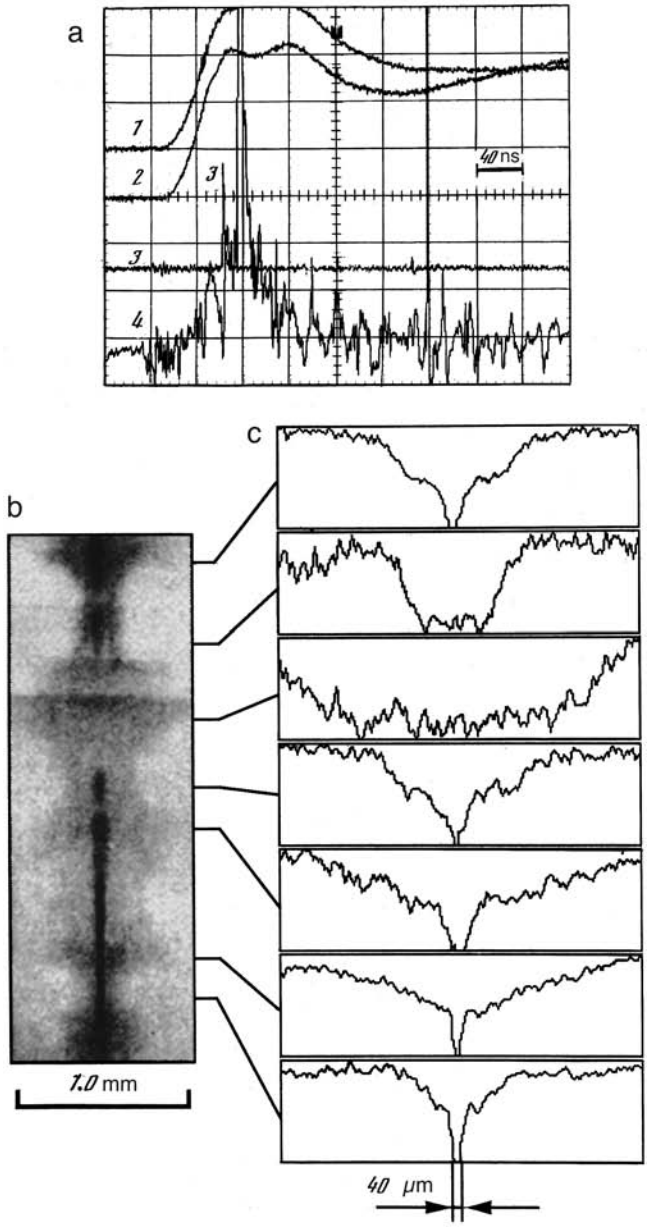

FIG. 8. Explosion of a pair of close wires in an x-ray pulse from an X pinch: a — currents through the diode and signals from the x-ray photodiodes (*1* — X pinch, *2* — current through the wire, *3* — X-pinch radiation intensity, *4* — intensity of the radiation from the wires), b — photograph, c — densitometer traces in different sections.

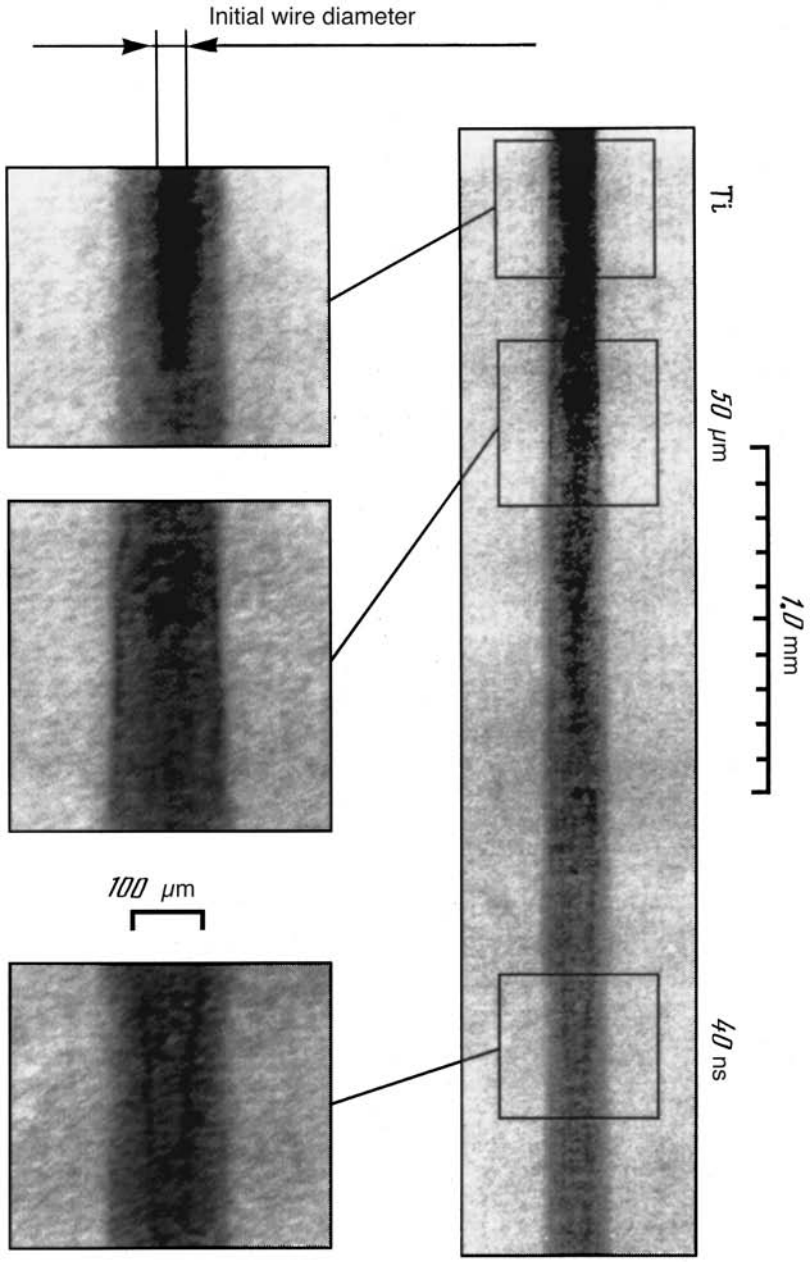

FIG. 9. X-ray schlieren image of the explosion of a titanium wire with visible internal structure of the core.

sive core. The durations of these processes are $\simeq 10$ ns.

Here it is possible to make use of a similar problem from gas dynamics — the analogy with one version of the decay of an arbitrary discontinuity,[9] where instead of a shock wave arriving from the corona, new shock waves — a transmitted wave moving to the core and backward reflected wave – are formed at the interface, and the interface becomes a contact discontinuity moving toward the axis. This process is probably observed in the form of extended double-emission structures in the pinhle photographs in Fig. 3. On account of the characteristic features of the plasma compression wave, a wave of radiative and electronic heat conduction propagates first into the interior of the dense core, and the shock (which is much narrower in the dense core medium than in the corona) is at first still too weak and slow to be at all visible in the overall picture. However, over the long time during which the transformation of waves occurs the continuing growth of the current has an effect on the boundary, and because of nonlinearity the refracted wave forms more rapidly. This shock wave, entraining and heating the core matter, is gradually focused on the axis. The shock wave is assisted in this process by the reflected wave in the plasma corona, which in turn compresses the plasma behind it. As a result of this focusing, the maximum of T_e moves onto the axis, while the heated core becomes the main carrier of the current flowing through the plasma and continues to heat up even more as it grows.

Here it is necessary to recall how the instabilities grow. Evidently, we observe them in Fig. 9. As the corona plasma is compressed, as usual, the sausage instability responsible for the formation of constrictions appears. The existence of a fairly sharp core–corona boundary makes the buildup of the

associated hydrodynamic instabilities of the Rayleigh-Taylor, Richtmeier–Meshkov, Kelvin–Helmholtz, and other types possible. The first of these instabilities is due to the accelerated motion, occurring under the conditions of current growth, of the matter near the boundary located between the transmitted and reflected shock waves. The form of the fronts of these waves is modulated by the sausage instability, and focusing of the compression wave on the axis of the core occurs nonsimultaneously in accordance with this modulation.

If the nonlinear development of Rayleigh-Taylor instability is sufficiently strong, jets of coronal plasma are directed toward the axis. This effect should be most highly developed near the maximum of the current in the regions where the shock front reaches the axis. Here the flow of the hot plasma in the jets relative to the less mobile core matter can give rise to Kelvin–Helmholtz instability, whose development in turn engenders growing vortices and bursts of hydrodynamic turbulence. As a result, spots of mixed matter of core and corona matter are formed and cluster inside the constrictions of the plasma corona. This affects the above-described concentration of current in the core region near the axis. Thus, in the process of the evolution of the miniature clusters of hot, dense plasma the dynamo effect can be expected to engender near them a series of induction loops of intense magnetic fields. Ultimately, the large constrictions will fragment into groups of small luminous spots, forming a fine structure similar to that observed in Figs. 5 and 6.

The observed facts presented above support the arguments made, but the picture drawn here still contains much that is hypothetical. The numerical simulation results presented in Ref. 10 provide additional confirmation of this picture. The observation of hydrodynamic instability of the core–corona boundary was reported in a recent work.[11]

5. CONCLUSIONS

The results presented reveal the specific nature of the explosion of wires, which sharply distinguishes such explosions from other types of fast Z pinches—vacuum spark, gas leak-in, pinching of laser plasma, capillary discharge. It was determined that the characteristic features are due to the presence of a dense, cold core which is formed in the process of incomplete vaporization of the wire material in the course of the electrical explosion and exists for a long time during the discharge. It was shown experimentally not only that a core exists but also that a dynamical structure arises and influences the formation of the fine structure of the hot spots. As a result, the standard models that ignore the heterogeneous distribution of matter over the cross section of the plasma column are hardly capable of describing correctly the complicated dynamics of exploding wires and need to be modified.

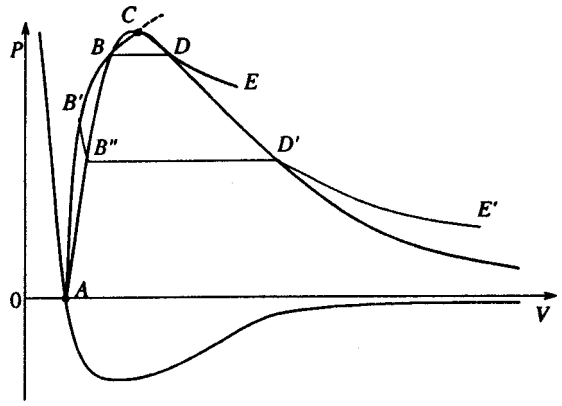

FIG. 10. $P-V$ diagram of a rapid electric explosion of a metal.

This work was supported in part by the Sandia National Laboratory (Albuquerque, USA) under contract AJ-6400.

*)E-mail: ivanenkv@sci.lebedev.ru

[1] S. M. Zakharov, G. V. Ivanenkov, A. A. Kolomenskiĭ, S. A. Pikuz, and A. I. Samokhin, Fiz. Plazmy **9**, 469 (1983) [Sov. J. Plasma Phys. **9**, 271 (1983)].
[2] L. E. Aranchuk, S. L. Bogolyubskiĭ, G. S. Volkov *et al.*, Fiz. Plazmy **112**, 1324 (1986) [Sov. J. Plasma Phys. **12**, 765 (1986)].
[3] D. H. Kalantar and D. A. Hammer, Phys. Rev. Lett. **71**, 3806 (1993).
[4] T. A. Shelkovenko, S. A. Pikuz, A. R. Mingaleev, and D. A. Hammer, Bull. Am. Phys. Soc. **42**, 2051 (1997).
[5] N. A. Bobrova, T. L. Razinkova, and P. V. Sasorov, Fiz. Plazmy **14**, 1053 (1988) [Sov. J. Plasma Phys. **14**, 617 (1988)].
[6] G. V. Ivanenkov, A. R. Mingaleev, T. A. Novikova *et al.*, Fiz. Plazmy Zh. Tekh. Fiz. **65**(4), 40 (1995) [Tech. Phys. **40**, 312 (1995)].
[7] A. Ya. Faenov, S. A. Pikuz, A. I. Erko *et al.*, Phys. Scr. **50**, 333 (1994).
[8] S. A. Pikuz, T. A. Shelkovenko, V. M. Romanova *et al.*, Zh. Éksp. Teor. Fiz. **112**, 894 (1997) [JETP **85**, 484 (1997)].
[9] Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Academic Press, N. Y., 1966, 1967, Vols. 1, 2 [Russian original, Nauka, Moscow, 1966].
[10] G. V. Ivanenkov, A. R. Mingaleev, S. A. Pikuz *et al.* in *Proc. 4-th Int. Conf.* on *''Dense Z-pinches,''Vancouver*, AIP, New York, 1997, p. 253.
[11] S. Yu. Gus'kov, G. V. Ivanenkov, A. R. Mingaleev *et al.*, JETP Lett. **67**, 559 (1998).

Translated by M. E. Alferieff

Three-particle recombination of non-hydrogen-like ions in the presence of plasma microfields

M. Yu. Romanovskiĭ*)

Institute of General Physics, Russian Academy of Sciences 117942 Moscow, Russia
(Submitted 30 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1230–1241 (October 1998)

The effect of electric and magnetic plasma microfields on three-particle electron–ion recombination via the highly excited states of a hydrogen-like ion is studied. It is shown that electric microfields impede this process, and at some electron temperature it ceases for sufficiently high plasma density. Magnetic microfields speed up recombination via low-lying states only negligibly. The rate of such recombination into non-hydrogen-like ion states is comparatively higher than for the equivalent hydrogen-like ion states. © *1998 American Institute of Physics.* [S1063-7761(98)00710-0]

1. INTRODUCTION

There are two types of three-particle electron–ion recombination. The first process, which is related to impact ionization by the principle of detailed balance, consists of an ion capturing a free electron directly into a low-lying state (including the ground state). The second process occurs as follows:[1] An ion captures a free electron into a highly excited (Rydberg) state. It is obvious that the probability of such a process is high, but ionization of this electron from the highly excited state likewise occurs very rapidly. However, the probability of a transition of an electron weakly bound in a Rydberg state downward along the energy scale (i.e., into more strongly bound states) under the action of impacts from free electrons is still higher than the probability of a transition upwards.[1] The process is reminiscent of diffusion along the spectrum of Rydberg states or diffusion in energy, since the density of Rydberg states is high and one can switch to a continuous variable (energy). This second kind of three-particle recombination was studied initially by Pitaevskiĭ for recombination in weakly ionized plasma[1] and later extended[2] to the case of a relatively cold dense plasma (only hydrogen-like ions were studied).

These two types of three-particle recombination are related as follows: For $\hbar v_T/e^2>1$ (where e is the electron charge and v_T is the average electron thermal velocity) the rate of the first process in plasmas with electron density up to 10^{18} cm^{-3} is higher, which corresponds to an electron temperature of approximately 30 eV. At lower temperatures the second process predominates. This is why the rate of the latter recombination process enters in all calculations, for example, of the recombination scheme of an x-ray laser.[3,4] The problem is, however, that as the plasma density increases, the random amplitudes of the electric and magnetic microfields increase. The latter have a weaker effect on recombination, and in plasma they are usually weaker than the electric microfields. The electric microfields simply destroy the structure of highly excited levels of ions in plasma (the Inglis–Teller effect) and the channel for such three-particle recombination closes. The problem is to investigate how the rate of this three-particle recombination process decreases as a function of the electric microfields and the fields (or equivalently, the plasma densities) for which the channel for three-particle recombination via highly excited states can be regarded as closed.

Another interesting problem is to analyze recombination in non-hydrogen-like ions (this problem has not previously been investigated systematically[5]). It is obvious that a recombining electron should ''feel'' the presence of a strongly charged nucleus because the electron wave function is not zero at the location of the nucleus. For this reason, substantial differences should be expected from three-particle recombination into hydrogen-like ion states.

2. EFFECT OF ELECTRIC AND MAGNETIC MICROFIELDS ON THREE-PARTICLE ELECTRON–ION RECOMBINATION VIA HIGHLY EXCITED STATES

The Pitaevskiĭ theory must be extended to the case of an electric field (see the similar extension for radiative[6] and dielectronic[7–9] recombinations). Since the recombination process is quite slow, only the ion component of the microfield influences it, while the electron component is averaged. It is obvious that only the electron component can give a magnetic microfield of any significant magnitude, so that one would think that the effect of this field on recombination can be neglected. However, in a nonequilibrium plasma the electron magnetic microfield can be quite pervasive, so that we shall still analyze its effect.

First, we shall find the laws governing the dependence of the three-particle recombination rate on the comparatively weak microfields. The field strength at which the rate changes substantially (for example, vanishes, although the field in this case falls outside the limits of the initial approximation), and hence the value of the plasma density, will critically separate the two recombination processes—only the channel by means of which free electrons are captured into low-lying states will remain in a high-density plasma.

Therefore the quantities determining recombination via highly excited states must be modified for the case when a

field is present. There is only one such quantity, the mean-squared energy that a recombining electron gives up to free electrons per unit time. It can be calculated by two methods: purely classically[10] or quantum-mechanically.[11] The latter method gives a somewhat higher value.[11] However, we do not need to determine the absolute value of the recombination rate; it is sufficient to find the relative value (as a function of the field). Thus, at first glance both computational methods are suitable, since the classical calculation of the state of a highly excited electron is completely correct (here the quasiclassical approximation is very close to the purely classical approximation).

In the classical approach the mean-square energy which a recombining electron gives up to free electrons per unit time can be calculated as follows:

$$\left\langle \frac{\partial(\delta E)^2}{\partial t} \right\rangle_0 = \int n_e (\Delta E)^2 |v - v_1| f_0(E) F_1(E_1) \times d\Gamma \, d\sigma \, dE_1 . \quad (1)$$

Here v is the velocity of a ''finite'' (recombining) electron, v_1 is the velocity of a free electron with which the finite electron collides and to which it gives up energy, n_e is the electron density, m is the electron mass, $\Delta E = m(v^2 - v_1^2)/2$, $d\sigma$ is the Rutherford differential scattering cross section, $d\Gamma = dv_x dv_y dv_z dx dy dz$ is the differential phase volume,

$$F_1(E_1) = \left(\frac{m}{2\pi T} \right)^{3/2} \exp\left(-\frac{E_1}{T} \right)$$

is the Boltzmann distribution function over energy E_1, and $f_0(E)$ is the energy distribution function of the finite electron. If the latter function is normalized to one electron,[2] then

$$f_0(E) = A\delta\left(E + \frac{Ze^2}{r} - \frac{mv^2}{2} \right), \quad (2)$$

where Z is the ion charge (A is introduced below). The quantity (1) is very simply modified for the case of an electric microfield (similarly for a magnetic field): The Stark term $e(\mathscr{E}\cdot\mathbf{r})$, where $\mathscr{E}$ is the electric microfield (it enters only in the expression for $f_0(E)$), must be inserted in the δ function and the expression obtained must be expanded in powers of the field and the first nonvanishing term taken.

It is found, however, that this first nonvanishing term is quadratic in the field and does not describe a linear Stark effect (the same is true for the Zeeman effect with the magnetic microfield). This is physically correct only for an electron microfield which is rapidly time-averaged, but for a stationary (on the scale of the problem) ion microfield this is incorrect and such linear effects are present (for example, in the hydrogen atom). Therefore the effect of the microfields on the three-particle recombination process cannot be calculated classically, and quantum calculations must be used.

As is well known (see, for example, Refs. 12 and 13) the Schrödinger equation for a hydrogen-like ion in a constant electric field admits separation of variables in parabolic coordinates and introduction of parabolic quantum numbers n_1 and n_2 such that their sum gives a principal quantum number n which is the same as the principal number in spherical coordinates. Then, in the case of a transition from a state with principal quantum number $n+1$ to a state n,

$$\Delta E = E_n - E_{n-1} = \frac{I_H Z^2}{n^2} - \frac{I_H Z^2}{(n+1)^2} - \frac{3\mathscr{E} r_B e}{2Z} n(n_1 - n_2) + \frac{3\mathscr{E} r_B e}{2Z}(n+1)(n_1 - n_2 - 1). \quad (3)$$

Here I_H is the ionization potential of hydrogen and r_B is the Bohr radius. For $n_2 \gg n_1$ (this condition always holds in sufficiently strong microfields) the expression (3) can be simplified:

$$\delta E \simeq \frac{2I_H Z^2}{n^3} - \frac{3\mathscr{E} r_B e}{2Z} n. \quad (3a)$$

The second negative term shows that the energy difference decreases in the presence of a field. We obtain for the square of ΔE

$$(\delta E)^2 = \frac{4I_H^2 Z^4}{n^6}\left(1 - \frac{3Z\mathscr{E} e^3}{2E^2} \right). \quad (3b)$$

The rate of change of this quantity can be expressed in terms of the collisional transition rate between the levels $n+1$ and n.[11] It is easy to see that

$$\left\langle \frac{\partial(\delta E)^2}{\partial t} \right\rangle = \left\langle \frac{\partial(\delta E)^2}{\partial t} \right\rangle_0 \left(1 - \frac{3Ze^3\mathscr{E}}{2E^2} \right). \quad (4)$$

The recombination rate b is determined by the expression[2]

$$b = \left[(2\pi m T)^{3/2} \int_{-\infty}^{0} \frac{2Ae^{E/T}}{\langle \partial(\delta E)^2/\partial t \rangle} dE \right]^{-1}. \quad (5)$$

Here A is the normalization constant of the function f_0. Just as for the classical calculation of $\langle \partial(\delta E)^2/\partial t \rangle$, the coefficient A contains terms which are proportional only to the square of $\mathscr{E}$:

$$A(\mathscr{E}, E) = E^{5/2}/\sqrt{2}\,\pi^3 Z^3 m^{3/2}[1 + O(Ze^3\mathscr{E}/E^2)^2]. \quad (5a)$$

Therefore it is sufficient to use the expression given in Ref. 2 or 10 for A. Using $\langle \partial(\delta E)^2/\partial t \rangle_0$ we now obtain

$$b = Z^3 \pi^{3/2} e^6 n_e^2 \left\{ 4T^{3/2} \int_{-\infty}^{0} e^{E/T} \left[\left\langle \frac{\partial(\delta E)^2}{\partial t} \right\rangle_0 \times \left(1 - \frac{3Z\mathscr{E} e^3}{2E^2} \right) \right]^{-1} dE \right\}^{-1}. \quad (5b)$$

Performing the elementary integration we obtain

$$b = b_0(1 - 2Ze^3\mathscr{E}/T^2). \quad (6)$$

Here b_0 is the recombination rate from Ref. 2, i.e., the rate that is always used in calculations of three-particle electron–ion recombination.

Now, knowing the physical principles by means of which electric microfields influence the rate of three-particle recombination via highly excited states, we can extend the perturbation theory to the case of arbitrary fields. In view of what was said above the classical approach cannot be used

here.[1,2] Therefore we shall use a quantum analysis, determining the recombination rate as the electron flux density, due to collisional deactivation transitions (into highly excited states), into the ion. This approach was demonstrated in Ref. 11. We shall employ immediately an expression determining the dependence of the flux density j_n through the level n on the population density n_n of this level. Treating the principal quantum number n as a continuous variable, we have, on account of the high density of highly excited states,[11]

$$j_n = -\alpha_{n-1,n} n_e \left[\frac{\partial n_n}{\partial n} + \left(\frac{\beta_{n,n-1}}{\alpha_{n-1,n}} - 1 \right) n_n \right]. \tag{7}$$

Here $\alpha_{n-1,n}$ is the rate of collisional excitation of an electron from level $n-1$ to level n, $\beta_{n,n-1}$ is the rate of the reverse process (i.e., the desired recombination rate; for the transitions $n \to n-2$, $n \to n-3$, and so on, see Ref. 11; their total contribution is less than the factor 1.08). The principle of detailed balance holds for such an elementary transition $n \to n-1$. The Rydberg levels satisfy the requirement

$$\delta E = E_n - E_{n-1} \ll T, \quad n \gg 1.$$

In this approximation, with the exception of the restrictions indicated above, there are no restrictions on the magnitude of the field $\mathscr{E}$. Therefore

$$\frac{\beta_{n,n-1}}{\alpha_{n-1,n}} = \frac{n_{n-1}}{n_n} = \frac{(n-1)^2}{n^2} e^{\delta E/T}$$

$$\simeq 1 + \frac{2I_{\mathrm{H}} Z^2}{Tn^3} - \frac{2}{n} - \frac{3\mathscr{E} r_B e}{2ZT} n$$

(see Eq. (3a)). Substituting this expression into Eq. (7) we obtain a linear differential equation for n_n:

$$\frac{\partial n_n}{\partial n} + \left[\frac{2I_{\mathrm{H}} Z^2}{Tn^3} - \frac{2}{n} - \frac{3\mathscr{E} r_B e}{2ZT} n \right] n_n = -\frac{j_n}{\alpha_{n-1,n} n_e}. \tag{8}$$

Let

$$F = -\frac{I_{\mathrm{H}} Z^2}{Tn^2} - \ln(n^2) - \frac{3\mathscr{E} r_B e}{4ZT} n^2$$

$$= -\frac{E_n}{T} - \ln(n^2) - \frac{3\mathscr{E} Z e^3}{8E_n T}.$$

The solution of Eq. (8) is

$$n_n = e^{(-F)} \left[C_1 - \int_{n_0}^{n} \frac{j_n}{\alpha_{n-1,n} N_e} e^{(F)} dn \right].$$

The constants C_1 and n_0 must be determined from the boundary conditions. For large n (in the presence of a field there is a maximum bound state $n_{\max}$; the number $n_{\max}$ must satisfy only the condition which was imposed above, $n_{\max} \gg 1$) the Boltzmann–Saha condition is satisfied:[11]

$$n_n \to n^2 \exp\left(\frac{E_n}{T} \right) \left(\frac{\hbar^2}{2\pi m T} \right)^{3/2} n_e n_i.$$

Here n_i is the ion density. Then $C_1 = 0$ and

$$\exp\left(\frac{3EZe^3}{8TE_{n_{max}}} \right) \int_{n_0}^{n_{max}} \frac{j_n}{\alpha_{n-1,n} N_e n^2} \exp\left[-\frac{I_{\mathrm{H}} Z^2}{Tn^2} - \frac{3\mathscr{E} r_B e}{4ZT} n^2 \right] dn = -\left(\frac{\hbar^2}{2\pi m T} \right)^{3/2} n_e n_i. \tag{9}$$

The number n_0 can be set equal to zero, just as in Ref. 11, since the integral obtained here is a typical integral that can be calculated by the steepest-descent method. The preexponential expression in the integrand can be removed from the integral for values of n determined by the maximum of the argument of the exponential function. In our case this maximum is reached for $n = n_{\max}$, and it is sufficient to remove the quantity $j_{n_{\max}}$ from the integral (for large n the quantity $\alpha_{n-1,n} n^2$ does not depend on n (Ref. 11)). For sufficiently large $n_{\max}$ it can likewise be replaced in the upper limit of integration by infinity, since the error so introduced will be exponentially small because of the second term in the exponential in Eq. (9). Now, switching from integration over the quantum number to integration over energy, we can return to the form (5) (see Refs. 1 and 2):

$$n_e j_{n_{\max}} = b = b_0 \exp\left(-\frac{3\mathscr{E} Z e^3}{8TE_{n_{\max}}} \right) \int_{-\infty}^{0} |E|^{3/2} \exp\left(\frac{E}{T} + \frac{3\mathscr{E} Z e^3}{8ET} \right) dE \left[\int_{-\infty}^{0} |E|^{3/2} \exp\left(\frac{E}{T} \right) dE \right]^{-1}. \tag{10}$$

Finally, the expression obtained must be averaged over the distribution of the energy $E_{n_{\max}}$, for which, just as for E_n, the Boltzmann law is valid. Setting $q = 3Z\mathscr{E} e^3/8T^2$ and calculating the required integrals we have

$$b = b_0 \frac{2\sqrt{q}\, K_1(2\sqrt{q}) \exp(2\sqrt{q})}{1 + 2\sqrt{q} + 4q/3}, \tag{11}$$

where $K_1(x)$ is a modified Bessel function of order 1. This expression is limited with respect to the magnitude of the field only in the sense that the condition $n \gg 1$ holds, i.e., its applicability is wider than that of the expression (5) (we recall that the latter expression is applicable only for $ZEe^3/T^2 \ll 1$). As $E \to 0$ the quantity q also approaches zero and $2\sqrt{q} K_1(2\sqrt{q}) \simeq 1 + O(q)$, so that when the numerator and denominator in Eq. (11) are expanded in a series the term of order $\sqrt{q}$ cancels, and the leading term of the expansion in powers of the field will be (essentially) linear, just as in Eq. (6),

$$b \simeq b_0 \{1 - q[\ln(1/q) + 17/6 - 2\gamma]\}. \tag{11a}$$

Here γ is the logarithm of the Euler constant. However, instead of the 2 multiplying the expression $Z\mathscr{E} e^3/T^2$ in Eq. (6), the same coefficient of proportionality in the expansion of the general expression becomes logarithmic; it depends weakly on the field.

We shall now make some estimates for hydrogen ($Z = 1$). Let the electron density equal the ion density, $n_e = n_i = 10^{18}$ cm^{-3}, $T = 1$ eV; we take for $\mathscr{E}$ the average ion microfield for the Holtsmark distribution $\mathscr{E}_{\mathrm{H}} = 8.8 e n_i^{2/3}$. Then $q = 0.0626$ and $b \simeq b_0(1 - 0.279) = 0.721 b_0$. We note

that a calculation using the simple formula (6) gives a somewhat lower value $b \simeq 0.665 b_0$. Therefore the recombination rate decreases by almost a factor of 1.5 even for such densities! Let us now estimate the decrease in the recombination rate for the experimental conditions of Ref. 14 and Ref. 15 for which recombination occurred into a state of hydrogen-like lithium, $Z=3$, $n_i \simeq 1.67\times 10^{18}$ cm^{-3}, and the average electron temperature T was estimated as 3 eV. Then $b = b_0(1-0.153) = 0.847 b_0$, i.e. approximately 15% lower. The formula (6) gives virtually the same result for b: $b_0(1-0.157) \simeq 0.843 b_0$.

In the limit $q \to \infty$ we have

$$2\sqrt{q}\,K_1(2\sqrt{q})\exp(2\sqrt{q}) \to \sqrt{\pi\sqrt{q}},$$

$$b \to 3 b_0 \sqrt{\pi}\,(3Z\mathscr{E}e^3/8T^2)^{-3/4}/4.$$

If a decrease of the three-particle electron–ion recombination rate by a factor e is taken as the condition for closure of this channel, then closure occurs when

$$Ze^3\mathscr{E}_H/T^2 \simeq 1.034. \tag{12}$$

It is easy to see that it corresponds to the criterion $\Gamma = Z^2e^2/Td_i > 1$ for a noniedal plasma (d_i is the average distance between ions: $\mathscr{E}_H = 8.8 e n_i^{2/3} = 3.88 e/d_i^2$, so that $Ze^3\mathscr{E}_H/T^2 \simeq 3.277\Gamma^2 \geqslant 1$ or $\Gamma \geqslant 0.55$. Therefore the three-particle recombination channel (according to Pitaevskiĭ) under discussion is essentially closed in a nonideal plasma.

Let us now examine briefly the effect of a magnetic microfield on this recombination rate. The Rydberg levels are split by the Zeeman effect. In this case an electron is much more likely to be captured into states with negative energy of smaller absolute magnitude. The analog of the expression (3) for the energy difference between neighboring states $n+1$ and n will be[1]

$$\Delta E \simeq \frac{2I_{\rm H}Z^2}{n^3} + \frac{\hbar e\mathscr{H}}{mc} = \frac{2I_{\rm H}Z^2}{n^3}\left(1 + \frac{\hbar n^3 e\mathscr{H}}{2Z^2 I_{\rm H} mc}\right)$$
$$= \frac{2I_{\rm H}Z^2}{n^3}\left(1 + \frac{Z\sqrt{I_{\rm H}}\,e\mathscr{H}\hbar}{2mc|E|^{3/2}}\right). \tag{13}$$

Here $\mathscr{H}$ is the microfield produced by the plasma electrons, in which recombination corresponds to the conditions of quasistationariness. The expression for the relative change in energy as a function of the microfield actually does not contain $\hbar$, i.e., it is a classical number:

$$\frac{Z\sqrt{I_{\rm H}}\,e\mathscr{H}\hbar}{2mc|E|^{3/2}} = \frac{Ze^3\mathscr{H}}{2\sqrt{2m}\,c|E|^{3/2}}.$$

Therefore the relative increment in the energy is less by the factor $\sqrt{|E|/32mc^2}$ than the corresponding increment for an electric microfield even if both microfields have the same amplitude, which happens only under certain (extreme) conditions. The recombination rate in a magnetic microfield can be written exactly by analogy to the electric microfield:

$$b = Z^3\pi^{3/2}e^6\left\{4T^{3/2}\int_{-\infty}^{0} e^{E/T}\left[\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle_0 \times\left(1+\frac{Ze^3\mathscr{H}}{2\sqrt{2m}\,cE^{3/2}}\right)\right]^{-1} dE\right\}^{-1}. \tag{14}$$

After integrating we obtain

$$b = b_0\left(1 + \frac{\sqrt{2}\,Ze^3\mathscr{H}}{3\sqrt{\pi mc^2T^3}}\right). \tag{15}$$

We shall now make estimates for hydrogen ($Z=1$) in the case that the plasma is produced by a short intense laser pulse. The electron temperature T_e of the plasma reaches 1 keV. Then we have $\mathscr{H} = 5.2\mathscr{H}_H$, the average magnetic microfield for the Holtsmark distribution, where $\mathscr{H}_H = 8.8 e n_i^{2/3}\sqrt{2T_e/mc^2}$, $n_e = n_i = 10^{18}$ cm^{-3}, and finally $b = b_0(1+0.376\times 10^{-10})$, i.e., the changes induced in the recombination rate by the magnetic microfield are very small (since we assumed the electron temperature to be high). At low temperature $T_e = 1$ eV and $\mathscr{H} = \mathscr{H}_H$ the increment is four orders of magnitude larger but still small compared with the decrease in the electric microfield.

In summary, the magnetic microfield in itself cannot strongly influence three-particle electron ion recombination via highly excited states. However, it can (somewhat) speed-up this recombination for high electron temperatures in the plasma. The electric microfield, on the other hand, can substantially decrease the rate of three-particle electron–ion recombination via highly excited states of the ion even with charge densities in the plasma $\sim 10^{18}$ cm^{-3} and comparatively low temperatures of the order of several electron volts. In a nonideal plasma this three-particle recombination channel closes completely.

3. THREE-PARTICLE RECOMBINATION IN STATES OF NON-HYDROGEN-LIKE IONS

One would think that as soon as diffusion of the recombining electron along highly excited states becomes the bottleneck that determines the rate of the recombination process under study the non-hydrogen-like nature of the ion should be of no consequence, since the arrangement of the Rydberg levels of any ion is very close to the arrangement of the levels of the equivalent hydrogen-like ion. However, this difference still exists, the recombining electron should feel the presence of the strongly charged nucleus simply because the electron wave function at the location of the nucleus is different from zero. An electron is seemingly additionally (compared with the situation in the equivalent hydrogen-like ion) attracted to the nucleus, i.e., the recombination rate in a non-hydrogen-like ion should be higher than the recombination rate in the equivalent hydrogen-like ion.

Thus, we must take account of the influence of the nucleus on the shift of the energy levels. This can be done by the Thomas–Fermi method modified for ions.[16,17]

Let the nuclear charge in the problem at hand be N, and let the ion charge once again be Z. The Thomas–Fermi equation for ions is

$$\Delta(\phi-B)=\frac{4e}{3\pi\hbar^2}[2me(\phi-B)]^{3/2}. \tag{16}$$

Here ϕ is the desired potential, the constant B is determined from the boundary conditions (for a neutral atom $B=0$). Switching to dimensionless quantities Φ and x

$$e(\phi-B)=\frac{Ne^2\Phi}{r}, \quad r=\frac{xr_B(3\pi)^{2/3}}{2\cdot 4^{2/3}N^{1/3}}, \tag{17}$$

we obtain the Thomas–Fermi equation in the standard form

$$\sqrt{x}\,\Phi''=\Phi^{3/2} \tag{17a}$$

with boundary conditions ($r=R$, $x=x_0$)

$$\Phi(x_0)=0, \quad \Phi(0)=0, \quad x_0\Phi'(x_0)=-N/(N-Z). \tag{18}$$

We now perform Bohr–Sommerfeld quantization, treating the difference between the real potential in which the recombining electron moves and the potential of the equivalent hydrogen-like ion as a correction. The Bohr–Sommerfeld conditions are

$$\pi\hbar\left(n_r+\frac{1}{2}\right)=\sqrt{2m}\int_b^a dr \times\sqrt{-|E|+\frac{e\varphi(r)}{r}-\frac{\hbar^2 l(l+1)}{2mr^2}}. \tag{19}$$

Here $\varphi(r)=Ze\Phi(r)$ and l is the orbital quantum number. We rewrite $\varphi(r)$ in the form $\varphi(r)=(N-Z)[1+f(r)]$, where

$$f(r)=\begin{cases} Z/(N-Z), & r=0,\\ 0, & r=R,\end{cases}$$

a and b are classical turning points

$$a,b=\frac{e^2(N-Z)}{|E|}\left[1\pm\sqrt{1-\frac{2|E|\hbar^2 l(l+1)}{me^4(N-Z)^2}}\right].$$

For $a\gg R$ and $b\ll R$ we can simplify the radicand in Eq. (19) as

$$\sqrt{(a-r)(r-b)}\to\sqrt{ar}\,.$$

In this case the Bohr–Sommerfeld conditions reduce to the expression

$$\sqrt{\frac{I_{\rm H}^{N-Z}}{E}}=n_r+\frac{1}{2}+\sqrt{l(l+1)} -\frac{\sqrt{m}\,e^2(N-Z)}{\pi\hbar\sqrt{2b|E|}}\int_b^R\frac{f(r)dr}{\sqrt{r}}\,, \tag{20}$$

and for the inner turning point

$$b\simeq l(l+1)r_B^{N-Z}.$$

Here r_B^{N-Z} is the Bohr radius of the equivalent hydrogen-like ion, $r_B^{N-Z}=r_B/(N-Z)$. Collecting all quantum numbers into a single principal quantum number we obtain from Eq. (20)

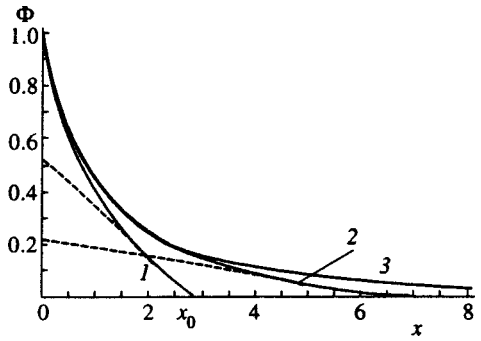

FIG. 1. Solutions of the Thomas–Fermi equation (17a) for two types of ions (curves *1*, *2*) and a neutral atom (curve *3*): *1* — $Z=0.5N$; *2* — $Z=0.21N$. The tangents to the curves at $\Phi=0$ intercept on the ordinate the quantity Z/N.[12]

$$\sqrt{\frac{I_{\rm H}^{N-Z}}{E}}=n-\frac{1}{2\pi\hbar\sqrt{r_B^{N-Z}}}\int_{r_B^{Z-N}}^R\frac{f(r)dr}{\sqrt{r}}\,. \tag{21}$$

Finally, we make a decisive simplification, taking only the first nonzero term in the Taylor series expansion of $f(r)$ at $r\simeq R$. It is obvious that on account of Eq. (17) or (17a) the second derivative here equals zero and this series starts with the cubic terms:

$$f(r)=aZ^2(R-r)^3/(N-Z)\beta^2.$$

Here $\beta=(3\pi/4)^{2/3}r_B/2\simeq 0.8853r_B$. The formula (1) makes it possible to obtain an expression for quantum defects. For the S states we have $l=0$ and

$$\sqrt{\frac{I_{\rm H}^{N-Z}}{E}}=n_0-\frac{16\sqrt{0.8853}}{35\pi}\sqrt{N-Z}\,N^{5/6}x_0^{7/2}=n_0-\delta_0. \tag{22}$$

If the ion charge is taken to be half the nuclear charge $Z=N/2$, then $x_0=2.75$ holds (see Fig. 1) and the quantum defect δ_0 of these states will depend on the nuclear charge as

$$\delta_0\simeq 0.161N^{1/3}.$$

For example, let us examine how the theory developed above describes the ionization potential of the helium atom—the standard test for all such calculations. Here $N=2$ and the ionization energy is

$$I_{\rm He}=I_{\rm H}/(1-\delta_0)^2\simeq 22.8\ \text{eV}, \tag{23}$$

which is 1.8 eV less than the true value. This accuracy is adequate for our purposes, since we did not intend to study deep-lying states.

We shall study recombination into a state of a non-hydrogen-like ion just as we did above, introducing the quantities A and $\langle\partial(\delta E)^2/\partial t\rangle$ and calculating their increments:

$$A=A_0+\Delta A,$$

$$\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle=\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle_0+\Delta\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle.$$

The normalization condition is now

$$A(E)\int \delta\left[-|E|-\frac{p^2}{2m}+(N-Z)e^2\frac{1+f(r)}{r}\right]d\Gamma=1, \qquad (24)$$

or after expanding in a series in powers of Er_B/e^2

$$\frac{1}{A}=\frac{1}{A_0}+4\pi^2\sqrt{(2m)^3(N-Z)e^2}\int_0^R\sqrt{r^3}\,f(r)dr$$

$$\simeq\frac{1}{A}\left[1+0.026\sqrt{\frac{x_0^{11}Z(Er_B/e^2)^5}{N-Z}}\right]. \qquad (24a)$$

For transitions $\Delta l=1$ we obtain from Eq. (22)

$$\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle=\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle_0(1+\delta_l-\delta_{l-1})$$

$$\simeq\left\langle\frac{\partial(\delta E)^2}{\partial t}\right\rangle_0\left[1+\frac{N}{\pi(Z-N)}\right]. \qquad (25)$$

For $Z<N-Z$ the final expression for the recombination rate of a non-hydrogen-like ion, taking account of the thermal velocity distribution of the electrons, is

$$b=b_0\left[1+\frac{Z}{\pi(N-Z)}+0.265\sqrt{\frac{x_0^{11}N(Tr_B/e^2)^5}{N-Z}}\right]. \qquad (26)$$

Here b_0 is the recombination rate for the corresponding equivalent hydrogen-like ion. The second term in brackets is the leading-order correction. For sufficiently strong ionization (but for $Z<N-Z$) the recombination rate of a real ion can be tens of percent higher than the recombination rate of the corresponding equivalent hydrogen-like ion. This is important for interpretation of experiments on recombination pumping of x-ray lasers, especially when the plasma is produced by short pulses of a strong electromagnetic field.[14,15,18]

4. CONCLUSIONS

In summary, the random electric microfield of a plasma decreases the rate of three-particle electron–ion recombination via highly excited states of a hydrogen-like ion even with charge densities $\sim 10^{18}$ cm^{-3} in the plasma and comparatively low temperatures on the order of several electron volts. As plasma density increases further with the plasma temperature remaining constant, this microfield essentially closes this recombination channel completely, and recombination into strongly bound states of the ion becomes the only possibility. The criterion determining whether or not the first process is forbidden is approximately the same as the criterion for a nonideal plasma, i.e., in a nonideal plasma only the latter three-particle recombination process remains.

The magnetic microfield in itself cannot strongly influence the process of three-particle electron–ion recombination via highly excited states. However, it can speed-up this recombination somewhat at high electron temperatures in the plasma.

The recombination rate for a real ion can be tens of percent higher than that of the corresponding equivalent hydrogen-like ion. This is important for the interpretation of experiments on recombination pumping of x-ray lasers, especially when the plasma is produced by a short pulse of a strong electromagnetic field.

I thank W. Ebeling, A. Förster, D. Beule, and V. P. Shevel'ko for valuable discussions of the results obtained. This work was supported by INTAS (Grant 94-1937), the Russian Fund for Fundamental Research jointly with the German Scientific and Technical Society (Grant No. 98-02-04116), and the A. Humboldt Foundation.

*)E-mail: slon@kapella.gpi.ru

1)Most transitions taken into account in the calculations occur, just as in Ref. 11, with a unit change of the magnetic quantum number and with the orbital quantum number decreasing by unity (see Ref. 13), so that the energy difference in the presence of a magnetic field increases and the recombination rate in a magnetic microfield increases.

[1] L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **42**, 1326 (1962) [Sov. Phys. JETP **15**, 919 (1962)].
[2] A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **46**, 1281 (1964) [Sov. Phys. JETP **19**, 870 (1964)].
[3] E. Ya. Kononov and K. N. Koshelev, Kvant. Elektron. **1**, 2411 (1974) [Sov. J. Quantum Electron. **4**, 1340 (1975)].
[4] A. V. Borovskiĭ and A. L. Galkin, *Laser Physics* [in Russian], Izdat, Moscow, 1995.
[5] Y. Hahn in *Atomic and Molecular Processes in Fusion Edge Plasmas*, edited by R. Yanev, Plenum Press, New York, 1995.
[6] J. Li and Y. Hahn, Phys. Rev. E **48**, 2934 (1993).
[7] A. Mueller, D. S. Belic, B. D. DePaola *et al.*, Phys. Rev. Lett. **56**, 127 (1986).
[8] V. L. Jacobs, J. Davis, and P. S. Kepple, Phys. Rev. Lett. **37**, 1390 (1976).
[9] K. LaGattuta and Y. Hahn, Phys. Rev. Lett. **51**, 558 (1983).
[10] A. V. Gurevich, Geomagn. Aeron. Geomagn. Aeronom. **4**, 3 (1964).
[11] Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Academic Press, N. Y., 1966, 1967, Vols. 1, 2 [Russian original Nauka, Moscow, 1966].
[12] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, N. Y., 1977 [Russian original, Nauka, Moscow, 1968].
[13] H. Bethe and E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, Academic Press, N. Y., 1957 [Russian translation, Fizmatlit, Moscow, 1960].
[14] D. C. Eder, P. Amendt, L. B. DaSilva *et al.*, Phys. Plasmas **1**, 1744 (1993).
[15] Y. Nagata, K. Midorikawa, M. Obara *et al.*, Phys. Rev. Lett. **71**, 3774 (1993).
[16] P. Gombás, *Die Statistische Theorie des Atoms und Ihre Anwendungen*, Springer-Verlag, Vienna, 1949 [Russian translation, Inostr. Lit., Moscow, 1951].
[17] H. J. Brudner and S. Borowitz, Phys. Rev. **120**, 2053 (1960).
[18] B. N. Chichkov, A. Egbert, H. Eichmann *et al.*, Phys. Rev. A **52**, 1629 (1995).

Translated by M. E. Alferieff

Shock compression and thermodynamics of highly nonideal metallic plasma

V. K. Gryaznov*) and V. E. Fortov

Institute of Chemical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

M. V. Zhernokletov, G. V. Simakov, R. F. Trunin, and L. I. Trusov

All-Russia Research Institute of Experimental Physics, 607200 Sarov, Nizhnii Novgorod Region, Russia

I. L. Iosilevskiĭ†)

Moscow Institute of Physics and Technology, 141700 Dolgoprudnyi, Moscow Region, Russia
(Submitted 12 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1242–1265 (October 1998)

Using experimental data on compression and heating of dense metallic plasma by powerful shock waves, we have analyzed the effect of strong Coulomb interaction on both discrete and continuum bands of energy spectrum, the role of short-range repulsion, and the effect of degeneracy on the equation of state for a dense, nonideal metallic plasma. Explosive devices have been used to produce plasma for which the degree of ionization, nonideal parameter, and degeneracy varied over wide ranges. In order to increase effects of irreversible energy dissipation, metal targets of low densities have been used. Thermodynamic measurements have been compared to theoretical models taking into account Coulomb interaction, short-range repulsion, and degeneracy of electrons. The plasma models have been shown to be applicable to the equilibrium properties of multiply ionized plasma in a wide region of the phase diagram characterized by extremely high parameters [$T \geqslant 10^4$ K, $P \geqslant 10$ GPa, and $\rho = (0.1-1)\rho_0$], which is beyond the traditional domain of plasma physics. © *1998 American Institute of Physics.* [S1063-7761(98)00810-5]

1. INTRODUCTION

To date it has been found that the physical properties of plasmas can be described theoretically in the limiting cases of high temperatures and/or ultrahigh densities, when the interaction between plasma particles is weak, which allows researchers to use classical (Boltzmann statistics)[1] or quasiclassical (Fermi statistics) self-consistent field methods.[2] The region of intermediate parameters is characterized by highly uncertain theoretical predictions, whereas it is of great practical interest[3] in view of applications to power generation (inertial thermonuclear synthesis, MHD generators, etc.), astrophysics (structure of giant planets, brown dwarfs, etc.), and other applications to specific areas. For this reason, one of the most interesting and complicated problems of modern plasma physics is the investigation of plasma properties under conditions of strong interparticle interaction, which makes difficult application of conventional theoretical techniques (such as the perturbation theory, computer simulations, etc.) and requires extremely thermal energy high densities to be generated in physical experiments.[4] The experimental search for phase transitions in nonideal plasma predicted by heuristic models and detected recently in dust and colloidal plasmas[5] is of special interest.

Until now the bulk of experimental data concerning properties of strongly nonideal plasma has been obtained using dynamic techniques,[4,6] when materials are compressed and irreversibly heated by powerful shock waves. Explosive devices and pneumatically driven projectiles in such experiments allow researchers to test and analyze theoretical models describing thermodynamic, electrodynamic, and optical properties of nonideal plasma composed of cesium, noble gases, and hydrogen under highly collisional conditions ($\Gamma_D = \sqrt{4\pi(e^2/k_BT)^3 \Sigma n_\alpha Z_\alpha} = 0.1-5$ is the Coulomb interaction parameter), when $n_e \lambda\!\!\!^{-3}_e \leqslant 0.1$ holds (here $\lambda\!\!\!^{-}_e = \sqrt{2\pi\hbar^2/m_e k_B T}$ is the electron de Broglie wavelength), the electron gas statistics is Boltzmann, and the ionization degree of such a plasma $\alpha = n_e/(n_a + n_i)$ is within 1–2. Plasma beyond this region, namely, a multiply ionized and partly degenerate plasma, can be described by using experimental data on compression of solid and porous materials by shocks with pressure amplitudes of hundreds of thousands to millions of atmospheres. Thus far a considerable amount of experimental data concerning dynamic compression of metals (see the compendia in Refs. 7–9 and references therein) by shocks generated by chemical[6,10,11] and nuclear[12] explosives, pneumatically driven projectiles,[13] and in recent time by concentrated laser,[14] X-ray,[15] and ion[16] beams has been accumulated.

The data on shock compression, supplemented with measurements of unloading adiabats in shock-compressed metals, form the basis for constructing equations of state,[17] which is done by selecting optimal constants in functional thermodynamic relationships derived from simplified thermodynamic models. At the same time, even under moderate pressures of 100 to 200 GPa metals melt, and under an in-

creasing pressure they are sequentially ionized by temperature and pressure. Thus a dense, disordered system of multiply ionized particles, i.e., an electron–ion plasma with a complicated spectrum of intense interactions among its particles, is generated.

For this reason, the shock-compressed metal plasma appears to be of interest for testing theoretical models of plasma, in view of both the search for phase transitions in plasma[18] and the analysis of various models of nonideal features of plasmas in discrete and continuum bands of their energy spectra, and the effects of bound electrons and their statistical properties on plasma thermodynamics at high energy densities. The aim of these studies is, in reality, to extend the application domain of plasma models[19] to the region of densities corresponding to condensed states and of megabar pressures, which has not been studied by traditional methods of plasma physics until recently and where the researchers have used either semiempirical approximate equations of state[17] or far-fetched extrapolations of quasiclassical approximations.[20] Moreover, thermodynamic measurements of this kind in the region of the metal–dielectric transition might be used in testing the hypothesis[21] of a relation between metallization and the first-order phase transition in disordered materials.

This paper presents experimental data on generation of nonideal plasma of copper, nickel, and iron through compression and irreversible heating of porous samples exposed to powerful shock fronts in either plane or converging configuration. Measurements of shock velocities in tested and reference samples allow us to calculate, using conservation of mass, momentum, and energy, thermodynamic parameters of shock-compressed plasma and compare them to results of theoretical models of multicomponent nonideal plasmas.

2. GENERATION AND DIAGNOSTIC OF PLASMA

The interesting range of parameters of strongly nonideal plasmas corresponds to material densities lower than the densities ρ_0 of solids and energies per particle higher than the binding energies of atoms and molecules in solids ($\sim$1 eV per particle). In generating such states of metals, we used shock compression of finely dispersed (porous) metals, which enhanced the effects of irreversible energy dissipation on the shock discontinuity front and in higher plasma overheats. Powders of metals containing at least 99.5% of the host material were tested. The samples were cylindrical pellets with a diameter-to-thickness ratio higher than 2.5 and the required density. They were pressed from powders or fabricated from metal hydrides using a dedicated technique. The high diameter-to-thickness ratio was needed to limit effects of perturbations generated on the sample side surface on the shock front.

Samples of relatively low porosities, $1<m=\rho_0/\rho_{00}<4$ (where ρ_0 is the initial density of the solid metal and ρ_{00} is the density of a sample), were manufactured by pressing powders in special molds or, for $m\geqslant 4$, immediately in a measuring cell. The powder particles were smaller than 100 μm and in some experiments were varied to estimate the effect of the grain size, which proved to be negligible in our experiments. A higher porosity ($4<m<8$) was achieved in samples fabricated from hydrides of these metals and subsequently dehydrated. Samples of the highest porosity ($m>8$) were fabricated from fine powders with grain sizes of several (2–3) hundreds of angströms obtained by stripping very fine particles from surfaces of melted metals by a high-speed helium jet. The oxide content in the powders was checked. In all cases it was a fraction of one percent. The size of powder particles was selected taking into account the condition of their uniform heating. This circumstance and some technological conditions limited the porosity in our experiments to $m\leqslant 20$.

The dynamic diagnostic methods of shock-compressed states rely on the conservation of mass, momentum, and energy on the front of a plane stationary discontinuity[10]:

$$V=\rho^{-1}=\frac{D-U}{\rho_{00}D},\quad P=\rho_{00}DU,\quad E=E_0+\frac{U^2}{2}. \qquad (1)$$

Measurements of shock D and particle U velocities in experiments allow the researchers to calculate thermodynamic quantities, namely, the pressure P, specific volume V, and specific internal energy E of a shock-compressed material.

Shock velocities in porous targets shaped as pellets of thickness 3–4 mm were measured using electrical contacts to detect shock fronts. At pressures beyond 10 GPa, we used electrically insulated pins with a diameter of 0.14 mm fabricated from the PÉL-14 enamel-coated wire. At lower pressures, when such shock detectors are not reliable, measurements were performed using piezoceramic probes. The explosive measuring devices and configuration of shock detectors placed on samples allowed us to measure the shock velocity with an uncertainty of 1–1.5%, moreover, we could check the shock front symmetry in samples using the time differences among signals from different probes placed on the same level (on the same plane of a sample).

In our experiments, we used explosive generators of three types. The first includes so-called contact devices,[22] shown schematically in Fig. 1, which were used in experiments at low compression. After the plane detonation front emerges at the free surface of an explosive, explosion products ''softly'' expand across an air gap, hit against a metallic striker plate, and generate a quasi-stationary shock wave in it. In this process, the striker velocity becomes approximately constant. By varying the composition and thickness (mass) of the explosive charge, one can measure parameters of shock-compressed porous materials over a fairly wide pressure range.

Shocks in shields of devices of another type were generated by hitting them with relatively thin metal (aluminum, iron) plates of thickness 1.5–4.0 mm accelerated to velocities of 4.0 to 6.5 km/s. A diagram of such a device is given in Fig. 2. In this device, explosion products drive an aluminum striker[23] of thickness 4 mm pressed into a steel disk of the same thickness. A detonation wave arriving on the metal surface generates in the steel disk a pressure higher than in aluminum, thereby generating an additional thrust on the edges of the aluminum disk and equalizing the flying striker. In other devices of this type, steel plates with thicknesses of 1.5 to 2.2 mm were used. The pressure in such devices is

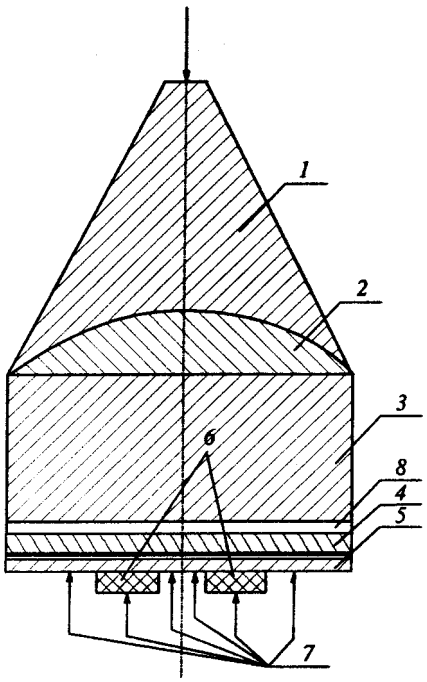

FIG. 1. Diagram of the facility with contact striker: *1* — high-explosive lens; *2* — inertial equilibrating support; *3* — cylindrical explosive charges; *4* — striker; *5* — shield; *6* — tested samples; *7* — electric pins; *8* — 5-mm air gap.

approximately twice that in plane contact devices (Figs. 1 and 2). The highest pressures could be generated in facilities with spherical converging strikers (''shells'') accelerated by explosion products of a converging detonation wave.[12,24] Figure 3 shows a diagram of such a cumulative device. The striker in this case is a steel shell about 3 mm thick, whose

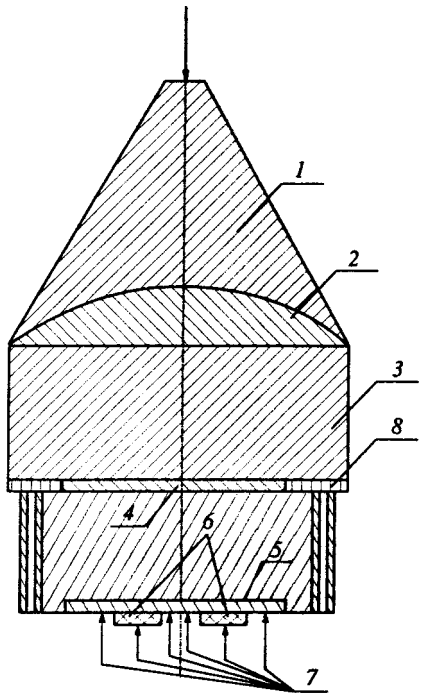

FIG. 2. Diagram of the projectile generator of shock waves: *1–7* — the same as in Fig. 1; *8* — steel disk.

velocity gradually increases as it converges to the center. By placing targets with the sample to be tested at various distances from the center, one can generate various pressures in them. The specific feature of these devices is the time dependence of the propagating shock wave parameters owing to the spherical symmetry. This circumstance impelled us to introduce small (about 1%) corrections due to the difference between the shock convergence parameters in the shield and sample.

In all the devices discussed above, the shields were made from aluminum. This material was selected because its Hugoniot in $P-U$ coordinates (where the problem of decay of the discontinuity at the shield–sample interface is solved by the reflection technique[25]) is close to those of porous metals and, as was shown by Bugaeva *et al.*,[26] the specular image of the shock Hugoniot in the $P-U$ plane can be used in calculating compression parameters in such powders. This makes calculation of the particle velocity behind the shock front and other shock compression parameters notably easier, more reliable, and less uncertain. In accordance with our data reduction routine, each value of the shock velocity measured in each experiment is the mean of three to eight independent measurements (the number of experiments at lower pressures was larger). Therefore the relative rms error of the shock velocity in all sets of measurements was less than 1–1.5%. The uncertainty of the particle velocity in the sample materials was in the same range. The error in the determining the amount of compression

$$\Delta\sigma=\Delta(\rho/\rho_{00})=\sigma(m\sigma-1)(|\Delta D/D|+|\Delta U/U|)$$

increases with m and σ and is considerably higher at maximum m, although this error was partially compensated in this region because the compression satisfied $\sigma<1$.

Nonetheless, judging by the good agreement among numerous measurements for various materials, the most probable experimental values were found with fair precision. This is also true for the region of lowest pressures, where measurements are especially susceptible to errors in D and U.

3. EXPERIMENTAL DATA AND THEIR ANALYSIS

The purpose of our measurements was to determine experimentally the compressibilities of porous samples in the region of plasma states that had not been investigated earlier in metals: for nickel, we have studied the region of highest possible porosities ($m=15$ and 20) and shock pressure beyond 50 GPa, which are considerably higher than in earlier experiments;[27,28] for copper, we have obtained data at porosity $m=10$, which is also higher than in earlier publications.[27,29] In addition, we have studied samples with porosity $m=7.2$ and particle sizes of 100 Å for comparison to previous measurements[27] performed at particle sizes of about 100 μm. The measurements of iron compression at $m=5$, 10, and 20 have been compared to and analyzed in combination with the earlier data.[27]

Before proceeding to the analysis of the experimental results, we recall that calibrations were performed earlier[27,28] in order to check whether the experimental data are affected by the sample humidity, thickness, presence of air inside

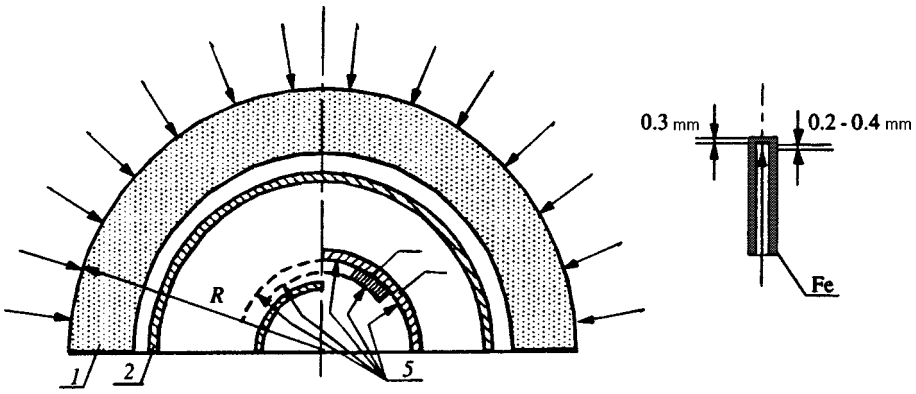

FIG. 3. Diagram of the cumulative shock generator of spherical geometry: *1* — hemispherical high-explosive charge; *2* — steel shell; *3* — aluminum shield; *4* — tested samples; *5* — electric pins.

them, and particle sizes. In those experiments, data obtained on different samples were compared: powders baked in isolation from the atmosphere and powders fabricated in air; samples from which air was pumped out; samples whose thicknesses differed by an order of magnitude; samples whose particle dimensions varied by a factor of 20 to 30. It was found that the data obtained in testing experiments were equal to those obtained in the main series of experiments within the measurement error. Keeping in mind these results, we limited our testing experiments to the investigation of the grain size effect. All new data are listed in Table I. The data for nickel and copper are compared to earlier measurements[27,28] in the $D-U$ coordinates in Fig. 4.

The comparison between different experiments shows that measurement data for a copper plasma obtained for the samples with grain sizes differing by a factor of 10000 (at $m=7.2$) coincide. Hence follows the conclusion that the shock front width, which depends, possibly, on the grain size, has little effect on measurements, and their interpretation is straightforward. The new experimental points for nickel are located on extrapolations of Hugoniots obtained at lower pressures, which is quite natural. The set of measurements given in the graphs support the conclusions given in Ref. 28, which apply, as one can see, not only to nickel, but also to copper, iron, and, possibly, to all the materials studied.

The common properties are the following:

1. The initial sections of the Hugoniots form fans of straight lines with different slopes converging to points (or a narrow interval ΔD) on the ordinate ($U=0$) at a distance of 100–200 m/s from the origin. This value may correspond to the speed of sound in air reduced by the presence of metal particles, which is plausible for Hugoniots in the case of high porosity. This hypothesis runs into difficulties when we analyze Hugoniots $D(U)$ in the case of low porosity, whose initial sections interpolated towards the ordinate also have this property.

2. The slope dD/dU of initial sections of Hugoniots varies over a wide range between approximately 3 (nickel, $m=1.1$) and 1.1 ($m=20$). The slope of the Hugoniot sections corresponding to higher pressures changes two times (this was observed in experiments with Cu and Ni at $m=10$ and 20): first in the regions of relatively small shock velocities $D<8$ km/s (in nickel it changes from 1.7 for low σ to 1.4 for higher values), then the slope becomes approximately equal to the theoretical value of 1.2–1.3 calculated by the Thomas–Fermi self-consistent field model.[20]

3. The change in the dD/dU slope on the initial section means, physically, that in the process of compression all pores have been filled, and the sample density is close (for the given porosity) to the crystal density of the material. A further increase in the pressure should compress the material

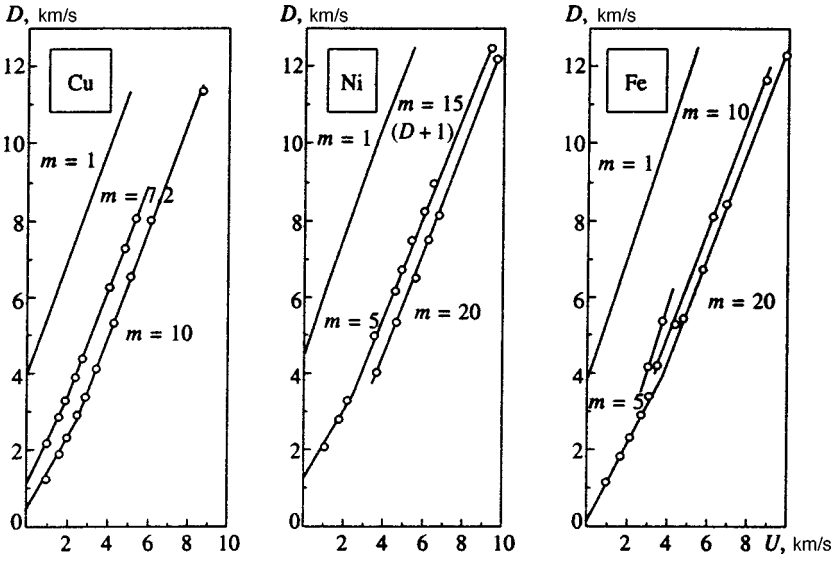

FIG. 4. $D-U$ Hugoniots of copper, nickel, and iron.

TABLE I.

Experiment						Theory [Eqs. (15)–(24)]		
M	D, km/s	U, km/s	P, kbar	ρ, g/cm^3	H, kJ/g	ρ, g/cm^3	T, K	H, kJ/g
				Nickel				
15	11.5	9.36	637	3.18	56.65	3.08	53300	56.97
20	11.9	9.74	516	2.42	61.6	2.48	54000	61.3
				Copper				
7.2	1.21	1.00	15	7.15	−4.57	9.61	4680	−4.6
7.2	1.88	1.58	37	7.86	−3.55	8.71	5796	−3.58
7.2	2.3	1.92	55	7.58	−2.7	8.03	6803	−2.72
7.2	2.91	2.40	87	7.08	−1.158	6.99	8835	−1.15
7.2	3.40	2.75	120	6.49	0.482	6.18	11248	0.527
7.2	5.25	4.04	263	5.38	7.767	5.27	19616	7.81
7.2	6.33	4.79	376	5.09	1.36	5.26	24511	1.35
7.2	7.09	5.31	467	4.94	18.3	5.27	28065	18.0
10	1.21	1.01	10.9	5.30	−4.57	9.57	4680	−4.61
10	1.90	1.64	28	6.36	−3.5	8.52	5880	−3.55
10	2.30	2.00	41	6.92	−2.69	7.78	6860	−2.72
10	2.91	2.52	65	6.66	−1.15	6.58	8900	−1.15
10	3.40	2.90	88	6.01	0.379	5.68	11150	0.422
10	4.15	3.44	127	5.21	3.05	4.86	14830	3.14
10	5.35	4.29	205	4.51	8.47	4.47	20462	8.49
10	6.58	5.08	293	3.92	14.9	4.39	25800	14.5
10	8.02	6.10	437	3.73	25.1	4.35	32880	24.2
10	11.33	8.69	880	3.83	55.5	4.26	51700	54.9
	Experiment					Theory [Eqs. (2)–(14)]		
M	D, km/s	U, km/s	P, kbar	ρ, g/cm^3	H, kJ/g	ρ, g/cm^3	T, K	H, kJ/g
				Iron				
5	4.16	3.09	202	6.10	0.662	6.09	11640	0.628
5	5.38	3.81	322	5.38	5.82	5.41	17400	5.70
10	4.18	3.50	115	4.83	1.09	4.68	13530	1.13
10	5.35	4.39	184	4.37	6.4	3.87	19940	6.67
10	6.62	5.20	270	3.66	13.5	3.59	25900	13.5
10	8.12	6.23	397	3.38	23.7	3.50	33150	23.5
20	1.15	1.03	4.6	3.58	−6.78	6.35	4860	−6.81
20	1.90	1.73	13	4.44	−5.63	7.54	5740	−5.68
20	2.36	2.19	20	5.37	−4.7	6.89	6700	−4.73
20	2.93	2.73	31	5.60	−3.2	5.91	8290	−3.22
20	3.40	3.15	42	5.33	−1.68	5.00	9990	−1.66
20	5.45	4.79	102	3.24	7.14	2.68	19920	7.47
20	6.75	5.73	152	2.59	14.9	2.38	26290	15.1
20	8.44	6.89	228	2.13	27.0	2.25	35150	26.7

along low-compressibility branches of Hugoniots. Depending on the porosity, the slope $dP/d\sigma$ of the compression curves should change from positive values (for $m<2$) through zero (at $m=2-3$) to negative values ($m>3$). All these types of Hugoniots were observed in experiments with iron, copper, and nickel.

4. THERMODYNAMICS OF NONIDEAL METALLIC PLASMAS: COMPARISON WITH EXPERIMENTAL DATA

Before proceeding to the analysis of experimental data, note that their physical interpretation is rather difficult. In dynamic experiments[3,6] the thermodynamic characteristic directly derived from measurements is typically the so-called caloric equation of state which determines the internal energy as a function of thermodynamically conjugated variables, namely the pressure and specific volume, $E(P,V)$. On the other hand, the traditional physical interpretation of nonideal plasma is based on the temperature and related nonideal parameters, electron degeneracy, degree of ionization, etc., thus it requires the knowledge of function $T(P,V)$. The temperature can be derived from the empirical caloric equation of state using Zel'dovich's idea.[30] This procedure was previously performed for cesium plasma.[31,32] Although this approach shows promise at the present time, it cannot be fully implemented in the case of shock-compressed porous materials discussed in this paper. Therefore it seems reasonable to separate the two stages: the first is analysis of directly measured quantities without calculations based on physical models, the second is analysis using specific models of plasma.

In characterizing experimental results as a whole, it is

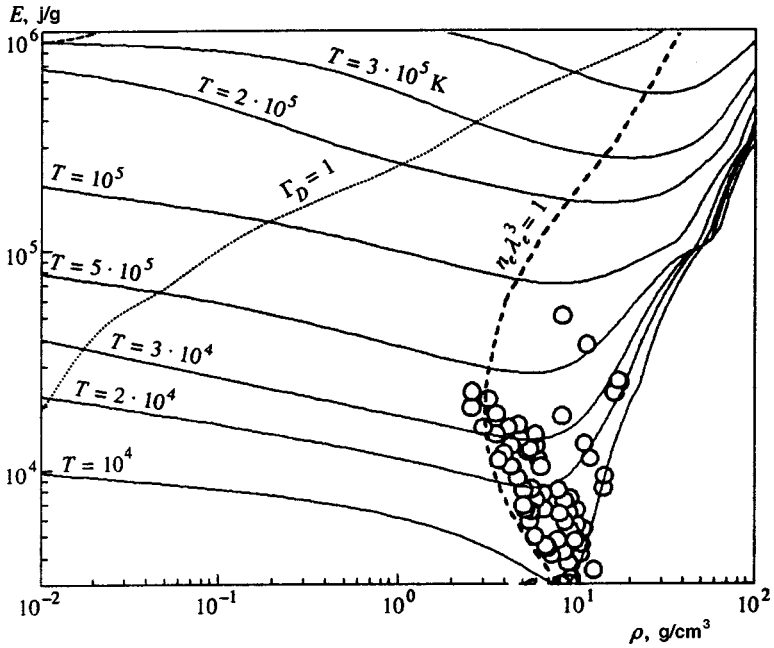

FIG. 5. Density–energy diagram for nickel plasma.[34] The graph shows isotherms and experimental points derived from shock compression measurements of solid and porous nickel samples.[4,28] Lines of constant Coulomb collisionality parameter [$\Gamma_D = \sqrt{4\pi(e^2/k_BT)^3 \Sigma n_\alpha Z_\alpha^2}$] and electron degeneracy parameter $n_e\lambda_e^3$ are given in the diagram.

useful to consider the internal energy E–density ρ diagram supplemented with isotherms of the material calculated using a specific model.[33] Figure 5 shows the $E-\rho$ diagram for nickel.[34] It clearly shows that the entire phase diagram of the material (diagrams for other materials are similar) can be divided into two regions with radically different shapes of thermodynamic functions. The major part is the region of relatively low-density ($\rho \ll \rho_0$) ''gaseous'' plasma, which is characterized by two properties, namely, the gradual decrease in the energy in the process of isothermal compression and clearly seen so-called ''shell oscillations'' of all thermodynamic functions over wide intervals of densities (see for details Ref. 35). For $\rho > \rho_0$ the system is characterized by an abrupt increase in the energy and generalized compressibility factor $Z(n_{\rm nucl}, n_e, T) = P/P^{\rm ideal}$, which is traditionally interpreted as ''ionization by pressure.'' In the high-density limit, the compression leads to the region of states adequately described by the model of mobile nuclei in a weakly nonideal gas of degenerate electrons. The thermodynamics in this region has been fairly accurately described using the well developed technique of cell representations.[2,36–38]

Between the two regions described above, there is a transitional region characterized by minimal values of the internal energy and compressibility factor, and by maximal deviation from the ''weakly nonideal'' condition. The depths and positions of minima on the curves determined by the thermal and caloric equations of state characterize the region where the uncertainty of our knowledge of thermodynamic properties of hot compressed matter is the greatest. The unique feature of shock-compressed porous samples is that they allow us to obtain information about dense, highly nonideal plasma in this interesting region, which is most difficult for interpretation.

The chemical model[1,39–43] is based on the representation of the free energy of a quasi-neutral mixture of electrons, ions, atoms, and molecules as a sum of ideal-gas terms $F_{i,e}^0$ for all components and the terms responsible for all types of interaction among plasma particles, which are treated separately in this model:

$$F \equiv \sum F_i^0 + F_e^0 + F_{ii,ie,ee,\ldots}^* . \tag{2}$$

The statistics of atoms and ions is Boltzmann, and their contribution is

$$F_i^0 = \sum_j N_j k_B T \left(\ln \frac{n_j \lambda_j^3}{Q_j} - 1 \right). \tag{3}$$

Here k_B is the Boltzmann constant, Q_j is the full partition function of atoms or ions of the jth sort, λ_j is the thermal de Broglie wavelength of particles of species j

4.1. Electron degeneracy

The effects of electron degeneracy are important in this region of the phase diagram since the degeneracy parameter of electrons can be up to several units:

$$n_e\lambda_e^3 \approx 1, \quad \lambda_e^2 \equiv 2\pi\hbar^2/m_e k_B T.$$

In the quasi-chemical representation, which separates electrons into two types, namely free and bound, the effect of degeneracy is manifested primarily in the ideal-gas term since it radically changes the density dependence of the pressure and chemical potential. The main effect of the electron degeneracy on the mechanism for nonideal behavior is that it reduces the electrostatic screening by electrons as their degeneracy factor increases. In the limiting case, when the compression is extremely high,[2] electrons are excluded from this mechanism, and the screening of ion charges is described by the so-called ion mixture model, a version of the one-component plasma model. In this study, the electron degeneracy is taken into account mostly by including this effect in the ideal-gas term of the free energy:

$$F_e^0=\frac{2Vk_BT}{\lambda_e^3\sqrt{\pi}}\left[\frac{\mu_e}{k_BT}I_{1/2}\left(\frac{\mu_e}{k_BT}\right)-\frac{2}{3}I_{3/2}\left(\frac{\mu_e}{k_BT}\right)\right], \quad (4)$$

$$\frac{P_e^0}{n_ek_BT}=\frac{2}{3}\frac{E_e^0}{N_ekT}=\frac{2}{3}\frac{I_{3/2}(\mu_e/k_BT)}{I_{1/2}(\mu_e/k_BT)}. \quad (5)$$

Here the electron density n_e and chemical potential μ_e are related by the equation

$$n_e\lambda_e^3=\frac{2}{\sqrt{\pi}}I_{1/2}\left(\frac{\mu_e}{k_BT}\right), \quad I_t(x)=\int_0^\infty\frac{y^t dy}{1+\exp(y-t)}. \quad (6)$$

In the chemical model of plasma, electron degeneracy mainly causes an effective shift of the ionization equilibrium toward low ionization, and also generates an additional term in the equation of state describing an effective repulsion.

4.2. Nonideal effects

The analysis of nonideal effects in this study was performed in two stages. On the first stage, we employed the SAHA-3 computer code, which had been extensively used in applications.[1,34,44–48] It involves the minimum number of steps consistent with achieving satisfactory agreement within experimental uncertainty with data on shock compression of various materials, such as noble gases, cesium, high-porosity metallic samples, etc. Some simplifications of this approach are needed in order to calculate the equation of state together with the gasdynamic calculations, which require a lot of auxiliary calculations.

This simplified approach relies on the following approximations:

1. The effect of electron degeneracy is included only in the ideal-gas term.

2. The effect of the Coulomb nonideal behavior is described by the so-called Debye (ring) approximation for a large canonical ensemble.[44,49]

3. In the calculation of excitation partition functions for atoms and ions, only their ground states are taken into account.

4. The intense short-range repulsion among atoms and ions is taken into account through the approximation of hard spherical shells with essentially different sizes of atoms and ions with different ionization numbers.

5. Atoms and ions are assumed to be ''permeable'' for electrons, i.e., the latter are not affected by the approximation of hard spheres and their density is not affected by the total volume of ions and atoms.

6. The approximation takes into account an additional short-range attraction among atoms and ions, which effectively describes the binding energy of condensed materials.

One advantage of the suggested simplified approach in terms of numerical calculations is that in principle it does not require taking into account the Coulomb nonideal behavior and the partition functions of the excitation, and the system cannot spontaneously lose its thermodynamic stability (the matrix $\|\partial\mu_i/\partial n_j\|$ is positive definite at all compression factors of plasma[1,44]). Special calculations taking into account the ion nonideal behavior described by the ion mixture model[50] and the electron nonideal behavior by the model of interacting electron[51] (it was described by Iosilevskii[52] as a coupled model of a one-component plasma and applied to the thermodynamics of mixed hydrogen and helium plasma[53]) showed that the range of parameters achieved in shock compression of porous samples of most metals are in the region of an anomaly similar to a phase transition (van der Waals loops) predicted by this approximation. The shape and position of this feature essentially depend on the maximum ionization degree allowed by the calculation technique. Such sensitivity of the equation of state to the choice of the approximation selected for describing the Coulomb collisionality is typical for most of the Coulomb collision models suggested in literature for this region of parameters.

4.3. Coulomb interaction

The Coulomb interaction was taken into account in the Debye approximation for a macrocanonical ensemble[49] (in the case of multiple ionization, see Ref. 44):

$$\frac{\Omega}{Vk_BT}\equiv\frac{F-\Sigma N_j\mu_j}{Vk_BT}\equiv\frac{P}{k_BT}$$
$$=\sum_\alpha n_\alpha-\frac{\tilde{\kappa}_D^3}{24\pi}=\sum_\alpha\left[n_\alpha-\frac{\tilde{\Gamma}}{6}\frac{n_\alpha z_\alpha^2}{1+z_\alpha^2\tilde{\Gamma}_D/2}\right]. \quad (7)$$

Here the modified Coulomb collision parameter $\tilde{\Gamma}_D$ expressed in terms of activities is given by the equation

$$\tilde{\Gamma}_D^2=\left(\frac{e^2}{k_BT\tilde{r}_D}\right)^2=4\pi\left(\frac{e^2}{k_BT}\right)^3\sum_\alpha\frac{z_\alpha^2n_\alpha}{1+z_\alpha^2\tilde{\Gamma}_D/2}, \quad (8)$$

where $\tilde{r}_D=1/\tilde{\kappa}_D$ is the modified Debye screening radius. Note that the collision parameter is different from the conventional Debye collision parameter Γ_D:

$$\Gamma_D^2\equiv4\pi(e^2/k_BT)^3\sum n_\alpha z_\alpha^2.$$

This approximation is equivalent to the classical Debye–Hückel approximation in the limit $\Gamma_D\to0$ and differs by the notably smaller corrections in the region of moderate and strong nonideal behavior ($\Gamma_D\geqslant1$).

4.4. Short-range repulsion among ions and atoms

The effect of overlap between electron shells of atoms and ions at high compression degrees leads to a strong short-range repulsion among heavy particles. This effect is taken into account by the model of a mixture of hard spheres. In our numerical calculations, we used the Mansoori formula[54]

$$\frac{\Delta F_{HSM}}{\Sigma_iN_ik_BT}\equiv f_{HSM}(\nu)=X\frac{\nu}{(1-\nu)^2}+3Y\frac{\nu}{1-\nu}$$
$$+(X+1)\ln(1-\nu), \quad (9)$$

$$\nu\equiv\frac{4\pi}{3}n\overline{r^3}, \quad \overline{r^k}\equiv\sum_i n_ir_i^k\Big/\sum n_i, \quad k=1,2,3,$$

$$X=(\overline{r^2})^3(\overline{r^3})^{-2}, \quad Y=\overline{r^2}\,\overline{r}(\overline{r^3})^{-1}. \quad (10)$$

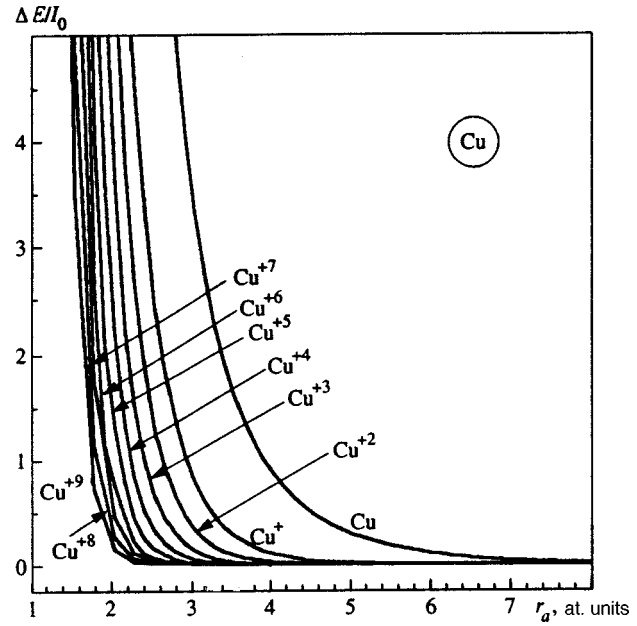

FIG. 6. Shifts of the ground state energies of copper atoms and ions calculated by the Hartree–Fock method.[56,57]

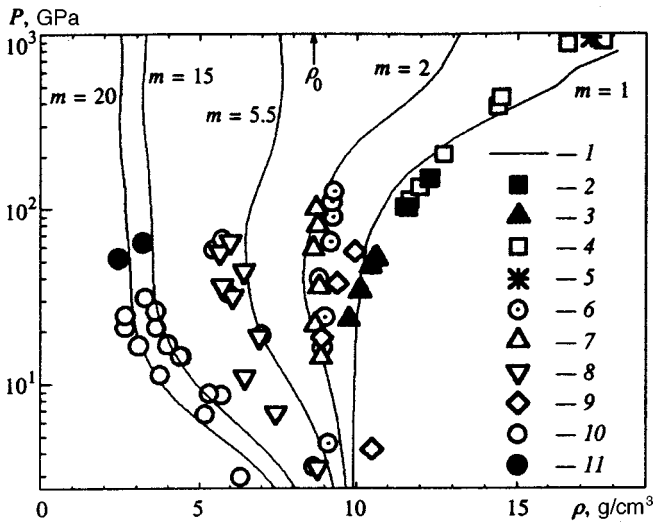

FIG. 7. Hugoniots of porous nickel: *1* — calculations by Eqs. (2)–(14) with the sum in Eq. (14) performed over all heavy particles; *2–11* — experimental data: *2* — Ref. 62; *3* — Ref. 63; *4* — Ref. 64; *5* — Ref. 29; *6* — Ref. 27, $m=2$; *7* — Ref. 27, $m=2.32$; *8* — Ref. 27, $m=5.62$; *9* — Ref. 65; *10* — Ref. 28; *11* — this work.

The contributions of this repulsion to the pressure, energy, and chemical potential derive from Eqs. (9) and (10):

$$\frac{\Delta P_{\rm HSM}}{\Sigma_i n_i k_{\rm B} T}=\frac{\partial f_{\rm HSM}(\nu)}{\partial \nu}, \quad \frac{\Delta \mu_i}{k_{\rm B}T}=f_{\rm HSM}(\nu)+\sum_j n_j \frac{\partial f_{\rm HSM}(\nu)}{\partial n_i},$$

$$\Delta E_{\rm HSM}\equiv 0. \tag{11}$$

In calculating the radii r_j of particles, we used two procedures. The first is based on the so-called confined atom approximation.[1,41] An atom (or ion) is placed in a spherical cell with hard walls, and its electronic structure is calculated by the Hartree–Fock method[55] at a variable cell radius.[56,57] Calculations by this model are given in Fig. 6, which shows energy shifts for the ground states of copper atoms and ions as functions of the cell radius.

Further, the ''effective'' cell radius is calculated using the relation

$$\Delta E(r_i)=\text{const}\cdot I_i,$$

where $\Delta E(r_i)$ is the ground state energy shift, I_i is the corresponding ionization potential, and r_i is the atom cell radius.

A simpler procedure[48,58,59] is based on the assumption that the atom's structure is hydrogen-like. In this case, each ion is characterized by a size related to the ionization potential:

$$r_i\approx r_0[(Z_i+1)I_0]/I_i, \tag{12}$$

where r_0 and I_0 are the atomic radius and ionization potential, r_i and I_i are these parameters for the ith ion, and Z_i is the ion charge.

In real calculations, both these procedures were used only in determination of the ratios between the atomic radius and those of ions, whereas the basic atomic radius was determined using the Ashcroft–Lekner rule,[60] which prescribes determination of the optimum radius of hard spheres corresponding to the normal density from a fixed packing parameter

$$\nu\equiv 4\pi\sum n_i r_i^3/3\simeq 0.45. \tag{13}$$

The radius calculated in this manner leads to the best agreement between the first maximum of the pair correlation function for the system of hard spheres and experimental data obtained for a set of normal metals.[60]

4.5. Additional attraction

Calculations based on the approximation (2)–(13)[34,48,58,59] produced satisfactory agreement with experimental Hugoniots for metals at relatively high shock velocities, i.e., at high pressures and temperatures. The short-range repulsion is of fundamental importance for this agreement between theory and experiment. Our calculations, however, demonstrate that there is a region of relatively low pressures[58,59] corresponding to low shock velocities where the approximation (2)–(13) cannot yield solutions of the Rankin–Hugoniot equation (1) for $D(U)$ with any set of radii $\{r_i\}$. The reason is that the approximation (2)–(13) does not take into account the binding energy responsible for the existence of condensed states of materials. In order to improve the accuracy of extrapolation to the region of low shock velocities, we supplemented Eqs. (2)–(13) with terms effectively taking into account this binding energy in the form

$$\Delta F=\Delta E=-A\left(\sum N_i\right)^{1+\delta} V^{-\delta}, \quad \Delta P=\delta(\Delta E/V), \tag{14}$$

$$\Delta\mu_i=-(1+\delta)V^{-\delta}\left(\sum N_i\right)^{\delta}, \quad A,\delta=\text{const}.$$

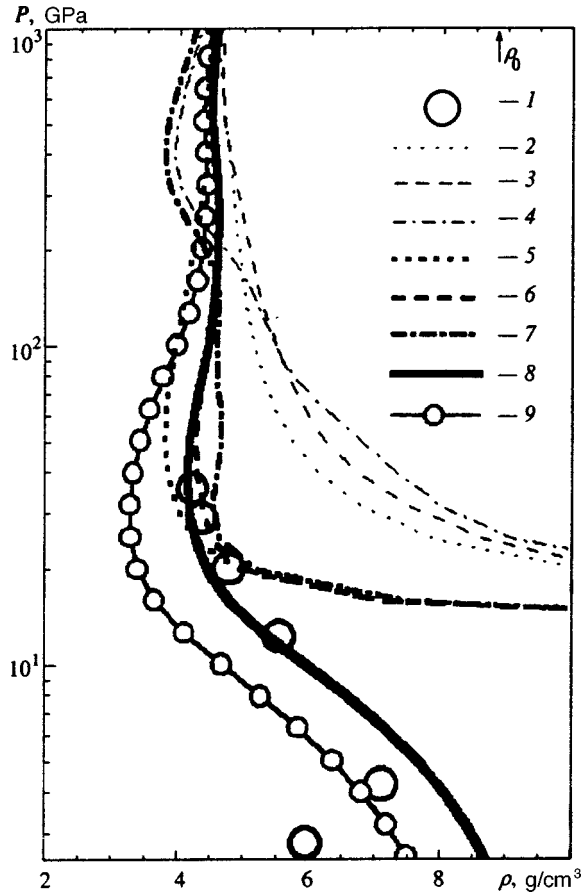

FIG. 8. Hugoniots of porous nickel at $m=10$. Calculations by several versions of the model are compared: *1* — experimental data; *2* — ideal plasma with atoms in ground states; *3* — same as *2* — plus Coulomb interaction; *4* — same as *3* — plus partition functions calculated by the Planck–Larkin method; *5*, *6*, *7* — same as *2*, *3*, *4* — plus effects of hard spherical shells; *8* — calculations by Eqs. (2)–(14) at $\delta=1$, but sums over atoms are performed in Eq. (14); *9* — same as *8*, but shell radii 20% larger.

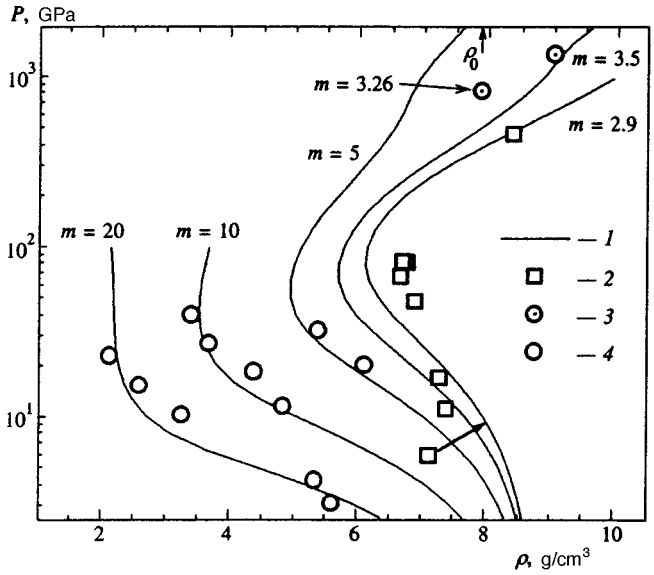

FIG. 9. Hugoniots of porous iron: *1* — calculations by Eqs. (1)–(14), summation over all heavy particles in Eq. (14); *2* — experimental data from Ref. 27 ($m=2.9$); *3* — experimental data from Ref. 27; *4* — this work.

These corrections are independent of temperature. The choice of $\delta=1$ corresponds to the traditional van der Waals approximation. It is generally accepted that $\delta=1/3$ best describes the ''metallic'' (plasma) type of chemical bond in condensed materials. In this case, the sum is performed over all heavy particles, so that corrections (14) do not shift equilibrium ionization parameters. According to Likalter,[61] this value of δ corresponds to the form of attraction in ''extended'' metals, i.e., metals at intermediate densities corresponding to the critical point on the gas–liquid line in the phase diagram.

4.6. Comparison between experimental data and calculations

Hugoniots of porous nickel, copper, and iron calculated using the model described by Eqs. (2)–(14) are plotted in Figs. 7–10. Figure 7 shows Hugoniots of porous nickel calculated earlier[34,48,58,59] and compared to experimental data available at that time.[4] These data have been supplemented with our recent measurements at porosities $m=15$ and 20. It is noteworthy that our results are in good agreement with previously available calculations. Figure 7 also compares Hugoniots for materials of low porosity (including the Hugoniot of the solid material). All these data have been summarized in one graph in order to demonstrate that, even in the simplified form given above, the chemical model of plasma provides a satisfactory description of the entire set of experimental data for nickel Hugoniots.

Figure 8 shows testing calculations performed with the aim of illustrating the effect of various components taken separately on the thermodynamic calculations based on the model (2)–(14). We stress once again that the shapes of Hugoniots in the range of relatively low shock velocities indicate the necessity of taking into account both the intense repulsion and effective attraction between heavy particles in the chemical model.

Calculations are compared with new experimental data for iron plasma in Fig. 9. Note that the choice of repulsion (inherent particle dimensions) and attraction parameters of the model was based on the same scheme as previously. Table II lists radii of atoms and ions for copper, nickel, and iron used in our calculations. Our results clearly indicate that, whereas the agreement between experiment and theory is satisfactory for both nickel and iron at high porosities (and the largest achievable porosity factors of condensed metals), it degrades gradually as the plasma density increases in experiments with samples of lower porosity.

Calculations and experimental data for copper plasma are compared in Fig. 10. The results of earlier experiments[34,48,58,59] are supplemented with new measurements. One can clearly see that the agreement between theory and experiment is satisfactory. At the same time, the discrepancy between calculations and shock measurements of the Hugoniot at $m=10$ in its upper section is remarkable. In analyzing this discrepancy, one should take into account the fact that the theoretical model describing interactions in the system is, obviously, oversimplified, and its results are highly sensitive to changes in the set of particle sizes selected in our calcu-

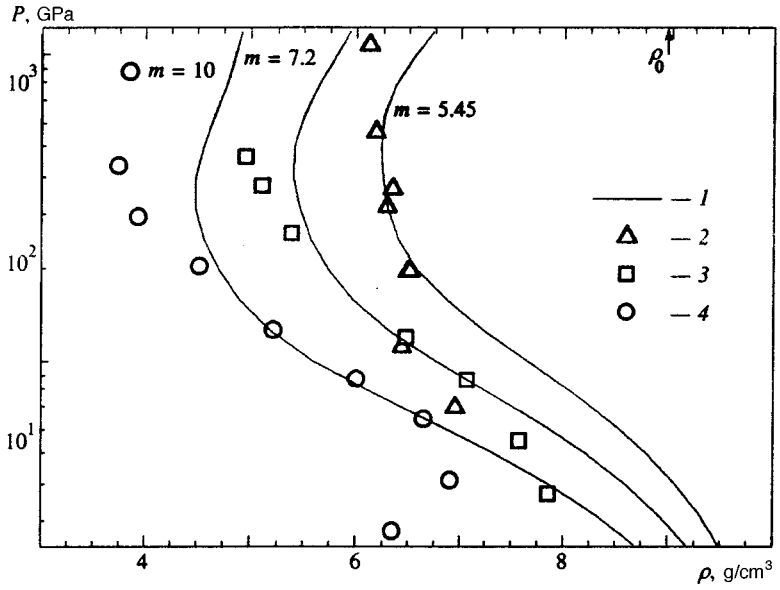

FIG. 10. Hugoniots of porous copper: *1* — calculations by Eqs. (1)–(14), summation over all heavy particles in Eq. (14); *2* — experimental data from Ref. 27; *3* — experimental data from Ref. 27; *4* — experimental data of this work.

lations. This is supported by additional calculations performed with slightly changed (increased) particle radii, which are also plotted in Fig. 10.

These supplementary calculations indicate that the discrepancy between the experimental data and previous calculations is within the natural uncertainty of the chemical model in the range of plasma parameters under consideration. At the same time, if we take into account the features of copper Hugoniots measured at $m=3$ and $m=4$, which can be described only by using a model equation of state with an effectively higher ''rigidity'' than in the model with a set of fixed particle sizes selected in earlier studies[34,48,58,59] (these features were detected earlier[46,48] and analyzed in detail[66]), we can conclude that the model description of thermodynamic properties of shock-compressed plasmas is far from universal, and individual properties of specific metals should be taken into account more accurately than is suggested by the simplified calculation technique described in this paper.

4.7. Comparison with pseudopotential calculations

Along with the approximation (2)–(14), in our calculations of thermodynamic parameters of plasmas generated by shock compression of porous samples we also used a model taking into account collisional effects and based on one version of the pseudopotential method (for details see Ref. 52). Its underlying idea is to describe the effects of Coulomb collisionality in terms of free charges interacting through certain effective potentials. This approach was used with success in various studies (Refs. 40, 67, and others). The basic postulates of one version of this approach, which has been used in our work, are formulated as follows.

1. The depth of the electron–ion pseudopotential depends on the position ε of a conventional boundary, which is, generally speaking, variable and separates free and bound states of each electron–ion pair. At the same time, this boundary limits the bound states taken into account in calculating the excitation partition function of atoms and ions. The ion–ion and electron–electron interactions are still described using the Coulomb shape of potential:

$$\Phi^*_{ie}(r)=-\frac{Z_i e^2}{r}\,(1-e^{-r/\sigma_{ie}}),$$

$$\sigma_{ie}\equiv\sigma_{ie}(n,T),\quad \Phi^*_{\alpha\alpha}(r)=\frac{Z_\alpha Z_\alpha e^2}{r},\quad \alpha=i,e. \qquad (15)$$

2. The approximation is constructed on the level of pair correlation functions $F_{ab}(r)$, whose form[68,69] at long distances is the same as in the high-temperature limit of weak collisionality for a system with interaction potential (15):[70]

$$F_{ei}(r)=1+\Psi_{ei}\frac{e^{-p_i r}-e^{-\kappa r}}{r}. \qquad (16)$$

3. At short distances, the ion–ion and electron–electron correlation functions are modified in order to get rid of the flaw inherent in linearized (''ring'') approximations[70] (like the Debye–Hückel approximation), namely, the so-called negative probabilities (here i and j are ion indices):

$$F_{ij}=1-B_{ij}\frac{e^{-\kappa r}}{r},\quad r\geqslant R_{ij},\quad F_{ij}\simeq 0,\quad r\leqslant R_{ij}. \qquad (17)$$

4. The parameters of correlation functions (16) and (17) as functions of interparticle interaction potentials are determined with due account of the local electric neutrality condition (18),[71] which is unrelated to weak collisionality. In addition, an approximate relation between the amplitude of electron–ion correlations (screening cloud) and pseudopotential depth $\Phi^*_{ei}(r=0)$ (19) is introduced:

TABLE II.

Z_i	0	1	2	3	4	5	6	7
Cu	2.00	1.700	1.55	1.40	1.25	1.10	0.95	0.8
Ni	2.00	1.684	1.27	1.10	1.00	0.84	0.80	—
Fe	2.27	2.00	1.58	1.18	1.075	0.978	0.904	—

$$\int \left\{ n_{ei} e[F_{ei}(r)-1] + \sum n_{ij} e Z_j [F_{ij}(r)-1] \right\} d\mathbf{r} = -Z_i, \tag{18}$$

$$F_{ei}(0) \equiv 1 + \Psi_{ei}(0) \approx \beta \Phi^*_{ei}(0), \quad F_{ii}(0) \approx 0. \tag{19}$$

5. The calculation of the screening parameter κ should take into account gradual exclusion of electrons from screening as their degree of degeneracy increases, so that in the limit of their full degeneracy the screening is fully controlled by the ion–ion correlations:

$$\kappa^2 = 4\pi e^2 \left(\frac{\partial \mu_e^0}{\partial n_e} + \sum_i \frac{\partial \mu_i^0}{\partial n_i} \right). \tag{20}$$

6. Corrections to thermodynamic quantities should include the shift $\Delta E_{\rm kin}$ of the mean kinetic energy of free charges derived from the virial theorem:

$$\Delta PV = (1/3)(2\Delta E - \Delta E_{\rm pot}) = (1/3)(2\Delta E_{\rm kin} + \Delta E_{\rm pot}), \tag{21}$$

$$\Delta E = V \int \left\{ \sum_i n_i n_e F_{ei}(r) \Phi^*_{ei}(r) + \sum_{i,j} n_i n_j F_{ij}(r) \Phi^*_{ij}(r) \right\} d\mathbf{r}, \tag{22}$$

$$\Delta E_{\rm pot} = -V \int \left\{ \sum_i n_i n_e F_{ei}(r) \frac{Z_i e^2}{r} - \sum_{i,j} n_i n_j F_{ij}(r) \frac{Z_i Z_j e^2}{r} \right\} d\mathbf{r}, \tag{23}$$

$$\Delta \mu_i = \int \left\{ n_e F_{ei}(r) \Phi^*_{ei}(r) + \sum_j n_j F_{ij}(r) \Phi^*_{ij}(r) \right\} d\mathbf{r} \tag{24}$$

(here $\Delta E_{\rm pot}$ and ΔE are corrections to the potential and full internal energies, μ_i and μ_e are chemical potentials of free charges). Note that in the region $\Gamma_{\rm D} \sim 1$ the approximation (15)–(24) is incompatible with the traditional relationship between corrections to the pressure and internal energy of free charges, $\Delta E = 3\Delta PV$. In comparison with many Coulomb approximations suggested in literature, this approximation leads to an effect that is equivalent to an additional electron–ion repulsion,[72] which is a function of the collisionality parameter $\Gamma_{\rm D}$.

Calculations of shock compression parameters for copper and nickel based on model (15)–(24) are listed in Table I for comparison with direct experimental measurements. Specific calculations were performed with the help of a complicated procedure in which the boundary ε separating free and bound states of electron–ion pairs was selected. This boundary corresponds to the limit of $-k_{\rm B}T$ at high temperatures and at the average distance between heavy particles at moderate temperatures and high compression factors. This energy ε, as was noted above, marked the boundary of bound states taken into account in calculations of excitation partition functions for atoms and ions.

The problem of taking into account bound states under conditions of partially ionized strongly nonideal plasma has a long history, and it has not been fully resolved by this time. Results of a set of earlier experiments[32,45,73–75] led the researchers to a conclusion (see a detailed discussion in Ref. 1, Ch. 4) about an unexpectedly small contribution of excited states to the caloric equation of state in comparison with results of some models that had been set forth by that time. There are publications which theoretically justify the low contribution of excited states.[76] A comparison between experimental results discussed here and calculations by the two models does not help in resolving this issue because, given the real accuracy of our measurements, it is difficult to separate and estimate the effect of excited states of atoms and ions on the full equation of states for the plasmas studied in the reported work.

Another problem on which the theory of nonideal plasma is traditionally focused is the existence of the so-called plasma phase transition different from the conventional gas–liquid transition in metals.[18,40] Our experiments have not detected any gas-dynamic anomalies that could be associated with an unknown phase transition in the system.[6,3,39] It is noteworthy that both approximations, (2)–(14) and (15)–(24), used in our calculations also have not produced any indications of such anomalies.

5. CONCLUSION

Using porous samples of copper, nickel, and iron with densities a factor of 10 to 20 lower than their normal densities and recently developed generators of shock waves of spherical configurations, we could penetrate into the region of high temperatures $T > 10^4$ K, densities $\rho = (0.1-1)\rho_0$, and pressures of several tens of gigapascal, which had been little investigated until recently.

It seems interesting to measure in future the compressibility of superporous metallic samples with initial densities $(0.05-0.1)$ g/cm^3. Such conditions can be created using metal vapors. Although this experiment is technically difficult, it seems quite feasible.

The chemical model of plasma is sufficiently versatile to provide a satisfactory description of metal plasmas in the regions of condensed matter densities nontraditional for this model and megabar pressures, when we have to deal with several mechanisms of strong correlations among plasma particles at the same time, which are caused by both degeneracy and strong collisionality due to numerous modes of effective interactions among plasma particles.

Our results indicate that the plasma equation of states derived from the chemical model can have sufficient accuracy even if we use such an approximation that does not adequately take into account details of the electronic structure and the character of electron localization in the plasma.

The work was supported by the Russian Fund for Fundamental Research (Grants 95-02-03886, 97-02-17339, 97-02-17340, and 96-02-18832).

*)E-mail: grvk@ficp.ac.ru
†)E-mail: ilios@orc.ru

[1] V. K. Gryaznov, I. L. Iosilevskiĭ, Yu. G. Krasnikov, N. I. Kuznetsova, V. I. Kucherenko, G. B. Lappo, B. N. Lomakin, G. A. Pavlov, É. E. Son, and V. E. Fortov, in *Thermodynamic Properties of Working Media in Gaseous Phase Nuclear Reactors* [in Russian], ed. by V. M. Ievlev, Atomizdat, Moscow (1980).
[2] D. A. Kirzhnits, Yu. E. Lozovik, and G. V. Shpatakovskaya, Usp. Fiz. Nauk **117**, 3 (1975) [Sov. Phys. Usp. **18**, 649 (1975)].
[3] V. E. Fortov and I. T. Yakubov, *Physics of Nonideal Plasma*, Hemisphere, New York, (1990); [Russ. original] Chernogolovka (1984).
[4] *Shock Compression of Condensed Materials*, Cambridge University Press, Cambridge (1998); All-Russia Institute for Experimental Physics, Sarov (1992).
[5] A. P. Nefedov, O. F. Petrov, and V. E. Fortov, Usp. Fiz. Nauk **167**, 1215 (1997) [Phys. Usp. **40**, 1163 (1997)].
[6] V. E. Fortov, Usp. Fiz. Nauk **138**, 361 (1982) [Sov. Phys. Usp. **25**, 781 (1982)].
[7] *Compendium of Shock Wave Data*, ed. by Van Thiel, Livermore Lawrence Laboratory Report, UCRL 50-108, Vol. 1–3 (1977).
[8] *LASL Shock Hugoniot Data*, ed. by S. P. Marsh, University of California Press, Berkley-LA-London (1980).
[9] M. V. Zhernokletov, V. N. Zubarev, R. F. Trunin, and V. E. Fortov, *Experimental Data on Shock Compressibility and Adiabatic Expansion of Condensed Materials at High Energy Densities* [in Russian], Chernogolovka (1996).
[10] L. V. Al'tshuler, Usp. Fiz. Nauk **15**, 197 (1965) [Sov. Phys. Usp. **8**, 52 (1965)].
[11] L. V. Al'tshuler, A. V. Bushman, M. V. Zhernokletov *et al.*, Zh. Éksp. Teor. Fiz. **78**, 741 (1980) [Sov. Phys. JETP **51**, 373 (1980)].
[12] L. V. Al'tshuler, R. F. Trunin, K. K. Krupnikov, and N. V. Panov, Usp. Fiz. Nauk **166**, 575 (1996) [Phys. Usp. **39**, 539 (1996)].
[13] A. H. Jones, W. H. Isbell, and C. J. Maiden, J. Appl. Phys. **37**, 3493 (1966).
[14] S. I. Anisimov, A. M. Prokhorov, and V. E. Fortov, Usp. Fiz. Nauk **142**, 395 (1984) [Sov. Phys. Usp. **27**, 181 (1984)].
[15] V. Fortov, M. Lebedev, K. Dyabilin, O. Vorobiev, V. Smirnov, and E. Grabovskij, in *Shock Compression of Condensed Matter-1995*, ed. by S. C. Schmidt and W. C. Tao, AIP Conf. Proc. **370**, 1255 (1996).
[16] K. Baumung, J. H. Bluhm, B. Goel, P. Hoppe, H. U. Karow, D. Rusch, V. E. Fortov, G. I. Kanel, S. V. Razorenov, A. V. Utkin, and O. Yu. Vorobjev, Laser Part. Beams **14**, 181 (1996).
[17] A. V. Bushman, I. V. Lomonosov, and V. E. Fortov, *Equations of State of Metals at High Energy Density* [in Russian], Chernogolovka (1992).
[18] G. É. Norman and A. N. Starostin, Teplofiz. Vys. Temp. **8**, 413 (1970).
[19] V. E. Fortov and I. T. Yakubov, *Nonideal Plasma*, Énergoatomizdat, Moscow (1994).
[20] N. N. Kalitkin and L. V. Kuz'mina, Preprint No. 35, Institute of Applied Mathematics, USSR Academy of Sciences, Moscow (1975).
[21] L. D. Landau and Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **14**, 32 (1944).
[22] L. V. Al'tshuler, M. N. Pavlovskiĭ, L. V. Kuleshova, and G. V. Simakov, Fiz. Tverd. Tela **5**, 279 (1965) [Sov. Phys. Solid State **5**, 203 (1965)].
[23] L. V. Al'tshuler, S. B. Kormer, A. A. Bakanova, and R. F. Trunin, Zh. Éksp. Teor. Fiz. **38**, 790 (1960) [Sov. Phys. JETP **11**, 573 (1960)].
[24] L. V. Al'tshuler, R. F. Trunin, K. K. Krupnikov, and N. V. Pavlov, Usp. Fiz. Nauk **15**, 197 (1965) [Sov. Phys. Usp. **8**, 52 (1965)].
[25] L. V. Al'tshuler, K. K. Krupnikov, B. N. Ledenev *et al.*, Zh. Éksp. Teor. Fiz. **34**, 866 (1958); **34**, 874 (1958) [Sov. Phys. JETP **7**, 600 (1958); **7**, 606 (1958)].
[26] A. V. Bugaeva, A. A. Evstigneev, and R. F. Trunin, Teplofiz. Vys. Temp. **34**, 684 (1996).
[27] R. F. Trunin, G. V. Simakov, Yu. N. Sutulov, A. V. Medvedev, B. D. Rogozkin, and Yu. E. Fedorov, Zh. Éksp. Teor. Fiz. **96**, 1024 (1989) [Sov. Phys. JETP **69**, 580 (1989)].
[28] R. F. Trunin and G. F. Simakov, Zh. Éksp. Teor. Fiz. **103**, 2180 (1993) [JETP **76**, 1090 (1993)].
[29] S. B. Kormer, A. I. Funtikov, V. D. Urlin, and A. N. Kolesnikova, Zh. Éksp. Teor. Fiz. **42**, 686 (1962) [Sov. Phys. JETP **15**, 477 (1962)].
[30] Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **32**, 1577 (1957) [Sov. Phys. JETP **5**, 1282 (1957)].
[31] B. N. Lomakin and V. E. Fortov, Zh. Éksp. Teor. Fiz. **63**, 92 (1972) [Sov. Phys. JETP **36**, 48 (1972)].
[32] A. V. Bushman, B. N. Lomakin, V. A. Sechenov, V. E. Fortov, O. E. Shchekotov, and I. I. Sharipdzhanov, Zh. Éksp. Teor. Fiz. **69**, 1524 (1975) [Sov. Phys. JETP **42**, 5828 (1976)].
[33] L. V. Al'tshuler, S. E. Brusnikin, and A. S. Marchenko, Teplofiz. Vys. Temp. **27**, 636 (1989).
[34] V. K. Gryaznov, I. L. Iosilevski, and V. E. Fortov, in *Physics of Strongly Coupled Plasmas*, ed. by W. D. Kraeft and M. Schlanges, World Scientific, Singapore (1996), p. 351.
[35] I. L. Iosilevskiĭ and V. K. Gryaznov, Teplofiz. Vys. Temp. **19**, 1121 (1981).
[36] G. V. Sin'ko, Chislennye Metody Mekhaniki Sploshnoi Sredy **10**, 124 (1978); Teplofiz. Vys. Temp. **21**, 1041 (1983).
[37] A. F. Nikiforov, V. G. Novikov, and V. B. Uvarov, in *Topics of Nuclear Sciences and Technology. Methods and Codes for Numerical Solutions of Problems of Mathematical Physics*, No. 4 (1979), p. 16.
[38] G. V. Shpatakovskaya, Preprint No. 54, Insitute of Applied Mathematics, USSR Academy of Sciences Moscow (1975); Teplofiz. Vys. Temp. **27**, 677 (1989).
[39] Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Processes*, Academic, New York (1968); [Russ. original] Nauka, Moscow (1968).
[40] W. Ebeling, W.-D. Kräft, and D. Kremp, *Theory of Bound States and Ionization Equilibrium in Plasmas and Solids*, Academie Verlag, Berlin (1976).
[41] W. Ebeling, A. Förster, V. Fortov, V. Gryaznov, and A. Polishchuk, *Thermophysical Properties of Hot Dense Plasmas*, Teubner, Stuttgart–Leipzig (1991).
[42] N. N. Kalitkin, I. V. Ritus, and A. M. Mironov, Preprint No. 46, Insitute of Applied Mathematics, USSR Academy of Sciences, Moscow (1983).
[43] B. N. Bazylev, F. N. Borovik, G. S. Romanov, and G. A. Vergunova, Kvant. Élektr. **13**, 1981 (1986) [Sov. J. Quantum Electron. **16**, 1308 (1986)].
[44] V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, Zh. Prikl. Mekh. Tekh. Fiz., No. 3, 70 (1973).
[45] V. K. Gryaznov, M. V. Zhernokletov, I. L. Iosilevskiĭ, V. N. Zubarev, and V. E. Fortov, Zh. Éksp. Teor. Fiz. **78**, 573 (1980) [Sov. Phys. JETP **51**, 288 (1980)].
[46] V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, Pis'ma Zh. Tekh. Fiz. **22**(8), 1376 (1982) [sic].
[47] V. K. Gryaznov and V. E. Fortov, Teplofiz. Vys. Temp. **25**, 1208 (1987).
[48] V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, in *Equations of State*, ed. by V. E. Fortov, Moscow (1995), p. 38.
[49] A. A. Likal'ter, Zh. Éksp. Teor. Fiz. **56**, 240 (1969) [Sov. Phys. JETP **29**, 133 (1969)].
[50] M. Baus and J. P. Hansen, Phys. Rep. **59**, 1 (1980).
[51] S. Ichimaru, H. Iyetomi, and S. Tanaka, Phys. Rep. **149**, 91 (1987).
[52] I. L. Iosilevskiĭ, Teplofiz. Vys. Temp. **19**, 1121 (1981).
[53] T. Kahlbaum and A. Förster, Fluid Phase Equilibria **76**, 71 (1992).
[54] C. F. Mansoori, V. Carnahan, K. E. Starling, and T. W. Leland, J. Chem. Phys. **54**, 1523 (1971).
[55] D. Hartree, *The Calculation of Atomic Structures*, Wiley, N. Y. (1957).
[56] V. K. Gryaznov, PhD Thesis, Institute of Chemical Physcis, USSR Academy of Sciences, Chernogolovka (1981).
[57] A. N. Ivanova and V. K. Gryaznov, *Hartree-Fock Calculations of Atoms* [in Russian], Report of Institute of Chemical Physics, USSR Academy of Sciences, Chernogolovka (1975).
[58] V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, in *Physics and Technology of Plasma*, Belarus State University, Minsk (1994), p. 1.
[59] V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, in *Physics of Low-Temperature Plasma*, Petrozavodsk (1995), p. 105.
[60] N. W. Ashcroft and J. Lekner, Phys. Rev. **145**, 83 (1966).
[61] A. A. Likal'ter, Dokl. Akad. Nauk SSSR **259**, 96 (1981) [Sov. Phys. Dokl. **26**, 676 (1981)]; Usp. Fiz. Nauk **162**, 119 (1992) [Sov. Phys. Usp. **35**, 591 (1992)].
[62] R. G. McQueen and S. P. Marsh, J. Appl. Phys. **31**, 1253 (1960).
[63] J. M. Walsh, M. H. Rice, R. G. McQueen, and F. L. Yarger, Phys. Rev. **108**, 169 (1957).
[64] L. V. Al'tshuler, A. A. Bakanova, and R. F. Trunin, Zh. Éksp. Teor. Fiz. **42**, 91 (1962) [Sov. Phys. JETP **15**, 65 (1962)].
[65] Yu. L. Alekseev, B. P. Ratnikov, and A. P. Rybakov, Zh. Prikl. Mekh. Tekh. Fiz. No. 2, 101 (1971).
[66] *Shock Waves and Extremal States of Matter*, ed. by R. F. Trunin and V. E. Fortov (1998) (in press).
[67] B. V. Zelener, G. E. Norman, and V. S. Filinov, Teplofiz. Vys. Temp. **11**, 922 (1973); **12**, 267 (1974); **13**, 712, 913 (1975).

[68] V. K. Gryaznov and I. L. Iosilevskiĭ, Chislennye Metody Mekhaniki Sploshnoi Sredy **4**, 166 (1973).
[69] I. L. Iosilevskiĭ, Teplofiz. Vys. Temp. **18**, 447 (1980).
[70] A. E. Glauberman and I. R. Yukhnovskiĭ, Zh. Éksp. Teor. Fiz. **22**, 562 (1952).
[71] F. Stillinger and R. Lowett, J. Chem. Phys. **49**, 1991 (1968).
[72] T. P. Wright and O. Theimer, Phys. Fluids **13**, 895 (1970).
[73] B. N. Lomakin and V. E. Fortov, Zh. Éksp. Teor. Fiz. **63**, 92 (1972) [Sov. Phys. JETP **36**, 48 (1972)].
[74] I. Ya. Dikhter and V. Ya. Zeigarnik, Teplofiz. Vys. Temp. **15**, 471 (1977).
[75] V. E. Fortov, A. A. Leont'ev, A. N. Dremin, and V. K. Gryaznov, Zh. Éksp. Teor. Fiz. **71**, 225 (1976) [Sov. Phys. JETP **44**, 116 (1976)].
[76] A. S. Kaklyugin and G. É. Norman, Teplofiz. Vys. Temp. **25**, 209 (1987).

Translation provided by the Russian Editorial office.

Spin vortices and stationary spin flows in a normal Fermi liquid

P. L. Krotkov and V. P. Mineev*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 19 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1266–1283 (October 1998)

We solve the equations for the collisionless spin dynamics of a normal Fermi liquid, which describe structures resembling spin vortices coherently precessing in a uniform magnetic field. We examine their stability and relaxation, and consider various regimes of stationary magnetization transport along a channel. © *1998 American Institute of Physics.* [S1063-7761(98)00910-X]

1. INTRODUCTION

A two-domain spin structure precessing coherently in a weakly nonuniform magnetic field has been predicted[1] in the collisionless region in a normal Fermi liquid and observed experimentally.[2] The lifetime of this structure, in which the orientation of the magnetization varies smoothly from parallel to antiparallel to the external magnetic field, significantly exceeds the precession dephasing time in a nonuniform field.

Similar coherently precessing, inhomogeneous distributions of the magnetization have previously been found in superfluid ^{3}He-B,[3,4] where coherently precessing quantum spin vortices were also discovered along with them, and the flow of spin currents along a channel accompanied by precession phase slippage has been investigated.[5–10] The development of a theory for the analogous phenomena in a normal Fermi liquid would be of unquestionable interest. The present paper examines solutions of the equations for the collisionless spin dynamics of a normal Fermi liquid, which describe stationary spin flows.

Coherently precessing structures in the form of spin vortices are found in a uniform magnetic field. The vortex wall thickness is determined not only by the field strength and the parameters of the Fermi liquid, but also by the absolute difference of the precession and Larmor frequencies. When these frequencies are equal, the distribution of spin in a vortex corresponds to the familiar Belavin–Polyakov skyrmion solution.[11]

In addition, the steady flow of a spin current along a channel is investigated in this paper.

The present paper is organized as follows. The equations for the collisionless spin dynamics of a normal Fermi liquid are written out in a form convenient for subsequent discussion in Sec. 2. The vortex solutions of these equations are found, and their stability and relaxation are investigated in Sec. 3. Section 4 is devoted to the flow of a spin current along a channel. The principal results are briefly discussed in the Conclusion.

2. BASIC EQUATIONS

A complete set of equations for the spin dynamics of a normal Fermi liquid in a magnetic field $\mathbf{H}(\mathbf{r},t)$ in terms of macroscopic quantities, viz., the quasiparticle spin density $\mathbf{S}(\mathbf{r},t)$ and quasiparticle spin current density $\mathbf{J}_i(\mathbf{r},t)$, was obtained by Leggett[12] from the kinetic equation for the spin-vector part of the quasiparticle distribution function $\boldsymbol{\nu}_{\mathbf{k}}(\mathbf{r},t)$. These equations have the form

$$\left(\frac{\partial}{\partial t}+\boldsymbol{\omega}_L\times\right)\mathbf{S}+\nabla_i\mathbf{J}_i=0, \tag{1}$$

$$\left(\frac{\partial}{\partial t}+\boldsymbol{\omega}_L\times\right)\mathbf{J}_i+\frac{w^2}{3}\nabla_i\left(\mathbf{S}-\frac{\chi_n}{\gamma^2}\,\boldsymbol{\omega}_L\right)+\kappa\frac{\gamma^2}{\chi_n}\mathbf{S}\times\mathbf{J}_i=-\frac{\mathbf{J}_i}{\tau_1}. \tag{2}$$

Here χ_n is the magnetic susceptibility of the Fermi liquid, γ is the gyromagnetic ratio for ^{3}He nuclei, $\boldsymbol{\omega}_L(\mathbf{r},t)=\gamma\mathbf{H}(\mathbf{r},t)$ is the Larmor frequency, $w^2=v_F^2(1+F_0^a)(1+F_1^a/3)$, $\kappa=(F_1^a/3-F_0^a)/(1+F_0^a)$, F_0^a and F_1^a are the coefficients of the expansion of the antisymmetric part of the Fermi-liquid quasiparticle interaction in spherical harmonics, v_F is the Fermi velocity, $\tau_1=\tau/(1+F_1^a/3)$, and τ is the quasiparticle free flight time.

The parentheses in the first terms of Eqs. (1) and (2) contain the total time derivative in the local coordinate system, which rotates about the direction of the external magnetic field at the Larmor frequency. The special role played by this frame derives from the free spin's fixed orientation relative to it. The second term in Eq. (2) describes a torque proportional to the gradient of the deviation of the spin density from the local equilibrium value $\chi_n\boldsymbol{\omega}_L/\gamma^2$. The third term in (2) is specific to Fermi liquids. Its relative contribution is not small compared to the Fermi-liquid interaction force. The latter is represented in the equation by κ, which subsumes the constants F_0^a and F_1^a. Physically, this term represents an additional torque, which acts on the current owing to the molecular magnetic field, even in a coordinate system rotating at the local Larmor frequency.

The domain of applicability of Eqs. (1) and (2) is restricted by the requirement of sufficiently slow spatial variation of the quasiparticle distribution. If the characteristic scale of the spatial inhomogeneity of $\boldsymbol{\nu}_{\mathbf{k}}(\mathbf{r},t)$ is denoted by ξ, this condition (see Ref. 12) can be written in the form

$$\xi\gg\min\left\{l,\frac{v_F}{\omega_m}\right\}. \tag{3}$$

Here $l=v_F\tau$ is the quasiparticle mean free path, and $\omega_m=\omega_L\kappa/(1+F_1^a/3)$ is the frequency corresponding to the molecular field.

Apart from the hydrodynamic region $\xi\gg l$ for $l<v_F/\omega_m$, the equations describing the evolution of the spin density and the current density, i.e., the equations for the zeroth and first spherical harmonics $\boldsymbol{\nu}_{\mathbf{k}}(\mathbf{r},t)$, also split off from the equations for the higher harmonics in the case $\xi\gg v_F/\omega_m$ with $l>v_F/\omega_m$ (which is equivalent to $\omega_m\tau>1$). The latter circumstance enables us to use them to investigate the spin dynamics of normal Fermi liquids in the collisionless regime $\omega_L\tau>1$. We note that the proportionality coefficient between the molecular field ω_m and the external field ω_L is $\kappa/(1+F_1^a/3)\approx 2$ for normal ^{3}He and ≈ 0.036 for a saturated ^{3}He–^{4}He solution at zero pressure (see Ref. 2). In the collisionless region the unusual third term in Eq. (2) becomes significant. The results of the present work pertain specifically to this region.

The boundary condition usually chosen for Eqs. (1) and (2) is that no spin current flow through the wall of the vessel containing the Fermi liquid:

$$\mathbf{J}_i n_i = 0, \tag{4}$$

where n_i is the ith component of the unit normal to the vessel surface.[1)] To fix ideas, let the external field be parallel to the $\hat{z}$ axis. Under such a boundary condition, it follows from the continuity equation (1) that the total longitudinal magnetization $\int S_z\, d\mathbf{r}$ is conserved.

To simplify the equations, we adopt a system of units in which $\chi_n=\gamma^2$. We consider a homogeneous external field, $\nabla\boldsymbol{\omega}_L=0$, and assume motions in the reference system, which rotates at the precession frequency, to be fairly slow: $\tau\delta\omega\sim\tau_1\delta\omega\ll 1$. Then the time derivative in this system is approximately zero, and the first term in Eq. (2) can be discarded. Solving the resulting equation for $\mathbf{J}_i$, we obtain an expression for the spin current:

$$\mathbf{J}_i\simeq -\frac{w^2\tau_1/3}{1+(\kappa S\tau_1)^2}[\nabla_i\mathbf{S}+\kappa\tau_1\nabla_i\mathbf{S}\times\mathbf{S}+(\kappa\tau_1)^2\mathbf{S}(\mathbf{S}\nabla_i\mathbf{S})]. \tag{5}$$

In the collisionless region each successive term in square brackets is $\kappa\omega_L\tau_1$ times its predecessor. However, the last term is anomalously small. In fact, it is exactly zero for a homogeneous spatial distribution of the absolute value of the magnetization. At the same time, as the estimate in Sec. 2 of Ref. 13 shows, the characteristic time to smooth out the inhomogeneity of the S^2 distribution is of order $\xi^2/w^2\tau_1$. This time is small compared to the reciprocal characteristic frequency $\delta\omega^{-1}$, since in our case $\xi^2\sim w^2/\kappa\omega_L\delta\omega$ (see below). We therefore assume that S^2 is constant. Thus, the first two terms are important. As we have already mentioned, the first term describes the ordinary diffusion current, while the second is a nondissipative current. Moreover, unless the second term is anomalously small, the first term can be neglected. Then the expression for the current takes the form

$$\mathbf{J}_i\simeq\frac{w^2}{3\kappa S^2}\mathbf{S}\times\nabla_i\mathbf{S}. \tag{6}$$

Substituting this expression into Eq. (1) for the evolution of the spin, a single vector equation remains:

$$\frac{\partial\mathbf{S}}{\partial t}=\mathbf{S}\times\boldsymbol{\omega}_L-\frac{w^2}{3\kappa S^2}\mathbf{S}\times\nabla^2\mathbf{S}. \tag{7}$$

This is the equation of interest to us. Along with the already mentioned total longitudinal magnetization $\int S_z\, d\mathbf{r}$, the absolute value of the magnetization S is also its integral. Therefore, in solving it, we use the natural parametrization of $\mathbf{S}$ in terms of the spherical coordinates α and β: $\mathbf{S}=S\hat{\mathbf{S}}$

$$\hat{\mathbf{S}}=\begin{pmatrix}\sin\beta\cos\alpha\\ \sin\beta\sin\alpha\\ \cos\beta\end{pmatrix}. \tag{8}$$

We are interested in the coherently precessing solutions of this equation. However, it is more convenient to transform to spherical coordinates not directly in (7), but by noting (see Ref. 13) that (7) is actually the Landau–Lifshitz equation with a negative coefficient of the gradient term (when $\kappa>0$). As we know, this equation is the Hamilton–Jacobi equation obtained from the Hamiltonian

$$H=\int d\mathbf{r}\left[\frac{(\mathbf{S}-\boldsymbol{\omega}_L)^2}{2}-\frac{w^2}{6\kappa S^2}(\nabla_i\mathbf{S})^2\right] \tag{9}$$

with the usual commutation relation for the spin:

$$[S_\alpha(\mathbf{r}),S_\beta(\mathbf{r}')]=ie_{\alpha\beta\gamma}S_\gamma(\mathbf{r})\,\delta(\mathbf{r}-\mathbf{r}').$$

Introducing the Lagrange multiplier $\boldsymbol{\omega}_P\|\hat{z}$, which takes the conservation of the total longitudinal magnetization into account (equivalent to transforming to a reference system that precesses at that frequency), and bearing in mind the fixed absolute value of the spin (to within constant terms), we have

$$H=\int d\mathbf{r}\left[(\boldsymbol{\omega}_P-\boldsymbol{\omega}_L)\mathbf{S}-\frac{w^2}{6\kappa S^2}(\nabla_i\mathbf{S})^2\right], \tag{10}$$

or in spherical coordinates

$$H=\int d\mathbf{r}\left[(\omega_P-\omega_L)S\cos\beta-\frac{w^2}{6\kappa}((\nabla_i\beta)^2+(\nabla_i\alpha)^2\sin^2\beta)\right]. \tag{11}$$

The equations for α and β, which specify spin distributions that are time-independent in the reference system precessing at $\boldsymbol{\omega}_P$, are found by varying this Hamiltonian. We note that $|\omega_P-\omega_L|$ plays the role of the frequency $\delta\omega$ of motion in the Larmor system.

3. SPIN VORTEX

We seek axisymmetric solutions

$$\beta=\beta(\rho),\quad \alpha=\alpha(\varphi).$$

Setting the variational derivative of the functional (11) with respect to α equal to zero yields

$$\alpha''=0, \tag{12}$$

whereupon

$$\alpha(\varphi)=N\varphi+\alpha_0 \tag{13}$$

with the integer circulation quantum number $N=\alpha'=\text{const}$, which follows from the single-valuedness of $\alpha(\varphi)$ modulo $2\pi N$ under a rotation (variation of φ) by 2π. In this case the gradient of α is

$$\nabla_i\alpha=\frac{N}{\rho}\hat{\varphi}_i. \tag{14}$$

Here and below $\hat{\varphi}$, $\hat{\rho}$, and $\hat{z}$ are the unit vectors of a cylindrical coordinate system.

Varying the Hamiltonian (11) with respect to β yields a differential equation for $\beta(\rho)$,

$$-(\omega_P-\omega_L)S\sin\beta+\frac{w^2}{3\kappa}\left(\nabla\beta-\frac{N^2}{\rho^2}\sin\beta\cos\beta\right)=0, \tag{15}$$

which, after rewriting the Laplacian in cylindrical coordinates, can be brought to the form

$$\beta''+\frac{\beta'}{\rho}-\frac{N^2}{\rho^2}\sin\beta\cos\beta-\frac{\operatorname{sgn}(\omega_P-\omega_L)}{\xi^2}\sin\beta=0, \tag{16}$$

where we have introduced the characteristic length of the problem ξ,

$$\xi=\sqrt{\frac{w^2}{3\kappa S|\omega_P-\omega_L|}}. \tag{17}$$

The boundary conditions on Eq. (16) can be obtained as follows. In the axisymmetric case the expression for the current (6) takes the form

$$\mathbf{J}_i=\frac{w^2}{3\kappa}\left[\beta'\begin{pmatrix}-\sin\alpha\\ \cos\alpha\\ 0\end{pmatrix}\hat{\rho}_i+\frac{N}{\rho}\begin{pmatrix}-\sin\beta\cos\beta\cos\alpha\\ -\sin\beta\cos\beta\sin\alpha\\ \sin^2\beta\end{pmatrix}\hat{\varphi}_i\right], \tag{18}$$

where we put $\alpha'=N$. We make the simplifying assumption that the vessel containing the Fermi liquid is a circular cylinder with generatrix parallel to the z axis and a base of radius R. The condition (4) of vanishing current (6) through the vessel walls holds identically in this case at the ends of the cylinder, and it yields a boundary condition on the lateral surface:

$$\beta'|_{\rho=R}=0. \tag{19}$$

When $N=0$, from (13) we obtain $\alpha=\alpha_0$, i.e., all spins lie in a single plane parallel to the $\hat{z}$ axis. The z axis is the only preferred axis in the problem, so it follows from symmetry arguments that the functional (11) can only have a local extremum when $\beta(\rho)\equiv 0$ or $\beta(\rho)\equiv\pi$. The former value is stable, the latter is not.

Now consider $N\neq 0$. We are interested in solutions with no singularities at $\rho=0$. For this to occur when $N\neq 0$, $\beta(0)$ must equal 0 or π. Equation (16) is clearly invariant under the transformations

$$\begin{aligned}&(\omega_P-\omega_L)\to-(\omega_P-\omega_L),\\ &\beta\to\pi-\beta.\end{aligned} \tag{20}$$

Instead of the two asymptotes $\beta(0)=\{0,\pi\}$, we can consider one. Specifically, we assume that the boundary condition for $\rho=0$ has the form

$$\beta(0)=0. \tag{21}$$

Thus, the problem has been reduced to solving the differential equation (16) with boundary conditions (19) and (21). The problem contains N only in the form N^2, and N can be assumed to be a positive integer.

We first consider an infinite vessel: $R\to\infty$. In this case the boundary condition (19) can be replaced by

$$\beta'\to 0|_{\rho\to\infty}. \tag{19$'$}$$

When $\xi=\infty$ $(\omega_P=\omega_L)$, Eq. (16) has a gauge-invariant solution, which satisfies the conditions (19$'$) and (21) and is identical to the Belavin–Polyakov skyrmion,[11] and the topological mapping index turns out to equal the circulation quantum number:

$$\beta(\rho)=\arccos\frac{1-A\rho^{2N}}{1+A\rho^{2N}}\sim\begin{cases}A\rho^N, & \rho\to 0,\\ \pi-A\rho^{-N}, & \rho\to\infty,\end{cases} \tag{22}$$

where $A>0$ is an arbitrary constant. As can be seen, the skyrmion is characterized by a spin flip from the equilibrium direction with respect to $\boldsymbol{\omega}_L$ at zero to the antiequilibrium direction at infinity.[2)]

If ξ is finite, it is possible to obtain analytically only the asymptotic dependence of $\beta(\rho)$ as $\rho\to 0$ and $\rho\to\infty$, which must be matched numerically in the intermediate region. The feasibility of matching the two asymptotes implies the existence of the corresponding solution.

When $\xi\neq\infty$ $(\omega_P\neq\omega_L)$, the differential equation (16) can be reduced by the replacement $r=\rho/\xi$ to an equation for $\beta(r)$ of the form

$$\beta''+\frac{\beta'}{r}-\frac{N^2}{r^2}\sin\beta\cos\beta-\operatorname{sgn}(\omega_P-\omega_L)\sin\beta=0. \tag{23}$$

The expansion of this equation near zero yields the linearized equation

$$\beta''+\frac{\beta'}{r}-\left(\frac{N^2}{r^2}+\operatorname{sgn}(\omega_P-\omega_L)\right)\beta=0, \tag{24}$$

which is the ordinary (when $\omega_P<\omega_L$) or modified (when $\omega_P>\omega_L$) Bessel equation. Its solution is a linear combination of two linearly independent functions, which can be chosen so that (for $N\neq 0$) one of them diverges as r^{-N} when $r\to 0$, while the other tends to zero as r^N. Since β is bounded,

$$\beta\in[0,\pi], \tag{25}$$

only the nondivergent term must remain, and for the solutions of interest we obtain the asymptote

$$\beta\sim Ar^N|_{r\to 0}, \tag{26}$$

where $A>0$.

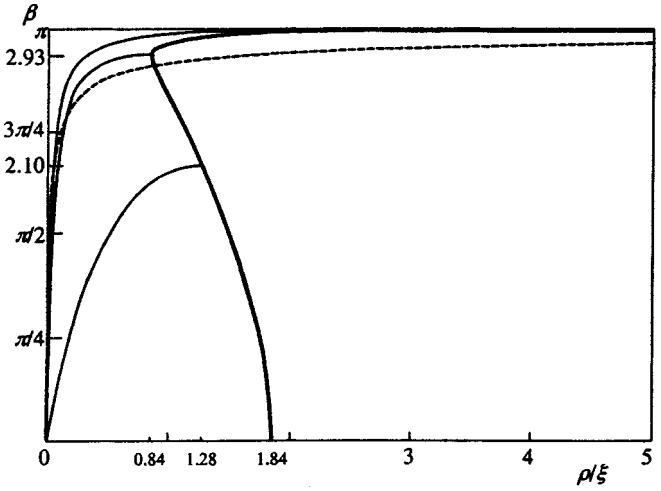

FIG. 1. Dependence of the deviation angle of the magnetization on the dimensionless distance to the axis for three spin vortices with circulation quantum number $N=1$ (light solid lines): upper line — the vortex (28); lower line — ''standing wave'' with vanishing total longitudinal component of the magnetization direction; middle line — ''standing wave'' corresponding to the minimum effective radius of the vessel; heavy line — locus of points for the value of β at the vessel boundary for vortex solutions with $N=1$; dashed line — the Belavin–Polyakov skyrmion (22) with the same value of the derivative at zero as the vortex (28).

Numerical solution of the differential equation (23) with initial condition (26) shows that, depending on the value of A, the function $\beta(r)$ asymptotically approaches either zero or π at infinity (we discard the asymptotes $\beta \to 2\pi, 3\pi, \ldots$, which do not satisfy the constraint (25)). An analogous examination of the linearized equations shows that nondivergent asymptotes are feasible as $r\to\infty$:

$$\beta\sim\begin{cases} Be^{-r}/\sqrt{r}, & \omega_P>\omega_L, \\ \pi-Be^{-r}/\sqrt{r}, & \omega_P<\omega_L, \end{cases} \tag{27}$$

where $B>0$. The constants A in (26) and B in (27), of course, can be determined only from the matching condition (26) to (27). As the numerical solution shows, matching is possible for $\omega_P<\omega_L$, but not for $\omega_P>\omega_L$.[3)] Thus, when $\omega_P<\omega_L$, there can be structures with spin flip in the interval from zero to infinity and the asymptotes

$$\beta(r)\sim\begin{cases} Ar^N, & r\to 0, \\ \pi-Be^{-r}/\sqrt{r}, & r\to\infty. \end{cases} \tag{28}$$

This structure resembles the skyrmion (22), but is characterized by an exponential approach to a spatially homogeneous distribution as $r\to\infty$. For $N=1$ it takes the form shown in Fig. 1. For comparison, the dashed line depicts the skyrmion (22) with the same asymptote at zero.

Apart from the asymptote (27), the theory of Bessel functions tells us that other asymptotes for $r\to\infty$ take the form of damped oscillations about 0, π, 2π, ..., which of course do not satisfy the constraint (25) and should be discarded. The numerical solution shows that when $\omega_P>\omega_L$, the asymptote (26) approaches a solution that oscillates about $\beta=\pi$ for all A.

The situation is more complicated when $\omega_P<\omega_L$. For a certain $A=A_0$ the solution approaches the exponential damping (27), for $A>A_0$ it approaches oscillations about 2π, and for $A<A_0$ it approaches oscillations about zero. Thus, there are no solutions for $A>A_0$. For $A<A_0$, however, a size effect is possible: the function $\beta(r)$, upon leaving the origin, reaches the point r_0 of the first maximum, $\beta(r_0)\in(0,\pi)$, then decreases to the first minimum, $\beta\in(-\pi,0)$, and finally oscillates about zero, successively passing though maxima and minimum of decreasing absolute value. If a wall is located at r_0, i.e., if $r_0=R/\xi$, such a solution will satisfy the boundary conditions (19) and (21) and not violate the constraint (25). Such a structure is a standing wave.

As $A\to A_0$, the position r_0 of the first maximum tends to infinity, and as $A\to 0$ the solution of Eq. (23), which deviates only slightly from zero everywhere, transforms, as we have already mentioned, into the Bessel function (of the first kind) of order N, and accordingly the point r_0 of the first maximum tends to a root of the derivative of the Bessel function: the numerical value is 1.84118 for vortices with $N=1$, 3.05424 for $N=2$, and 4.20119 for $N=3$. The value of $\beta(r)$ at the point r_0 of the first maximum varies from π for $A=A_0$ to 0 for $A=0$. The set of points $\beta(r_0)$ obtained by numerical methods is depicted in Fig. 1 by the thick line. The figure also shows two standing wave structures.

As can be seen from the figure, the set of points $\beta(r_0)$ describes the continuous transformation of a spin vortex as the effective size of the vessel R/ξ decreases from infinity. There is a minimum value $r_{0m}\approx 0.84$, which determines the minimum vessel radius required to produce a similar structure. The value of $\beta(r_{0m})$ is roughly 2.93 rad. Thus, the spin vortex (28) is the limiting case of a ''standing wave'' in a vessel of infinite size. From here on, we therefore call both structures spin vortices without distinguishing between them.

The initial conditions of an experiment set the total longitudinal magnetization $S\int d\mathbf{r}\cos\beta$, which subsequently does not vary, in accordance with Leggett's equations. Each spin vortex is uniquely characterized by one of two quantities: 1) the total longitudinal component of the magnetization direction $\int d\mathbf{r}\cos\beta$, or 2) the angle $\beta(R/\xi)$ that the spins make with the vessel boundary. The initial absolute value of the magnetization S can be assumed to be the equilibrium value of ω_L. Thus, the initial conditions fix the final value of $\beta(R/\xi)$ (see Fig. 1). For the structure depicted in Fig. 1 with $\beta(R/\xi)\approx 2.10$, the total longitudinal component of the magnetization direction $\int d\mathbf{r}\cos\beta$ vanishes. For vortices with $\beta(R/\xi)<2.10$, the total longitudinal component of the magnetization direction is greater than zero, while for $\beta(R/\xi)>2.10$ it is less than zero.

As we have already noted, when $\omega_P>\omega_L$, all structures shown in Fig. 1 can be mapped symmetrically about the straight line $\beta=\pi/2$.

Note that the existence of a minimum value r_{0m} implies that for a given real vessel radius R, the formation of a spin vortex is possible for an arbitrary value of the difference $|\omega_P-\omega_L|$ greater than its minimum value, which, as follows from the definition of ξ, is

$$|\omega_P-\omega_L|_{\min}=\frac{w^2}{3\kappa\omega_L}\left(\frac{r_{0m}}{R}\right)^2.$$

Since the applicability of the present theory is restricted by

the condition $|\omega_P-\omega_L|\tau_1\ll 1$, in this approximation the only legitimate claim is that such a structure exists at large enough R.

3.1. Stability

In the axisymmetric case the Hamiltonian (11) can easily be brought to the form

$$H=-\frac{w^2}{3\kappa}\int_0^{R/\xi} r\,dr\int_0^{2\pi} d\varphi\left[\frac{\beta'^2}{2}+\frac{1}{2}\left(\frac{\alpha'}{r}\right)^2 \times\sin^2\beta-\mathrm{sgn}(\omega_P-\omega_L)\cos\beta\right]. \tag{29}$$

Its second variation is

$$\delta^2 H=-\frac{w^2}{3\kappa}\int r\,dr\int d\varphi\left[(\delta\beta')^2+\left(\frac{\alpha'}{r}\right)^2 \times\cos 2\beta(\delta\beta)^2+2\frac{\alpha'\delta\alpha'}{r^2}\sin 2\beta\,\delta\beta+\left(\frac{\delta\alpha'}{r}\right)^2 \times\sin^2\beta+\mathrm{sgn}(\omega_P-\omega_L)\cos\beta(\delta\beta)^2\right]. \tag{30}$$

Integrating by parts, we write

$$\int d\varphi\,\alpha'\,\delta\alpha'=-\int d\varphi\,\alpha''\delta\alpha,$$

$$\int r\,dr\,(\delta\beta')^2=-\int dr\,(r\,\delta\beta')'\,\delta\beta =-\int r\,dr\,(\delta\beta''+\delta\beta'/r)\,\delta\beta.$$

The structures found above are described by the differential equations (12) and (23). Varying the second equation and substituting it into the expression for the second variation of the energy, we finally obtain

$$\delta^2 H=-\frac{w^2}{3\kappa}\int d^2r\left(\frac{\delta\alpha'}{r}\right)^2\sin^2\beta. \tag{31}$$

Here $\int d^2r=\int r\,dr\int d\varphi$. Thus, all structures described are local maxima (when $\kappa>0$ and, accordingly, local minima when $\kappa<0$) of the energy functional. Therefore, by virtue of energy conservation they are all stable.

3.2. Relaxation

It was shown in Ref. 1 for the functions '$\mathbf{S}$ and '$\mathbf{J}_i$, which are solutions of Leggett's equations, that the collision term for the spin current in Eq. (2) leads to relaxation behavior in a closed volume of a Fermi liquid:

$$\frac{d}{dt}\int d\mathbf{r}\left[\frac{(\mathbf{S}-\boldsymbol{\omega}_L)^2}{2}+\frac{3\mathbf{J}_i^2}{2w^2}\right]=-\frac{3}{w^2\tau_1}\int d\mathbf{r}\,\mathbf{J}_i^2. \tag{32}$$

Because of the conservation of the total longitudinal magnetization, the derivative $d\int(\boldsymbol{\omega}_L\mathbf{S})\,d\mathbf{r}/dt$ on the left-hand side of this equation vanishes (as of course does $d\boldsymbol{\omega}_L^2/dt$).

For a qualitative discussion of the relaxation of the structure it is sufficient to substitute the forms of $\mathbf{S}(\mathbf{r})$ and $\mathbf{J}_i(\mathbf{r})$ obtained under the assumption of stationarity, i.e., $S(\mathbf{r})=\mathrm{const}$ and the expression (6), into this equation. Some relatively simple calculations show that this yields

$$\frac{3}{w^2}\mathbf{J}_i^2=\frac{w^2}{3\kappa^2 S^2}(\nabla_i\mathbf{S})^2,$$

i.e., the current term is proportional to the inhomogeneity energy of the distribution of $\mathbf{S}$. This quantity can be estimated from the definition (17) of the characteristic scale ξ to be $\sim w^2/3\kappa^2\xi^2\sim|\omega_P-\omega_L|S/\kappa$. Therefore, because the absolute value of the magnetization at the start of relaxation is approximately the equilibrium value ($S\sim\omega_L$) and, in addition, $\kappa\omega_L\tau_1\gg 1$ (the collisionless regime) and $|\omega_P-\omega_L|\tau_1\ll 1$, the current term on the left-hand side can be neglected in comparison to S^2. Thus, the bulk relaxation reduces to damping of the absolute value of the magnetization, which is given by the small right-hand side:

$$\frac{dS^2}{dt}=-\frac{2w^2}{3\kappa^2\tau_1}\frac{1}{\pi R^2}\int d^2r(\nabla_i\hat{\mathbf{S}})^2. \tag{33}$$

Next, plugging in known dependences, we write

$$\int d^2r(\nabla_i\hat{\mathbf{S}})^2=\int_0^R\rho\,d\rho\int_0^{2\pi}d\varphi\,(\beta'^2+\alpha'^2\sin^2\beta) =2\pi\int_0^{R/\xi} r\,dr\left(\beta'^2+\frac{\sin^2\beta}{r^2}N^2\right),$$

where we have transformed from ρ to $r=\rho/\xi$. With allowance for the fact that $\beta'=0$ at the vessel boundary, integrating β'^2 by parts and then substituting the expression (23) finally yields

$$\frac{dS}{dt}=-\frac{2|\omega_P-\omega_L|}{\kappa\tau_1}C, \tag{34}$$

where C is a dimensionless integral that depends only on the actual form taken by the vortex:

$$C=\frac{1}{\pi(R/\xi)^2}\int_0^{R/\xi}2\pi r\,dr\left[\frac{N^2}{r^2}\sin\beta(\sin\beta-\beta\cos\beta) -\beta\,\mathrm{sgn}(\omega_P-\omega_L)\sin\beta\right]. \tag{35}$$

Here, in the limit $\beta\to 0$ we can expand in the small angle β. The expansion up to the quadratic term can be expressed in terms of the total longitudinal component of the magnetization direction averaged over the vessel:

$$C\approx\frac{\int\beta^2\,dr}{\int d^2r}\approx -2+\frac{2\int d^2r\cos\beta}{\int d^2r},$$

which is compared to β.

It can therefore be assume that C is negligible in the general case, so that it does not disrupt the slowness of the

relaxation process, i.e., the time during which S decreases linearly to zero: $\kappa\omega_L\tau_1/2|\omega_P-\omega_L|\geqslant\tau_1$. We note that this time is large by virtue of the condition $|\omega_P-\omega_L|\tau_1\leqslant 1$ for applicability of the theory under consideration. Since the total longitudinal magnetization $S\int d\mathbf{r}\cos\beta$ is an integral of Leggett's equations, a slow decrease in S leads to an increase (also slow) in the absolute value of the total longitudinal component of the magnetization direction $\int d\mathbf{r}\cos\beta$.

The stability of the family of vortex solutions and the slowness of their relaxation enable us to advance the intuitively simple hypothesis that the overall form of the solution does not vary during relaxation. Specifically, the vortex distribution in the class of solutions consisting of the family just indicated transforms during the relaxation process, so that a distribution specified by the instantaneous value of the total longitudinal component of the magnetization direction is realized at any given time. Since, however, a vortex with a fixed total longitudinal component of the magnetization direction can form only if the effective radius of the vessel R/ξ is fixed, variation of the total component of the magnetization direction during relaxation among the solutions of the family must lead to fine tuning of the precession frequency ω_P in such a manner that the characteristic scale ξ (17) maintains the ''correct'' effective radius.

Figure 1 depicts a vortex with $\beta(R/\xi)\approx 2.10$ and a vanishing total longitudinal component of the magnetization direction. This vortex structure is a limiting structure in the sense that vortices with $\beta(R/\xi)<2.10$ (and a total longitudinal component of the magnetization direction greater than zero) relax in such a case with a decrease in $\beta(R/\xi)$, i.e., to the spatially homogeneous distribution $\beta\equiv 0$, while vortices with $\beta(R/\xi)>2.10$ (and a total longitudinal component of the magnetization direction less than zero) relax with an increase in $\beta(R/\xi)$, i.e., to the limiting distribution (28). In the former case the effective radius $R/\xi\sim|\omega_P-\omega_L|^{1/2}$ increases to ≈ 1.84 (the precession frequency ω_P decreases[4]), and in the latter case the effective radius first decreases to ≈ 0.84 and then increases (ω_P, accordingly, first increases and then decreases).

When $\omega_P>\omega_L$ the behavior is different: for vortices with a total longitudinal component of the magnetization direction greater than zero the frequency ω_P first decreases and then increases, while for vortices with a total longitudinal component of the magnetization direction the frequency ω_P increases monotonically.

4. ONE-DIMENSIONAL FLOW

We consider the stationary flow of spin along a long ($L\gg\xi$), thin (transverse dimension $a\ll\xi$) channel of constant cross section oriented perpendicular to a magnetic field. In such a geometry it can be assumed that flow is one-dimensional and that α and β vary only in the direction parallel to the channel length:

$$\beta=\beta(y),\quad \alpha=\alpha(y).$$

Variation of the functional (11) with respect to α yields $(\alpha'\sin^2\beta)'=0$, whence we have

$$\alpha'\sin^2\beta=h=\text{const}. \tag{36}$$

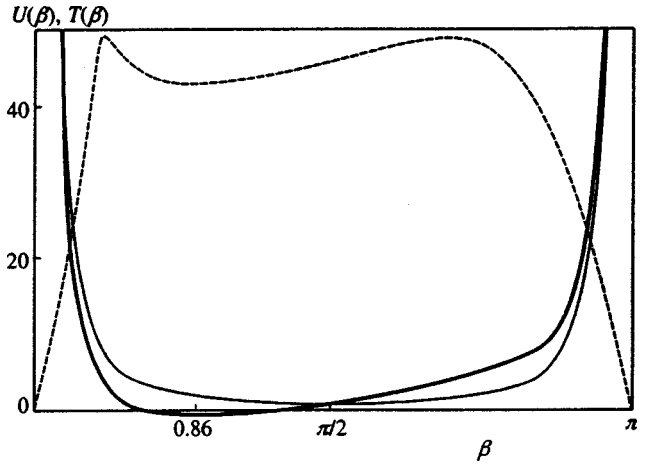

FIG. 2. Effective Newtonian potential for the deviation angle of the magnetization in one-dimensional spin transport, and the period of motion in that potential: light solid line — potential for $\omega_P=\omega_L$, i.e., the first term in (40); heavy solid line — potential for $\omega_P\neq\omega_L$ (for the special case $\xi=1$, $h^2=1/2$); dashed line — dependence of the period on the initial turning point (in arbitrary units).

It can be seen from this expression that if $\alpha'(y)=0$ or $\sin\beta(y)=0$ at some point, then $h=0$ and $\alpha'(y)\equiv 0$ or $\sin\beta(y)\equiv 0$. The case $\alpha'\not\equiv 0$ is of interest; therefore, we conclude from (36) that either $\sin\beta(y)\equiv 0$ or $\sin\beta(y)\neq 0$ at any y and

$$\alpha'=h/\sin^2\beta. \tag{37}$$

Variation of the functional (11) with respect to β yields

$$\beta''-(\alpha')^2\sin\beta\cos\beta-\frac{\operatorname{sgn}(\omega_P-\omega_L)}{\xi^2}\sin\beta=0. \tag{38}$$

The case $\sin\beta(y)\equiv 0$ is a solution of the problem, and given (37), substitution of the expression for α' yields an equation that describes $\beta(y)$ (and $\alpha(y)$ by virtue of (37)) along the channel:

$$\beta''-\frac{h^2\cos\beta}{\sin^3\beta}-\frac{\operatorname{sgn}(\omega_P-\omega_L)}{\xi^2}\sin\beta=0. \tag{39}$$

This is Newton's equation of motion for a particle of unit mass in the potential

$$U(\beta)=\frac{h^2}{2\sin^2\beta}+\frac{\operatorname{sgn}(\omega_P-\omega_L)}{\xi^2}\cos\beta, \tag{40}$$

in which the coordinate y plays the role of time. This potential is depicted for $\xi=1$ and $h^2=1/2$ in Fig. 2. Its minimum is at $\cos\beta=-\operatorname{sgn}(\omega_P-\omega_L)\sin^4\beta/(h\xi)^2$. Introducing the notation $\sin^2\beta=z$ and $(h\xi)^4=b$, we obtain $z^4=b(1-z)$, which has only one root for any $b>0$ in the interval $z\in(0,1)$. Equation (39) is invariant under the transformation (20). For definiteness, we take $\omega_P\leqslant\omega_L$.

The expression (6) for the nondissipative current in the one-dimensional case (the only nonvanishing orbital component is directed along the channel, so we omit the subscript) takes the form

$$\mathbf{J}=\frac{w^2}{3\kappa}\left[\begin{pmatrix}-\sin\alpha\\ \cos\alpha\\ 0\end{pmatrix}\beta'+\begin{pmatrix}-\sin\beta\cos\beta\cos\alpha\\ -\sin\beta\cos\beta\sin\alpha\\ \sin^2\beta\end{pmatrix}\alpha'\right]. \tag{41}$$

It can be seen from this expression that

$$J^z=\frac{w^2}{3\kappa}\alpha'\sin^2\beta=\frac{w^2}{3\kappa}h=\text{const},$$

$$J\equiv|\mathbf{J}|=\frac{w^2}{3\kappa}\sqrt{\beta'^2+\alpha'^2\sin^2\beta}.$$

We assume that the ends of the channel are connected to two reservoirs containing a Fermi liquid. We further assume that the spins are oriented at the same angle β to the magnetic field in the two reservoirs. In addition, the value of β does not depend on y in the reservoirs, so the boundary condition for Eq. (39) has the form

$$\beta(0)=\beta(L),\quad \beta'(0)=\beta'(L)=0. \tag{42}$$

With this boundary condition the analogy to classical motion acquires the simplest possible meaning: the initial deviation angle β_0 can be determined experimentally, and by virtue of (42), the initial ''kinetic energy'' β'^2 is zero, i.e., ''motion'' starts from the turning point of the potential $U(\beta_0)$, and by virtue of (42) should also end at it. In other words, an integer number of ''periods'' of $\beta(y)$ should fit into length L.

We first consider $\omega_P=\omega_L$ $(\xi=\infty)$. Then $U(\beta)=h^2/2\sin^2\beta$ and $\beta'=\sqrt{2[U(\beta_0)-U(\beta)]}$. We can write

$$dy=\frac{d\beta}{\beta'}=\frac{\sin\beta_0\sin\beta\ d\beta}{h\sqrt{x_0^2-x^2}}=-\frac{\sin\beta_0}{h}\frac{dx}{\sqrt{x_0^2-x^2}}, \tag{43}$$

where $x=\cos\beta$ and $x_0=\cos\beta_0$. The half-period of motion in this case is

$$T=\int dy=\frac{\sin\beta_0}{h}\arcsin\frac{x}{x_0}\Big|_{-x_0}^{x_0}=\pi\frac{\sin\beta_0}{h}. \tag{44}$$

The phase difference accrued during this time is

$$\Delta\alpha_T=\int\alpha'\,dy=\sin\beta_0\int_{-x_0}^{x_0}\frac{dx}{(1-x^2)\sqrt{x_0^2-x^2}}=\pi, \tag{45}$$

i.e., the phase difference does not depend on h.

Consequently, the condition (42) leads to $L=2nT$, where $n=1,2,\ldots$, hence

$$J^z\equiv\frac{w^2}{3\kappa}h=\frac{w^2}{3\kappa}\frac{2\pi n\sin\beta_0}{L}$$

for a phase difference $\Delta\alpha=2\pi n$ between the ends of the channel. As can be seen from the solution, in such a problem with the boundary condition (42) no stationary current-carrying states appear for any other phase difference. We note that in this case $J=h/\sin\beta_0=\text{const}$.

When $\omega_P\neq\omega_L$ $(\xi\neq\infty)$ the equation $U(\beta)=U(\beta_0)$ $(\beta\neq\beta_0)$ for the turning points reduces to the quadratic equation $x^2+ax+ax_0-1=0$ with roots

$$x_{1,2}=\frac{-a\pm\sqrt{a^2-4ax_0+4}}{2},\quad \text{where}\quad a=\frac{h^2\xi^2}{2(1-x_0^2)}. \tag{46}$$

If $x_0\in(-1,1)$, these roots satisfy $x_1\in(-1,1)$, $x_2<-1$. Then $x_1=x_0$, as it should, if β_0 yields the minimum of $U(\beta)$. We assume everywhere below that $x_1<x_0$, i.e., that β_0 is located to the right of the minimum, and integrate over the range from x_1 to x_0. Otherwise, the replacement $x_1\leftrightarrow x_0$ must be made in all formulas.

Thus, we can write

$$dy=\frac{d\beta}{\beta'}=-\frac{\xi dx}{\sqrt{2(x_0-x)(x-x_1)(x-x_2)}}, \tag{47}$$

and the formulas for the period and the phase difference can be reduced by the replacement $t=\sqrt{(x_0-x)(x_0-x_1)}$ to complete elliptic integrals in the canonical Legendre form (see, for example, Ref. 14):[5)]

$$T=\xi\int_{x_1}^{x_0}\frac{dx}{\sqrt{2(x_0-x)(x-x_1)(x-x_2)}}=\sqrt{\frac{2}{x_0-x_2}}K(k), \tag{48}$$

$$\Delta\alpha_T=h\xi\int_{x_1}^{x_0}\frac{dx}{(1-x^2)\sqrt{2(x_0-x)(x-x_1)(x-x_2)}}=\frac{h\xi}{\sqrt{2(x_0-x_2)}}\left[\frac{1}{1+x_0}\Pi(k,m)+\frac{1}{1-x_0}\Pi(k,p)\right], \tag{49}$$

where $\Pi(k,p)\equiv\Pi(\pi/2,k,p)$ and

$$k=\sqrt{\frac{x_0-x_1}{x_0-x_2}},\quad p=\frac{x_0-x_1}{1-x_0},\quad m=-\frac{x_0-x_1}{1+x_0}.$$

We note that in this notation the absolute value of the current is $J=(w^2/3\kappa)\times(\sqrt{2}/\xi)\sqrt{x-x_3}$, where we have introduced the additional notation $x_3=x_0-a$, and that the incoming current (the current at the entrance to the channel) and the outgoing current (the current at the exit from the channel), i.e., when $x=x_0$, are both $h/\sin\beta_0$. The mean absolute value of the current flowing through the channel in this case is

$$\langle J\rangle=\frac{w^2}{3\kappa}\frac{2n}{L}\int_{x_1}^{x_0}\frac{\sqrt{x-x_3}\,dx}{\sqrt{(x_0-x)(x-x_1)(x-x_2)}}=-\frac{w^2}{3\kappa}\frac{2n}{L}\frac{2(x_1-x_3)}{\sqrt{(x_0-x_3)(x_1-x_2)}}\times\Pi\left(\sqrt{\frac{x_0-x_1}{x_0-x_3}\frac{x_3-x_2}{x_1-x_2}},\frac{x_0-x_1}{x_0-x_3}\right). \tag{50}$$

Although we cannot write an explicit expression for $h(\Delta\alpha)$ in this case, the basic features of current flow remain the same as when $\omega_P=\omega_L$. Specifically, as can be seen from Fig. 3a, the condition $L=2nT$ for fixed L yields a discrete series of possible values of h, each of which corresponds to

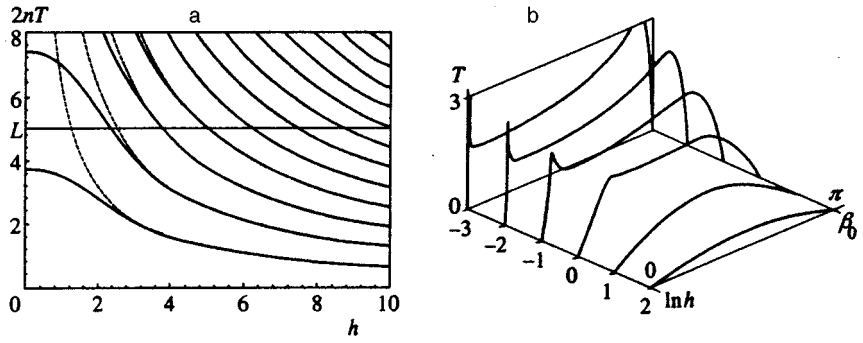

FIG. 3. a) Multiples of the half-period of motion T in the potential (40) as a function of h for $n=1,2,\ldots$ and the fixed initial angle $\beta_0=\pi/2$; b) dependence of the half-period on β_0 for several values of h. It is assumed that $\xi=1$. The dependence on β_0 is sinusoidal for large h, as when $\omega_P=\omega_L$, and tends to zero as $1/h$; at small h the half-period has a local minimum at the minimum of the potential (40) and two local maxima, one of which tends to zero, while the other tends to π with decreasing h. As can be seen in (a), if the channel length L is fixed, then $L=2nT$ holds for an infinite sequence of values of h, and by $h>3$ $T(h)$ is faithfully described by $T=\pi\sin\beta_0/h$ (the corresponding multiples are shown as dashed lines in (a)) and $h=2\pi n\sin\beta_0/L$, as when $\omega_P=\omega_L$.

its own number n of complete periods that fit into the channel length, and its own phase difference $\Delta\alpha$ between the channel ends.

In the limit $h\xi\gg 1$, the second term in the potential $U(\beta)$ can clearly be neglected. Then the problem reduces to the simpler case with $\omega_P=\omega_L$.

The situation in which the value of β is constant along the channel, i.e., $\beta'=0$, is special. Such a situation arises if β corresponds to the minimum of the potential (40) (this point is special, because motion starting there is aperiodic). In this case it follows from (37) that the derivative $\alpha'=g=\text{const}$ is constant along the channel. Accordingly, the phase difference between the ends is $\Delta\alpha=gL$.

The condition for a potential minimum in this notation takes the form

$$\cos\beta=-\frac{\operatorname{sgn}(\omega_P-\omega_L)}{(g\xi)^2}.$$

It follows from the range of the cosine that such a solution exists only when $g\xi\geqslant 1$. Otherwise, only $\sin\beta\equiv 0$ is possible. Thus, the current depends on the gradient g of the azimuthal angle α:

$$J^z=\begin{cases}0, & g\xi\leqslant 1,\\ \dfrac{w^2}{3\kappa}g\left(1-\dfrac{1}{(g\xi)^4}\right), & g\xi>1,\end{cases}\tag{51}$$

$$|\mathbf{J}|=\begin{cases}0, & g\xi\leqslant 1,\\ \dfrac{w^2}{3\kappa}g\sqrt{1-\dfrac{1}{(g\xi)^4}}, & g\xi>1.\end{cases}\tag{52}$$

This dependence is shown in Fig. 4.

For any value of the gradient, the spins in the channel are oriented at a fixed angle β_0 to the magnetic field; that angle is 0 for $g\xi<1$ and tends to $\pi/2$ as $g\xi\to\infty$:

$$\beta_0=\begin{cases}0, & g\xi\leqslant 1,\\ \arccos\dfrac{1}{g^2\xi^2}, & g\xi>1.\end{cases}\tag{53}$$

5. CONCLUSION

In this paper, we have investigated axisymmetric and one-dimensional quasisteady solutions of the equations for the spin dynamics of a normal Fermi liquid in a uniform magnetic field in the collisionless regime. Structures resembling a spin vortex with an integer circulation number that are stable against small perturbations show up in the axisymmetric case. In each such structure, magnetization parallel or antiparallel to the field at the vortex axis deviates smoothly with increasing distance from the axis, up to some angle at the vessel boundary. This angle, as well as the exact dependence of the deviation of the magnetization on the distance to the axis, are determined by the total longitudinal component of the magnetization direction $\int d\mathbf{r}\cos\beta$. The initial conditions of the experiment determine the total longitudinal magnetization $S\int d\mathbf{r}\cos\beta$, whose value does not vary during further evolution of the system, in accordance with Leggett's equations. At the beginning of the experiment, S can be assumed to take the equilibrium value $\chi_n\omega_L/\gamma^2$.

The requirement that no spin current flow through the chamber walls (when its size is fixed) maps every deviation angle at the lateral surface of the vessel to a value of the difference between the precession frequency and the Larmor

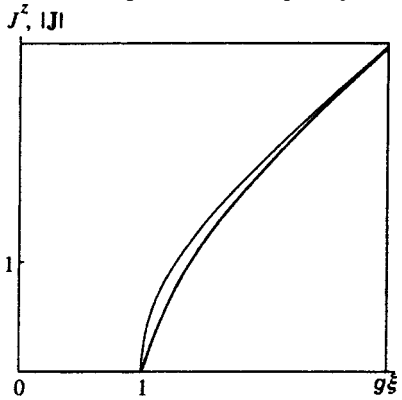

FIG. 4. Dependence of the current through the channel (in units of $w^2/3\kappa\xi$) on the phase difference at the channel ends when the deviation of the magnetization corresponds to the minimum of the potential (38): light line — absolute value of the current, heavy line — component along the field. At large values of the phase gradient $g\xi$, both dependences become asymptotically linear.

frequency in such a manner that the required effective (i.e., in units of the characteristic length ξ) vessel radius is ensured. The effective vessel radius needed to form a vortex has a lower bound ≈ 0.84 (for a vortex with circulation number $N=1$). As the effective radius tends to infinity, the deviation angle at the lateral surface tends to π, and equal precession and Larmor frequencies correspond to the magnetization distribution described by the Belavin–Polyakov skyrmion solution.

We seek solutions in the quasistationary approximation, with slow motion relative to the Larmor reference frame, i.e., in the limit in which the absolute value of the difference between the precession frequency and the Larmor frequency is small.

The collision integral in Leggett's equation for the spin current leads to relaxation of the absolute value of the magnetization S. The quasistationary approximation can then be equivalently reformulated as the condition that this relaxation be slow. The latter circumstance in conjunction with the stability of the entire family of vortex solutions suggests in this case that relaxation will be described by solutions in this family, i.e., the distribution of spin in the system at any instant will correspond to a vortex with a value of $\int d\mathbf{r}\cos\beta$ such that $S\int d\mathbf{r}\cos\beta$ remains constant, as required by Leggett's equations. Since, however, each vortex is uniquely characterized not only by the total longitudinal component of the magnetization direction, but also by the difference between its precession frequency and the Larmor frequency (see above), in this case relaxation should lead to variation of the precession frequency with time.

Stationary magnetization transport along a thin channel connecting two reservoirs with a Fermi liquid has also been investigated. It has been found that the dependence of the polar angle on distance along the channel is described in this case by Newton's equations of motion for a particle in a potential well with a single minimum. Distance along the channel plays the role of time. The spin current through the channel, the ''phase'' difference (i.e., the difference between the azimuthal angles) at the ends of the channel, as well as the exact form of the potential well, depend on the single parameter h. When there are definite fixed boundary conditions at the ends of the channel, the parametric dependence of the current through the channel on the phase difference between the ends is obtained.

The boundary conditions chosen in this paper correspond to a situation in which motion in the potential well must begin and end at the same turning point. In this case an integer number of periods of motion fit into the channel length. Therefore, if the polar angle β_0 in the reservoirs is fixed, current-carrying states can appear only at certain discrete values of the phase difference, which form an infinite sequence, each with its own value of the current. In the general case the current cannot be expressed analytically as an explicit function of the phase difference. However, the qualitative features of this function are the same as in the simpler case (which is also the limiting case for large h) of equal precession and Larmor frequencies, in which a phase difference that is a multiple of 2π is needed for current to flow. In this case

$$J^z \equiv \frac{w^2}{3\kappa}h = \frac{w^2}{3\kappa}\,\frac{2\pi n\sin\beta_0}{L}.$$

Only when β_0 corresponds to the potential minimum (for a given h) is there no such approach. In this case the motion is aperiodic, i.e., the polar angle is constant along the channel (and coincides with the polar angle in the reservoirs). It also turns out in this case that the production of a current-carrying state is possible only for phase gradients greater than the reciprocal of the characteristic scale ξ^{-1}.

This work was performed with partial financial support from the Statistical Physics Program of the Ministry of Sciences of the Russian Federation, by Grant Nos. 96-02-16041b and 96-15-96632 from the Russian Fund for Fundamental Research (State Program for Support of Scientific Seminars), and the INTAS Program (Grant 96-0610).

*E-mail: mineev@landau.ac.ru

[1)]Vectors in spin space are denoted by bold letters, and the components of vectors in orbital space are denoted by a subscript.

[2)]By virtue of (20), there clearly also exists an inverse structure, i.e., a skyrmion for which β varies from π (at 0) to 0 (at ∞).

[3)]Of course, because of invariance under the transformation (20), it is clear for $\omega_P > \omega_L$ that the asymptote (27) at infinity matches the asymptote $\pi - Ar^N$ near zero.

[4)]Recall that we assume $\omega_P < \omega_L$.

[5)]Because of the various existing definitions of p, we write out the integral of the third kind in explicit form:

$$\Pi(\varphi,k,p)=\int_0^{\sin\varphi}\frac{dt}{\sqrt{1-t^2}\sqrt{1-k^2t^2}(1+pt^2)}.$$

[1] V. V. Dmitriev and I. A. Fomin, JETP Lett. **59**, 378 (1994).
[2] V. V. Dmitriev, S. R. Zakazov, and V. V. Moroz, JETP Lett. **61**, 324 (1995).
[3] A. S. Borovik-Romanov, Yu. M. Bun'kov, V. V. Dmitriev *et al.*, Zh. Éksp. Teor. Fiz. **88**, 2025 (1985) [Sov. Phys. JETP **61**, 1199 (1985)].
[4] I. A. Fomin, Zh. Éksp. Teor. Fiz. **88**, 2039 (1985) [Sov. Phys. JETP **61**, 2039 (1985)].
[5] A. S. Borovik-Romanov, Yu. M. Bun'kov, V. V. Dmitriev *et al.*, JETP Lett. **45**, 124 (1987).
[6] A. S. Borovik-Romanov, Yu. M. Bunkov, V. V. Dmitriev *et al.*, Jpn. J. Appl. Phys., Part 1 **26**, Suppl. 26-3, 175 (1987).
[7] A. S. Borovik-Romanov, Yu. M. Bun'kov, A. de Vaard *et al.*, JETP Lett. **47**, 478 (1988).
[8] A. S. Borovik-Romanov, Yu. M. Bunkov, V. V. Dmitriev *et al.*, Physica B **165–166**, 649 (1990).
[9] I. A. Fomin, JETP Lett. **45**, 136 (1987).
[10] I. A. Fomin, Zh. Éksp. Teor. Fiz. **94**(6), 112 (1988) [Sov. Phys. JETP **67**, 1148 (1988)].
[11] A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 245 (1975).
[12] A. J. Leggett, J. Phys. C **3**, 448 (1970).
[13] Yu. G. Makhlin and V. P. Mineev, Zh. Éksp. Teor. Fiz. **109**, 441 (1996) [JETP **82**, 234 (1996)].
[14] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions*, Dover, New York (1979).

Translated by P. Shelnitz

Continuum description of anomalous diffusion on a comb structure

I. A. Lubashevskiĭ*) and A. A. Zemlyanov

General Physics Institute, Russian Academy of Sciences, 117942 Moscow, Russia; M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 3 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 1284–1312 (October 1998)

Anomalous diffusion on a comb structure consisting of a one-dimensional backbone and lateral branches (teeth) of random length is considered. A well-defined classification of the trajectories of random walks reduces the original problem to an analysis of classical diffusion on the backbone, where, however, the time of this process is a random quantity. Its distribution is dictated by the properties of the random walks of the diffusing particles on the teeth. The feasibility of applying mean-field theory in such a model is demonstrated, and the equation for the Green's function with a partial derivative of fractional order is obtained. The characteristic features of the propagation of particles on a comb structure are analyzed. We obtain a model of an effective homogeneous medium in which diffusion is described by an equation with a fractional derivative with respect to time and an initial condition that is an integral of fractional order. © *1998 American Institute of Physics.* [S1063-7761(98)01010-5]

1. INTRODUCTION

Diffusion in heterogeneous media containing inhomogeneities of various scales, particularly on fractals, has anomalous properties (for reviews see Refs. 1–3). For example, in such media without a preferred direction, the mean-square displacement of a Brownian particle $\langle x^2\rangle$ depends on the time t as $\langle x^2\rangle \propto t^{\alpha}$, where $\alpha \neq 1$. A self-consistent, general description of anomalous transport is presently lacking, so various models of inhomogeneous media are investigated individually. Typical examples of such media include percolation clusters and clusters that emerge during diffusion-controlled aggregation.

From the standpoint of diffusive transport, such a cluster can be represented as consisting of an infinitely extended backbone and a large number of lateral branches. The random walk of a diffusing particle on the backbone of the cluster largely determines its movement in space, while the lateral branches serve as specific traps where the particle spends most of its time (see, for example, Refs. 1–3). The degree of branching of the backbone is small,[4] so the model of a comb structure with lateral branches (teeth) of random length (Fig. 1) was proposed to describe diffusion in such systems in Refs. 5 and 6. If the properties of percolation clusters are taken into account, distributions for which all moments of the tooth length distribution diverge can also be considered.

Diffusion along the backbone of a comb structure with teeth of identical length is described by the classical law $\langle x^2\rangle \propto t$, if the length of the branches is small, and is anomalous, $\langle x^2\rangle \propto t^{1/2}$, if their length is infinite.[7,8] For a comb structure with teeth of random length it is natural to expect that the function $\langle x^2(t)\rangle$ will also be a power function with an exponent $\alpha \in (1/2,1)$. In fact, in Refs. 9 and 10 the properties of anomalous transport were in fact investigated in the classical mean-field approximation. In particular, it was found that $\alpha = \gamma/2$ for a power-law distribution $f(l) \propto l^{-\gamma}$ of the tooth length l, with γ in the range $1 < \gamma < 2$. If all moments of $f(l)$ converge, diffusion on the comb structure can be described by classical laws.[11] The properties of such diffusion under the influence of a topological drift field were investigated in a similar approximation in Refs. 12 and 13.

The value of α for exponential and power-law distributions of the lengths of the lateral branches with and without a drift field along the backbone was found by isolating the leading terms with respect to t in the lattice description of Refs. 14 and 15. In particular, in the presence of a drift field, $\alpha = \gamma$ for such a power-law distribution of the tooth lengths. This problem was studied in the presence of a drift field along the teeth in Ref. 16. The Green's function in the absence of a drift field was also found for the lattice model, but without any justification for the modification of the mean-field approximation.[17]

In this paper we develop a continuum description of diffusive transport on a comb structure. Using a well-defined classification of random-walk trajectories, we reduce the original problem to a classical description of random walks on the backbone, where, however, the time of each random walk is a random quantity. Its distribution is dictated by the random walks of diffusing particles on the teeth. We ultimately obtain a model of a homogeneous medium with anomalous properties. In particular, diffusion in this medium is described by an equation with a fractional derivative with respect to time and a corresponding initial condition that is an integral of fractional order.

2. MODEL

Let the x axis represent the backbone of the comb structure under consideration (Fig. 1). The teeth $\{i\}$ $(i \in \mathbb{Z})$ are attached perpendicularly to it (parallel to the y axis). They

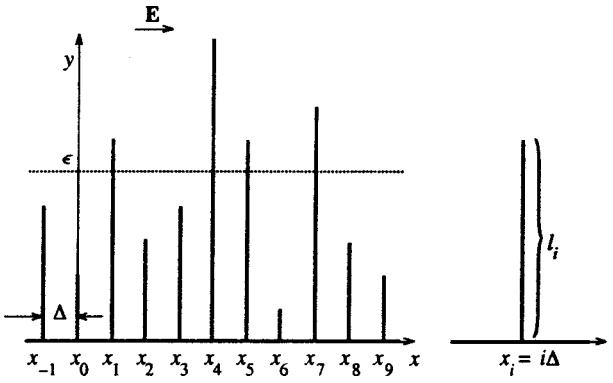

FIG. 1. Model of a comb structure: **E**—drift field; points—the ''observation line'' $y=\epsilon$; x_i—tooth coordinates; l_i – tooth lengths.

are separated by a distance Δ, and their lengths $\{l_i\}$ are independent random quantities. Their distribution $f(l)$ is described by the power law

$$f(l)=(\gamma-1)\frac{l_0^{\gamma-1}}{l^{\gamma}}, \tag{2.1}$$

with exponent $\gamma\in(1,2)$, and $l_0\sim\Delta$ is the minimum tooth length. The function (2.1) satisfies the following normalization and divergence conditions for all moments:

$$\int_{l_0}^{\infty} f(l)\,dl=1, \quad \int_{l_0}^{\infty} l^n f(l)\,dl=\infty, \quad n=1,2,\ldots .$$

As will be seen below, these conditions ensure the feasibility of describing anomalous diffusion in terms of a homogeneous medium.

The diffusion law for Brownian particles on such a structure is given by the following equations: for the backbone, away from the points of tooth attachment $\{x_i=i\Delta\}$, i.e., at $x\neq x_i$,

$$\frac{\partial c}{\partial t}=D\frac{\partial^2 c}{\partial x^2}-\mu E\frac{\partial c}{\partial x}, \tag{2.2}$$

and for the ith tooth

$$\frac{\partial C_i}{\partial t}=D\frac{\partial^2 C_i}{\partial y^2}. \tag{2.3}$$

Here $c(x,t)$ is the concentration of diffusing particles on the backbone, $C_i(y,t)$ is their concentration on tooth i, D is the diffusion coefficient, μ is the mobility, and, in addition, we assume that a uniform field E acts along the x axis. At the points of attachment $\{x_i\}$ the concentration of particles and their fluxes along the backbone and the teeth

$$j=-D\frac{\partial c}{\partial x}+\mu Ec, \quad J_i=-D\frac{\partial C_i}{\partial y},$$

respectively, satisfy the continuity equations

$$c(x_i,t)=C_i(0,t), \quad j|_{x=x_i-0}=j|_{x=x_i+0}+J|_{y=+0}. \tag{2.4}$$

The condition on the reflection of particles at the ends of the teeth is

$$\left.\frac{\partial C_i}{\partial y}\right|_{y=l_i}=0. \tag{2.5}$$

Equations (2.2)–(2.5) comprise a complete microscopic description of diffusion on a comb structure.

To justify going to the continuum limit, Δ must be a small quantity. Specifically, we assume that

$$\Delta^2\ll Dt \quad \text{and} \quad \mu E\Delta\ll D. \tag{2.6}$$

The first inequality means that the characteristic diffusion length at times t considered here significantly exceeds the distance between teeth Δ, and the second means that the influence of the drift field E is small on scales of length Δ.

3. GREEN'S FUNCTION. CLASSIFICATION OF THE TRAJECTORIES OF DIFFUSING PARTICLES

For simplicity, we analyze the diffusion of particles on the backbone when all particles are localized on the backbone with density $c_0(x)$ at the initial time $t=0$. In this case the solution of the problem (2.2)–(2.5) reduces to finding the Green's function $G(x,x'|t)$, which specifies the concentration $c(x,t)$ at time t:

$$c(x,t)=\int_{-\infty}^{\infty} G(x,x'|t)c_0(x')\,dx'. \tag{3.1}$$

The Green's function

$$G(x,x'|t)=G(x,y,t|x',0)|_{y=0}$$

is the solution of the problem under consideration with a δ-function initial condition, and can be interpreted as the probability density of finding a diffusing particle at x at time t given that it was at x' at the initial time $t=0$. All possible particle trajectories that begin at x' and reach x after a time t contribute to this probability. In addition, they are independent because of the linearity of the system of equations (2.2)–(2.5). This enables us to represent the Green's function as an integral over the trajectories

$$G(x,x'|t)=\int_{\substack{x(0)=x'\\ y(0)=0}}^{\substack{x(t)=x\\ y(t)=0}} \mathscr{D}\{x(t'),y(t')\}\mathscr{P}_{\mathrm{tr}}, \tag{3.2}$$

where $\mathscr{P}_{\mathrm{tr}}$ is the statistical weight of the trajectory $\{x(t'),y(t')\}_{t'=0}^{t'=t}$ and $\mathscr{D}\{x(t'),y(t')\}$ is its Wiener measure.

Let us construct the following classification of trajectories (Fig. 2). We introduce an ''observation line,'' the horizontal straight line $y=\epsilon$, where ϵ satisfies

$$\Delta\sim l_0\ll\epsilon\ll\sqrt{Dt},\Delta\left(\frac{D}{\mu E\Delta}\right)^{2/\gamma} \tag{3.3}$$

(the meaning of the latter bound on ϵ will shortly become clear). The first segment of a trajectory, $P^{\mathrm{in}}\{x',i_1,\tau_1\}$, begins at x' and corresponds to random walks of a particle within the region $\Gamma_\epsilon=\{(x,y):y\in[0,\epsilon)\}$, i.e., without crossing the observation line. It ends at (x_{i_1},ϵ) at time τ_1, when the particle first crosses the observation line. The next segment of the trajectory, $P^{\mathrm{out}}\{i_1,t_1\}$, corresponds to random walks of the particle on tooth i_1 without touching the x axis. It ends at time t_1, when the particle first arrives at x_{i_1}. The

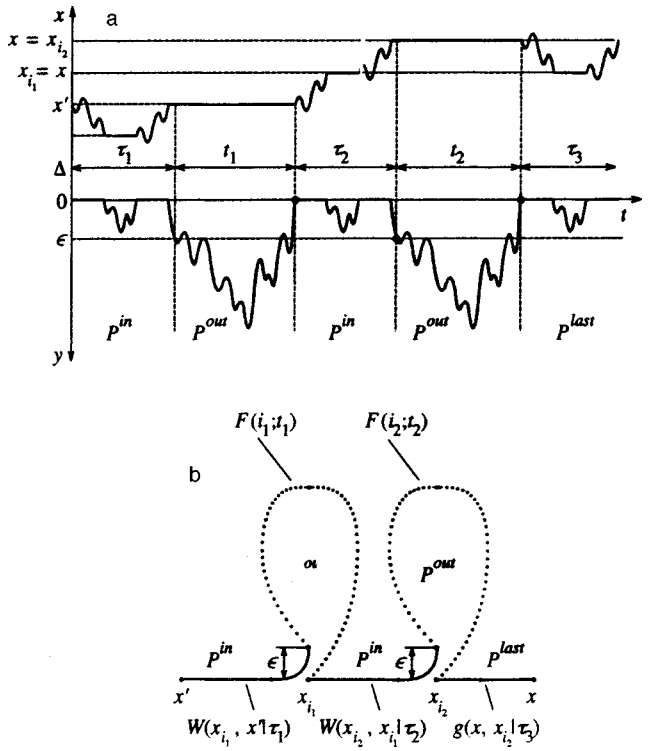

FIG. 2. Characteristic segments of the trajectory of a random walker that was at x' at time $t'=0$ and arrived at x at time t: a — schematic representation of this trajectory in the space $\{x,y\}$; b — the corresponding diagram. For simplicity, a trajectory containing five characteristic segments for a random walker on the backbone and on the teeth of a comb structure is shown.

subsequent structure of the trajectory is composed of $N-1$ successive repetitions of this pair of segments representing random walks on the comb structure within the region Γ_ϵ and on teeth without touching the backbone:

$$P^{\rm in}\{i_1,i_2,\tau_2\};\quad P^{\rm out}\{i_2,t_2\};\ \ldots;$$

$$P^{\rm in}\{i_{N-1},i_N,\tau_N\};\quad P^{\rm out}\{i_N,t_N\}$$

($P^{\rm in}\{i_{k-1},i_k,\tau_k\}\equiv P^{\rm in}\{x_{i_{k-1}},i_k,\tau_k\}$), with the exception of the last segment $P^{\rm last}\{i_N,x,\tau_{N+1}\}$. The last segment begins at x_{i_N} and represents a random walk of duration τ_{N+1} within the region Γ_ϵ, which ends at x on the x axis.

This classification enables us to perform the integration in (3.2) first, assuming that the parameters

$$N,\{i_k,\tau_k,t_k\}_{k=1}^{k=N},\quad \tau_{N+1}$$

are given, and then to integrate the result over these parameters (summing over i_k and N). We thereby obtain

$$G(x,x'|t)=g(x,x'|t)+\sum_{N=1}^{\infty}\int_0^t d\tau_{i_{N+1}}\sum_{i_N=-\infty}^{\infty} g(x,x_{i_N}|\tau_{i_{N+1}})\times\Phi_N(x_{i_N},x'|t-\tau_{i_{N+1}}),\qquad(3.4)$$

where

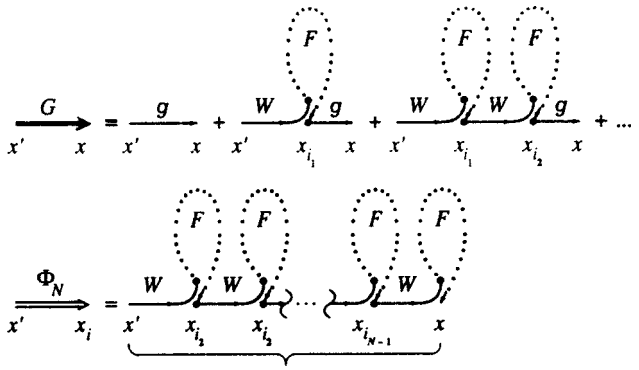

FIG. 3. Diagrams of Eqs. (3.4) and (3.5) for the Green's function $G(x,x')$ and diagrammatic representation of $\Phi(x_i,x)$ (the relations (3.5) and (3.7)). The bracket denotes the convolution $\{W*F\}$ of N factors.

$$\Phi_N(x_{i_N},x'|t)=\sum_{i_1=-\infty}^{\infty}\cdots\sum_{i_N=-\infty}^{\infty}\int_0^{\infty}\cdots\int_0^{\infty}\prod_{k=1}^{N}d\tau_k\,dt_k\times\delta\left[t-\sum_{j=1}^{N}(\tau_j+t_j)\right]W(x_{i_k},x_{i_{k-1}}|\tau_k)\times F(i_k,t_k).\qquad(3.5)$$

Here $g(x,x'|t)$ is the probability of arriving at x from x' in time t without touching the observation line. The analogous function $g(x,x_{i_N}|\tau_{N+1})$ describes the last segment of the trajectory for $N\geqslant 1$. The factor $W(x_{i_k},x_{i_{k-1}}|\tau_k)$ is the probability of first arriving at (x_{i_k},ϵ) on the observation line in time τ_k if the particle was at $x_{i_{k-1}}$ at the initial time. In $W(x_{i_1},x_{i_0}|\tau_1)$ the symbol x_{i_0} formally denotes the initial point of the trajectory: $x_{i_0}=x'$. The factor $F(i_k,t_k)$ is the probability of first arriving at the x axis in time t_k during a random walk on tooth i_k for a particle that was at $y=\epsilon$ at the initial time. These factors and trajectory segments are compared in Fig. 2a. Figure 3 presents Eqs. (3.4) and (3.5) in diagrammatic form.

Since the problem under consideration is translationally invariant with respect to time, it is convenient to take Laplace transforms:

$$(\ldots)(s)=\int_0^{\infty}dt'\,e^{-st'}(\ldots)(t'),$$

where we assume that the argument $s\sim 1/t$. This enables us to rewrite (3.4) and (3.5) in the form

$$G(x,x'|s)=g(x,x'|s)+\sum_{N=1}^{\infty}\sum_{i_N=-\infty}^{\infty}g(x,x_{i_N}|s)\Phi_N(x_{i_N},x'|s),\qquad(3.6)$$

$$\Phi_N(x_{i_N},x'|s)=\sum_{i_1=-\infty}^{\infty}\cdots\sum_{i_{N-1}=-\infty}^{\infty}\prod_{k=1}^{N}W(x_{i_k},x_{i_{k-1}}|s)F(i_k,s).\qquad(3.7)$$

To complete this description we need relations that specify the principal factors in (3.6) and (3.7); these form the subject of the remainder of this section.

3.1. Green's function for the elementary segment P^{out}

The general expression for the Laplace transform $F(i,s)$ of the probability of first arriving at the x axis in time t during a random walk on tooth i for a particle that was initially at $y=\epsilon$ is derived in Appendix A. Given the constraints on ϵ (see (3.3)), to first order in the small quantity $\epsilon\sqrt{s/D}$ this expression reduces to

$$F(i,s)=1-\epsilon\sqrt{\frac{s}{D}}\tanh\left(l_i\sqrt{\frac{s}{D}}\right), \tag{3.8}$$

which is the expression desired in this part of the paper.

3.2. Green's function for the elementary segment P^{in}

We first consider the Laplace transform $g(x,x'|s)$ of the probability of a particle arriving at x from x' in time t during a random walk within the region Γ_ϵ. Such a random walk is accompanied not only by displacement of the particle along the backbone, but also by visits to points on nearby teeth. This process is described by (2.2)–(2.5) and the vanishing boundary condition $C_i(\epsilon,s)=0$ at $y=\epsilon$ for teeth whose length $l_i>\epsilon$. In Appendix B (see Eq. (B10)) we use this description of $g(x,x'|s)$ to obtain

$$sg=D\frac{\partial^2 g}{\partial x^2}-\mu E\frac{\partial g}{\partial x}-\eta(x)g+\delta(x-x'), \tag{3.9}$$

where

$$\eta(x)=\sqrt{sD}\left[\sum_{i=-\infty}^{\infty}b_i\delta(x-x_i)\right] \tag{3.10}$$

describes the rate of escape of the random walkers from the backbone, and the coefficients b_i are random numbers given by (B8), which can be rewritten in the limit (3.3) as

$$b_i=\begin{cases}l_i\sqrt{\dfrac{s}{D}}, & l_i\leqslant\epsilon\\[2ex] \dfrac{1}{\epsilon}\sqrt{\dfrac{D}{s}}+\dfrac{\epsilon}{3}\sqrt{\dfrac{s}{D}}, & l_i>\epsilon.\end{cases} \tag{3.11}$$

Finally, to complete the problem we need an expression for the probability $W(x_i,x'|t)$ of first arriving at tooth i on the observation line $y=\epsilon$ in time t when the particle starts on the backbone at x'. As shown in Appendix B (see B12)), the Laplace transform of this function in the limit (3.3) is

$$W(x_i,x'|s)=\frac{D}{\epsilon}\left(1-\frac{\epsilon^2 s}{6D}\right)g(x_i,x'|s)\,\theta_i^\epsilon, \tag{3.12}$$

where $\theta_i^\epsilon=1$ for $l_i>\epsilon$ and $\theta_i^\epsilon=0$ for $l_i\leqslant\epsilon$.

The relations (3.6)–(3.9) and (3.12) comprise a complete description of the present diffusion problem on a comb structure for a given distribution of the tooth lengths $\{l_i\}$. Subsequent analysis will be given over to investigating typical properties of such diffusion with ensemble averaging over $\{l_i\}$.

As can be seen from the expressions obtained, the Green's function $g(x,x'|s)$ plays a significant role in the approach described here. Therefore, before proceeding to further derivations, we consider some salient properties.

We separate $g(x,x'|s)$ into a regular (i.e., averaged over the ensemble $\{l_i\}$) part $\overline{g}(x,x'|s)$ and a random component $\delta g(x,x'|s)$:

$$g(x,x'|s)=\overline{g}(x,x'|s)+\delta g(x,x'|s). \tag{3.13}$$

In the zeroth approximation with respect to Δ/ϵ the function $\overline{g}(x,x'|s)$ satisfies Eq. (3.9) after the replacement of $\eta(x)$ by its averaged value $\overline{\eta}$ and neglect of the small-amplitude variations of $\overline{g}(x,x'|s)$ over the distance Δ, due to the discrete positions of the teeth. In this approximation (see Appendix B) for $\overline{g}(x,x'|s)$ we have

$$s\Omega_\epsilon\overline{g}=D\frac{\partial^2\overline{g}}{\partial x^2}-\mu E\frac{\partial\overline{g}}{\partial x}-\frac{2D}{\mathscr{l}_\epsilon^2}\overline{g}+\delta(x-x'), \tag{3.14}$$

where

$$\Omega_\epsilon=1+\frac{(2\gamma-1)\,l_0}{3(2-\gamma)\,\Delta}\left(\frac{\epsilon}{l_0}\right)^{2-\gamma} \tag{3.15}$$

is the dimensionless capacity of the layer Γ_ϵ, and

$$\mathscr{l}_\epsilon=\sqrt{2\Delta l_0}\left(\frac{\epsilon}{l_0}\right)^{\gamma/2} \tag{3.16}$$

is the rms length of the elementary segment $P^{\rm in}$. It also follows from (3.14) that the mean duration τ_ϵ of $P^{\rm in}$ and the corresponding mean residence time τ_ϵ^b of a particle on the backbone are

$$\tau_\epsilon=\frac{\Omega_\epsilon\mathscr{l}_\epsilon^2}{2D}\simeq\frac{2\gamma-1}{3(2-\gamma)}\frac{\epsilon^2}{D},\quad \tau_\epsilon^b=\frac{\mathscr{l}_\epsilon^2}{2D}=\frac{\sqrt{\Delta l_0}}{\sqrt{2}D}\left(\frac{\epsilon}{l_0}\right)^{\gamma/2}. \tag{3.17}$$

In this approximation the random component $\delta g(x,x'|s)$ is given by the expression (Appendix B, Eq. (B29))

$$\delta g(x,x'|s)=-\sqrt{sD}\sum_{i=-\infty}^{\infty}\delta b_i\,\overline{g}(x-x_i|s)\,\overline{g}(x_i-x'|s), \tag{3.18}$$

where $\delta b_i=b_i-\langle b\rangle$ is the random component of the coefficient b_i. Then the ratio of the Fourier components δg_k and $\overline{g}_k$ of the functions δg and $\overline{g}$ can serve as a measure of the influence of random inhomogeneities of the geometry of the comb structure on the random walks of particles within the layer Γ_ϵ. According to Eq. (B33),

$$\frac{\langle|\delta g_k|^2\rangle}{|\overline{g}_k|^2}\sim\left(\frac{l_0}{\epsilon}\right)^{(2-\gamma)/2}\ll 1, \tag{3.19}$$

and thus when $\gamma<2$ and $\epsilon\gg l_0\sim\Delta$, $\delta g(x,x'|s)$ is small compared to the regular part $\overline{g}(x,x'|s)$, and in the first approximation the Green's function $g(x,x'|s)$ can be considered identical to its regular part. We note that this statement in fact follows from the condition that there be a large number of teeth with length $l_i>\epsilon$ over a characteristic length of $P^{\rm in}$. Specifically, the number of such teeth

$$N_\epsilon\sim\frac{\mathscr{l}_\epsilon}{\Delta}P_\epsilon\sim\left(\frac{\epsilon}{l_0}\right)^{(2-\gamma)/2}\gg 1, \tag{3.20}$$

where

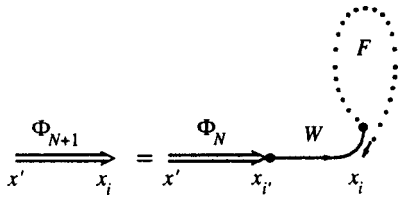

FIG. 4. Diagram of Eq. (4.1) for the Green's function $\Phi_N(x_i,x')$. The large points on the diagram denote summation and integration over the corresponding arguments.

$$P_\epsilon=\int_\epsilon^\infty dl\, f(l)=\left(\frac{l_0}{\epsilon}\right)^{\gamma-1} \tag{3.21}$$

is the probability of finding a tooth of length $l_i>\epsilon$.

By virtue of (3.3) and (3.17), a typical trajectory of a random walk of a diffusing particle includes a large number, $N\gg 1$, of the elementary segments P^{in} and P^{out}. Thus, after appropriate regularization, the Green's function $\Phi_N(x_i,x'|s)$ varies little in response to variations of the arguments x_i and x' on scales not only of order Δ, but also of order $\diagup\!\!\!\epsilon$. This enables us to regard the regularized Green's function $\Phi_N(x_i,x'|s)$ as a continuous function of its arguments, making it possible to model diffusion on a comb structure in terms of an effective homogeneous medium with anomalous properties. The next section is devoted to devising this regularization procedure and deriving the corresponding differential equation for the regularized Green's function.

4. FOKKER–PLANCK EQUATION FOR THE GREEN'S FUNCTION $\Phi_N(x,x'|s)$

Let us write (3.7) as a Smoluchowski–Kolmogorov equation, which relates the functions $\Phi_{N+1}(x,x'|s)$ and $\Phi_N(x,x'|s)$:

$$\Phi_{N+1}(x_i,x'|s)=F(i,s)\sum_{j=-\infty}^{\infty} W(x_i,x_j|s)\Phi_N(x_j,x'|s), \tag{4.1}$$

which corresponds to the diagram in Fig. 4. We next consider large values of the argument $N\gg 1$. In this case it is natural to expect that the functions $\Phi_{N+1}(x_i,x'|s)$ and $\Phi_N(x_i,x'|s)$ will differ only slightly, in which case we can consider the argument N to be continuous and put

$$\Phi_{N+1}(x_i,x'|s)-\Phi_N(x_i,x'|s)=\frac{\partial\Phi_N(x_i,x'|s)}{\partial N}.$$

However, spatial fluctuations of $\Phi_N(x_i,x'|s)$ remain significant when the coordinate x_i varies on scales of length Δ, since $\Phi_N(x_i,x'|s)\equiv 0$ when $l_i<\epsilon$ by virtue of (3.12). Nevertheless, these fluctuations do not play an appreciable role, since the relationship between the desired Green's function $G(x,x'|s)$ and $\Phi_N(x_i,x'|s)$ is given by the integral relation (3.6), and the kernel $g(x,x'|s)$ of that relation has a localization region whose width is of order $\diagup\!\!\!\epsilon$. The function $g(x,x'|s)$ varies weakly as a function of the argument x' on scales of length Δ, and its localization region contains a large number of teeth of length $l_i>\epsilon$ (see (3.20)). It would therefore be reasonable to smooth the function $\Phi_N(x_i,x'|s)$ with respect to x_i on scales smaller than $\diagup\!\!\!\epsilon$ and treat it as a continuous function of $x=x_i$, x', and $t_b\equiv N\tau_\epsilon^b$:

$$\Phi_N(x_i,x'|s)\to\Phi(x,x'|t_b,s).$$

The use of the argument t_b is justified because it is the mean residence time of a random walker on the backbone of the comb structure.

A Fourier transform-based smoothing procedure is devised in Appendix C for wave numbers $k\ll 1/\diagup\!\!\!\epsilon$. It is also shown there that by following Pontryagin's method,[18] the Smoluchowski–Kolmogorov equation (4.1) for $\Phi_N(x_i,x'|s)$ can be reduced to the following Fokker–Planck equation for $\Phi(x,x'|t_b,s)$:

$$\tau_\epsilon^b\frac{\partial\Phi}{\partial t_b}=\frac{\partial^2}{\partial x^2}[D_\epsilon(x)\Phi]-\frac{\partial}{\partial x}[\vartheta(x)\Phi]-\upsilon(x,s)\Phi. \tag{4.2}$$

The kinetic coefficients in this equation are

$$D_\epsilon(x)=\frac{1}{2}\sum_{j=-\infty}^{\infty}(x_j-x)^2W(x_j,x|0)=D\int_{-\infty}^{\infty}d\tilde{x}\,g(\tilde{x},x|0)$$
$$+\mu E\int_{-\infty}^{\infty}d\tilde{x}\,(\tilde{x}-x)g(\tilde{x},x|0), \tag{4.3}$$

$$\vartheta(x)=\sum_{j=-\infty}^{\infty}(x_j-x)W(x_j,x|0)=\mu E\int_{-\infty}^{\infty}d\tilde{x}\,g(\tilde{x},x|0), \tag{4.4}$$

$$\upsilon(x,s)=\epsilon\sqrt{\frac{s}{D}}\sum_{i=-\infty}^{\infty}\tanh\left(l_i\sqrt{\frac{s}{D}}\right)W(x_j,x|0). \tag{4.5}$$

For $N=1$, to a first approximation we can put

$$\Phi_1(x_i,x'|s)\approx W(x_i,x'|0).$$

In this case the width of the localization region of $\Phi_1(x_i,x'|s)$ is of the order of $\diagup\!\!\!\epsilon$, and by virtue of (B17),

$$\sum_{i=-\infty}^{\infty}\Phi_1(x_i,x'|s)\approx 1.$$

This enables us to supplement Eq. (4.2) with the formal initial condition

$$\Phi(x,x'|t_b,s)|_{t_b=0}=\Delta\,\delta(x-x'), \tag{4.6}$$

since in going over to the continuum limit we make the replacement

$$\sum_{i=-\infty}^{\infty}\to\frac{1}{\Delta}\int_{-\infty}^{\infty}dx_i.$$

The ensuing analysis will be based on the solution of Eq. (4.2), using the mean values of its kinetic coefficients because of the small magnitude of their random inhomogeneous perturbations. However, a comb structure is a low-dimensional system, so in general the mean-field approximation for Fokker–Planck type equations can break down under even the smallest of perturbations (see, for ex-

ample, Ref. 19). Therefore, use of the mean-field approximation to describe a given system requires special justification, which we provide in the next section.

4.1. Mean-field approximation

We can show that the influence of random fluctuations of the kinetic coefficients (4.3)–(4.5) is negligible under the strictures of (3.3). We apply the mixed Laplace and Fourier transformations

$$(\ldots)(\sigma)=\int_0^\infty dt_b\, e^{-\sigma t_b}(\ldots)(t_b),$$

$$(\ldots)(k)=\int_{-\infty}^\infty dx\, e^{-ikx}(\ldots)(x)$$

to Eq. (4.2). From (4.2) we then obtain the following equation for the transform $\Phi^*(k,x'|\sigma,s)$ of the Green's function $\Phi(x,x'|t_b,s)$:

$$L_k^*\Phi_k^*=-\frac{1}{2\pi}\int_{-\infty}^\infty dp\ \delta L_{k,k-p}^*\Phi_p^*+\Delta e^{-ikx'}, \tag{4.7}$$

where

$$L_k^*=\tau_\epsilon^b\sigma+D_\epsilon k^2+i\vartheta k+\upsilon;$$

$$\delta L_{k,p}^*=k^2\delta D_\epsilon(p)+ik\delta\vartheta(p)+\delta\upsilon(p);$$

D_ϵ, ϑ, and υ are the mean values of the corresponding kinetic coefficients, and $\delta D_\epsilon(p)$, $\delta\vartheta(p)$, and $\delta\upsilon(p)$ are the Fourier transforms of their fluctuation components.

It can be seen from (4.7) that in the zeroth approximation with respect to the fluctuation operator $\delta L_{k,p}^*$, the regular part $\overline{\Phi}_k^*$ of the transform $\Phi^*(k,x'|\sigma,s)$ of the Green's function is

$$\overline{\Phi}_k^*=\frac{\Delta e^{-ikx'}}{L_k^*} \tag{4.8}$$

and that in the first approximation its fluctuation component $\delta\Phi_k^*$ is

$$\delta\Phi_k^*=-\frac{e^{ikx'}}{2\pi\Delta}\,\overline{\Phi}_k^*\int_{-\infty}^\infty dp\ \delta L_{k,k-p}^*\overline{\Phi}_p^*. \tag{4.9}$$

When $t_b\gg\tau_\epsilon^b$ ($N\gg1$), the width ℓ_Φ of the localization region of $\Phi(x,x'|t_b,s)$ significantly exceeds the scale length ℓ_ϵ, i.e., the width of the localization region of $g(x_i,x|s)$: $\ell_\Phi\gg\ell_\epsilon$. Therefore, for Φ_k^* the significant values of the wave numbers lie in the range $|k|\lesssim1/\ell_\Phi\ll1/\ell_\epsilon$. As can be seen from (4.3)–(4.5), the characteristic correlation length of the kinetic coefficients $D_\epsilon(x)$, $\vartheta(x)$, and $\upsilon(x)$ is of order ℓ_ϵ. Therefore, in (4.9) it can be assumed that the random fluctuations $\delta D_\epsilon(x)$, $\delta\vartheta(x)$, and $\delta\upsilon(x)$ are δ-correlated. From (4.9) we can then assess the influence of random inhomogeneities of comb structure geometry on the Green's function $\Phi(x,x'|t_b,s)$:

$$\left\langle\left|\frac{\delta\Phi_k^*}{\overline{\Phi}_k^*}\right|^2\right\rangle=\frac{\Lambda}{2\pi}\int_{-\infty}^\infty dp\,\frac{1}{L_p^*L_{-p}^*}, \tag{4.10}$$

where

$$\Lambda=\int_{-\infty}^\infty dx'\langle\delta L_k^*(x)\,\delta L_{-k}^*(x')\rangle, \tag{4.11}$$

$$\delta L_k^*(x)=k^2\delta D_\epsilon(x)+ik\delta\vartheta(x)+\delta\upsilon(x). \tag{4.12}$$

The integral (4.10) is evaluated in Appendix D (see Eq. (D17)), and yields the estimate

$$\left\langle\left|\frac{\delta\Phi_k^*}{\overline{\Phi}_k^*}\right|^2\right\rangle\lesssim\left(l_0\sqrt{\frac{s}{D}}\right)^{(2-\gamma)/2}\ll1, \tag{4.13}$$

if the wavelength $1/k$ of the perturbation $\delta\Phi_k^*$ is not much greater than the typical thickness of the localization region of $\Phi(x,x'|t_b,s)$.

By virtue of (4.13), to a first approximation the transform Φ_k^* of the Green's function $\Phi(x,x'|t_b,s)$ is given by (4.8), i.e., we can use the mean-field approximation in solving Eq. (4.2). In other words, replacing the kinetic coefficients $D_\epsilon(x)$, $\vartheta(x)$, and $\upsilon(x,s)$ by their mean values D_ϵ, ϑ, and υ in this equation is justified. This enables us, first, to no longer distinguish between the Green's function $\Phi(x,x'|t_b,s)$ and its regular part. Second, using the expressions (D2) and (D7) obtained in Appendix D for D_ϵ, ϑ, and υ along with Eq. (3.17), we can write Eq. (4.2) in the form

$$\frac{\partial\Phi}{\partial t_b}=D\frac{\partial^2\Phi}{\partial x^2}-\mu E\frac{\partial\Phi}{\partial x}-\nu(s)\Phi, \tag{4.14}$$

where

$$\nu(s)=I_\gamma\frac{D}{\Delta l_0}\left(l_0\sqrt{\frac{s}{D}}\right)^\gamma, \tag{4.15}$$

I_γ is a constant of order unity, which is specified by (D6), and $1/\nu(s)$ plays the role of the characteristic lifetime of a diffusing particle on the comb structure backbone during a random walk over a time period $t\sim1/s$.

Equation (4.14) with initial condition (4.6) is the principal result of this section. Therefore, before moving on to the concluding portion of this paper, we give a qualitative explanation for the validity of the mean-field approximation in the system under consideration. We assume for simplicity that the drift field is weak. As follows from the results in Appendix D, random fluctuations of the kinetic coefficient $\upsilon(x,s)$ determine the spatial fluctuations of the Green's function $\Phi(x,x'|t_b,s)$ and are associated mainly with random walks of a particle on teeth whose length is of order $l_s=\sqrt{D/s}$. These fluctuations are small if the random walker encounters a large number N_s of such teeth during a time $t\sim1/s$. By virtue of (2.1), the fraction of teeth with length of order l_s is

$$P_s\sim\int_{l_s}^\infty f(l)\,dl\sim\left(l_0\sqrt{\frac{s}{D}}\right)^{\gamma-1},$$

and as follows from Eq. (4.14), the characteristic thickness l_Φ of the localization region of $\Phi(x,x'|t_b,s)$ is

$$l_\Phi \sim \sqrt{\frac{D}{\nu(s)}} \sim \Delta\left(l_0\sqrt{\frac{s}{D}}\right)^{-\gamma/2}.$$

Hence, for N_s we have

$$N_s \sim P_s \frac{L_s}{\Delta} \sim \left(l_0\sqrt{\frac{s}{D}}\right)^{-(2-\gamma)/2} \gg 1,$$

which is in fact the same condition as (4.13).

5. CONTINUUM DESCRIPTION OF DIFFUSION ON THE BACKBONE OF A COMB STRUCTURE

Returning to the original problem of calculating the Green's function $G(x,x'|s)$ of random walks on a comb structure, we see that the results obtained above enable us to transform Eq. (3.6) in the following manner. First, we change from summation over N and i to integration over $t_b = N\tau_b^\epsilon$, and let $\tilde{x} \to x_{i_N}$:

$$\sum_{N=1}^{\infty} \to \frac{1}{\tau_b^\epsilon}\int_0^\infty dt_b, \quad \sum_{i_N=-\infty}^{\infty} \to \frac{1}{\Delta}\int_{-\infty}^{\infty} d\tilde{x}.$$

Since $g(x,x_{i_N}|s)$ and the smoothed function $\Phi(x_{i_N},x'|t_b,s)$ scarcely vary on scales of length Δ, the main contribution to $G(x,x'|s)$ comes from $N \sim \tau_b^\epsilon/\nu(s) \sim (\epsilon\sqrt{s/D})^{-\gamma} \gg 1$, and by virtue of (4.6), the first term on the right-hand side of Eq. (3.6) can be written

$$g(x,x'|s) = \frac{1}{\Delta}\int_0^\infty d\tilde{x}\, g(x,\tilde{x}|s)\Phi(\tilde{x},x'|0,s).$$

We then put $s=0$ in $g(x,x_{i_N}|s)$ and integrate over $\tilde{x}$, ignoring the variations of $\Phi(\tilde{x},x'|t_b,s) \simeq \Phi(x,x'|t_b,s)$ in a neighborhood of thickness ℓ_ϵ about x; ℓ_ϵ is small in comparison to the thickness ℓ_Φ of the localization region of $\Phi(\tilde{x},x'|t_b,s)$ itself. Finally, also taking into account (3.17) and (D1) and the fact that under the present assumptions, $G(x,x'|s)$ and $\Phi(x,x'|t_b,s)$ depend only on the difference $x-x'$, we can write

$$G(x-x'|s) = \frac{1}{\Delta}\int_0^\infty dt_b\, \Phi(x-x'|t_b,s). \tag{5.1}$$

The relation (5.1), together with Eq. (4.2) and the boundary condition (4.6), shows that the Green's function $G(x-x'|s)$ satisfies

$$\nu(s)G = D\frac{\partial^2 G}{\partial x^2} - \mu E\frac{\partial G}{\partial x} + \delta(x-x'), \tag{5.2}$$

which is indeed a general solution of the present problem.

The latter equation yields a continuum description of anomalous diffusion along the backbone of a comb structure. We next analyze the specific properties of such diffusion in greater detail.

5.1. Continuum equation for anomalous diffusion

The set of relations that we have constructed, together with (3.1), enables us to write for the Laplace transform $c_L(x,s)$ of the concentration of diffusing particles $c(x,t)$

$$\nu(s)c_L = D\frac{\partial^2 c_L}{\partial x^2} - \mu E\frac{\partial c_L}{\partial x} + c_0, \tag{5.3}$$

where, as in Sec. 3, $c_0(x)$ is the concentration of particles at the initial time $t=0$. Using the concepts of fractional integrals and derivatives,[20] an analogy can be drawn, by virtue of (4.15), between this equation and the following equation of anomalous diffusion:

$$I_\gamma\frac{l_0}{\Delta}\left(\frac{D}{l_0^2}\right)^{(2-\gamma)/2}\frac{\partial^\alpha c}{\partial t^\alpha} = D\frac{\partial^2 c}{\partial x^2} - \mu E\frac{\partial c}{\partial x}. \tag{5.4}$$

Here the operator for a fractional derivative of degree $\alpha=\gamma/2$ (for $0<\alpha<1$) is given by[20]

$$\frac{\partial^\alpha c(t)}{\partial t^\alpha} = \frac{1}{\Gamma(1-\alpha)}\frac{\partial}{\partial t}\int_0^t dt'\frac{c(t')}{(t-t')^\alpha}, \tag{5.5}$$

where $\Gamma(\ldots)$ is the gamma function. Equation (5.4) must be supplemented by an ''initial'' condition, which describes the behavior of the concentration $c(x,t)$ at small t. However, the formal limiting transition $t \to +0$ in Eq. (5.4) is impossible, since this equation was obtained under the assumption that $t \gg \Delta^2/D$.

Nevertheless, there is a possibility for determining some limiting behavior of the concentration $c(x,t)$ while adhering to the adopted description.[1)] Specifically, in the model under consideration it is natural to assume that the initial distribution $c_0(x)$ varies only slightly on scales of length Δ. Therefore, after a time $t \gtrsim \Delta^2/D$ has elapsed, the concentration $c(x,t)$ will vary mainly as a consequence of the diffusion of particles along teeth, rather than along the backbone. In addition, its spatial fluctuations on scales of length Δ will also be insignificant despite the structural inhomogeneity of the system of teeth, as in fact follows from the estimates concluding Sec. 4.1. Such behavior of $c(x,t)$ at $t \gtrsim \Delta^2/D$ can be described by Eq. (5.3) if the first two terms on its right-hand side are discarded, i.e., if $\nu(s)c_L \simeq c_0$. The latter can also be written in terms of the integral of fractional order $\hat{\mathfrak{J}}^{1-\alpha}$:

$$I_\gamma\frac{l_0}{\Delta}\left(\frac{D}{l_0^2}\right)^{(2-\gamma)/2}\hat{\mathfrak{J}}^{1-\alpha}\{c(x,t)\} \simeq c_0(x), \tag{5.6}$$

where $t \gtrsim \Delta^2/D$, and the operator $\hat{\mathfrak{J}}^{1-\alpha}$ (for $0<1-\alpha<1$) is given by[20]

$$\hat{\mathfrak{J}}^{1-\alpha}\{c(x,t)\} = \frac{1}{\Gamma(1-\alpha)}\int_0^t dt'\frac{c(x,t')}{(t-t')^\alpha}.$$

The relation (5.6) describes particle conservation. In other words, it states that after a time $t \gtrsim \Delta^2/D$, the total number of particles (per unit length of the backbone) located on the backbone and the nearby teeth is equal to their initial concentration on a tooth. Of course, this condition is actually independent of t for $t \gtrsim \Delta^2/D$, and can serve as a formal initial condition for solving Eq. (5.4) in the continuum description:

$$I_\gamma\frac{l_0}{\Delta}\left(\frac{D}{l_0^2}\right)^{(2-\gamma)/2}\hat{\mathfrak{J}}^{1-\alpha}\{c(x,t)\}\big|_{t\to+0} = c_0(x). \tag{5.7}$$

It is noteworthy that with such an approach the concentration $c(x,t)$ does not tend to $c_0(x)$ as $t\to+0$, but instead, by virtue of (5.7), it is given by

$$c(x,t)\sim\frac{\Delta}{I_\gamma\Gamma(\gamma/2)l_0}\left(\frac{l_0^2}{Dt}\right)^{(2-\gamma)/2}c_0(x), \tag{5.8}$$

which has a singularity as $t\to+0$. The latter, however, is no more than a formality, since in calculating specific values of $c(x,t)$ we should at least take $t\gtrsim\Delta^2/D$. In that event (5.8) yields a reasonable result. In other words, in the continuum description of diffusion on a comb structure, scales lengths Δ and times Δ^2/D should be regarded as infinitesimal quantities, and in this case the limit $t\to+0$ does not imply violation of the condition $t\gtrsim\Delta^2/D$.

Equation (5.4) with the initial condition (5.7) also comprises a continuum description of anomalous diffusion in terms of equations with fractional partial derivatives. We note that direct application of the trap model to the description of anomalous transport also leads to an equation with fractional derivatives,[21,22] which differs in form, however, from the present one.

5.2. Dispersion relations

To analyze the propagation properties of random walkers on a comb structure, we calculate the time dependence of the mean displacement $\langle x(t)\rangle$ and the mean-square displacement $\langle x^2(t)\rangle$ of the random walkers along the backbone. By definition

$$\langle x(t)\rangle=\frac{\int_{-\infty}^{\infty}dx\,x\,G(x,t)}{\int_{-\infty}^{\infty}dx\,G(x,t)},\quad \langle x^2(t)\rangle=\frac{\int_{-\infty}^{\infty}dx\,x^2\,G(x,t)}{\int_{-\infty}^{\infty}dx\,G(x,t)}. \tag{5.9}$$

Equation (5.2) makes it possible to find the Laplace transforms of the given factors directly. Integrating (5.2) with weighting functions 1, x, and x^2, taking the inverse Laplace transform of the results, and substituting them into (5.9), we have

$$\langle x(t)\rangle=\frac{\mu E\lambda^2}{D}\left(\frac{Dt}{l_0^2}\right)^{\gamma/2}, \tag{5.10}$$

$$\langle x^2(t)\rangle=2\lambda^2\left(\frac{Dt}{l_0^2}\right)^{\gamma/2}+(1+\chi)\langle x(t)\rangle^2, \tag{5.11}$$

where

$$\lambda^2=\frac{\Gamma(\gamma/2)}{I_\gamma\Gamma(\gamma)}\Delta l_0,$$

$$\chi=\frac{2\Gamma^2(\gamma)}{\Gamma(\gamma/2)\Gamma(3\gamma/2)}-1\approx0.27(2-\gamma),$$

the latter estimate being good to within 0.004 for $1<\gamma<2$.

It follows, in particular, from (5.10) and (5.11) that for a fixed process duration t, the first or second term in the expression for $\langle x^2(t)\rangle$ dominates, depending on whether $E\ll E_c(t)$ or $E\gg E_c(t)$, where

$$E_c(t)=\frac{D}{\mu\lambda}\left(\frac{l_0^2}{Dt}\right)^{\gamma/4}. \tag{5.12}$$

In weak fields ($E\ll E_c(t)$), we have $\langle x^2(t)\rangle\propto t^{\gamma/2}$ with an exponent in the range $1/2<\gamma/2<1$, i.e., in this limit particle propagation on the comb structure is subdiffusive (in the terminology of Refs. 1–3). In addition, we have

$$\frac{\langle x(t)\rangle}{\langle x^2(t)\rangle}=\frac{\mu E}{2D}, \tag{5.13}$$

which can be regarded as a generalized Einstein equality for the relationship between regular and random motions.

In strong fields ($E\gg E_c(t)$),

$$\frac{\langle x^2(t)\rangle-\langle x(t)\rangle^2}{\langle x(t)\rangle^2}=\chi\sim1, \tag{5.14}$$

which demonstrates that in anomalous diffusion, the drift field results not only in regular displacement of the diffusing particles, but also to somewhat efficient random motion. In this range of fields the latter effect is commensurate with the former, and can be regarded provisionally as field-induced diffusion. Such diffusive transport is characterized by a spreading of the ensemble of particles $\langle x^2(t)\rangle-\langle x(t)\rangle^2\propto t^\gamma$ with an exponent in the range $1<\gamma<2$, and can be relegated on the basis of this phenomenon to the superdiffusion class.[1–3]

We note that superdiffusion is usually associated with the Lévy flight model, and can be described by a diffusion equation with fractional spatial derivatives, rather than temporal, as a consequence of the divergence of the moments for the distribution of jumping distances of the particles after a single displacement event to neighboring sites (Ref. 21).[2)] We also note that field-induced diffusion also appears during grain-boundary diffusion in polycrystals and in crystals with dislocations.[23–25]

As might be expected, in the present case with $1<\gamma<2$, the asymptote (5.10) and the leading term in the asymptote (5.11) are consistent with the results of lattice models[14,15] in formal mean-field theories.[9,10] For $\gamma>2$ (where, for example, all except the first moments diverge), we encountered violation of the mean-field approximation for the Green's functions $g(x,x'|s)$ and $\Phi_N(x,x'|s)$. In other words, when $\gamma>2$, the distribution of the diffusing particles along the backbone can become highly inhomogeneous because of the random geometry of the comb structure. This case must therefore be examined in isolation, and the applicability of specific varieties of mean-field theory requires separate investigation.

To better describe the physical import of anomalous diffusion, we conclude with a qualitative derivation of the dispersion relations using simple estimates, which are essentially akin to the approach in Ref. 9. Let $\overline{l}(t)$ be the mean penetration depth of particles diffusing along the teeth at time t:

$$\overline{l}(t)\sim\int_{l_0}^{\sqrt{Dt}}dl\,lf(l)\sim\frac{Dt}{l_0}\left(\frac{l_0}{\sqrt{Dt}}\right)^\gamma.$$

Then the mean time t_b that a random walker spends on the backbone is

$$t_b \sim \frac{\Delta}{l} t \sim \frac{\Delta l_0}{D}\left(\frac{Dt}{l_0^2}\right)^{\gamma/2}.$$

At that time, the mean distance $\langle x(t)\rangle$ and the mean-square distance $\langle x^2(t)\rangle$ traversed by the particles along a tooth are

$$\langle x(t)\rangle \sim \mu E t_b \sim \frac{\mu E\sqrt{\Delta l_0}}{D}\left(\frac{Dt}{l_0^2}\right)^{\gamma/2},$$

$$\langle x^2(t)\rangle \sim D t_b + (\mu E)^2 t_b^2 \sim \Delta l_0\left(\frac{Dt}{l_0^2}\right)^{\gamma/2} + \Delta l_0\left(\frac{\mu E\sqrt{\Delta l_0}}{D}\right)^2\left(\frac{Dt}{l_0^2}\right)^{\gamma},$$

which are the same, in order of magnitude, as the rigorous results (5.10) and (5.11).

This work was carried out with partial support from the Russian Fund for Fundamental Research (Grant No. 96-02-17576).

APPENDIX A: GREEN'S FUNCTION FOR FIRST ARRIVING AT THE BACKBONE DURING RANDOM WALKS ON A TOOTH

As follows from the general properties of random walks (see, for example, Ref. 26), the Laplace transform $F(i,s|y)$ of the probability of first arriving at the backbone, given that the random walker was initially located at y on tooth i, satisfies

$$sF = D\frac{\partial^2 F}{\partial y^2} \tag{A1}$$

with boundary conditions

$$F|_{y=0} = 1, \quad \left.\frac{\partial F}{\partial y}\right|_{y=l_i} = 0. \tag{A2}$$

Hence,

$$F(i,s|y) = \frac{\cosh[(l_i - y)\sqrt{s/D}]}{\cosh(l_i\sqrt{s/D})}. \tag{A3}$$

The desired factor is $F(i,s) = F(i,s|\epsilon)$. As can be seen from (3.3), when $y = \epsilon$ the parameter $y\sqrt{s/D}$ can be assumed to be small. Then, to terms linear in y, from (A3) we find

$$F(i,s|y) = 1 - y\sqrt{\frac{s}{D}}\tanh\left(l_i\sqrt{\frac{s}{D}}\right) + O(y^2), \tag{A4}$$

which leads directly to (3.8) for $y = \epsilon$.

APPENDIX B: GREEN'S FUNCTION OF RANDOM WALKS WITHIN THE REGION Γ_ϵ

1. General relations

The random walks of a diffusing particle on a comb structure can be characterized by two types of functions if its trajectory is localized in the region Γ_ϵ. One is the probability density $g(x,t|x',0)$ at time t of finding the particle on the backbone at x after starting at x'. The second is the set of probability densities $\{\varphi_i(y,t)\}$ of finding it on the teeth $\{i\}$ at time t. These functions are interrelated by Eqs. (2.2)–(2.5), so their Laplace transforms satisfy the following equations. For the function $g(x,x'|s)$ not at the points $\{x_i\}$ of attachment of the teeth to the backbone, i.e., at $x \neq x_i$ ($i = 0, \pm 1, \pm 2, \ldots$), we have

$$sg = D\frac{\partial^2 g}{\partial x^2} - \mu E\frac{\partial g}{\partial x} + \delta^*(x - x'), \tag{B1}$$

where $\delta^*(x - x')$ is an ordinary δ function for $x' \neq x_i$, and for $x' = x_i$ we take the following regularization rule:

$$\delta^*(x - x_i) = \lim_{x' \to x_i - 0} \delta(x - x').$$

The Laplace transforms $\{\varphi_i(y,s)\}$ of the probabilities of finding a random walker on the teeth $\{i\}$ satisfy

$$s\varphi_i = D\frac{\partial^2 \varphi_i}{\partial y^2} \tag{B2}$$

in the range $0 < y < \min(\epsilon, l_i)$, with boundary conditions

$$\left.\frac{\partial \varphi i}{\partial y}\right|_{y=l_i} = 0, \quad l_i \leqslant \epsilon \tag{B3}$$

$$\varphi_i|_{y=\epsilon} = 0, \quad l_i > \epsilon. \tag{B4}$$

We note that the latter vanishing boundary condition on the functions $\{\varphi_i(y,s)\}$ reflects the fact that we are considering particle trajectories that do not leave the region Γ_ϵ during random particle walks on the teeth. The following continuity conditions hold at the points of attachment of the teeth to the backbone $\{(x_i, y = 0)\}$:

$$\varphi_i(0,s) = g(x_i, x'|s), \tag{B5}$$

$$J_i|_{y=0} = j|_{x=x_i-0} - j|_{x=x_i+0}, \tag{B6}$$

where

$$J_i = -D\frac{\partial \varphi_i}{\partial y}, \quad j = -D\frac{\partial g}{\partial x} + \mu E g.$$

The solution of Eq. (B2) with boundary conditions (B3)–(B5) is

$$\varphi_i(y,s) = g(x_i, x'|s)\left[\cosh\left(y\sqrt{\frac{s}{D}}\right) - b_i\sinh\left(y\sqrt{\frac{s}{D}}\right)\right], \tag{B7}$$

where

$$b_i = b(l_i) = \begin{cases} \tanh(l_i\sqrt{s/D}), & l_i \leqslant \epsilon, \\ \coth(\epsilon\sqrt{s/D}), & l_i > \epsilon. \end{cases} \tag{B8}$$

The expression (B8) enables us to represent the boundary condition (B6) in the form

$$j|_{x=x_i+0} - j|_{x=x_i-0} = D\left.\frac{\partial \varphi_i}{\partial y}\right|_{y=0} = -\sqrt{sD}\,b_i g(x_i, x'|s). \tag{B9}$$

The relation (B9) shows that Eq. (B1) can be continued without singularities over the entire x axis provided that δ-function power sinks $\sqrt{sD}b_i g$ are introduced at the points $\{x_i\}$. In other words, for the Green's function we can write the following equation, which is defined over the entire x axis:

$$sg = D\frac{\partial^2 g}{\partial x^2} - \mu E\frac{\partial g}{\partial x} - \sqrt{sD}\left[\sum_{i=-\infty}^{\infty} b_i \delta(x-x_i)\right] g + \delta^*(x-x'). \tag{B10}$$

In this equation there is no reason to distinguish between $\delta^*(x-x')$ and the ordinary delta function $\delta(x-x')$. Hence Eq. (3.9) follows directly.

The random walk of a particle within the region Γ_ϵ can also be characterized by the probability of first arriving at its boundary, i.e., the observation line $y=\epsilon$. For the theory presented here we need the Laplace transform $W(x_i,x'|s)$ of the probability density of a random walker first arriving at the observation line at tooth i and time t after starting at x' on the backbone. As follows from the general properties of random walks (see for example, Ref. 26), the function $W(x_i,x'|s)$ and the probability flux density J_i are related by

$$W(x_i,x'|s) = J_i|_{y=\epsilon} = -D\left.\frac{\partial \varphi_i}{\partial y}\right|_{y=\epsilon}, \tag{B11}$$

of course, under the condition $l_i > \epsilon$. Substituting (B7) into (B11), we obtain the desired expression,

$$W(x_i,x'|s) = \sqrt{sD}\left[\sinh\left(\epsilon\sqrt{\frac{s}{D}}\right)\right]^{-1} g(x_i,x'|s)\,\theta_i^\epsilon, \tag{B12}$$

where $\theta_i^\epsilon = 1$ for $l_i > \epsilon$ and $\theta_i^\epsilon = 0$ for $l_i \leqslant \epsilon$.

The quantity $\epsilon\sqrt{s/D}$ can be considered a small parameter in the limit (3.3). Then, retaining the first two terms of the expansion in $\epsilon\sqrt{s/D}$ in (B12), we obtain (3.12).

Some general identities, which establish the relationship between $g(x_i,x'|s)$ and $W(x_i,x'|s)$, will also be useful below. To derive them, on the right-hand side of Eq. (B10) we distinguish between contributions of lateral teeth with lengths $l_i \leqslant \epsilon$ and $l_i > \epsilon$, and allowing for (B12), we rewrite this equation in the limit $\epsilon \ll \sqrt{D/s}$:

$$\sum_{i=-\infty}^{\infty} \delta(x-x_i) W(x_i,x'|s) = D\frac{\partial^2 g}{\partial x^2} - \mu E\frac{\partial g}{\partial x} - s\left[1 + \sum_{i=-\infty}^{\infty} \delta(x-x_i) \times \left(\frac{1}{2}\epsilon\theta_i^\epsilon + l_i(1-\theta_i^\epsilon)\right)\right] g + \delta(x-x'). \tag{B13}$$

Integrating this relation over x from $-\infty$ to ∞ with the weighting functions 1, $(x-x')$, and $(x-x')^2$, we obtain

$$\sum_{i=-\infty}^{\infty} W(x_i,x'|s) = 1 - s\left[\int_{-\infty}^{\infty} dx\, g(x,x'|s) + \sum_{i=-\infty}^{\infty}\left(\frac{1}{2}\epsilon\theta_i^\epsilon + l_i(1-\theta_i^\epsilon)\right) g(x_i,x'|s)\right], \tag{B14}$$

$$\sum_{i=-\infty}^{\infty} (x_i - x') W(x_i,x'|s) = \mu E\int_{-\infty}^{\infty} dx\, g(x,x'|s) - s \times\left[\int_{-\infty}^{\infty} dx\,(x-x') g(x,x'|s) + \sum_{i=-\infty}^{\infty} (x_i-x') \times\left(\frac{1}{2}\epsilon\theta_i^\epsilon + l_i(1-\theta_i^\epsilon)\right) g(x_i,x'|s)\right], \tag{B15}$$

$$\sum_{i=-\infty}^{\infty} (x_i - x')^2 W(x_i,x'|s) = 2D\int_{-\infty}^{\infty} dx\, g(x,x'|s) + 2\mu E \times\int_{-\infty}^{\infty} dx\,(x-x') g(x,x'|s) - s\left[\int_{-\infty}^{\infty} dx\,(x - x')^2 g(x,x'|s) + \sum_{i=-\infty}^{\infty} (x_i-x')^2\left(\frac{1}{2}\epsilon\theta_i^\epsilon + l_i(1 - \theta_i^\epsilon)\right) g(x_i,x'|s)\right]. \tag{B16}$$

In particular, (B14) yields the useful identity

$$\sum_{i=-\infty}^{\infty} W(x_i,x'|0) = 1, \tag{B17}$$

which reflects the fact that a particle starting on the backbone must eventually cross the boundary $y=\epsilon$ of the region Γ_ϵ.

In the next part of this appendix we examine specific statistical properties of random walks within the region Γ_ϵ.

2. Statistical properties of the Green's function $g(x,x'|s)$

We separate the regular and fluctuation components of the Green's function $g(x,x'|s)$:

$$g(x,x'|s) = \overline{g}(x,x'|s) + \delta g(x,x'|s). \tag{B18}$$

In accordance with the ensuing results, the fluctuation component δg is a small perturbation. It enables us to solve Eq. (3.9) (or (B10)) using perturbation theory. In this case, averaging (3.9) over the ensemble $\{l_i\}$, we find

$$s\overline{g}=D\frac{\partial^2\overline{g}}{\partial x^2}-\mu E\frac{\partial\overline{g}}{\partial x}-\sqrt{sD}b \times\sum_{j=-\infty}^{\infty}\overline{g}\,\delta(x-x_j)+\delta(x-x'), \tag{B19}$$

where b is the mean value of $\{b_j\}$:

$$b=\int_{l_0}^{\infty}dl\,f(l)b(l) =\frac{2\gamma-1}{3(2-\gamma)}\sqrt{\frac{s}{D}}l_0^{\gamma-1}\epsilon^{2-\gamma}+\sqrt{\frac{D}{s}}l_0^{\gamma-1}\epsilon^{-\gamma} \tag{B20}$$

by virtue of (2.1), (B8), and the inequality (3.3). To transform the discrete description of the influence of the teeth on random walks within the region Γ_ϵ into a continuum description, we take Fourier transforms in Eq. (B19). Then, for the Fourier transform $\overline{g}_k$ of $\overline{g}(x,x'|s)$ we have

$$\overline{g}_k(x'|s)=\int_{-\infty}^{\infty}dx\,\overline{g}(x,x'|s)e^{-ikx}, \tag{B21}$$

Eq. (B19) takes the form

$$L_k\overline{g}_k+\sqrt{sD}\,\frac{b}{\Delta}\sum_{n=-\infty}^{\infty}{}'\;\overline{g}_{k+K_n}=e^{-ikx'}, \tag{B22}$$

where

$$L_k=s+Dk^2+i\mu Ek+\sqrt{sD}\,\frac{b}{\Delta}, \tag{B23}$$

the $K_n=2\pi n/\Delta$ are Brillouin numbers (the n are integers) and the prime on the summation sign means that the term with $n=0$ has been omitted. In the present problem we are interested in scale lengths much greater than Δ. This enables us to restrict attention to wave numbers k that are small compared to $1/\Delta$. It then follows from (B22) that the correction $\delta\overline{g}_k$ to the zeroth approximation for $\overline{g}_k$ in the small parameter $k\Delta\ll 1$ is

$$\left|\frac{\delta\overline{g}_k}{\overline{g}_k}\right|\lesssim\frac{\sqrt{sD}b}{L_{K_1}\Delta}\ll 1$$

in the limit (3.3). We can thus disregard the second term on the left-hand side of Eq. (B22), and thereby obtain

$$L_k\overline{g}_k=e^{-ikx'}. \tag{B24}$$

Hence, in particular, we find that in the present approximation the regular part $\overline{g}(x-x'|s)$ of the Green's function depends only on the difference between x and x', and that its Fourier transform is

$$\overline{g}_k=\frac{e^{-ikx'}}{s+Dk^2+i\mu Ek+\sqrt{sD}b/\Delta}. \tag{B25}$$

Taking the inverse Fourier transforms, we obtain

$$s\overline{g}=D\frac{\partial^2\overline{g}}{\partial x^2}-\mu E\frac{\partial\overline{g}}{\partial x}-\sqrt{sD}\frac{b}{\Delta}\overline{g}+\delta(x-x'). \tag{B26}$$

Substituting (B20) into (B26) yields Eq. (3.14). In particular, this form of the equation turns out to be useful when $s=0$. By virtue of (B8) and (B20) we have

$$\frac{D}{\Delta l_0}\left(\frac{l_0}{\epsilon}\right)^{\gamma}\overline{g}+\mu E\frac{\partial\overline{g}}{\partial x}-D\frac{\partial^2\overline{g}}{\partial x^2}=\delta(x-x'). \tag{B27}$$

In this same approximation for the fluctuation component $\delta g(x,x'|s)$, Eq. (3.9) (or (B10)) takes the form

$$\left(s-D\frac{\partial^2}{\partial x^2}+\mu E\frac{\partial}{\partial x}+\sqrt{sD}\frac{b}{\Delta}\right)\delta g(x,x'|s) =-\sqrt{sD}\sum_{i=-\infty}^{\infty}\delta b_i\,\delta(x-x_i)\overline{g}(x_i-x'|s), \tag{B28}$$

where $\delta b_i=b_i-b$. Transforming Eq. (B28) and allowing for (B26), we obtain (3.18):

$$\delta g(x,x'|s)=-\sqrt{sD}\sum_{i=-\infty}^{\infty}\delta b_i\overline{g}(x-x_i|s)\overline{g}(x_i-x'|s), \tag{B29}$$

or for the Fourier transform δg_k (which can be specified by the same relation as (B21)),

$$\delta g_k(x'|s)=-\frac{\sqrt{sD}}{L_k}\frac{1}{2\pi}\int_{-\infty}^{\infty}dp\,\frac{1}{L_p}\sum_{j=-\infty}^{\infty}\delta b_j \times\exp[-i(k-p)x_j-ipx']. \tag{B30}$$

In particular, an estimate of the fluctuation component δg follows from (B30). Specifically, considering only small values of the wave number $|k|\ll 1/\Delta$, we find

$$\frac{\langle|\delta g_k|^2\rangle}{|g_k|^2}=\frac{sD}{2\pi\Delta}(\delta b)^2\int_{-\infty}^{\infty}dp\,\frac{1}{|L_p|^2}, \tag{B31}$$

where $(\delta b)^2$ is the variance of the random components $\{\delta b_j\}$, i.e.,

$$(\delta b)^2=\int_{l_0}^{\infty}dl\,f(l)[b(l)-b]^2\simeq b^2\left(\frac{\epsilon}{l_0}\right)^{\gamma-1} \tag{B32}$$

in the zeroth approximation with respect to the small parameter $\epsilon\sqrt{s/D}$. Then, setting $L_p=L_p|_{E=0}$ in (B31) (making the right-hand side an overestimate) and allowing for (B20) and (B32), we have

$$\frac{\langle|\delta g_k|^2\rangle}{|g_k|^2}\simeq\frac{1}{4}\sqrt{\frac{\Delta}{l_0}}\left(\frac{l_0}{\epsilon}\right)^{(2-\gamma)/2}, \tag{B33}$$

whence follows the estimate (3.19).

APPENDIX C: PROCEDURE FOR SMOOTHING THE FUNCTION $\Phi_N(x_j,x'|s)$ AND DERIVATION OF THE FOKKER–PLANCK EQUATION (4.2)

Using the periodic Born–Kármán conditions, we write the Fourier transform of $\Phi_N(x_j,x'|s)$ in the form

$$\Phi_N(k_m)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{j=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\Phi_N(x_j,x'|s)\exp(-ik_mx_j), \tag{C1}$$

where $\mathfrak{M}\to\infty$ is the (even) integer number of teeth, and the wave number k takes values $k_m=2\pi m/\Delta\mathfrak{M}$ for $m=0,\pm1,\pm2,\ldots,\pm(\mathfrak{M}/2-1),-\mathfrak{M}/2$. Then the inverse transform is

$$\Phi_N(x_j,x'|s)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{m=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\Phi_N(k_m)\exp(ik_mx_j). \quad \text{(C2)}$$

We define the procedure for smoothing the Green's function

$$\Phi_N(x_j,x'|s)\to\Phi(x,x'|N,s)$$

by the condition that only values of the wave number k_m that do not exceed $1/\ell_\epsilon$, i.e., $|k_m\ell_\epsilon|\lesssim1$, are taken into account in (C2):

$$\Phi(x,x'|N,s)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{|k_m\ell_\epsilon|\lesssim1}\Phi_N(k_m)\exp(ik_mx_j). \quad \text{(C3)}$$

As will be seen in the ensuing analysis, the immediate source of random fluctuations of $\Phi_N(x_j,x'|s)$ corresponds to wave numbers $k_m\gg1/\ell_\epsilon$, and the width of the localization region of $\Phi(x,x'|N,s)$ is estimated to be $\ell_\epsilon\sqrt{N}$. Random perturbations can be interpreted as δ-correlated sources. Hence it also follows that in (C3) it is in fact sufficient to consider the region $|k_m\ell_\epsilon|\ll1$.

We obtain the equation for $\Phi(x,x'|N,s)$ from (4.1), following Pontryagin's method.[18] Specifically, taking Fourier transforms of the right- and left-hand sides of Eq. (4.1), we have

$$\Phi_{N+1}(k)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{j=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\exp(-ikx_j)F(j,s)\times\sum_{j'=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}W(x_j,x_{j'}|s)\Phi_N(x_{j'},x'|s),$$

whence, by virtue of (C2),

$$\Phi_{N+1}(k)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{m=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\Pi(k,k-p_m)\Phi_N(p_m), \quad \text{(C4)}$$

where

$$\Pi(k,k-p_m)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{j,j'=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\exp[-ik(x_j-x_{j'})-i\times(k-p_m)x_{j'}]F(j,s)W(x_j,x_{j'}|s). \quad \text{(C5)}$$

Since the localization region of $W(x_j,x_{j'}|s)$ is the same as the localization region of the Green's function $g(x_j,x_{j'}|s)$ (see (B12)), if we expand the exponential factor in powers of $k(x_j-x_{j'})$ in the limit $|k\ell_\epsilon|\ll1$, we can rewrite (C5) in the form

$$\Pi(k,k-p_m)\simeq1-\upsilon^F(k-p_m)-ik\vartheta^F(k-p_m)-k^2D_\epsilon^F(k-p_m), \quad \text{(C6)}$$

where

$$\upsilon^F(k-p_m)=1-\frac{1}{\sqrt{\mathfrak{M}}}\sum_{j,j'=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\exp[-i(k-p_m)x_{j'}]F(j,s)W(x_j,x_{j'}|s), \quad \text{(C7)}$$

$$\vartheta^F(k-p_m)=\frac{1}{\sqrt{\mathfrak{M}}}\sum_{j,j'=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}(x_j-x_{j'})\exp[-i(k-p_m)x_{j'}]F(j,s)W(x_j,x_{j'}|s), \quad \text{(C8)}$$

$$D_\epsilon^F(k-p_m)=\frac{1}{2\sqrt{\mathfrak{M}}}\sum_{j,j'=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}(x_j-x_{j'})^2\exp[-i(k-p_m)x_{j'}]F(j,s)W(x_j,x_{j'}|s) \quad \text{(C9)}$$

are in fact the Fourier transforms of the functions

$$\upsilon(x_{j'}|s)=1-\sum_{j=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}F(j,s)W(x_j,x_{j'}|s), \quad \text{(C10)}$$

$$\vartheta(x_{j'}|s)=\sum_{j=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}(x_j-x_{j'})F(j,s)W(x_j,x_{j'}|s), \quad \text{(C11)}$$

$$D_\epsilon(x_{j'}|s)=\sum_{j=-\mathfrak{M}/2}^{\mathfrak{M}/2-1}\frac{1}{2}(x_j-x_{j'})^2F(j,s)W(x_j,x_{j'}|s). \quad \text{(C12)}$$

As can be seen from the following, the values of the kinetic coefficients $\upsilon^F(k)$, $\vartheta^F(k)$, and $D_\epsilon^F(k)$ are small when $|k|\gg1/\ell_\epsilon$. Therefore, first, in Eqs. (C10)–(C12) the argument $x_{j'}$ can be considered continuous. Second, in (C4) it can be assumed that $|p_m|\lesssim1/\ell_\epsilon$, since $\Phi_N(p_m)$ is appreciable when either $|p_m|\ll1/\ell_\epsilon$ or $|p_m|\gg1/\ell_\epsilon$. Noting that we have the Fourier transform of a product of functions, we can go from Eq. (C4) to the equation for the smoothed Green's function $\Phi(x,x'|N,s)$. Then, taking (C6) into account and setting

$$\Phi(x,x'|N+1,s)-\Phi(x,x'|N,s)\simeq\frac{\partial\Phi(x,x'|N,s)}{\partial N},$$

we obtain for $\Phi(x,x'|N,s)$ an equation of the Fokker–Planck type:

$$\frac{\partial\Phi}{\partial N}=\frac{\partial^2}{\partial x^2}[D_\epsilon(x|s)\Phi]-\frac{\partial}{\partial x}[\vartheta(x|s)\Phi]-\upsilon(x|s)\Phi. \quad \text{(C13)}$$

As can be seen from the structure of Eq. (C13), the behavior of $\Phi(x,x'|N,s)$ as a function of s is dominated by the last term on the right-hand side. The dependence of the kinetic coefficients ϑ^F and D_ϵ on s merely leads to a minor renormalization of the width of the localization region of the Green's function $\Phi(x,x'|N,s)$, while the dependence of υ on s determines the typical number N of elementary segments $\{P^{\rm in},P^{\rm out}\}$ of the trajectory of a random walker at time $t\sim1/s$. This enables us to set $s=0$ in (C11) and (C12), which by virtue of (B15) and (B16) yields

$$D_\epsilon(x_{j'}|0)=\sum_{-\infty}^{+\infty}\frac{1}{2}(x_j-x_{j'})^2W(x_j,x_{j'}|0)$$

$$=D\int_{-\infty}^{\infty}dx\,g(x,x_{j'}|0)+\mu E$$

$$\times\int_{-\infty}^{\infty}dx\,(x-x_{j'})g(x,x_{j'}|0), \tag{C14}$$

$$\vartheta(x_{j'}|0)=\sum_{-\infty}^{\infty}(x_j-x_{j'})W(x_j,x_{j'}|0)$$

$$=\mu E\int_{-\infty}^{\infty}dx\,g(x,x_{j'}|0). \tag{C15}$$

The expression (C10) can also be simplified by virtue of (B14) and (3.8). It can be shown that the dependence of $F(j,s)$ on s is stronger than the corresponding dependence of $W(x_j,x_{j'}|s)$ by a factor $(D/s\epsilon^2)^{(2-\gamma)/2}\gg 1$. Hence

$$\upsilon(x_{j'}|s)\simeq\epsilon\sqrt{\frac{s}{D}}\sum_{i=-\infty}^{\infty}\tanh\left(l_i\sqrt{\frac{s}{D}}\right)W(x_j,x_{j'}|0). \tag{C16}$$

The relations (C14)–(C16) yield the desired general expressions (4.5)–(4.3) for the kinetic coefficients of Eq. (4.2).

APPENDIX D: MEAN VALUES AND INTENSITY OF THE RANDOM FLUCTUATIONS OF THE KINETIC COEFFICIENTS D_ϵ, ϑ, AND υ

We first calculate the mean values D_ϵ, ϑ, and $\upsilon(s)$ of the kinetic coefficients $D_\epsilon(x)$, $\vartheta(x)$, and $\upsilon(x,s)$. By virtue of Eq. (B27) and the definitions (3.16) and (3.21) for the regular part of the Green's function $\overline{g}(x,x')\equiv\overline{g}(x,x'|0)$, we have

$$\int_{-\infty}^{\infty}dx\,\overline{g}(x,x')=\frac{\mathscr{l}_\epsilon^2}{2D},$$

$$\int_{-\infty}^{\infty}dx\,(x-x')\overline{g}(x,x')=\mu E\left(\frac{\mathscr{l}_\epsilon^2}{2D}\right)^2. \tag{D1}$$

These relations, along with (4.3), (4.4), and the inequality (3.3), yield the mean values of the kinetic coefficients:

$$D_\epsilon\simeq\frac{\mathscr{l}_\epsilon^2}{2},\quad \vartheta=\frac{\mu E\mathscr{l}_\epsilon}{D}\frac{\mathscr{l}_\epsilon}{2}. \tag{D2}$$

To calculate the mean value of $\upsilon(x,s)$, we denote the set of teeth whose length exceeds ϵ by $\{i\}_\epsilon=\{i:\ l_i>\epsilon\}$. Then we treat a realization of the set of tooth lengths that exceed ϵ, $\{l_i\}_\epsilon\equiv\{l_i:l_i>\epsilon\}$, as a realization of given specific values, given the occurrence of event $\{i\}_\epsilon$. In other words, we represent the probability $f(l)$ of the event $l>\epsilon$ as the product of the probability P_ϵ that the tooth length exceeds ϵ and the conditional probability $f_\epsilon(l)$ of the realization of that value:

$$f(l)=P_\epsilon f_\epsilon(l). \tag{D3}$$

Hence, with Eq. (2.1) and (3.21),

$$f_\epsilon(l)=(\gamma-1)\frac{\epsilon^{\gamma-1}}{l^\gamma}. \tag{D4}$$

Then to calculate $\upsilon(s)$ we can first average over the conditional distribution of tooth lengths and then with respect to the realization of the condition that the teeth under consideration are longer than ϵ. In the first step Eq. (4.5) yields

$$\langle\upsilon(x|s)\rangle_1=\epsilon\sqrt{\frac{s}{D}}\left\langle\tanh\left(l\sqrt{\frac{s}{D}}\right)\right\rangle_\epsilon\sum_{i=-\infty}^{\infty}W(x_j,x_{j'}|0),$$

since $W(x_j,x_{j'}|0)$ does not depend on the actual values of $\{l_i>\epsilon\}$. Here $\langle\ldots\rangle_\epsilon$ denotes averaging over the conditional realizations of the tooth lengths $\{l_i\}_\epsilon$, particularly in the limit $\epsilon\sqrt{s/D}\ll 1$:

$$\left\langle\tanh\left(l\sqrt{\frac{s}{D}}\right)\right\rangle=\int_\epsilon^\infty dl\,f_\epsilon(l)\tanh\left(l\sqrt{\frac{s}{D}}\right)$$

$$\simeq\left(\epsilon\sqrt{\frac{s}{D}}\right)^{\gamma-1}I_\gamma, \tag{D5}$$

where (see, for example, Ref. 27)

$$I_\gamma=(\gamma-1)\int_0^\infty d\xi\frac{\tanh\xi}{\xi^\gamma}$$

$$=(\gamma-1)2^\gamma(2^\gamma-1)\Gamma(1-\gamma)\zeta(1-\gamma), \tag{D6}$$

and $\zeta(\ldots)$ is the Riemann zeta function. By virtue of the identity (B17), this yields the desired expression,

$$\upsilon(s)=I_\gamma\left(\epsilon\sqrt{\frac{s}{D}}\right)^\gamma, \tag{D7}$$

since $\langle\upsilon(x|s)\rangle_1$ is independent of the specific realization of the event $\{i\}_\epsilon$.

Furthermore, using these relations, we obtain for the integral

$$\int_{-\infty}^{\infty}dp\frac{1}{|L_p^*|^2}\simeq\frac{\pi}{2I_\gamma^{3/2}\sqrt{\Delta l_0}}\left(\frac{l_0}{\epsilon}\right)^{\gamma/2}$$

$$\times\left(\epsilon\sqrt{\frac{s}{D}}\right)^{-3\gamma/2}\left(1+\frac{E^2}{E_c^2}\right)^{-1/2}, \tag{D8}$$

where $L_p^*=k^2D_\epsilon+ik\vartheta+\upsilon(s)$ and

$$E_c=2\sqrt{I_\gamma}\frac{D}{\sqrt{\Delta l_0}\mu}\left(l_0\sqrt{\frac{s}{D}}\right)^{\gamma/2} \tag{D9}$$

is the critical value of the drift field, which for a given value of s (i.e., for a given duration of the diffusion process $t\sim 1/s$) divides the possible values of the drift field into weak ($E\ll E_c$) and strong ($E\gg E_c$). (The definition of E_c and (5.12) are essentially identical.)

Proceeding to a calculation of the total amplitude Λ of random fluctuations of these kinetic coefficients, we note that the preceding property enables us to treat the fluctuations $\delta D_\epsilon(x)$ and $\delta\vartheta(x)$ of two of the coefficients and the fluctuations $\delta\upsilon(x,s)$ independently of one another, since $\delta D_\epsilon(x)$

and $\delta\vartheta(x)$ likewise do not depend on the actual values of the lengths $\{l_i\}_\epsilon$. Therefore, (4.11) can be rewritten in the form

$$\Lambda=\Lambda_g+\Lambda_v, \tag{D10}$$

where

$$\Lambda_g=\int_{-\infty}^{\infty}dx'\,\langle[k^2\delta D_\epsilon(x)+ik\delta\vartheta(x)] \times[k^2\delta D_\epsilon(x')-ik\delta\vartheta(x')]\rangle, \tag{D11}$$

$$\Lambda_v=\int_{-\infty}^{\infty}dx'\,\langle\delta v(x)\,\delta v(x')\rangle. \tag{D12}$$

Taking (3.18), (D1), and (D2) into account, we can write

$$k^2\delta D_\epsilon(x)+ik\delta\vartheta(x)=-\frac{D}{\epsilon}\sum_{j=-\infty}^{\infty}\delta\theta_i\overline{g}(x_i-x)\{D_\epsilon k^2+\vartheta[ik+k^2(x_i-x))\}.$$

Then, substituting this expression into (D11) and integrating, we find

$$\Lambda_g=\Delta I_\gamma^2\left(\epsilon\sqrt{\frac{s}{D}}\right)^{2\gamma}\left(\frac{\epsilon}{l_0}\right)^{\gamma-1}\Psi(k), \tag{D13}$$

where

$$\Psi(k)=\frac{(kl_\Phi)^2}{1+4(E/E_c)^2}\left[\frac{(kl_\Phi)^2}{1+4(E/E_c)^2}+4(E/E_c)^2\right] \tag{D14}$$

and l_Φ is the characteristic width of the localization region of the Green's function $\Phi(x,x'|t_b,s)$, which is defined to be half the corresponding standard deviation $\sqrt{\langle x^2\rangle}$; specifically,

$$l_\Phi^2=\frac{D_\epsilon}{v}+\frac{\vartheta^2}{v^2}. \tag{D15}$$

We note that in the range of wave numbers $k\lesssim 1/l_\Phi$ that dominate the Green's function $\Phi(x,x'|t_b,s)$, the factor $\Psi(k)\lesssim 1$. Substituting (4.5) into (D12) and averaging first with respect to the conditional realization of the values $\{l_i\}_\epsilon$ and then over the random quantities θ_i^ϵ and over the variations of the Green's function $g(x,x')$, we obtain

$$\Lambda_v=\Delta\tilde{I}_\gamma\left(\epsilon\sqrt{\frac{s}{D}}\right)^{\gamma+1}\left(\frac{\epsilon}{l_0}\right)^{\gamma-1}, \tag{D16}$$

where

$$\tilde{I}_\gamma=(\gamma-1)\int_0^\infty d\xi\,\frac{\tanh^2\xi}{\xi^\gamma}.$$

The relations (D8), (D13), and (D16) lead to the following estimate for the relative intensity (4.10) of the random fluctuations of the Green's function $\Phi(x,x'|t_b,s)$:

$$\left\langle\left|\frac{\delta\Phi_k^*}{\overline{\Phi}_k^*}\right|^2\right\rangle=\frac{1}{4}\tilde{I}_\gamma I_\gamma^{-3/2}\sqrt{\frac{\Delta}{l_0}}\left(l_0\sqrt{\frac{s}{D}}\right)^{(2-\gamma)/2}\left(1+\frac{E^2}{E_c^2}\right)^{-1/2} \times\left[1+\frac{I_\gamma^2}{\tilde{I}_\gamma}\left(\epsilon\sqrt{\frac{s}{D}}\right)^{\gamma-1}\Psi(k)\right], \tag{D17}$$

yielding the estimate (4.13) for small to moderate values of kl_Φ.

*)E-mail: ialub@fpl.gpi.ru

1)Note that a different approach to assigning the initial conditions was developed in Refs. 21 and 22. It is based on constructing the crossover between a small neighborhood of $t=0$ and the domain of applicability of a partial differential equation with respect to time.

2)Superdiffusive transport is also encountered in problems on the diffusion of a passive scalar during the laminar motion of fluids (see, for example, Refs. 22).

[1] E. W. Montroll and M. F. Shlesinger, in *Studies in Statistical Mechanics, Vol. 11*, J. Leibowitz and E. W. Montroll (eds.), North-Holland, Amsterdam (1984), p. 1.
[2] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).
[3] M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
[4] S. Havlin, in *Fractals in Physics: Proceedings of the Sixth International Symposium on Fractals in Physics, ICTP, Trieste, Italy, July 9–12, 1985*, L. Pietronero and E. Tosatti (eds.), North-Holland, Amsterdam–New York (1986).
[5] S. R. White and M. Barma, J. Phys. A **17**, 2995 (1984).
[6] Y. Gefen and I. Goldhirsch, J. Phys. A **18**, 1037 (1985).
[7] G. H. Weiss and S. Havlin, Physica A **134**, 474 (1986).
[8] V. E. Arkhincheev and É. M. Baskin, Zh. Éksp. Teor. Fiz. **100**, 292 (1991) [Sov. Phys. JETP **73**, 161 (1991)].
[9] S. Havlin, J. E. Kiefer, and G. H. Weiss, Phys. Rev. A **36**, 1403 (1987).
[10] K. P. N. Murthy and K. W. Kehr, Phys. Rev. A **40**, 2082 (1989).
[11] S. Revathi, V. Balakrishnan, S. Lakshmibala, and K. P. N. Murthy, Phys. Rev. E **54**, 2298 (1996).
[12] S. Havlin, A. Bunde, H. E. Stanley, and D. Movshovitz, J. Phys. A **19**, L693 (1986).
[13] A. Bunde, S. Havlin, H. E. Stanley, B. Trus, and G. H. Weiss, Phys. Rev. B **34**, 8129 (1986).
[14] C. Aslangul and N. Pottier, Physica A **203**, 533 (1993).
[15] N. Pottier, Nuovo Cimento D **16** (Ser. 1), 1265 (1994).
[16] N. Pottier, Physica A **216**, 1 (1995).
[17] V. Balakrishnan and Van den Broeck, Physica A **217**, 1 (1995).
[18] Yu. B. Rumer and M. Sh. Ryvkin, *Thermodynamics, Statistical Physics, and Kinetics*, Mir, Moscow (1980).
[19] J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems*, Cambridge University Press, Cambridge–New York (1979).
[20] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach, Chur–Philadelphia (1993).
[21] K. V. Chukbar, Zh. Éksp. Teor. Fiz. **108**, 1875 (1995) [JETP **81**, 1025 (1995)].
[22] K. V. Chukbar, Zh. Éksp. Teor. Fiz. **109**, 1335 (1996) [JETP **82**, 719 (1996)].
[23] I. A. Lubashevskiĭ and V. L. Alatortsev, Metallofizika (Kiev) **11**, 63 (1989).
[24] I. A. Lubashevsky, V. L. Alatortsev, and A. G. Keijan, Physica A **193**, 259 (1993).
[25] I. A. Lubashevskiĭ, Khim. Fiz. **9**, 272 (1990).
[26] C. W. Gardiner, *Springer Series in Synergetics, Vol. 13: Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 2nd. ed., Springer-Verlag, Berlin (1983).
[27] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series, Vol. 1: Elementary Functions*, Gordon and Breach, New York (1986).

Translated by P. Shelnitz

Kinetics of crystal growth in superfluid helium at high temperatures

L. A. Maksimov and V. L. Tsymbalenko*)

Russian Science Center ''Kurchatov Institute,'' 123182 Moscow, Russia
(Submitted 30 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 1313–1328 (October 1998)

The growth rate of ^{4}He crystals from superfluid is measured in the temperature range 1.2–1.75 K at supersaturations up to 40 mbar. The growth rate is observed to decrease at high supersaturations: above 5 mbar in the bcc phase and above 20 mbar in the hcp phase. The temperature dependence of the kinetic growth factor K is measured in the low-supersaturation limit. In the vicinity of the superfluid transition the kinetic growth factor exhibits critical behavior: $K \propto (T_\lambda - T)^\varepsilon$ with the exponent $\varepsilon = 0.743 \pm 0.123$. A jump in the growth factor is observed at the bcc–hcp transition point. The crystal growth kinetics problem is solved in the hydrodynamic approximation, explaining both the temperature behavior of K and the existence of the jump in the modification of the crystal structure. © *1998 American Institute of Physics.* [S1063-7761(98)01110-X]

1. INTRODUCTION

The growth kinetics of ^{4}He crystals has been the object of intensive research ever since Andreev and Parshin[1] theoretically predicted the quantum state of atomically rough surfaces of helium crystals and since Keshishev, Parshin, and Babkin[2] experimentally observed one of the consequences of this phenomenon: crystallization waves. So far the kinetic growth factor K has been measured over a broad range of temperatures from 1.6 K to 0.05 K (Refs. 3–5). Of primary interest are low temperatures, at which the surface mobility is governed by the scattering by quasiparticles by elementary surface defects, or steps. Phonons are prevalent below 0.6 K, and the kinetic growth factor K obeys a power law: $K \propto T^{-4}$. Rotons play the dominant role at higher temperatures, and the kinetic growth factor follows an exponential law: $K \propto \exp(\Delta_r/T)$, where Δ_r is the roton gap.[1,6] However, the experimental work of Bodensohn, Nicolai, and Leiderer[3] has shown that a departure from a simple exponential dependence is observed above 1.4 K. For example, according to data at a temperature of 1.6 K, the growth rate V is approximately 2.5 times lower than the value expected from the extrapolation of high-temperature data. In our experiments[7] we have found that the growth rate at 1.746 K is already a factor of seven smaller than the extrapolated value. Bodensohn *et al.*[3] have advanced the hypothesis that the departure is caused by the dependence of the roton gap on the pressure and the temperature.[8]

In this paper we report an experimental investigation of the growth of ^{4}He crystals in the high-temperature range (1.2–1.75 K) at supersaturations up to 40 mbar, which exceeds the previously investigated range by three orders of magnitude. We have measured the kinetic growth factor for the first time in the vicinity of the superfluid transition point, $0.03\ \mathrm{K} \leqslant T_\lambda - T \leqslant 0.2\ \mathrm{K}$, and have found the corresponding crystal critical index. By extending the supersaturation interval we have been able to detect nonlinearity in the dependence of the growth rate on the supersaturation, i.e., the activation of a new dissipative channel in crystal growth.

To account for the experimental $K(T)$ dependence, we solve the problem of quasistationary crystal growth in the hydrodynamic approximation. Local equilibrium is at the crystal-liquid interface is rapidly established in this temperature range, while in the immediate vicinity of the interface the chemical potential μ of the liquid is equal to the chemical potential μ' of the crystal (throughout the article unprimed symbols refer to the liquid, and primed symbols refer to the crystal), and the Kapitza jump is small, $T \simeq T'$. In these boundary conditions we find the main departure from the low-temperature (below 0.6 K) case treated by the theory developed in Andreev and Knizhnik's work.[9] Andreev and Knizhnik have calculated the kinetic growth factor K in the low-temperature range, where crystallization waves are slightly damped. According to experimental data, this temperature range lies below 0.6 K. At such temperatures the factor K exhibits strong dispersion, and its poles specify the spectrum and damping of crystallization waves.

The growth of a crystal as a whole is dictated by the long-wavelength limit. If the crystal radius R is smaller than the mean free path λ, as is the case at temperatures below 0.5 K, the factor K is described by the equations for the ballistic regime (Sec. 3 in Ref. 9). In the limit $\lambda \ll R$ the equations in the first two sections of Ref. 9 must be used. The following boundary condition on the surface of the crystal is used in Ref. 9: $v_n = V$ (v_n is the velocity of the normal component), which is invalid at high temperatures. At temperatures of 1.2–1.7 K the crystal growth rate is limited by slow heat-conduction processes and viscosity in a finite surface layer of the liquid above the crystal boundary, where relaxation to quasistationary equilibrium of the supercooled superfluid at supercritical pressure takes place [$\mu(\infty) \neq \mu$, $T(\infty) \neq T'$]. It can be shown [see (22)] that the mean free paths of elementary excitations are small in comparison with the thickness of this layer. The problem of calculating the kinetic growth factor K can therefore be solved on the basis of superfluid hydrodynamics[10] with boundary conditions stipulating local

equilibrium on the surface of the crystal. In Sec. IV we derive an expression [Eq. (33)] for the factor K in terms of the thermal conductivity and viscosity of the superfluid. This expression satisfactorily describes the salient features of the experimental results.

2. EXPERIMENTAL PROCEDURE

The procedure is based on measurement of the pressure in the container during the growth of a crystal after its nucleation in a metastable liquid. The container is cooled at a constant rate until a critical solid-phase nucleus forms in the liquid. As the crystal grows, the pressure in the container drops from the maximum value to the phase-equilibrium pressure. The pressure drops at a rate proportional to the rate of change of the volume of the crystal and is related to the kinetic growth factor K. Consequently, the growth rate V can be calculated by measuring the time variation of the pressure $p(t)$, and the factor K can be determined when $V \propto \Delta p$.

This procedure can extend the range of supersaturations by three orders of magnitude from the characteristic values for previous experiments, 0.01–0.1 mbar, to values of 15–40 mbar. The pressure is recorded with high time resolution, i.e., it is possible to investigate the growth kinetics in time periods from seconds to fractions of a millisecond. The procedure is subject to limitations, first, by the probabilistic nature of the way in which a critical nucleus is formed. It emerges on the inner surface of the container when the supersaturation reaches a level determined by the properties of the surface at the nucleation site. In other words, the starting pressure is a property of the container surface and cannot be altered in the course of the experiment. We treated the container with hydrogen to maximize the supersaturation (see Ref. 11 for details).

A second limitation is associated with the supersaturation technique. The pressure was measured by means of a capacitative transducer formed by the container wall and an electrode. The transducer was used to measure pressure near the crystal when the growth time t_{gr} was much greater than the transit time τ_1 of first sound in the container, $t_{gr} \gg \tau_1 \sim 30\,\mu$s. It will be shown below that the calculation of the kinetic growth factor requires that the helium temperature be measured during the growth process. Based on the following considerations, we chose not to measure the temperature directly: Like the pressure sensor, a thermometer shows the temperature at the boundary of the crystal when the crystal growth time is much greater than the transit time τ_2 of second sound, $t_{gr} \gg \tau_2 \sim 0.5\,\mu$s. The simultaneous satisfaction of these two conditions implies that the helium pressure and temperature are constant throughout the volume of the container and that crystal growth is quasistationary. It has been shown[11] that the temperature variation in this case is related one-to-one to the pressure variation and can be calculated from the function $p(t)$, so that the measurements can be confined to the pressure only. These two conditions place a lower temperature limit on the applicability of the given procedure, $T \geqslant 1.2$ K.

The helium crystals were grown in a stainless steel container having an internal volume $\mathscr{V}_0 = 2$ cm^3; see Fig. 1. The

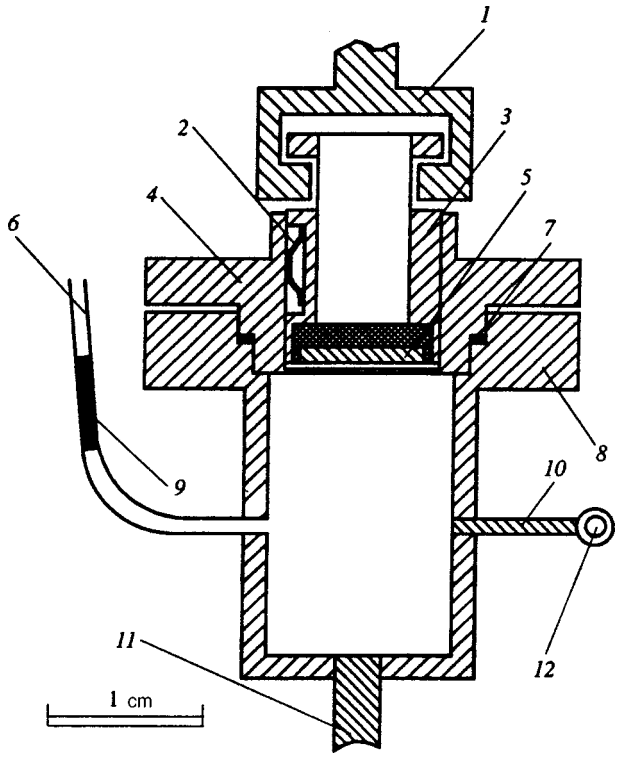

FIG. 1. Structure of the container. The numbers are explained in the text.

container consists of an outer casing 8 with a sealed-in copper cold conductor *11*, which leads to a ^{3}He reservoir, and a flange 4, which forms the base of the capacitative pressure transducer. The helium temperature is monitored by a resistance thermometer *12* soldered to a copper cold conductor 10, which is in direct contact with superfluid helium. An indium seal *7* maintains the vacuum tightness of the container casing and the flange.

The capacitative pressure transducer is formed by the web of the flange and an electrode *5*. The resulting capacitor is connected to an induction coil mounted in a superconducting baffle to reduce radiation losses and to raise the Q of the circuit. This circuit is weakly coupled (1:100) by means of a coaxial lead to a self-excited oscillator located on the lid of a cryostat at room temperature. The Q of the circuit at helium temperature is ~ 1000, which ensures frequency stability within limits of the order of 1 Hz during a measurement time of 1 s at an oscillator frequency of approximately 7 MHz. This procedure can be used to measure the pressure in relative frequency units, i.e., it is necessary to calibrate the measurement system against a pressure standard. Our standard was a manometer, whose calibration in the range of 26–30 atm was verified on the helium solidification curve. It will be shown below that the deviation of the pressure in the container during crystal growth from the phase-equilibrium pressure was less than 0.05 atm, which is much lower than the phase-equilibrium pressure ≈ 25 atm *per se*; it was therefore sufficient to determine the transducer sensitivity df/dp at the measurement site. For this purpose the function $f(p)$ was measured in each series of experiments in the pressure range 1–30 atm at a temperature ≈ 2.5 K, and the result was used to determine the differential sensitivity of the trans-

ducer. This procedure had to be performed after every repositioning of the transducer electrode *5* (see below). The transducer sensitivity was determined within approximately 3% error limits.

The superfluid helium was cooled at a constant rate until a critical nucleus formed and crystal growth began. The container was connected to the outer volume by a capillary tube *6* (see Fig. 1) having a length ≈ 2 m and an inside diameter of 0.15 mm. The total volume of the filling capillary tube ($\approx 0.03\,\mathrm{cm}^3$) was much smaller than the volume of the container. The first series of measurements was performed with this capillary tube and a pressure transducer whose web had a thickness of 0.55 mm at a diameter of 10 mm. The transducer sensitivity was equal to 4.1 kHz/bar. The measurements were performed at time increments of 1.024 ms. The relaxation time of the pressure in the system was $\tau_{\mathrm{rel}} = 2.76$ s. It will be shown below that this time is comparable with the crystal growth time at temperatures close to the superfluid transition line, so that corrections must be introduced to adjust to the flow of liquid along the capillary tube.

To diminish the influence of this effect, in the second series of experiments a choke *9* was inserted in the helium feed line to increase the pressure relaxation time in the system to 37 s. In addition, the sensitivity of the transducer was enhanced by decreasing the thickness of the flange web to 0.27 mm and accurately positioning the electrode at a minimum distance from the web. In assembling the device, it was difficult to set the optimum web-electrode gap at room temperature on account of, for example, heat shrinkage during cooling and deformation of the web as the pressure required for crystallization was created in the container. For these reasons the optimum gap was set directly in the experiment at helium temperature. The transducer electrode was mounted on a cylindrical ring *3* (see Fig. 1), which was inserted in the channel of the flange *4* and was held fast against its surface by a spring *2*. Two symmetrical slots were formed in the upper part of the ring for the insertion of a yoke *1*, which was connected to the stem of a piezoelectric driver (not shown in the figure). Piezoelectric drivers of this kind are used in a tunneling microscope to move a specimen in 0.1–1-μm steps (Ref. 12).[1)]

Once the necessary pressure had been established in the container, prior to the beginning of a series of experiments the electrode was positioned just short of contact with the web of the flange. The direction of motion of the piezoelectric driver stem was then reversed, and the necessary gap was set, having been determined from the oscillator frequency. The yoke was then moved forward to eliminate mechanical contact between the container and the piezoelectric driver, thereby significantly reducing the influence of mechanical vibrations on the transducer. The time resolution was improved to 0.25 ms in the second series of experiments.

3. PROCESSING OF THE MEASUREMENT RESULTS

To calculate the kinetic growth factor K, it is necessary to know the growth rate V and the difference in the chemical potentials $\Delta\mu = \mu - \mu'$:

$$V = K\Delta\mu. \tag{1}$$

The difference in the chemical potentials is given by the expression

$$\Delta\mu = -(s - s')\Delta T + \frac{\rho' - \rho}{\rho\rho'}\Delta p, \tag{2}$$

where s is the entropy per unit mass, $\Delta p = p - p_0$ and $\Delta T = T - T_0$ are the deviations of the pressure and the temperature from their equilibrium values p_0 and T_0. In our experiments the mass of the helium in the container was constant during crystal growth. In this case, as mentioned, the temperature variation during growth is related to the variation of the pressure by an equation derived in Ref. 12:

$$\Delta T = -\Delta p\,\frac{k\,dp/dT - \beta}{\rho C_p - \beta\,dp/dT}, \tag{3}$$

where dp/dT is the slope of the phase-equilibrium curve, β is the coefficient of thermal expansion, k is the compressibility, and C_p is the heat capacity at constant pressure. Taking Eq. (3) into account, we write Eq. (2) in the form

$$\Delta\mu = \frac{\rho' - \rho}{\rho\rho'}Dp, \quad Dp = \alpha(T)\Delta p,$$

$$\alpha(T) = 1 + \frac{dp}{dT}\,\frac{k\,dp/dT - \beta}{\rho C_p/T - \beta\,dp/dT}, \tag{4}$$

where the measure of nonequilibrium Dp is expressed in units of pressure. The correction for temperature variations is a maximum at high temperatures and can attain a value $\alpha \approx 1.25$. We note that the absolute temperature variations of the liquid helium during crystal growth did not exceed ~ 1 mK.

The volume of the crystal $\mathscr{V}_{\mathrm{crys}}$ is related to the pressure drop during growth by the equation

$$\mathscr{V}_{\mathrm{crys}} = \mathscr{V}_0\,\frac{\rho}{\Delta\rho}\,k[p(t=0) - p(t)], \tag{5}$$

where $t = 0$ is the crystal nucleation time. For initial supersaturations $\Delta p_{\mathrm{init}} = 15$–$40$ mbar the final volume of the crystal was 2–4 mm^3, i.e., the final radius of the crystal was approximately one millimeter. To relate the rate of change of the volume to the linear growth rate of the crystal, we need to make an assumption about the shape of the crystal, which is determined by the anisotropy of the kinetic growth factor and the contact of the crystal with the container wall. The hydrostatic pressure gradient does not introduce a significant contribution, because the pressure difference over the width of the crystal (≈ 0.02 mbar) is much smaller than the supersaturations investigated in the present study. The pressure increment due to the curvature of the crystal surface is also much smaller than the supersaturation.

The kinetic growth factor is isotropic in the region where the cubic phase exists. In the hexagonal phase, on the other hand, Bodensohn *et al.*[3] have observed anisotropy of the growth factor at $T = 1.35$ K with rate ratios $V_{\langle 1120\rangle} : V_{\langle 1010\rangle} : V_{\langle 0001\rangle} = 2.8 : 2.5 : 1$. However, these results are contradicted by reported measurements of the kinetic growth factor[13,14] using crystallization waves, where such pronounced anisotropy of the growth of atomically rough

surfaces has not been observed (the linear decay of K along the [0001] face is a geometrical effect[15]). One of the present authors[16] has conducted a special investigation of the growth anisotropy on the basis of the from of crystal growth, i.e., by a procedure similar to that of Bodensohn *et al.*[3] In all experiments with a freely growing helium crystal (see the series of photographs of a growing crystal at $T=1.285$ K in Ref. 16, Fig. 2), the large (1:3) anisotropy between the direction of the C_6 axis and the perpendicular direction was not noted, whereas the small (~10%) anisotropy in the basal plane observed by Tsymbalenko[16] and also by Bodensohn *et al.*,[3] can also be seen. It is therefore justified to regard the kinetic growth factor as being almost isotropic in the hcp phase as well.

During isotropic growth the crystal is in the shape of a spherical segment at the wall, and its volume is related to the radius of the segment R by the equation

$$\mathscr{V}_{\text{crys}}=\frac{\pi}{3}R^3(2-3\cos\theta+\cos^3\theta), \tag{6}$$

where θ is the contact angle. With allowance for Eqs. (5) and (6) the crystal growth rate is given by the equation

$$V=\frac{dR}{dt}=\gamma\left(\mathscr{V}_0\frac{\rho}{\rho'-\rho}k\Delta p_{\text{init}}\right)^{1/3}\frac{d}{dt}\left(\frac{p_{\text{init}}-p(t)}{\Delta p_{\text{init}}}\right)^{1/3},$$
$$\gamma=\left[\frac{3}{\pi(2-3\cos\theta+\cos^3\theta}\right]^{1/3}. \tag{7}$$

The geometrical factor γ for a sphere ($\theta=\pi$) is approximately equal to 0.62. According to visual observations, the contact angle between solid helium and stainless steel has values in the interval 90–120°, which corresponds to values of the geometrical factor 0.66–0.78. If the crystal forms in corners of the container, the expression for the geometrical factor is enormously complicated, but its value is still close to that given by Eq. (7) for large contact angles. We note that for a crystal of finite radius ~1 mm the area near the angles amounts to ≈8% of the total inside area of the container, i.e., the probability of crystal formation in the corners is small.

The error of determination of the volume is a maximum at the beginning of growth, when the pressure differs very little from the initial pressure. The relative error of determination of the supersaturation increases at the end of growth. In view of these considerations we have chosen to process the part of the pressure-drop curve in the interval $(0.05–0.95)\Delta p_{\text{init}}$. The total error of the growth rate in the first series of experiments was 10–20% and increased to 50% at the lowest temperatures. The error was almost halved in the second series of experiments.

4. RESULTS OF THE MEASUREMENTS AND DISCUSSION

In both series of measurements the initial saturation Dp_0 in the cubic phase was 15–18 mbar. In the hexagonal phase we found $Dp_0=8–10$ mbar in the first series of measurements, and $Dp_0=30–40$ mbar in the second series. Typical experimental curves of the pressure drop in the container are shown in Fig. 2. Clearly, the crystal growth time near the superfluid transition line is of the order of 0.5 s and decreases rapidly as the temperature drops to values of the order of a few milliseconds. The results of processing these curves according to Eqs. (4) and (7) are shown in Fig. 3. The kinetic growth factor is determined from the initial part of the $V(Dp)$ curve, where the growth rate is proportional to the supersaturation within the known error limits.

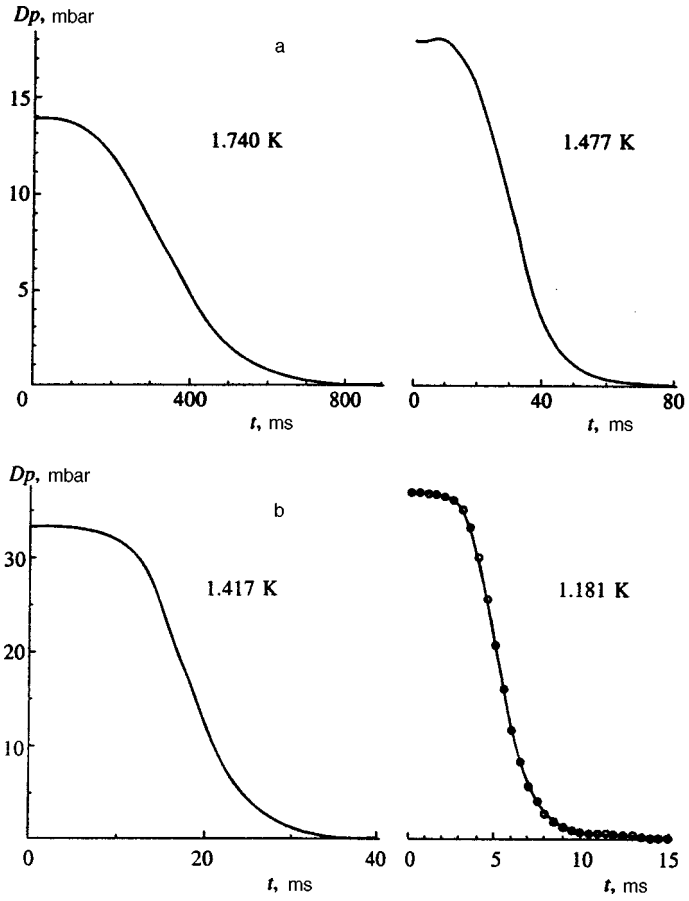

FIG. 2. Experimental plots of the pressure drop in the container during crystal growth: a) bcc phase; b) hcp phase.

Temperature curves of K obtained in both series of experiments with the parameter $\gamma=1.2$ are shown in Fig. 4. We see that the values of the kinetic growth factor obtained in different series for various initial supersaturations agree within the error limits. The temperature variation of K agrees with the results of Refs. 3 and 16. Also visible in Fig. 4 is the previously noted[3,7] deviation of the temperature variation of K from a simple exponential law above 1.5 K. And it is evident from Fig. 4 that up to temperatures of the order of 1.65 K the slope of $\ln(K)$ as a function of $1/T$ in the bcc phase does not deviate too noticeably from the slope in the hcp phase. Above this temperature the slope increases considerably, probably by virtue of proximity to the λ transition. Figure 5 shows the behavior of the kinetic growth factor in this temperature range. Within the error limits the temperature dependence of K agrees satisfactorily with the relation $K\propto(T_\lambda-T)^\varepsilon$, where $\varepsilon=0.743\pm0.123$. We emphasize that this equation merely serves to illustrate the temperature behavior in the interval close to the superfluid transition in the measured temperature range and does not purport to describe

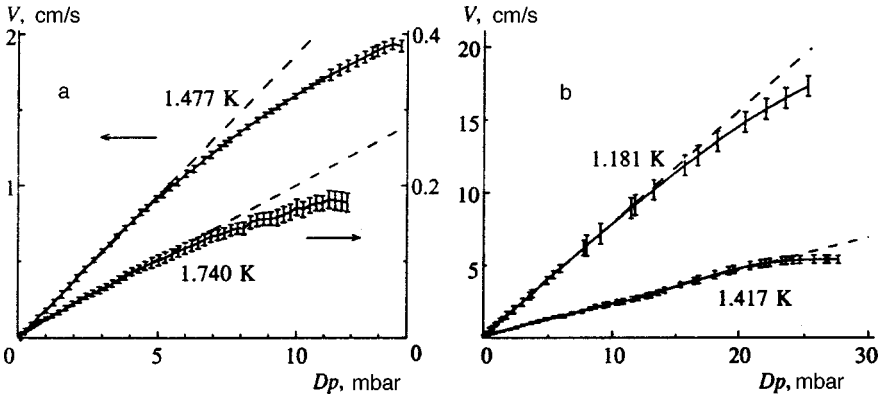

FIG. 3. Growth rate versus supersaturation: a) bcc phase; b) hcp phase. The graphs clearly reveal the reduction of the growth rate above 5 mbar in the cubic phase; in the hexagonal phase the growth rate is proportional to the supersaturation up to 25 mbar.

the temperature dependence of K beyond the limits of this interval.

The extrapolation of K to the bcc-hcp transition point yields $1/K_{\rm bcc}=9.9\pm0.86$ m/s on the cubic phase side and $1/K_{\rm hcp}=11.8\pm2.09$ m/s on the hexagonal phase side. The transition from the bcc to the hcp phase is therefore accompanied by a jump of the kinetic growth factor.

In the bcc phase (see Fig. 3), beginning with supersaturations ≈5 mbar, a deviation from linearity is observed, whereas in the hexagonal phase the nonlinearity is much weaker, and the deviation becomes appreciable above ≈25 mbar. To take nonlinearity into account, we add to Eq. (1) the next term of the expansion of the function $V(Dp)$ in powers of Dp in the form

$$V=K\frac{\rho'-\rho}{\rho\rho'}Dp\left(1-\frac{Dp}{G}\right). \tag{8}$$

In the bcc phase the parameter G has the value $G=38.6\pm2.7$ mbar and does not depend on the temperature. This means that the form of the function $V(Dp)$ remains invariant when the growth rate increases by an order of magnitude. For the hcp phase, increasing errors make it difficult to draw a straightforward conclusion.

The kinetics of the growth of a ^{4}He crystal is attributable to two processes: 1) a fast, microscopic mechanism of transfer of an atom from the liquid to the solid phase across the atomically rough surface; 2) slow mass and heat transfer through the surrounding medium.

The total difference in the chemical potentials $\Delta\mu$ [see Eq. (1)] is the sum of the jump of the chemical potential at the boundary of the crystal and its change in the bulk of the liquid. The total reciprocal of the kinetic growth factor can then be written as the sum of the reciprocals of the surface and bulk factors:

$$\frac{1}{K}=\frac{1}{K_{\rm surf}}+\frac{1}{K_{\rm bulk}}. \tag{9}$$

Inasmuch as surface processes are much faster than bulk processes, we have $K_{\rm surf}\gg K_{\rm bulk}$, so that crystal growth is governed by transport in the liquid. Experimental facts such as the isotropy of the growth factor and the closeness of the numerical values in both phases imply that the influence of crystal structure on the growth kinetics is weak at low supersaturations. Consequently, the main dissipative processes governing crystal growth take place in the liquid. The kinetic growth factor will be calculated Sec. 5 below on this assumption. There the calculations, which are carried out in the

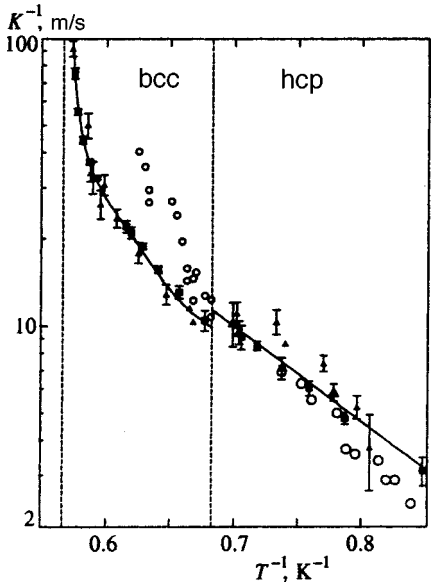

FIG. 4. Kinetic growth factor vs temperature. The light circles represent the results of Ref. 3, and the light triangles give data from Ref. 16. The dark triangles and squares represent our results in the first and second series of experiments, respectively. The dashed vertical lines separate the regions of the regions of the bcc and hcp phases.

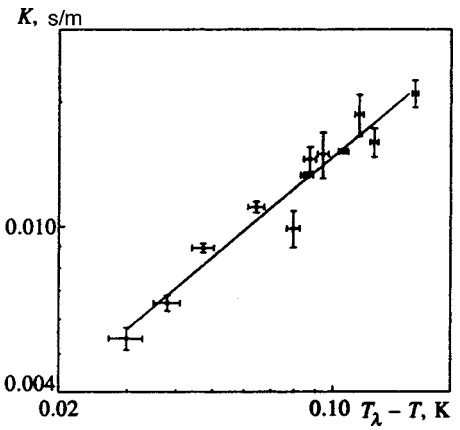

FIG. 5. Temperature behavior of the kinetic growth factor near the superfluid transition.

linear approximation, cannot reflect the experimentally observed behavior of the growth rate at high supersaturations.

We have therefore inquired into the legitimacy of disregarding squared terms in the exact superfluid hydrodynamic equations. Numerical estimates show that the squared terms are small over the entire temperature range at all growth rates, i.e., the observed nonlinear effect cannot be described by including them. Moreover, in experiment we see that the nonlinearity is related to the crystal structure of the solid phase. This relationship supports the hypothesis that the slowing of the growth rate is associated with the microscopic structure of the sample surface and, as a consequence, with nonlinear processes of the step kinetics on the atomically rough surface. Hence it follows that $K_{\rm surf}$ plays an important role at high supersaturations. The theory of crystal growth at high supersaturations will not be discussed in the present article.

5. CALCULATION OF THE KINETIC GROWTH FACTOR

We consider one-dimensional, quasistationary crystal growth in a coordinate system attached to the crystal-liquid interface. At high temperatures, contrary to the low-temperature case, the heat of crystallization q plays an important role:

$$q=T(s-s')=T\,\frac{\rho'-\rho}{\rho\rho'}\,\frac{dp}{dT}. \tag{10}$$

The liberation of heat on the atomically rough surface of the crystal ($x=0$) during its growth induces rapid relaxation to local equilibrium between the crystal and the boundary layer of the liquid:

$$T'=T(0),\quad p'=p(0),\quad \mu'=\mu(0). \tag{11}$$

We therefore disregard the Kapitza jump; according to experimental data,[17] it is very small on an atomically rough surface at high temperatures. Local equilibrium at the boundary of the crystal upsets thermal equilibrium in the space between the surface of the crystal and the walls of the container:

$$\delta p=p(0)-p,\quad \delta T=T(0)-T,\quad \delta\mu=\mu(0)-\mu, \tag{12}$$

From now on T, p, and μ denote the values of the thermodynamic variables far from the crystal nucleus. Here, in addition to the unavoidable flow of liquid as a whole, we also encounter relative viscous flow of the liquid components and heat flux away from the boundary of the crystal. These dissipative fluxes limit the crystal growth rate and, in the final analysis, determine its value. The temperature and pressure differences (12) can be related to the degree of nonequilibrium (2) by the equation

$$\Delta\mu=\frac{\rho'-\rho}{\rho\rho'}\left(\frac{dp}{dT}\,\delta T-\delta p\right). \tag{13}$$

Here $\Delta\mu=\mu(T,p)-\mu'(T,p)$ is the degree of nonequilibrium of the system.

The process of crystal growth can be regarded as quasistationary when the transit times of first and second sound from the crystal to the container wall are much shorter than the crystal growth time. This condition is consistent with the domain of applicability of our measurement procedure and confines the given investigation to the temperature range $1.2\ \mathrm{K}<T<T_\lambda=1.763\ \mathrm{K}$.

In a coordinate system affixed to the crystal boundary the growth kinetics is described by the system of static equations of superfluid hydrodynamics[10] linearized with respect to small deviations from equilibrium:

$$\frac{dj}{dx}=0,\quad j=\rho_s v_s+\rho_n v_n, \tag{14}$$

$$\frac{dp}{dx}=\left(\frac{4}{3}\,\eta+\zeta_2-\rho\zeta_1\right)\frac{d^2v_n}{dx^2}. \tag{15}$$

$$\frac{dQ}{dx}=0,\quad Q=\mu j+T\rho s v_n-\kappa\frac{dT}{dx}, \tag{16}$$

$$\frac{d\mu}{dx}=(\zeta_1-\rho\zeta_3)\frac{d^2v_n}{dx^2}. \tag{17}$$

We recall that ρ_s, ρ_n and v_s, v_n are the densities and velocities of the superfluid and normal components ($\rho=\rho_s+\rho_n$), j and Q are the mass and energy flux densities of the liquid η, ζ_1, ζ_2, and ζ_3 are the corresponding viscosity coefficients, and κ is the thermal conductivity of the liquid. The thermal conductivity of the crystal on the melting curve is small[18] and will be disregarded. This assumption will be justified by estimates below.

We note that the trivial solution of the system (14)–(17) when all the gradients are zero does not satisfy the boundary conditions, which state that the values of the pressure, temperature, and chemical potential on the crystal surface and in the depth of the liquid differ and are related by Eq. (13). To find a nontrivial solution, we invoke the thermodynamic relation

$$d\mu=-s\,dT+dp/\rho \tag{18}$$

and then use (15) and (17) to relate the temperature gradient to the second derivative of the velocity of the normal component:

$$\frac{dT}{dx}=\zeta_0\frac{d^2v_n}{dx^2},\quad \zeta_0=\frac{4}{3}\,\eta+\zeta_2+\rho^2\zeta_3-2\rho\zeta_1. \tag{19}$$

From Eqs. (16) and (19) we find an equation for the velocity of the normal component:

$$T\rho s v_n-\frac{\kappa\zeta_0}{\rho s}\,\frac{d^2v_n}{dx^2}=Q-\mu j. \tag{20}$$

It follows from the laws of conservation of mass (14) and energy (16) that the right-hand side of Eq. (20) is constant. From this result we obtain the law governing the decay of the velocity of the normal component:

$$\frac{dv_n}{dx}=C\exp\left(-\frac{x}{L}\right),\quad L=\frac{\sqrt{\kappa T\zeta_0}}{T\rho s}. \tag{21}$$

The thickness of the transition layer in this temperature range is of order $L\sim10^{-6}$ cm. The radius of the crystal is much greater than the thickness of the transition layer, justifying a one-dimensional approach to the problem. For the

hydrodynamic approach to be valid, it is necessary to satisfy the inequality $\lambda < L$, where λ is the mean free path of elementary excitations in the superfluid. Rotons play the principal role in the given temperature range.[10] Determining the roton mean free path from the viscosity,[10] we obtain the following expression for the ratio λ/L:

$$\frac{\lambda}{L} \leqslant \frac{15}{\sqrt{20/3}} \frac{T\rho s}{\rho_s v_r v_c}, \tag{22}$$

where v_r is the roton thermal velocity, and v_c is the critical velocity in helium. In the temperature interval 1.2–1.7 K we have the numerical relation $\lambda/L \leqslant 0.1$–0.2, i.e., the inequality is satisfied, and the hydrodynamic approach is legitimate.

Substituting Eq. (21) into (15) and (19), we determine the behavior of the pressure and the temperature in the transition layer:

$$\delta p(x) = \eta_0 C \exp\left(-\frac{x}{L}\right), \quad \eta_0 = \frac{4}{3}\eta + \zeta_2 - \rho\zeta_1, \tag{23}$$

$$\delta T = \frac{\zeta_0}{\rho s} C \exp\left(-\frac{x}{L}\right). \tag{24}$$

The factor C is readily determined from Eq. (13):

$$C = A\Delta\mu, \quad A^{-1} = \frac{\rho' - \rho}{\rho\rho'}\left(\frac{dp}{dT}\frac{\zeta_0}{\rho s} - \eta_0\right). \tag{25}$$

The coefficient A is positive at high temperatures. This means, as should be expected, that the temperature and the pressure in the liquid layer around a nucleus are higher than near the container walls. However, the magnitude of this effect is small when the derivative dp/dT is large:

$$\frac{\delta p(0)}{\Delta p} = \frac{\alpha\eta_0}{(dp/dT)(\zeta_0/s\rho) - \eta_0} \ll 1. \tag{26}$$

Thus, Eqs. (21)–(25) give the dependence of δp, δT, and dv_n/dx on the degree of nonequilibrium $\Delta\mu$. The fluxes j and Q are obtained form the law of continuity of the mass and energy fluxes at the crystal-liquid interface:

$$\begin{aligned} &j = -\rho' V = \rho_s v_s + \rho_n v_n, \\ &Q = j(Ts' + \mu') = \mu j + T s\rho v_n - \kappa\frac{dT}{dx}, \\ &Q - \mu j = jTs'. \end{aligned} \tag{27}$$

From Eqs. (14), (16), and (27) we readily find the velocities of the superfluid and normal components as functions of V and $\Delta\mu$ at the interface:

$$\begin{aligned} &v_n = -\frac{s'\rho'}{s\rho} V - LC\exp\left(-\frac{x}{L}\right), \\ &v_s = V\frac{\rho'}{\rho_s}\left(\frac{s'\rho_n}{s\rho} - 1\right) + \frac{\rho_n}{\rho_s} LC\exp\left(-\frac{x}{L}\right). \end{aligned} \tag{28}$$

Additional considerations must be brought into the picture to determine how δp, δT, and dv_n/dx are related to each other.

It has been shown[9] that the energy dissipated per unit time and per unit surface area for a planar crystal boundary growing with velocity V is given by

$$\frac{dE}{dt} = \rho'\Delta\mu V = \rho' K(\Delta\mu)^2. \tag{29}$$

On the other hand, the energy dissipated in a superconducting fluid is given by the integral of the dissipation function:[10]

$$\frac{dE}{dt} = \int R\, dx, \quad R = \left(\zeta_0 - \frac{4}{9}\eta\right)\left(\frac{dv_n}{dx}\right)^2 + \frac{\kappa}{T}\left(\frac{dT}{dx}\right)^2. \tag{30}$$

With allowance for Eqs. (21) and (24) the integral of the dissipation function is

$$\int R\, dx = LC^2\zeta_0\left(1 - \frac{2\eta}{9\zeta_0}\right). \tag{31}$$

Equating (29) and (31), we find

$$K = LA^2\frac{\zeta_0}{\rho'}\left(1 - \frac{2\eta}{9\zeta_0}\right). \tag{32}$$

Here we substitute Eq. (25) and reduce the result to a form suitable for analysis:

$$\begin{aligned} K^{-1} = &\frac{1}{Ts\rho\rho'}\left(T\frac{\rho' - \rho}{\rho}\right)^2\left(\frac{dp}{dT} - s\rho\frac{\eta_0}{\zeta_0}\right)^2 \\ &\times\sqrt{\frac{\zeta_0}{T\kappa}}\left(1 - \frac{2\eta}{9\zeta_0}\right)^{-1}. \end{aligned} \tag{33}$$

It is evident from this expression that in this approximation the kinetic growth factor is governed by the thermodynamic parameters of the liquid and solid phases and by the dissipation factors of the superfluid. The growth rate varies approximately as the inverse of the square of the heat of crystallization.

We now estimate the contribution of the thermal conductivity of the crystal to the kinetic growth factor. When the crystal grows at a constant rate at constant supersaturation, it is evident from Eq. (24) that the solid phase forms at the same temperature and, accordingly, there are no heat fluxes or additional dissipation associated with the thermal conductivity of the crystal. In a real experiment the supersaturation varies from the maximum Dp_0 at the instant of nucleation of the crystal to zero at the end of the growth period. Inasmuch as the temperature of the newly formed solid-phase layer, according to Eqs. (3) and (24), depends on the supersaturation, the crystal acquires a temperature gradient and, as a consequence, additional dissipation $E_{\rm crys}$ associated with the thermal conductivity, which lowers the crystal growth rate. According to Eq. (29), the total energy dissipation $E_{\rm sum}$ during the growth time $t_{\rm gr}$ is

$$E_{\rm sum} \sim \rho' K\Delta\mu^2 R^2 t_{\rm gr} \sim \left(\frac{\Delta\rho}{\rho}\right)^2\frac{1}{\rho'} KR^2 Dp_0^2 t_{\rm gr},$$

where R is the radius of the crystal. It follows from the experimental supersaturations and Eq. (5) that $R \sim 1$ mm. Equations (3) and (24) can be used to show that the initial temperature at the start of growth is higher than the final temperature by the amount $\Delta T_{\rm crys} \sim Dp_0/(dp/dT)$. The dis-

sipation $E_{\rm crys}$ associated with the thermal conductivity of the crystal can then be estimated from the expression

$$E_{\rm crys}\sim t_{\rm gr}\frac{\kappa'}{T}\int(\nabla T)^2 dV\sim t_{\rm gr}\frac{\kappa'}{T}\frac{Dp_0^2}{(dp/dT)^2}R.$$

The ratio of $E_{\rm crys}$ to the total dissipation is

$$\frac{E_{\rm crys}}{E_{\rm sum}}\sim\left(\frac{\rho}{\Delta\rho}\right)^2\frac{\kappa'}{T}\left(\frac{dp}{dT}\right)^{-2}\rho' K^{-1}\frac{1}{R}\sim 10^{-4},$$

which can be disregarded, as indeed we have assumed in our analysis.

We now analyze the contribution of the parameters to the kinetic growth factor. In Eq. (33) the last factor lies in the interval 1–1.2, because the inequality $2\eta/9\zeta_0<1/6$ holds. The ratio η_0/ζ_0 is of order unity. In light of these remarks we rewrite Eq. (33) in the form

$$K^{-1}\approx\frac{1}{Ts\rho\rho'}\left(T\frac{\rho'-\rho}{\rho}\right)^2\left(\frac{dp}{dT}-s\rho\right)^2\sqrt{\frac{\zeta_0}{T\kappa}}\,. \qquad (34)$$

The first three factors characterize the variation of K due to variations of the thermodynamic variables of helium. Their values on the melting curve are well known. The last factor is associated with the kinetic coefficients of the superfluid. Unfortunately, experimental data are unavailable on the quantities η, ζ_1, ζ_2, ζ_3, and κ on the melting curve in the given pressure and temperature ranges. The second viscosity ζ_2 is known to be one and a half orders higher than the first viscosity at the saturated vapor pressure. As the pressure increases, the first viscosity increases,[19] and measurements of the damping of first sound[20] imply that the second viscosity ζ_2 decreases. It follows from the Onsager relations for the kinetic coefficients[10] that the combination of the second viscosity coefficients in ζ_0 is positive. Summing up this information and making use of the expressions for the thermal conductivity and the viscosity in the roton domain, we infer that the factor associated with the dissipation factors of the liquid satisfy the inequality

$$\sqrt{\frac{\zeta_0}{T\kappa}}>\sqrt{\frac{4\eta/3}{T\kappa}}=\frac{2}{\sqrt{15}}\frac{1}{v_c}. \qquad (35)$$

We calculate the temperature dependence of the critical velocity at $p\sim 25$ atm from the values of the roton gap and the position of the roton minimum from Ref. 8. With allowance for the approximations and simplifications of our analysis Eq. (34) describes the temperature behavior of K, and the numerical value close in order of magnitude should be expected. We introduce a scale factor, $K_{\rm app}=mK_{\rm theor}$, to approximate the data. The result of the approximation using $m=0.2$ is shown in Fig. 6. Satisfactory agreement between calculations and experiment is observed. The decrease in K as the temperature increases is associated primarily with an increase in the slope of the phase-equilibrium curve. The variation of the roton gap, as postulated by Bodensohn *et al.*,[3] also contributes to the temperature dependence of K, but to a lesser degree, being proportional to the decrease in the critical velocity [see (35)]. At the bcc-hcp transition point

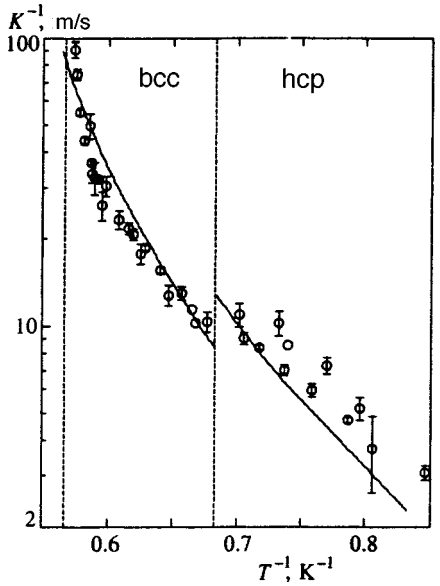

FIG. 6. Comparison of experimental data and the results of calculations according to Eq. (34).

the ratio of the growth factors is determined by the thermodynamic parameters and is equal to

$$\frac{K_{\rm bcc}}{K_{\rm hcp}}=\frac{\rho'_{\rm bcc}}{\rho'_{\rm hcp}}\left(\frac{\rho'_{\rm hcp}-\rho}{\rho'_{\rm bcc}-\rho}\right)^2\left[\frac{(dp/dT)_{\rm hcp}-s\rho}{(dp/dT)_{\rm bcc}-s\rho}\right]^2\approx 1.48. \qquad (36)$$

It is evident from Fig. 6 that the sign of the jump agrees with the experimental, and its magnitude is in satisfactory agreement with the experimental value.

We emphasize that the values of K have been calculated using the equations for the roton contribution to the kinetic coefficients and, for this reason, cannot describe the behavior of crystal growth in the critical region.

6. CONCLUSION

We have determined the temperature dependence of the kinetic growth factor by measuring the growth rate of a helium crystal at high temperatures. We have detected a jump of K at the bcc-hcp transition point and have determined the critical index in the behavior of the function $K(T)$ in the vicinity of the superfluid transition point. To explain these experimental results, we have analyzed the problem of quasistationary crystal growth in the hydrodynamic approximation. This model, based on the hypothesis that mass and heat transfer in the liquid play the principal role in the crystal growth kinetics, satisfactorily explains both the temperature behavior of the kinetic growth factor and the presence of the jump at the bcc-hcp transition point. The extension of the range of supersaturations to 40 mbar has exhibited the nonlinear character of the growth kinetics at high supersaturations.

This work has been supported by the Russian Fund for Fundamental Research (Grant N96-02-18511a) and the Netherlands Scientific Research Organization (Grant NWO-07-30-002).

*E-mail: vlt@isssph.kiae.ru

[1)]The authors are grateful to V. S. Édel'man for furnishing the piezoelectric driver.

[1] A. F. Andreev and A. Ya. Parshin, Zh. Éksp. Teor. Fiz. **75**, 1511 (1978) [Sov. Phys. JETP **48**, 763 (1978)].
[2] K. O. Keshishev, A. Ya. Parshin, and A. V. Babkin, JETP Lett. **30**, 56 (1979).
[3] J. Bodensohn, K. Nicolai, and P. Leiderer, Z. Phys. B **64**, 55 (1986).
[4] S. G. Lipson and E. Polturak, in *Progress in Low Temperature Physics*, Vol. 9, edited by D. F. Brewer (Elsevier, Amsterdam-New York, 1987).
[5] E. Rolley, E. Chevalier, C. Guthmann, and S. Balibar, Phys. Rev. Lett. **72**, 872 (1994).
[6] R. M. Bowley and D. O. Edwards, J. Phys. (Paris) **44**, 723 (1983).
[7] V. L. Tsymbalenko, Phys. Lett. A **211**, 177 (1996).
[8] O. W. Dietrich, E. H. Graf, C. H. Huang, and L. Passell, Phys. Rev. A **5**, 1377 (1972).
[9] A. F. Andreev and V. G. Knizhnik, Zh. Éksp. Teor. Fiz. **83**, 416 (1982) [Sov. Phys. JETP **56**, 226 (1982)].
[10] I. M. Khalatnikov, *Theory of Superfluidity* [in Russian] (Nauka, Moscow, 1971).
[11] V. L. Tsymbalenko, J. Low Temp. Phys. **88**, 55 (1992).
[12] I. N. Khlyustikov and V. S. Édel'man, Prib. Tekh. Éksp. **1**, 158 (1996).
[13] O. A. Andreeva, K. O. Keshishev, and S. Yu. Osip'yan, JETP Lett. **49**, 759 (1989).
[14] O. A. Andreeva and K. O. Keshishev, JETP Lett. **52**, 164 (1989).
[15] P. Nozieres and M. Uwaha, J. Phys. (Paris) **48**, 389 (1987).
[16] V. L. Tsymbalenko, Fiz. Nizk. Temp. **21**, 162 (1995) [Low Temp. Phys. **21**, 120 (1995)].
[17] H. J. Maris and T. E. Huber, J. Low Temp. Phys. **48**, 99 (1982); J. Low Temp. Phys. **48**, 463 (1982).
[18] A. A. Golub and S. V. Svatko, Fiz. Nizk. Temp. **6**, 957 (1980) [Sov. J. Low Temp. Phys. **6**, 465 (1980)].
[19] V. I. Sobolev and L. A. Pogorelov, in *Proceedings of the 17th International Conference on Low Temperature Physics LT-17*, edited by U. Eckern *et al.*, (North-Holland, Amsterdam-New York, 1984).
[20] K. Dransfeld, J. A. Newell, and J. Wilks, Proc. R. Soc. London, Ser. A **243**, 500 (1958).

Translated by James S. Wood

Exciton-photon interaction in low-dimensional semiconductor microcavities

A. I. Tartakovskiĭ*) and V. D. Kulakovskiĭ

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

Yu. I. Koval' and T. B. Borzenko

Institute of Problems in the Technology of Microelectronics and High-Purity Materials, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

A. Forchel and J. P. Reithmaier†)

Technische Physik, Universität Würzburg, D 97074 Würzburg, Germany
(Submitted 5 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1329–1345 (October 1998)

The structure of the photon states and dispersion of cavity polaritons in semiconductor microcavities with two-dimensional optical confinement (photon wires), fabricated from planar Bragg structures with a quantum well in the active layer, are investigated by measuring the angular dependence of the photoluminescence spectra. The size quantization of light due to the wavelength-commensurate lateral dimension of the cavity causes additional photon modes to appear. The dispersion of polaritons in photon wires is found to agree qualitatively with the prediction for wires having an ideal quantum well, for which the spectrum is formed by pairwise interaction between exciton and photon modes of like spatial symmetry. The weak influence of the exciton symmetry-breaking random potential in the quantum well indicates a mechanism of polariton production through light-induced collective exciton states. This phenomenon is possible because the light wavelength is large in comparison with the exciton radius and the dephasing time of the collective exciton state is long. © *1998 American Institute of Physics.* [S1063-7761(98)01210-4]

1. INTRODUCTION

Semiconductor microcavities with imbedded quantum wells are currently stimulating a growing interest among researchers. They are intersting in connection with the feasibility of monitoring and investigating the modification of exciton properties as a result of the mixing of exciton states with cavity optical modes. In particular, exciton-photon interaction leads to Rabi splitting and modification of the dispersion of exciton and photon modes.[1–5] The bulk of the experimental papers report the interaction of light and excitons in cavities that imbed one or more quantum wells situated at antinodes of the optical electromagnetic field in an active layer of thickness λ or $(3/2)\lambda$ contained between two Bragg mirrors.[6–8] This geometry can be regarded as two-dimensional (2D) both for excitons in a well and for light in a Fabry–Perot cavity. In a planar microcavity optical momentum is not conserved in the direction perpendicular to the planes of the mirrors, whereas in a plane parallel to the mirrors the quasimomentum is a ''good'' quantum number, and the photon mode has a completely defined dependence of the energy E on k. This dependence can be measured by recording the reflection, transmission, or luminescence signal at various angles ϕ relative to the plane of the sample.[7] The planar quasimomentum k is related to the light momentum in vacuum q by the equation

$$k = q \sin \phi. \tag{1}$$

Here the dispersion of light in the empty microcavity is described by the expression[5,7]

$$E_k = \sqrt{E_v^2 + \alpha k^2}, \tag{2}$$

where E_v is the energy of the vertical cavity mode, and the coefficient α depends on the effective dielectric permittivity. In a microcavity containing quantum wells the energy of the photon and exciton modes is modified by the interaction of excitons in the wells with light. In the presence of strong exciton–photon coupling the exciton and photon dispersion branches are repulsive and can be described on the basis of the polariton model. The dispersion of a cavity polariton was first measured in Ref. 7 from an analysis of the angular dependence of the energies of the luminescence peaks at low excitation densities.

Papers have also been published very recently, describing investigations of three-dimensionally (3D) optically confined semiconductor microcavities, i.e., photon dots.[9,10] In these structures, with dimensions of the order of a few micrometers, light has been quantized in one dimension by means of Bragg mirrors and in the other two dimensions by a large difference in the refractive indices at the semiconductor–vacuum boundary. The localization of light in such structures produces a discrete photon spectrum, where the energy distance between modes increases as the size of the dot decreases. The dispersion of photon modes does not occur in such a cavity.

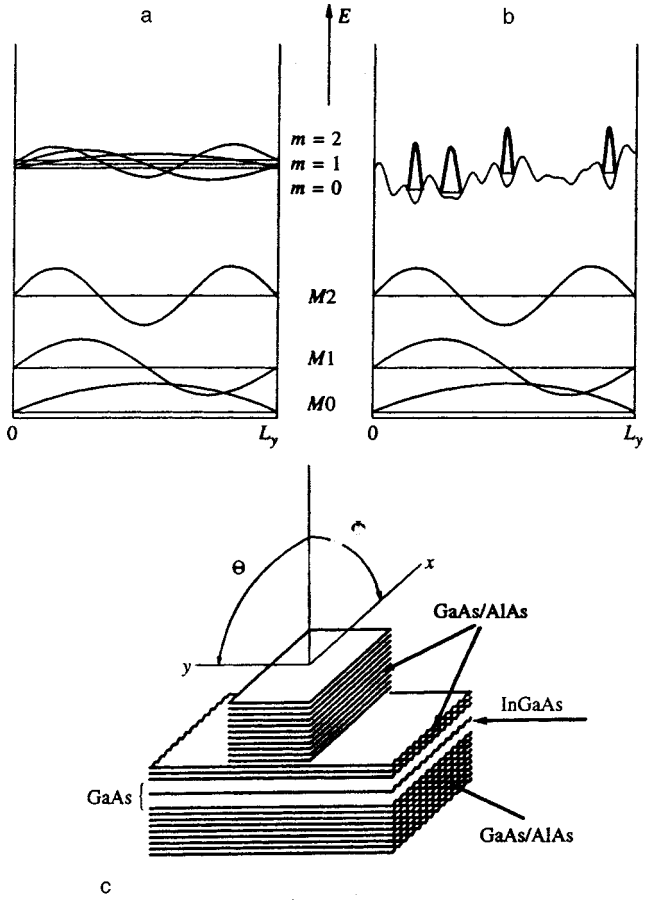

FIG. 1. a) Diagram of the exciton and photon states in photon wires with an ideal quantum well; b) the same with quantum-well localization of excitons by a random potential; c) schematic view of the sample, showing the planes in which the angles Φ and Θ are plotted.

In this paper we report an investigation of the dispersion of polaritons and the role of exciton–photon interaction in microcavities with 2D light quantization. Such cavities are aptly called photon wires. In photon wires 3D light confinement leads to the additional quantization of cavity photon states over and above the quantization in planar cavities, but in contrast with photon dots, all modes exhibit one-dimensional dispersion along the axis of the wire (x axis).

A diagram of the size-quantization levels in a microcavity of width L_y containing an ideal quantum well is shown in Fig. 1a. To simplify the diagram, only three exciton quantization levels ($m=0,1,2$) and three photon states ($M0, M1, M2$) situated below the exciton energy are shown in the figure. Under conditions such that the light wavelength is commensurate with the transverse dimensions of the cavity, the photon quantization energies attain several millielectron volts; on the other hand, owing to the large translational mass of the exciton, the energies of exciton states with different numbers m ($m=n-1$, where n is the number of the size-quantization level) differ only very slightly. Figure 1a also shows the exciton and photon wave functions corresponding to levels with different quantum numbers. The wave functions of excitons and photons with identical m have the same symmetry.

Exciton–photon interaction in one-dimensional microcavities with an ideal quantum well will necessarily lead to the mixing of photon and exciton states having the same spatial symmetry. This means that as the wave vector k_x increases, all photon modes whose energy at zero quasimomentum was lower than the exciton energy must make an anticrossing with an exciton mode of the same symmetry.[1)] For the case illustrated in Fig. 1a strong exciton–photon coupling results in the formation of three lower (Mm_L) and three upper (Mm_U) polariton modes.

In real quantum wells the localization of excitons by a random potential breaks the symmetry of the exciton states (Fig. 1b). The influence of a random potential on the formation and radiative properties of polaritons in planar microcavities has been a topic of debate up to this point.[11–13,15] In a system with broken spatial symmetry of the exciton states a light wave interacts with all excitons. In this case the properties of the polariton states depend on the dephasing time.[16] In photon wires, as opposed to planar microcavities, there are optical modes with different spatial symmetries. For this reason the investigation of polariton phenomena in low-dimensional microcavities offers deeper insight into the mechanism underlying the interaction of light with excitons in a system with disorder. The experimental results obtained in the present study suggest that the dispersion of polaritons in the investigated one-dimensional microcavities coincides with the expected dispersion for a resonator having an ideal quantum well. The reasons for the weak influence of exciton localization in quantum wells will be discussed below.

The article is organized as follows: In Sec. 2 we describe the method of preparation and the structure of the photon wires, along with the experimental technique; in Sec. 3 we discuss the structure, measured at high excitation density, of the optical states in photon wires in the presence of weak exciton–photon interaction; in Sec. 4 we describe the variation of the exciton–photon interaction as the density of excitons varies, along with the transition from the quantization of photon states to the quantization of polariton states; in Sec. 5 we discuss the dispersion of polaritons in photon wires in the presence of strong exciton–photon coupling, as determined experimentally at low excitation densities.

2. THE SAMPLE AND EXPERIMENTAL TECHNIQUE

For the measurements we used a microcavity structure grown by molecular-beam epitaxy. The active element of the microcavity comprises an $In_{0.14}Ga_{0.86}As$ quantum well of width 70 Å contained in a GaAs λ-layer between two Bragg mirrors. The active GaAs layer is wedge-shaped, making it possible (by exciting luminescence at different points of the sample) to vary the energy distance between the Fabry–Perot mode and the exciton level. The Bragg mirrors, having better than 99% reflectivity, consist of 17 and 21 AlAs/GaAs pairs (above and below the active layer, respectively), where each layer has a width equal to $\lambda/4$. Arrays of photon wires having widths from 4.2 μm to 6.2 μm were prepared by electron-beam lithography and ion-beam etching. A polymethyl methacrylate PMMA 950 K electronic resist of thickness 1.5 μm was deposited on the sample, and windows in the form of wires were opened in the resist by electron-beam

lithography. Next, an Al layer of thickness 1 μm was sprayed onto the entire sample, and explosion lithography was applied. The stenciled Al pattern then served as a mask for ion-beam etching. Etching was performed with 500-eV Ar ions at a current density of 0.1 mA/cm^2. Oblique low-temperature etching is known to have the potential benefit of reducing radiation damage to structures.[14] During etching, therefore, the sample was tilted at a 75° angle relative to the direction of the ion beam and was precisely oriented in such a way as to align the direction of incidence of the ions with the wires. The sample was kept at a temperature of 77 K during etching. Only the upper mirror of the microcavity etched, leaving two or three pairs of AlAs/GaAs layers between the untouched surfaces of the photon wires. A diagram of the sample is shown in Fig. 1c. Also shown in the figure are the planes in which were measured the angles Θ and Φ used in our description of the experimental results.

The dispersion of cavity polaritons was investigated by analyzing the angular dependence of the luminescence spectra. The sample was placed in an optical cryostat in helium vapor at a temperature $T=5$ K. Luminescence was excited by a HeNe or Ar^+ laser and was recorded by means of a 0.5 m monochromator and a nitrogen-cooled CCD camera. A laser beam modulator was used to avoid overheating of the sample at high excitation power.

3. PHOTON MODES IN A ONE-DIMENSIONAL MICROCAVITY

The quantization of light in a one-dimensional cavity causes a photon mode to split into states that have dispersion only in the direction of the axis of the wire (x axis). The photon mode energy Mm can be described by the expression[9]

$$E_m(k_x)=\sqrt{E_v^2+\frac{\hbar^2c^2}{\varepsilon_{\text{eff}}}\left[k_x^2+\frac{(m+1)^2\pi^2}{L_y^2}\right]}. \tag{3}$$

Here $m=0,1,\ldots$, L_y is the width of the wire, and ε_{eff} is the effective dielectric constant.[9] Calculations also show that the intensity maxima of the excited states of the one-dimensional photon system are detected at nonzero angles Θ because of the different symmetries of the field distribution in the cavity interior for different mode orders. In addition, it follows from these calculations that for $\Theta=0°$ mode $M0$ has an intensity maximum, and mode $M2$ has a local maximum.[15]

It has been shown[8] that for high excitation densities the mixing of exciton and photon modes in the cavity tends to be suppressed and does not influence the type of dispersion of the photon modes when the carrier density in the quantum well corresponds to the onset of an electron–hole plasma. Consequently, the energies and dispersion of the cavity photon modes for $k=0$ can be measured by analyzing the emission spectra recorded in the presence of a sufficiently high excitation density.

Figure 2 shows luminescence spectra recorded for wires of width 4.2 μm at $\Phi=0°$ and at various detection angles Θ for a high excitation density ($P=1600$ W/cm^2). For this excitation density the carrier density created in the quantum well is greater than 10^{11} cm^{-2}. At this density screening plays a significant part, while exciton effects and, accordingly, exciton–photon interaction tends to be suppressed. In Fig. 2 photon modes with different quantum numbers are seen in all the spectra as strong, narrow lines. The energy positions of the photon modes are indicated by vertical arrows. The wide line with a maximum in the vicinity of 1.413 eV, labeled $e-h$, corresponds to a recombination peak of a dense exciton system. As expected, the quantization of the system in a direction perpendicular to the axis of the wire renders the energies of the photon modes independent of the angle Θ. At $\Theta=0°$ only odd mode orders ($M1, M3, \ldots$) are strictly forbidden. Consequently, in addition to the ground photon state $M0$, a feature corresponding to mode $M2$ is also visible in the luminescence spectrum at $\Theta=0°$. As Θ increases, states corresponding to mode orders $m>0$ successively appear and disappear (as the angle is further increased) in the luminescence spectra. It is evident from the figure that optical modes up to $M4$ are clearly visible in the spectra for $\Theta\leqslant 23°$.

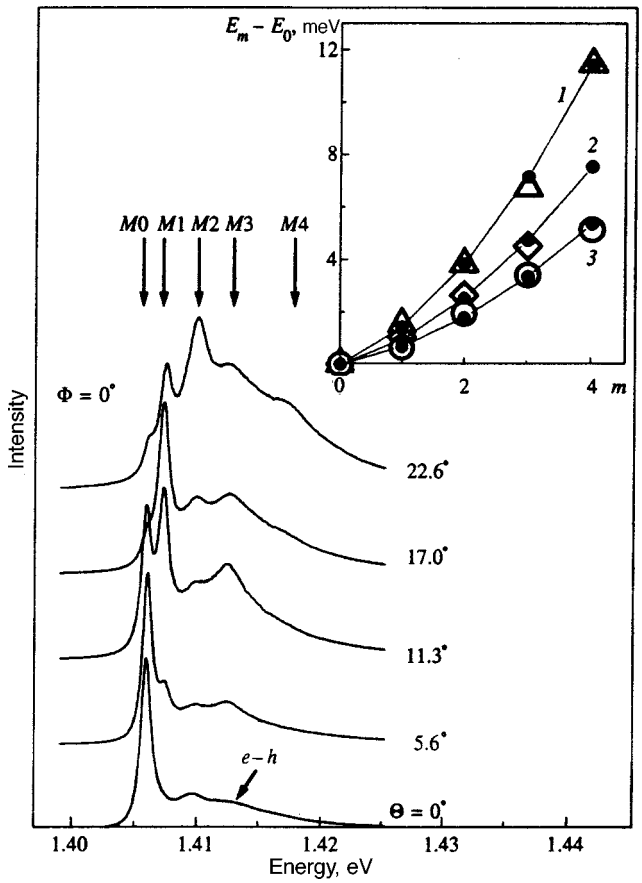

FIG. 2. Photoluminescence spectra for wires with $L_y=4.2$ μm, recorded for a high excitation density at $\Phi=0°$ and various angles Θ. The vertical arrows indicate the positions of modes with different quantum numbers m. The electron–hole recombination peak is labeled e-h. Inset: splitting $\Delta E_{m0}=E_m-E_0$ as a function of the mode order m for wires of various widths. The triangles correspond to splitting for a wire of width 4.2 μm, the diamonds correspond to 5.2 μm, and the circles correspond to 6.2 μm. Dark symbols, which are joined by smooth curves to aid visualization, represent the results of calculations according to Eq. (3) at $k_x=0$. Effective wire width d_{eff} used for curve fitting: *1)* 5.2 μm; *2)* 6.4 μm; *3)* 7.6 μm.

The inset to Fig. 2 shows the splittings $\Delta E_{m0}=E_m-E_0$ between optical modes, measured for photon wires of various widths. The triangles in the inset correspond to split-

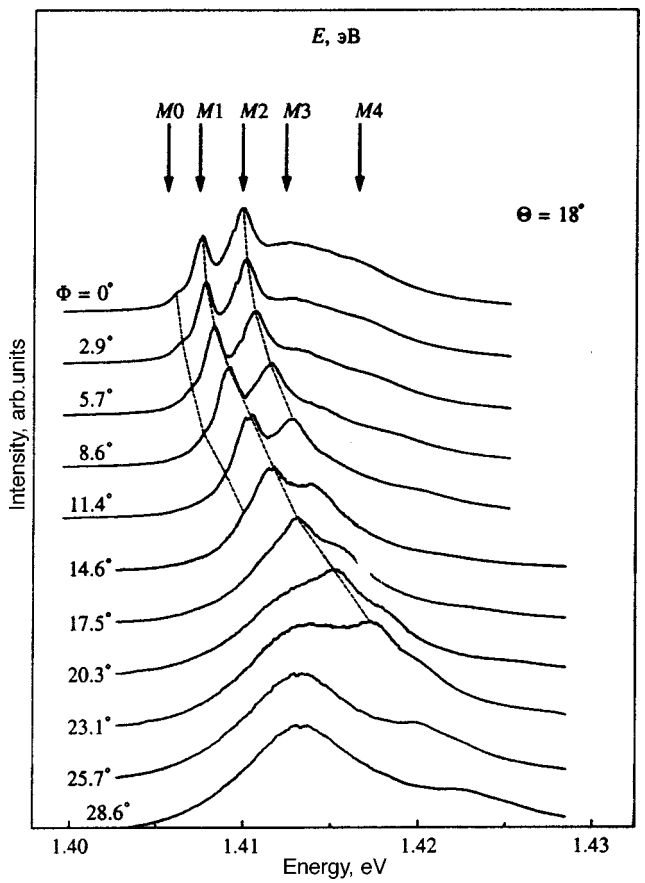

FIG. 3. Photoluminescence spectra for wires with $L_y = 4.2\ \mu$m, recorded for a high excitation density at $\Theta = 18°$ and various angles Φ. The vertical arrows indicate the mode positions for $\Phi = 0°$. The dashed curves represent the positions of modes $M0$, $M1$, and $M2$, as functions of the angle Φ.

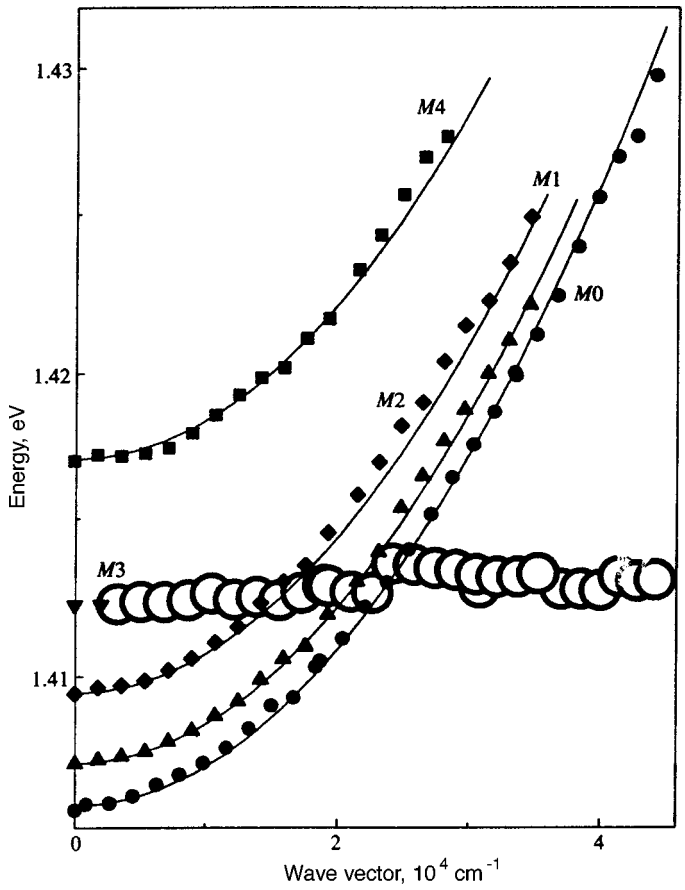

FIG. 4. Dispersion of photon modes of a wire with $L_y = 4.2\ \mu$m for a high excitation density. Dark symbols represent the measured photon mode positions. The light circles represent the position of the electron–hole recombination peak. The solid curves are calculated from Eq. (3).

tings for a wire of width 4.2 μm, the diamonds correspond to 5.2 μm, and the circles correspond to 6.2 μm. It is evident from the figure that the splitting of the photon modes decreases as the width of the wire increases. We use Eq. (3) with $k_x = 0$ to describe the splitting between optical modes. The role of the fitting parameter in this case is taken by the effective width of the wire d_{eff}. In the inset to Fig. 2 the calculations of ΔE_{m0} according to (3) are represented by dark symbols joined by solid lines. The value of d_{eff} used in the calculations is given for each curve. It is evident from the figure that the behavior of the experimentally measured dependence of the splitting on the width of the wire is qualitatively consistent with the calculations according to (3). However, the values of d_{eff} are found to be greater than the true widths of the photon wires. We identify this discrepancy primarily with the fact that the Bragg mirrors left unetched to the very end in the regions between the photon wires permit light to penetrate beyond the limits of the wires, and this process, in turn, causes the effective width of the cavity to increase.

In photon wires light is not quantized in the direction of the wire axis. The energy of the photon modes is therefore a monotonic function of the quasimomentum k_x. Figure 3 shows the luminescence spectra in wires of width 4.2 μm, recorded at $\Theta = 18°$ for various angles Φ. The spectra were recorded at a high excitation density $P = 1600\ \text{W/cm}^2$, such that exciton–photon interaction is suppressed. At $\Phi = 0°$ two strong, narrow lines corresponding to modes $M1$ and $M2$ are visible in the spectrum against the background of the broad electron–hole recombination line. All other modes are seen as weak features in the spectrum. The mode energies at $\Phi = 0°$ are indicated by vertical arrows. As Φ increases, all the photon modes shift toward the violet end. The dashed curves in Fig. 3 represent the positions of modes $M0$, $M1$, and $M2$ as the angle Φ is varied. Particularly noticeable in the figure is the absence of anomalies in the behavior of the energies of the optical modes when they intersect the $e-h$ recombination peak. This behavior is characteristic of the suppression of interaction between light and carriers in quantum wells. The dispersion of photon modes $M0-M4$ is shown in Fig. 4. It is extracted from the luminescence spectra measured at $\Theta = 0°$ and 18° over a wide range of angles Φ. The large open circles indicate the energy of the $e-h$ recombination peak. As expected, its energy does not depend on the quasimomentum. Dark symbols represent the measured dispersion of the photon modes. The energy of mode $M3$ is shown only for $k_x < 10^3\ \text{cm}^{-1}$, because for large angles Φ this line has poor resolution against the background of the rather strong, broad $e-h$ recombination line. The solid curves represent the photon mode dispersion calculated

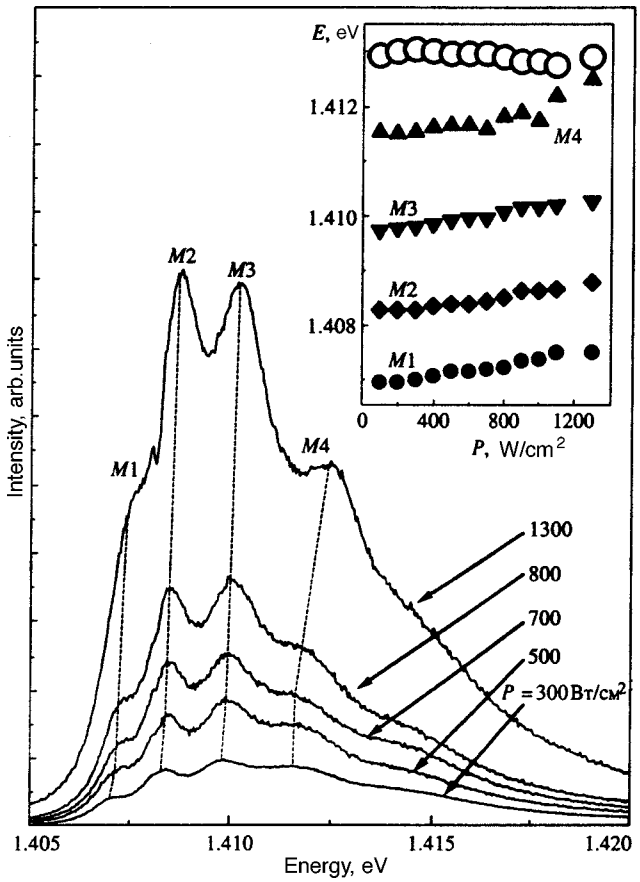

FIG. 5. Luminescence spectra of wires with $L_y = 6.2\ \mu$m, recorded for various excitation densities at $\Theta = 10°$ and $\Phi = 0°$. The dashed lines indicate the variation of the photon mode positions. Inset: energies of photon modes and the exciton peak (light circles) as functions of the excitation density.

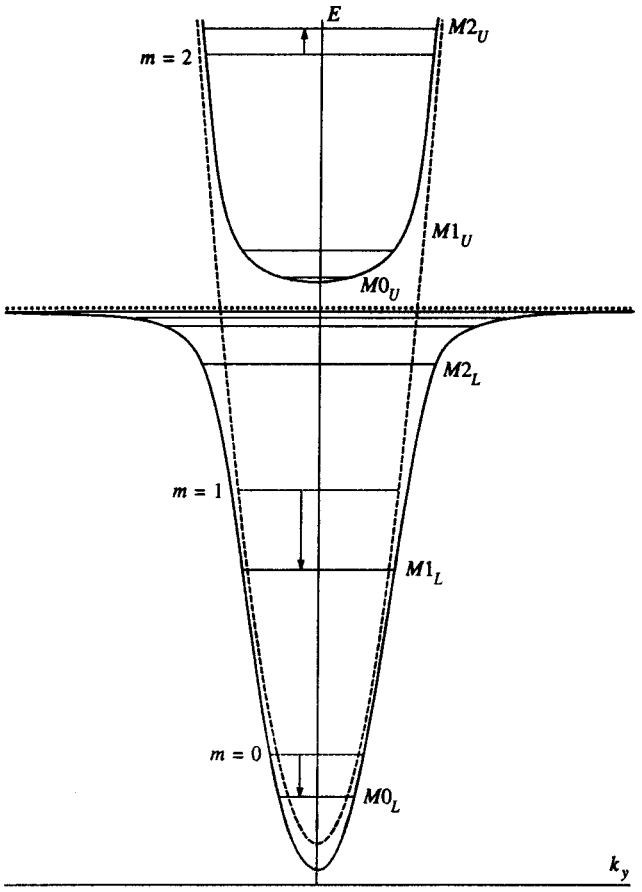

FIG. 6. Diagram of the quantization of polaritons in a photon wire (see the text).

from Eq. (3). The only fitting parameter used in the calculations is the dielectric permittivity. It is evident form the figure that the experimental curves are well described for $\varepsilon_{\text{eff}} = 10.9$. As expected, this value differs markedly from ε in the GaAs λ layer and is close to the average of the dielectric constants of AlAs (8.8) and GaAs (12.5). Indeed, owing to the low reflection coefficient at the boundary of the active layer, light penetrates the Bragg mirrors to a depth of several GaAs/AlAs layers.

4. VARIATION OF EXCITON–PHOTON INTERACTION AS THE DENSITY OF EXCITONS IS VARIED

Effects associated with exciton–photon interaction becomes more significant as the excitation density decreases. To illustrate the transition from the weak to the strong coupling regime, we refer to Fig. 5. It shows the luminescence spectra measured at $\Theta = 10°$ and $\Phi = 0°$ in wires of width 6.2 μm for various excitation densities. The angle $\Theta = 10°$ is chosen to achieve good resolution in the spectra simultaneously for all four optical modes $M1 - M4$, which are situated below the exciton energy of 1.413 eV. In Fig. 5 all four modes are seen as strong, narrow lines against the background of the weaker, broad exciton line. As the excitation density is decreased, all modes shift toward lower energies. The shifting of the peaks is reflected by the dashed curves in the figure. The inset to the figure shows the dependence of the mode energies on the excitation density.[2)] The repulsion of the cavity modes from the exciton term is evidence of the repulsion of exciton and photon states, i.e., strong coupling between excitons and photons in the microcavity.[3)] The most pronounced variation of the energy with decreasing excitation density is observed for $M4$, the mode closest to the excitation level. As the distance between the photon mode and the exciton peak increases, the degree of repulsion subsides considerably (from 1 meV for $M4$ to 0.5 meV for $M1$).

The experimental pattern of states in photon wires can be described in terms of quantized microcavity polaritons. The formation of polariton states in photon wires is illustrated in Fig. 6. We first consider a planar microcavity. If excitons and photons do not interact, they have a parabolic dispersion law. The dashed curve in Fig. 6 shows the photon dispersion, and the dashed curve represents the exciton energy level. The exciton energy depends weakly on k by virtue of the large effective exciton mass m_{ex} relative to the photon mass M_p: $m_{\text{ex}} \sim 10^5 M_p$ (Refs. 12 and 13). Exciton–photon interaction results in the production of a cavity polariton with a dispersion law described by the expression for two intersecting, interacting levels:[16]

$$E_{U,L}^2(k) = \frac{\Sigma^2 \pm \sqrt{\Sigma^4 - 4E_{\text{ex}}^2(k)E_p^2(k)}}{2}. \qquad (4)$$

Here $E_{U,L}(k)$ denotes the energies of the upper and lower branches of the polariton dispersion curve (with the plus or minus sign, respectively), $E_p(k)$ and $E_{ex}(k)$ are the photon and exciton mode energies without interaction, $\Sigma^2 = E_{ex}^2(k) + E_p^2(k) + \Omega^2$, and Ω is the splitting between the upper and lower branches at resonance. The dispersion of polaritons is represented by solid curves in Fig. 6. Far from resonance the lower branch of the polariton dispersion curve exhibits predominantly photon behavior for small k and predominantly exciton behavior for large k. The opposite pattern is encountered for the upper branch. For large k its dispersion is close to the dispersion of the photon mode without interaction. In a planar microcavity the dispersion of excitons, photons, and polaritons is independent of direction in the xy plane.

The confinement of light in the y direction by the semiconductor–vacuum boundaries (L_Y is of the order of a few μm) leads to quantization of the photon cavity mode. The lower quantized states ($m=0; 1, 2$) without exciton–photon interaction are represented by dashed horizontal lines in Fig. 6. The splitting of the exciton levels due to size quantization is very small and is therefore not shown in Fig. 6.

Polariton states corresponding to pair-interacting exciton and photon states with identical numbers m are formed in the strong exciton–photon coupling regime. In this case each m corresponds to two polariton states, one for the lower branch and one for the upper branch of the dispersion curve (solid horizontal lines labeled $Mm_{U,L}$) in Fig. 6.

It is evident from Fig. 6, consistent with experiment, that photon states situated below the exciton level ($m=0, 1$ in the figure) are shifted downward on the energy scale (indicated by arrows in the figure). The shift decreases when the number m increases. Figure 6 shows that the more pronounced decrease of the energy level with $m=2$ is readily explained within the context of polariton quantization by an increase in the effective mass of the polaritons as the exciton level is approached. It is also evident from Fig. 6 that lateral quantization of the lower polariton branch in photon wires should produce a large number of levels in the vicinity of exciton resonance. These levels, in contrast with those with $m=0, 1$, are excitonlike; the smallness of the splitting makes them spectrally unresolved.

We now consider the quantization of the upper polariton branch. It is evident from the figure that the energy of the polariton state $M2_U$, which is situated far from the exciton level, differs only slightly from the energy of the photon level with $m=2$ in the noninteracting system. The latter result is attributable to the large contribution of the photon component to this polariton state. States near the bottom of the upper polariton branch ($M0_U$, $M1_U$) are shifted slightly relative to the plateau as a result of the large polariton mass for small k. This behavior is explained by the large contribution of the exciton component to the polariton state. The smallness of the splitting makes it impossible to resolve states $M0_U$ and $M1_U$ in the luminescence spectrum for $k=0$, but it will be shown in the next section that they can be distinguished by analyzing the polariton behavior for large k_x.

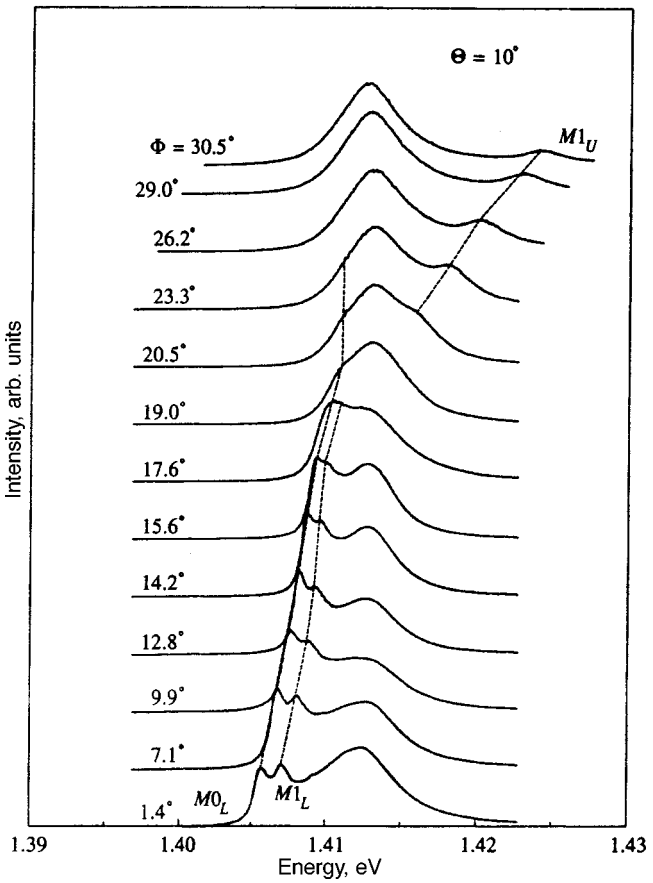

FIG. 7. Photoluminescence spectra for wires with $L_y = 4.2$ μm, recorded at $\Theta = 10°$ and various angles Φ. The dashed lines indicate the positions of modes $M0$ and $M1$. The subscripts L and U signify the lower and upper branches of polariton dispersion.

5. DISPERSION OF POLARITONS IN PHOTON WIRES IN THE PRESENCE OF STRONG EXCITON–PHOTON COUPLING

To investigate the dispersion of polaritons in photon wires we have conducted a series of luminescence measurements at a low excitation density $P = 300$ W/cm^2 in wires of width 4.2 μm at various angles Φ. Since different mode orders are dominant in the spectra for different angles Θ, we carried out the series of measurements at three angles $\Theta = 0°, 10°, 15°$. One series is represented in Fig. 7, which shows luminescence spectra recorded at $\Theta = 10°$ and several angles Φ. As in the case of high excitation density, the photon modes shift toward higher energies as Φ increases. At low excitation densities, however, the photon modes go through an anticrossing with the exciton peak. This feature differs qualitatively from the high-density case (Fig. 4), which does not have any anomalies in the region of intersection of the photon modes and the exciton level. Moreover, it is evident from Fig. 7 that the branches of modes $M0$ and $M1$ situated below the exciton energy tend to the exciton level. For large Φ a new line appears above the exciton peak; it can be ascribed to the upper branch of the polariton mode $M1$. The exciton peak (1.413 eV) scarcely alters its position as Φ increases, but this is not too surprising in view of the major contribution of excitons in the regions between wires

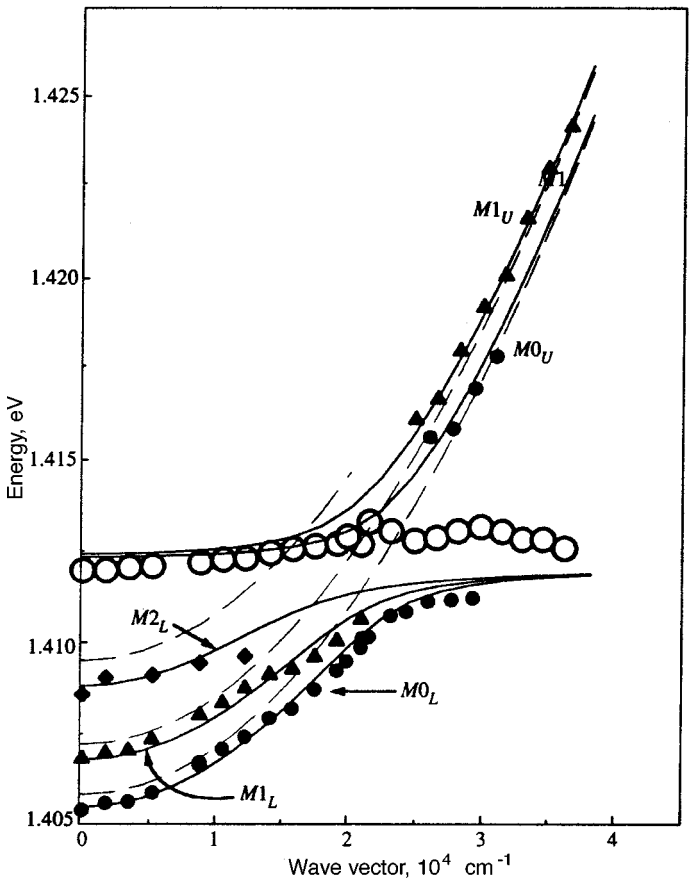

FIG. 8. Dispersion of polaritons in wires with $L_y = 4.2\ \mu$m, obtained from the luminescence spectra for a low excitation density. The subscripts L and U signify the lower and upper branches of polariton dispersion. Dark symbols represent the polariton mode energies. The solid curves are calculated from Eq. (4). The dashed curves represent the photon mode dispersion determined from measurements at a high excitation density.

to the luminescence signal. The formation of the polariton modes $M0_U$, $M0_L$, and $M2_L$ is also observed in the other series of measurements at $\Theta = 0°$ and 15°.

The observed anticrossing of exciton and photon states indicates the formation of microcavity polaritons. The polariton mode dispersion determined from an analysis of the energies of the spectral peaks in Fig. 7 is shown in Fig. 8. The corresponding mode energies measured at high excitation density, when the interaction of light with excitons is suppressed, are also shown (dashed curves) for comparison. A comparison of the mode energies at low and high excitation densities shows that exciton–photon mixing has the effect of shifting cavity modes with $E < E_{\rm ex}$ toward lower energies and shifting modes with $E > E_{\rm ex}$ toward higher energies. This behavior is natural when each mode interacts with the exciton state corresponding to it in terms of spatial symmetry. Consequently, to describe the experimental results for each pair $E_{mU,L}(k)$ we have used Eq. (4), replacing $E_p(k)$, $E_{\rm ex}(k)$, and Ω by their counterparts for the given mth mode: $E_m(k_x)$, $E_{{\rm ex},m}(k_x)$, and Ω_m. We obtain the photon mode energies $E_m(k_x)$ from the spectra at high energy density. We then disregard the exciton quantization energies, i.e., we assume that $E_{{\rm ex},m} = E_{\rm ex} = 1.413$ eV for all $m = 0, 1, 2$. For low mode orders m ($m = 0-2$) the splitting Ω_m in wide photon wires should not depend too strongly on the photon mode order, i.e., we can set $\Omega_0 = \Omega_1 = \Omega_2 = \Omega$. We are thus left with only the one fitting parameter Ω to describe several polariton branches in Eq. (4). Better agreement between the calculated curves and the experimental points is attained for $\Omega = 3$ meV. The corresponding fitted curves are shown as solid curves in Fig. 8. The magnitude of the interaction in the case of photon wires is found to be slightly lower than in the original planar microcavity. It has been determined from similar measurements of the luminescence spectra in a planar microcavity that $\Omega = 4$ meV. The reduction of Ω in wires is most likely attributable to lowering of the Q of the structure in etching of the planar microcavity.

The better agreement of the experimental polariton dispersion curves with those calculated on the assumption of interaction between exciton and photon states of like spatial symmetry indicates that exciton–photon interaction retains essentially the same character as in a microcavity with an ideal quantum well, despite the fact that potential fluctuations in a real quantum well lead to the localization of excitons and induce breaking of the symmetry of their states. In the structures we studied the localization range for excitons in a quantum well does not exceed $0.1-0.2\ \mu$m, whereas L_y is an order of magnitude higher: $4-6\ \mu$m. To explain the weak influence of a random potential for the indicated ratio of the characteristic dimensions, it must be assumed that in a photon wire an optical mode induces a collective coherent state of excitons of like symmetry. This state is constructed from a large number of localized and free exciton states. The interaction of light with such exciton modes results in the formation of coherent polariton states. This kind of interaction mechanism is possible by virtue of the fact that the light wavelength is much greater than the exciton radius, so that a great many exciton states are excited coherently at the same

time. It is obvious that the coherent state has a finite lifetime, owing to the scattering of excitons by the random potential and phonons and because of exciton–exciton collisions. The polariton dispersion observed by us in photon wires indicates that the dephasing time of polariton states is at least greater than the Rabi oscillation period (~ 0.4 ps). This conclusion is consistent with a measurement of the dephasing time of exciton states by the four-wave mixing method,[17], where a dephasing time of 12 ps has been determined for a low-density exciton system. The dephasing time decreases as the exciton density increases, accounting for our experimentally observed suppression of exciton–photon interaction at high excitation densities.

6. CONCLUSION

In 2D optically confined semiconductor microcavities we have investigated the structure of the states of the phonon system and the dispersion of cavity polaritons in the presence of strong exciton–photon coupling. Light is spatially confined in one dimension by a system of Bragg mirrors and in the other dimension by the semiconductor–vacuum boundary. The lateral confinement of the cavity leads to quantization of the vertical cavity mode. The additional photon states have a maximum in the luminescence spectrum at various values of the angle Θ. At low excitation densities exciton–photon coupling is fairly strong and results in the formation of cavity polaritons. The experimental polariton dispersion curves are in good agreement those calculated on the assumption that interaction takes place only between photon and exciton states of like spatial symmetry. Fluctuations of the random potential in the quantum wells break the spatial symmetry of the excitons but have little influence on the dispersion of polaritons. Consequently, polaritons are produced in microcavities through collective coherent exciton states induced by the optical electromagnetic field. The dephasing time of such states is much longer than the Rabi oscillation period. As the excitation density increases, the screening of exciton states and shortening of the dephasing time have the effect of suppressing the strong exciton–photon coupling regime.

We are grateful to M. Bayer, R. Sirus, S. Tikhodeev, N. Gippius, and A. Dzyubenko for profitable discussions and also to the International Association for the Promotion of Cooperation with Scientists from the Independent States of the Former Soviet Union (INTAS), the Russian Fund for Fundamental Research, and the ''Physics of Solid State Nanostructures'' Program for financial support.

*)E-mail: tarta@issp.ac.ru
†)E-mail: kulakovs@issp.ac.ru

1)In reality, the boundary conditions on the side walls of the cavity differ for excitons and light, so that orthogonality is lost between the wave functions of exciton and photon modes having different orders. As a consequence, the interaction of even (odd) exciton modes with all even (odd) photon modes is allowed. However, interaction between modes of like orders must be dominant.

2)The photon mode energies and the position of the exciton recombination peak, which is not resolved as a distinct band in the spectra of Fig. 5, have been determined by the application of a program for deconvolution of the spectral positions of peaks by fitting of the line shape.

3)The position of the exciton peak does not change, because the main contribution to the exciton luminescence signal is from excitons that recombine in the region between the wires. In fact, the dimensions of these regions exceeds the area of the photon wires, and the essentially completely etched Bragg mirrors are two or three orders of magnitude more transparent both to the exciting laser beam and to exciton radiation. It is obvious that exciton–photon interaction is not intensified by the cavity in the regions between wires, and the position of the exciton peak therefore remains unchanged in transition from the weak to the strong exciton–photon coupling regime.

[1] V. Savona, L. C. Andreane, P. Schwendimann, and A. Quattropani, Solid State Commun. **93**, 733 (1995).
[2] V. Savona and F. Tassone, Solid State Commun. **95**, 733 (1995).
[3] V. Savona, Z. Hradil, A. Quattropani, and P. Schwendimann, Phys. Rev. B **49**, 8774 (1994).
[4] S. Jorda, Phys. Rev. B **50**, 18690 (1994).
[5] S. Jorda, Phys. Rev. B **51**, 10185 (1995).
[6] R. Houdré, R. P. Stanley, U. Oesterle, M. Ilegems, and C. Weisbuch, Phys. Rev. B **49**, 16761 (1994).
[7] R. Houdré, C. Weisbuch, R. P. Stanley *et al.*, Phys. Rev. Lett. **73**, 2043 (1994).
[8] R. Houdré, J. L. Gibernon, P. Pelladini *et al.*, Phys. Rev. B **52**, 7810 (1995).
[9] J. P. Reithmaier, M. Rohner, H. Zull *et al.*, Phys. Rev. Lett. **78**, 378 (1997).
[10] B. Ohnesorge, M. Bayer, A. Forchel *et al.*, Phys. Rev. B **56**, 4367 (1997).
[11] R. Houdré, R. P. Stanley, and M. Ilegems, Phys. Rev. A **53**, 2711 (1996).
[12] D. M. Whittaker, P. Kinsler, T. A. Fisher *et al.*, Phys. Rev. Lett. **77**, 4792 (1996).
[13] V. Savona, C. Piermarochhi, A. Quattropani *et al.*, Phys. Rev. Lett. **78**, 4470 (1997).
[14] T. B. Borzenko, Y. I. Koval, L. V. Kulik, and V. V. Larionov, Appl. Phys. Lett. **70**, 2297 (1997).
[15] N. A. Gippius, S. G. Tikhodeev, and A. Forchel, in *Digests of the Third All-Russian Conference on the Physics of Semiconductors* [in Russian] (FIAN, Moscow, 1997).
[16] B. Sermage, S. Long, I. Abram *et al.*, Phys. Rev. B **53**, 16516 (1996).
[17] A. Honold, L. Schultheis, J. Kuhl, and C. W. Tu, Phys. Rev. B **40**, 6442 (1989).

Translated by James S. Wood

Short-range order in spin density wave antiferromagnets with chemical dimerization

M. Avignon

Russian Research Centre ''Kurchatov Institute'' 123182 Moscow, Russia

V. Men'shov*) and V. Tugushev

Laboratoire d'Études des Proprietés Electroniques des Solides, CNRS, 38042, BP166, Grenoble, France
(Submitted 5 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1346–1364 (October 1998)

The spin density wave model in a quasi-one-dimensional itinerant antiferromagnet with staggered potential at finite temperature is studied. Only short-range ordering exists in this system above the Néel temperature. The local-band theory of spin fluctuations is developed to calculate the spin density wave amplitude and the effective exchange integral. The one-electron spectrum and magnon spectrum are obtained in the short-range ordering regime. © *1998 American Institute of Physics.* [S1063-7761(98)01310-9]

1. INTRODUCTION

In the modern theory of itinerant magnetism it is assumed that the destruction of long-range magnetic order with increasing temperature is mainly due to spin density fluctuations,[1] which determine the phase transition temperature, changes in magnetic structure parameters below the transition point and short-range order. One of the most physically illustrative methods to account for spin fluctuations, known as the local band theory,[2,3] was developed for itinerant ferromagnets. In this theory, the magnetic moment of the unit cell is formed at a temperature considerably higher than the temperature for the onset of long-range order. Amplitude excitations of the spin density have characteristic frequencies of the order of the Stoner exchange splitting, and do not determine the temperature of the phase transition. Accordingly, transverse long-wavelength fluctuations of the spin density with low (of order of the spin-wave) frequencies play the dominant role in the production of ferromagnetic order. Thus, both the exchange (Stoner) splitting of electronic subbands and short-range ferromagnetic order (but not long-range) are conserved above the transition point.

In Ref. 4 local band theory was applied to investigate the thermodynamics of itinerant antiferromagnets with spin density waves (SDW). In the vicinity of the Néel point T_N, the thermally excited state of an antiferromagnet is formed by the electronic Bloch states with energies within $\sim T_N$ of the Fermi level ε_F, where $T_N \ll \Delta$ (Δ is the SDW amplitude). This assumption strongly reduces the configuration field of spin fluctuations, i.e., the wave packets are formed by Bloch functions with wave vectors $\mathbf{K}$ close to $\mathbf{G}/2$ ($\mathbf{G}$ being the reciprocal lattice vectors). As a result, long-wavelength transverse fluctuations of the magnetization with wave vectors $\mathbf{q}$ close to the vector of antiferromagnetic structure $\mathbf{Q}=\mathbf{G}/2$ (that is, the transverse spin density waves) will be the most relevant. Nevertheless, short-range order band structure similar (but not identical) to the long-range order structure is preserved. The quasimomentum and spin of a quasiparticle are ''almost good'' quantum numbers, at least as long as $|\mathbf{K}-\mathbf{G}/2|l>1$, where l is the characteristic correlation length of SDW fluctuations ($l\to\infty$, when $T\to T_N$).

It is well known that band structure properties play an important role in the formation and properties of SDW antiferromagnets.[5] In nesting-type models[6] the first well-known mechanism what leads to the suppression of SDW in the ground state is doping. Indeed, in this case, when particles are added to (or removed from) the system, the perfect nesting of the Fermi surface at the wave vector $\mathbf{Q}=\mathbf{G}/2$, of the antiferromagnetic structure, is spoiled. Another band structure mechanism responsible for the destruction of the SDW state is associated with the deformation of the electronic spectrum as a consequence of the onset of a charge density wave. It may be shown that in the absence of doping, spin and charge density waves can not coexist.

However, a SDW state can exist in an external staggered potential introduced into the system as so-called ''chemical dimerization''.[7] This system is of great interest, as it describes some real physical situations. For example, this is the case for reconstructed Si (111) and C (111) surfaces in which the crystal field of the bulk induces chemical dimerization at the surface, and SDW ordering appears in the surface band of itinerant electrons along the so-called π-chain.[8]

The ground-state properties of the SDW system with chemical dimerization are investigated in Ref. 7. Below we consider the thermodynamics of such a system in the short-range ordering (SRO) regime. We examine the influence of a staggered potential on the parameters of effective exchange, correlation length, electronic spectrum, and collective excitations of SDW phase with short-range order.

In Sec. 2 we introduce the model and briefly recall some results of Ref. 7 about the phase diagram of the system in the ground state. In Sec. 3 we introduce the techniques of the ''local-band'' spin-fluctuation theory specialized to the problem of itinerant antiferromagnets.

In Sec. 4 we calculate the effective exchange integral and the Néel temperature for our model. In Sec. 5 we develop the single-particle Green's function approach to calcu-

late the electronic spectrum and renormalization of the SDW amplitude in the SRO state.

In Sec. 6 we develop the two-particle Green's function approach to calculate the dynamical transverse susceptibility of the system.

In Sec. 7 we calculate the spectrum of low-frequency excitations (paramagnons) in the SRO state. Concluding remarks are found in Sec. 8.

2. SELF-CONSISTENCY EQUATION FOR THE SDW AMPLITUDE IN THE GROUND STATE IN THE HARTREE–FOCK APPROXIMATION

We consider the single-band model with Hamiltonian

$$H=H_0+H_{\rm int}, \tag{1}$$

$$H_0=\sum_{ij\sigma} t_{ij}C^+_{i\sigma}C_{j\sigma}+I\sum_{j\sigma}\exp(i\mathbf{QR}_j)n_{j\sigma}, \tag{2}$$

$$H_{\rm int}=U\sum_j\left(n_{j\uparrow}-\frac{\overline{n_j}}{2}\right)\left(n_{j\downarrow}-\frac{\overline{n_j}}{2}\right), \tag{3}$$

where $C^+_{i\sigma}$, $C_{i\sigma}$ are fermionic operators, t_{ij} is the nearest-neighbor hopping parameter, I is the amplitude of a staggered local potential produced by the crystal field, $\mathbf{Q}=\mathbf{G}/2$, and $\mathbf{R}$, is a lattice site. This form of the interaction term $H_{\rm int}$ is discussed in Ref. 7. Introducing the unit vector $\mathbf{e}_i$, which specifies the local orientation of the quantization axis at the ith site, we rewrite the term $H_{\rm int}$ in the equivalent form:

$$H_{\rm int}=U\sum_j\left[\left(\frac{n_j-\overline{n_j}}{2}\right)^2-(\mathbf{e}_j\mathbf{S}_j)^2\right], \tag{4}$$

where $n_j=n_{j\uparrow}+n_{j\downarrow}$ and $S_{jz}=(1/2)(n_{j\uparrow}-n_{j\downarrow})$ are the charge and spin densities, respectively. The representation (4) enables us, by means of the Hubbard-Stratonovich transformation, formally to reduce the initial many-particle problem (1) to a single-particle problem that involves the motion of an electron in the arbitrary scalar $x_j(\tau)$ and vector $\mathbf{y}_j(\tau)=\mathbf{e}_j(\tau)y_i(\tau)$ fields conjugate to the charge and spin densities, respectively.[1] In the ''static'' approximation, i.e., neglecting the dependence of x_i and y_i on the time τ, the partition function corresponding to the Hamiltonian (1) has the form

$$Z=Z_0\int Dx\ d\mathbf{y}\ \exp\left\{-\pi\sum_j(x_j^2+y_j^2)+\mathrm{Tr}\ \ln(1-gV)\right\}, \tag{5}$$

where

$$g^{\sigma\sigma'}_{ij}(\omega)=\frac{1}{N}\sum_{\mathbf{k},\mathbf{p}}\exp(i(\mathbf{kR}_i-\mathbf{pR}_j))\times\frac{(i\omega_n+\varepsilon_{\mathbf{k}})\delta_{\mathbf{kp}}+I\delta_{\mathbf{k},\mathbf{p}+\mathbf{Q}}}{(i\omega+\mu)^2-E^2_{\mathbf{k}}}\delta_{\sigma\sigma'} \tag{6}$$

is the Green's function of the noninteracting electrons with dispersion $E_{\mathbf{k}}=\pm(\varepsilon^2_{\mathbf{k}}+I^2)^{1/2}$, $\mathbf{k}$ is the quasimomentum, $\omega_n=\pi T(2n+1)$ is the frequency, μ is the chemical potential,

$$V_{j\sigma\sigma'}=(\pi TU)^{1/2}(ix_j\delta_{\sigma\sigma'}+\boldsymbol{\sigma}_{\sigma\sigma'}\mathbf{y}_j) \tag{7}$$

is a random potential, and Z_0 is the partition function of the noninteracting electrons. In expression (5) and in what follows, the symbol Tr denotes the sum over spin index j, and the frequency variable n.

We neglect fluctuations of the charge density x_j. In the saddle-point approximation for the variable x_j, one can obtain

$$\langle x_j\rangle\sim\langle n_j-\overline{n_j}\rangle=0.$$

Thus, in the Hamiltonian (1), there is no renormalization of the chemical potential μ and the staggered potential I (see Ref. 7).

We now turn to the problem of SDW antiferromagnetism in the model (1). We propose that the ground state of the system has a Néel structure. For $I=0$ we assume that the electronic spectrum has the ''nesting'' property:

$$\varepsilon_{\mathbf{k}}=-\varepsilon_{\mathbf{k}+\mathbf{Q}}. \tag{8}$$

We distinguish two alternating magnetic sublattices, for one of these we take the local quantization axis $\mathbf{e}_j$ to coincide in direction with the local magnetization of the site $\mathbf{S}_j$, while for the other we assume that the vectors $\mathbf{e}_j$ and $\mathbf{S}_j$, are antiparallel.

Thus,

$$\mathbf{e}_j\mathbf{S}_j=S_{jz}\exp(i\mathbf{QR}_j).$$

The vector field $\mathbf{y}_j$ is normalized in energy units:

$$\boldsymbol{\Delta}_j=(\pi UT)^{1/2}\mathbf{y}_j\exp(i\mathbf{QR}_j)=U\mathbf{S}_j\exp(i\mathbf{QR}_j). \tag{9}$$

In the Hartree–Fock approximation we can write the standard expression for the one-particle Green's function in the lineary polarized SWD structure $\boldsymbol{\Delta}_i=\Delta\mathbf{e}_z$:

$$^{0\sigma\sigma'}G_{\mathbf{kp}}(\omega_n)=\frac{(i\omega_n+\varepsilon_{\mathbf{k}}+\mu)\delta_{\mathbf{k},\mathbf{p}}+(\Delta\sigma^z_{\sigma\sigma'}+I\delta_{\sigma\sigma'})\delta_{\mathbf{k},\mathbf{p}+\mathbf{Q}}}{(i\omega_n+\mu)^2-E^2_{\mathbf{k}\sigma}}. \tag{10}$$

The spectrum of one-particle excitations has the form

$$\omega^\sigma=\pm E_{\mathbf{k}\sigma},\quad E_{\mathbf{k}\sigma}=[\varepsilon^2_{\mathbf{k}}+(I+\sigma\Delta)^2]^{1/2}. \tag{11}$$

It consists of four branches which-differ in their band indices ($\pm$ sign on the radical in (11)) and their spin projection σ along the vector $\mathbf{e}_z$.

The self-consistency equation for the spin density amplitude is found as a saddle-point condition for the field $\mathbf{y}_i$, and has the standard form:

$$\Delta=\frac{U}{2}T\sum_n\frac{1}{N}\sum_{\mathbf{k}}\sum_{\sigma\sigma'}\sigma^z_{\sigma\sigma'}G^{\sigma\sigma'}_{\mathbf{k},\mathbf{k}-\mathbf{Q}}(\omega_n). \tag{12}$$

We analyze Eq. (12) at zero temperature and in the absence of doping ($\mu=0$). We restrict attention to the case of a one-dimensional chain, for which $\varepsilon_{\mathbf{k}}=-2t\cos(ka)$ (a is the chain period). We can then rewrite (12):

$$\frac{\Delta}{U}=\frac{1}{4\pi t}\sum_{\sigma=\pm1}\sigma k_\sigma K(k_\sigma)(I+\sigma\Delta), \tag{13}$$

where $K(k_\sigma)$ is the complete elleptic integral of the first kind with modulus

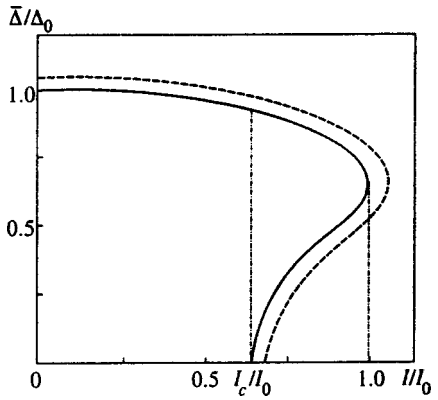

FIG. 1. $\bar{\Delta}(I)$ dependence for large (solid line) and short (dashed line) range order phases; limits of stability region $I_c<I<I_0$ (dash-dotted lines).

$$k_\sigma(\Delta)=\frac{2t}{\sqrt{(2t)^2+(I+\sigma\Delta)^2}}. \tag{14}$$

Integrating Eq. (13) with respect to Δ, one can obtain the thermodynamic potential $\Omega(\Delta)$ per site. The energy difference between the antiferromagnetic and paramagnetic phase (per lattice site) is

$$E(\Delta)=\Omega(\Delta)-\Omega(\Delta=0)$$
$$=\frac{\Delta^2}{U}-\frac{2t}{\pi}\sum_{\sigma=\pm1}\frac{E(k_\sigma)}{k_\sigma}+\frac{4t}{\pi}\frac{E(k_0)}{k_0}, \tag{15}$$

where $E(k_\sigma)$ is the complete elliptic integral of the second kind, and $k_0=k_\sigma(\Delta=0)$.

In the weak-coupling limit $(I^2,\Delta^2\ll(2t)^2)$, we can use the well-known expansions of the functions $K(k_\sigma)$ and $E(k_\sigma)$ near $k_\sigma=1$ (the so-called logarithmic approximations). For $I\ll\Delta_0$, where

$$\Delta_0=\Delta(I=0)=8t\ \exp[-2\pi t/U]$$

we obtain that

$$(\Delta(I)-\Delta_0)/\Delta_0=-(I/\Delta_0)^2.$$

It can easily be shown that for $I<I_c=\Delta_0/e$, there exists only one nontrivial solution of Eq. (9), which is a decreasing function of I. However, for $I>I_c$ a second solution appears, which increases with I. In the vicinity of I_c the dependence $\Delta(I)$ is given by

$$\Delta^2(I)=6I_c(I-I_c).$$

The numerical solution of Eq. (13) is obtained in Ref. 7 and plotted in Fig. 1 (solid line). It is clear that two nonzero solutions of Eq. (13) exist in the interval $I_c<I<I_0$, the first corresponding to a local minimum of the thermodynamic potential, and the second to a local maximum. The point $I=I_{eq}$ is the AFM-PM phase equilibrium point in the first-order phase transition, and I_0 and I_c are ''superheating'' and ''supercooling'' points for the AFM state.

Some analytic relations can be also obtained in the strong-coupling regime $(I,U\gg t)$, when $k_\sigma\to0$, as well numerical calculations for $\Delta(I)$ (see Ref. 7). Thus, mean field analysis shows that in a dimerized chain with a staggered potential, the transition to the AFM phase is of first order in both the weak-coupling and strong-coupling limits.

3. TRANSVERSE SDW FLUCTUATIONS AND THE EFFECTIVE HAMILTONIAN OF THE SHORT RANGE ORDER ANTIFERROMAGNET

According to current ideas in the theory of itinerant magnetism based on the spin-fluctuation approach, the temperature T_N^0 determined by Eq. (12) is not the actual temperature at which antiferromagnetic long-range order is established (the Néel point). The quantity T_N^0 gives some nominal upper bound on the temperature below which long wave length SDW fluctuations with amplitude weakly dependent on T are formed. Long-range order (i.e., the onset of mean magnetization of the sublattices) is described not by Eq. (12), but by other relations that can be derived in the context of the spin-fluctuation approach for $T<T_N^0$. Below we will be oriented to the simply and physically visual scheme,[2,3] in which the formation and breakdown of long-range order are associated mainly with transverse long wave length thermodynamic fluctuations.

Let us briefly recall a familiar series of techniques from the theory of spin fluctuations in itinerant antiferromagnets. These constitute a ''local band'' method used earlier[4,9] for antiferromagnets. We assume that for $T_N<T\ll T_N^0$, short-range order is present in the system, and the main contribution to the partition function comes from SDW configurations $\{\mathbf{\Delta}_i\}$ of Eq. (9) for which $\mathbf{\Delta}_i=\mathbf{e}_i\Delta$, and the vector $\mathbf{e}_i$ changes direction slightly between nearest-neighbor sites of the lattice I and j: $|\mathbf{e}_i-\mathbf{e}_j|\ll1$. We introduce a local coordinate system defined by the angles $\{\theta_i,\Phi_i\}$, which specify the direction of the local quantization axis $\mathbf{e}_i$ relative to the laboratory quantization axis. We introduce also the angle b_i, which describes the rotation of the spin-density vector about the $\mathbf{e}_i$ axis.

In the local coordinate system, the Hamiltonian (1) can be rewritten in the form[2]

$$H_{\rm loc}=H_0+H_1+H_2+H_{\rm int}, \tag{16}$$

$$H_1=-i\sum_{ij}\sum_{\sigma\sigma'}\tilde{t}_{ij}^{\sigma\sigma'}C_{i\sigma}^+C_{j\sigma'}, \tag{17}$$

$$\tilde{t}_{ij}^{\sigma\sigma'}=t_{ij}[\sigma_{\sigma\sigma'}^+a_{ij}^*+\sigma_{\sigma\sigma'}^-a_{ij}-\sigma_{\sigma\sigma'}^zg_{ij}], \tag{18}$$

$$H_2=\sum\tilde{\tilde{t}}_{ij}^{\sigma\sigma'}C_{i\sigma}^+C_{j\sigma'}, \tag{19}$$

$$\tilde{\tilde{t}}_{ij}^{\sigma\sigma'}=t_{ij}d_{ij}\delta_{\sigma\sigma'}. \tag{20}$$

The matrix elements a_{ij}, g_{ij} and d_{ij} can be written in the form

$$a_{ij}\approx\frac{1}{2}\exp(-\bar{b}_{ij})(\Phi_{ij}\sin(\bar{\Theta}_{ij}-i\Theta_{ij})), \tag{21}$$

$$g_{ij}\approx\frac{1}{2}(b_{ij}+\Phi_{ij}\cos\bar{\Theta}_{ij}), \tag{22}$$

$$d_{ij}\approx-\frac{1}{2}(|a_{ij}|^2+g_{ij}^2), \tag{23}$$

where $\Theta_{ij}=\Theta_i-\Theta_j$, $\overline{\Theta}_{ij}=(1/2)(\Theta_i+\Theta_j)$, and analogously for the angles Φ_i and b_i. In Eqs. (16)–(20), it is assumed that the operators $C_{i\sigma}$ and $C^+_{i\sigma}$ act in the local coordinate system, and the Hamiltonians H_0 and $H_{\rm int}$ have the same form as in (1). The matrix elements (21)–(23) have been written with allowance for the condition $|\mathbf{e}_i-\mathbf{e}_j|\ll 1$ for nearest neighbors. In the local coordinate system, the free energy of the system F can easily be represented as

$$F=-T\ \ln\ \mathrm{Tr}\ \exp\left(-\frac{1}{T}H_{\rm loc}\right)=F_0[\overline{\Delta}]+F_{SF}[\overline{\Delta},\{\mathbf{e}_j\}], \tag{24}$$

where the functional $F_0[\overline{\Delta}]$ can be reconstructed by integrating the self-consistency equation for the mean SDW amplitude $\Delta=\overline{\Delta}$. It is formally independent of the $\{\mathbf{e}_i\}$ directions, so $F_0[\overline{A}]$ is the free energy of the homogeneous SDW with some amplitude $\overline{\Delta}$. The term $F_{SF}[\overline{\Delta},\mathbf{e}_i]$ is an additional exchange energy of the antiferromagnet, associated with thermodynamic orientational disorder. The large-scale SDW configurations $\{\mathbf{e}_i\}$ in the functional F_{SF} can be represented as an expansion in the angles Θ_{ij}, Φ_{ij}, b_{ij}. Let us write the explicit expression for the functional F_{SF} to second order in H_1 and H_2:

$$F_{SF}=F_\perp+F_\parallel, \tag{25}$$

$$F_\perp=-\frac{1}{4}T\sum_n\left[\sum_{ij}\sum_\sigma \overset{0\sigma\sigma'}{G}_{ij}t_{ij}|a_{ij}|^2 + \sum_{ijlm}\sum_{\sigma\alpha\beta\gamma}\overset{0\sigma\alpha}{G}_{ij}t_{jl}(\sigma^+_{\alpha\beta}a^*_{jl} + \sigma^-_{\alpha\beta}a_{jl})\overset{0\beta\gamma}{G}_{lm}t_{mi}(\sigma^+_{\gamma\sigma}a^*_{mi}+\sigma^-_{\gamma\sigma}a_{mi})\right], \tag{26}$$

$$F_\parallel=-\frac{1}{4}T\sum_n\left[\sum_{ij}\sum_\sigma \overset{0\sigma\sigma}{G}_{ij}t_{ij}g^2_{ij} + \sum_{ijlm}\sum_{\sigma\alpha\beta\gamma}\overset{0\sigma\alpha}{G}_{ij}t_{jl}\sigma^z_{\alpha\beta}g_{ij}\overset{0\beta\gamma}{G}_{lm}t_{mi}\sigma^z_{\gamma\sigma}g_{mi}\right], \tag{27}$$

where $\overset{0}{G}$ is the Green function (10), but $\overline{\Delta}$ is not defined by Eq. (12) (see Sec. 5 below). It is noted in Ref. 2 that the structure of the term g_{ij} (22) is such that it leads to contributions to the energy F_{SF} of fourth and higher order in Φ_{ij}. In the long-wave limit, the first nonvanishing terms in the energy expansion of F_{SF} are of second order in Φ_{ij} and Θ_{ij} and are proportional to $a^*_{ij}a_{lm}$. We now take into account that in accordance with (21),

$$a^*_{ij}a_{lm}=\frac{1}{4}(\mathbf{e}_i-\mathbf{e}_j)(\mathbf{e}_l-\mathbf{e}_m). \tag{28}$$

Thus, in this limit the spin-fluctuation contribution to the free energy of the itinerant antiferromagnet has the form of the effective Heisenberg Hamiltonian for classical spins $S=1$,

$$F_{SF}=-\sum_{ij}J_{ij}\mathbf{e}_i\mathbf{e}_j. \tag{29}$$

In the next section, we analyze in detail the structure of the exchange integral J_{ij}. Here we emphasize that the calculation of the partition function reduces to an integration over the orientations of the random vector field $\{\mathbf{e}_i\}$, $|\mathbf{e}_i|=1$:

$$Z=Z_0\ \exp(-E_0/T)Z_{SF}, \tag{30}$$

$$Z_{SF}=\int \exp(-F_{SF}/T)D\{\mathbf{e}_i\}, \tag{31}$$

where F_{SF} is given in the form (29).

To find any physical characteristic of an itinerant antiferromagnet in the SRO region, it is necessary to average over the ensemble of SDW fluctuations, restricted in our approach to the set of configurations $\{\mathbf{e}_i\}$. If the quantity $A[\{\mathbf{e}_i\}]$ corresponds to any configuration $\{\mathbf{e}_i\}$, then the thermodynamic average $\langle A\rangle$ is calculated as a functional integral,

$$\langle A\rangle=Z^{-1}_{SF}\int A[\{\mathbf{e}_i\}]\exp(-F_{SF}[\overline{\Delta},\{\mathbf{e}_i\}]/T)D\{\mathbf{e}_i\}. \tag{32}$$

We calculate $\langle A\rangle$ in the ordinary Gaussian approximation for the distribution $\{\mathbf{e}_i\}$ of the vector orientation, whose effective Hamiltonian is of the quadratic form (29). In this approximation, all even spin correlators decouple into products of the pairwise correlators $\langle\mathbf{e}_i\ldots\mathbf{e}_j\rangle\to\langle\mathbf{e}_i\mathbf{e}_l\rangle\ldots\langle\mathbf{e}_m\mathbf{e}_j\rangle$, and all odd correlators into products of pairwise correlators and mean values $\langle\mathbf{e}_i\rangle$. In what follows we restrict our discussion solely to paramagnetic phase for which $\langle\mathbf{e}_i\rangle=0$ holds, and all odd correlators vanish.

To calculate any average $\langle A\rangle$, it is thus necessary to determine the amplitude and the pairwise correlator $f_{ij}=\langle\mathbf{e}_i\mathbf{e}_j\rangle$ and it is necessary, strictly speaking, to have a system of self-consistency equations for these quantities. Many qualitative results, however, can by obtained on the basis of general assumptions about the spatial and temperature dependence of f_{ij}. Thus, it is natural to assume that in the short-range order regime, with $T_N<T\ll T^0_N$, the Fourier component $f_{\mathbf{q}}$ has a sharp maximum at small values of the wave vectors $q\ll a^{-1}$, corresponding to the maximum of the spin correlator $\langle\mathbf{S}_{\mathbf{q}}\mathbf{S}_{-\mathbf{q}}\rangle$ at wave vectors $\mathbf{q}$ near the antiferromagnetic structure vector $\mathbf{Q}=\mathbf{G}/2$. The spin correlation radius $l(T)$ and the temperature dependence of $f_{\mathbf{q}}(T)$ can be estimated[4] on the basis of well-known results for the classical Heisenberg Hamiltonian.

4. EFFECTIVE EXCHANGE INTEGRAL AND THE NÉEL POINT

The exchange integral in (29) is given by

$$J_{ij}=\frac{1}{N^2}\sum_{\mathbf{pp}'}J_{\mathbf{pp}'}\ \exp(i(\mathbf{pR}_i-\mathbf{p}'\mathbf{R}_j)), \tag{33}$$

$$J_{\mathbf{pp}'}=I_{\mathbf{p}}\delta_{\mathbf{pp}'}+K_{\mathbf{p}}\delta_{\mathbf{p},\mathbf{p}'+\mathbf{Q}}, \tag{34}$$

$$I_{\mathbf{p}}=-\frac{T}{8N}\sum_{n\mathbf{k}}\left[\varepsilon_{\mathbf{k}-\mathbf{p}}(A_{\mathbf{k}}^{\uparrow}+A_{\mathbf{k}}^{\downarrow})+(\varepsilon_{\mathbf{k}}-\varepsilon_{\mathbf{k}-\mathbf{p}})^2(A_{\mathbf{k}}^{\uparrow}A_{\mathbf{k}-\mathbf{p}}^{\downarrow}-B_{\mathbf{k}}^{\uparrow}B_{\mathbf{k}-\mathbf{p}}^{\downarrow})\right], \tag{35}$$

$$K_{\mathbf{p}}=-\frac{T}{8N}\sum_{n\mathbf{k}}\left[\varepsilon_{\mathbf{k}-\mathbf{p}}(B_{\mathbf{k}}^{\uparrow}+B_{\mathbf{k}}^{\downarrow})+(\varepsilon_{\mathbf{k}}^2-\varepsilon_{\mathbf{k}-\mathbf{p}}^2)^2(A_{\mathbf{k}}^{\uparrow}B_{\mathbf{k}-\mathbf{p}}^{\downarrow}-B_{\mathbf{k}}^{\uparrow}A_{\mathbf{k}-\mathbf{p}}^{\downarrow})\right], \tag{36}$$

$$A_{\mathbf{k}}^{\sigma}=\frac{i\omega_n+\mu+\varepsilon_{\mathbf{k}}}{(i\omega_n+\mu)^2-E_{\mathbf{k}\sigma}^2},\quad B_{\mathbf{k}}^{\sigma}=\frac{I+\sigma\Delta}{(i\omega_n+\mu)^2-E_{\mathbf{k}\sigma}^2}. \tag{37}$$

Below, we analyze only the case $T=0$ and $\mu=0$ to obtain explicit expressions for J_{ij}. We also assume that to zeroth order in $(a/l)^{-1}\ll 1$, $\overline{\Delta}\approx\Delta$ from Eq. (12) and $(\Delta\pm I)^2\gg(a/l)^2t^2$. In this case we can convince ourselves that the ferromagnetic component $K_{\mathbf{p}}=0$ (since $\mu=0$), and that to calculate the Néel temperature we need the quantity $M_{\mathbf{p}}=I_{\mathbf{p}=0}-I_{\mathbf{p}}$. After some transformations, the latter can be written in the form

$$M_{\mathbf{p}}=\frac{\Delta^2}{8N}\sum_{\mathbf{k}}\frac{(\varepsilon_{\mathbf{k}}-\varepsilon_{\mathbf{k}-\mathbf{p}})^2}{E_{\mathbf{k}\uparrow}E_{\mathbf{k}-\mathbf{p}\downarrow}(E_{\mathbf{k}\uparrow}+E_{\mathbf{k}-\mathbf{p}\downarrow})}, \tag{38}$$

where the summation over the wave vector $\mathbf{k}$ extends over the first Brillouin zone. For the classical Heisenberg model with $S=1$, mean-field theory yields the phase-transition temperature

$$T_N^{MF}=\frac{2}{3}\frac{1}{N}\sum_{\mathbf{p}}M_{\mathbf{p}}. \tag{39}$$

The self-consistent contribution of magnetization fluctuations, evaluate 1 by means of the spin Green's functions method, reduces the Néel temperature in comparison with (39):

$$T_N^{SF}=\frac{2}{3}\left(\frac{1}{N}\sum_{\mathbf{p}}M_{\mathbf{p}}^{-1}\right)^{-1}. \tag{40}$$

Equation (40) was derived previously (for example, see Refs. 10 and 11). The difference between T_N^{MF} and T_N^{SF} becomes sizable in low-dimensional systems. To illustrate this fact, we specify the dispersion relation in the form

$$\varepsilon_{\mathbf{k}}=-\sum_{\alpha}2t_{\alpha}\cos(k_{\alpha}a_{\alpha}),\quad \alpha=x,y,z.$$

Above all, we consider the formal limit of strong interaction: $U,I\gg t$. If $(\Delta-I)^2\gg(2t)^2$, the reduced exchange integral acquires the simple form

$$M_{\mathbf{p}}=\frac{\Delta}{2(\Delta^2-I^2)}\sum_{\alpha}t_{\alpha}^2\sin^2\left(\frac{p_{\alpha}a_{\alpha}}{2}\right). \tag{41}$$

The Néel temperature in the mean-field approximation is

$$T_N^{MF}=\frac{\Delta}{6(\Delta^2-I^2)}\sum_{\alpha}t_{\alpha}^2 \tag{42}$$

and remains finite for all effective dimensions of the electronic system. Thermodynamic fluctuations lead to different behavior of the Néel point. For example, if $(t_y/t_x)^2\ll 1$, $(t_z/t_y)^2\ll 1$, we obtain

$$T_N^{SF}=T_N^{MF}\pi\frac{t_y}{t_x}\left(\ln\frac{8t_y}{t_z}\right)^{-1}. \tag{43}$$

Recall that for an isotropic situation $(t_x=t_y=t_z)$, $T_N^{SF}=T_N^{MF}/1.51$. The dependence of the temperature T_N^{SF} on the anisotropy parameter for a quasi-two-dimensional itinerant antiferromagnet was calculated in Ref. 4 in the case $t_y\equiv t_x\gg t_z$ (see Eq. (43)). In the limit $t/U\ll 1$, the temperature at which of the SDW amplitude Δ appears (given by Eq. (12)) is formally $T_N^0\sim U$, i.e., much higher than T_N^{SF}. Thus, over a wide range of temperatures above the Néel point, short-range magnetic order can be described by a model with fixed (independent of T) SDW amplitude $(T_N^{SF}<T\ll T_N^0)$. This order is characterized by the correlation length $l(T)$ of transverse fluctuations of the spin density. For the classical Heisenberg Hamiltonian in the one-dimensional case, the expression for $l(T)$ can be obtained by means of the renormalization-group method. For $T\ll J$,

$$l(T)\simeq aJ/T.$$

In the strong-coupling regime $(\Delta\pm I)^2\gg(2t)^2$, the exchange integral is

$$J\sim\frac{\Delta t^2}{|\Delta^2-I^2|}\sim\frac{t^2}{U}\ll t. \tag{44}$$

For $t\ll I\ll U$ we can write the explicit expression

$$J(I)=J_0+\frac{8t^2I^2}{U^3},\quad J_0=\frac{2t^2}{U}.$$

Thus, the staggered potential leads to an increase in correlation length. When the ratio U/t is not large, calculation of the function $M_{\mathbf{p}}$ is very complicated. However, only the behavior of the exchange integral at small values of $\mathbf{p}$ $(p_{\alpha}a_{\alpha}\ll 1)$ is of importance. Therefore, to estimate the transition temperature, we make use of a piecewise linear approximation to the function $M_{\mathbf{p}}$. For the one-dimensional chain we take

$$M_{\mathbf{p}}^{(1)}=\begin{cases}M(p_xa_x)^2, & |p_x|<p_0,\\ \tilde{M}, & |p_x|>p_0,\end{cases} \tag{45}$$

where

$$M=-\frac{\Delta t_x}{8\pi I}\sum_{\sigma}\sigma\frac{K(k_{\sigma})-E(k_{\sigma})}{k_{\sigma}}, \tag{46}$$

$$\tilde{M}=-\frac{\Delta t_x}{2\pi I}\sum_{\sigma}\sigma\frac{E(k_{\sigma})-(1-k_{\sigma})^2K(k_{\sigma})}{k_{\sigma}},$$

$$(p_0a)^2=\tilde{M}/M\ll 1. \tag{47}$$

For the quasi-one-dimensional case, up to the first-order terms in the overlap integral between chains, for which $\Delta^2,t_x^2\gg t_y^2\gg t_z^2$, we have

$$M_{\mathbf{p}}^{(2)}=M_{\mathbf{p}}^{(1)}+\frac{t_y^2\,\sin^2(p_y a_y/2)+t_z^2\,\sin^2(p_z a_z/2)}{8t_x^2}\,(\tilde{M}+4M). \tag{48}$$

In the weak-coupling limit, where $|I\pm\Delta|^2\ll(2t)^2$, one can obtain

$$T_N^{MF}=\frac{\Delta^2}{3\pi t}\ln\frac{8t}{\sqrt{|\Delta^2-I^2|}},\quad t=t_x, \tag{49}$$

$$T_N^{SF}=\left[(T_N^{MF})^{-1}+\frac{12\sqrt{2}I\,\ln(8t_y/t_z)}{\Delta t_y\,\ln|(\Delta+I)/(\Delta-I)|}\right]^{-1}. \tag{50}$$

Thus, long-range antiferromagnetic order appears at the temperature T_N^{SF}. With increasing temperature ($T_N^{SF}<T<T_N^0$), the system passes over the region of short-range order with the correlation length $l(T)\sim aJ/T$, where for the weak-coupling case ($U\ll t$) we can estimate

$$J\sim\frac{\Delta^2}{t}\ln\frac{8t}{\sqrt{|\Delta-I||\Delta+I|}}\sim\frac{\Delta^2}{U}\ll\Delta. \tag{51}$$

This estimate holds for all $I<I_c$ in the stability region of the phase diagram (see Fig. 1 and Ref. 7), where $1\gg|\Delta\pm I|/t\gg a/l$. Note that $T_N^0\ll U$ for weak-coupling case, and T_N^{MF} and T_N^{SF} are also small compared with T_N^0. The situation $I_c<I<I_0$ (the metastable region) is more difficult to analyze, since the ratio $|\Delta-I|/t$ may be of order a/l. Thus, the renormalization of $\overline{\Delta}(T)$ takes an important role in this case, and a $J(T)$ dependence arises. We are not interested here in the detailed expressions for $J(T)$ and T_N^{SF} in this special case. For $I\ll\Delta_0$ can easily obtain from (51) the dependence of J on the staggered potential I:

$$J(I)=J_0-\frac{4\pi}{U}I^2,\quad J_0=\frac{4\pi\Delta_0^2}{U}.$$

Thus, the correlation length $l(I)$ decreases with increasing I.

5. SINGLE-PARTICLE EXCITATIONS AND RENORMALIZATION OF SDW AMPLITUDE IN THE LOCAL HARTREE–FOCK APPROXIMATION

The scattering of itinerant electrons by a certain SDW fluctuation $\{\mathbf{e}_i\}$ above the Néel point $T_N\simeq T_N^{SF}$ is characterized by the single-particle Green's function $G_{ij}^{\sigma\sigma'}(\{\mathbf{e}_i\},t)$:

$$G=\overset{0}{G}+\overset{0}{G}(H_1+H_2)G. \tag{52}$$

Thermodynamic disorder enters into Eq. (52) through the Hamiltonians H_1 and H_2, and $\overset{0}{G}$ has the form (10).

The foregoing configurational integration procedure makes it possible to express the propagator $\langle G_{\mathbf{kp}}^{\sigma\sigma}(\{\mathbf{e}\},\omega)\rangle$ averaged over the $\{\mathbf{e}_i\}$ configurations; in the lowest order of perturbation theory,

$$\langle G^{-1}\rangle=\langle\overset{0-1}{G}\rangle-\Sigma,\quad \Sigma=\langle H_2\rangle+\langle H_1\overset{0}{G}H_1\rangle. \tag{53}$$

The self-energy part Σ is

$$\Sigma_{\mathbf{kp}}^{\sigma\sigma}=-\frac{1}{4}\,\delta_{\sigma\sigma'}\Big\{\delta_{\mathbf{kp}}\sum_{\mathbf{q}}\,[f_{\mathbf{q}}(\varepsilon_{\mathbf{k}}-\varepsilon_{\mathbf{k}-\mathbf{q}})+f_{\mathbf{q}}(\varepsilon_{\mathbf{k}}-\varepsilon_{\mathbf{k}-\mathbf{q}})^2\times A_{\mathbf{k}-\mathbf{q}}^{\sigma}]-\delta_{\mathbf{k},\mathbf{p}+\mathbf{Q}}\sum_{\mathbf{q}}\,f_{\mathbf{q}}(\varepsilon_{\mathbf{k}}-\varepsilon_{\mathbf{k}-\mathbf{q}})^2B_{\mathbf{k}-\mathbf{p}}^{-\sigma}\Big\}. \tag{54}$$

Since the structure factor of the spin correlations $f_{\mathbf{q}}$ is assumed to be nonzero in the small interval of quasimomenta near $\mathbf{q}=0$, the slowly varying factors (in comparison to $f_{\mathbf{q}}$) on the right-hand side of Eq. (54) can be expanded in series near the point $\mathbf{q}=0$. Retaining the lowest-order terms in this expansion in $\mathbf{q}$, we obtain an expression for the averaged Green's function:

$$\langle G_{\mathbf{kp}}^{\sigma\sigma'}(\omega)\rangle=\{\delta_{\mathbf{kp}}\delta_{\sigma\sigma'}[\omega^+\varepsilon_{\mathbf{k}}-\nu_{\mathbf{k}}-F_{\mathbf{k}}^2A_{\mathbf{k}+\mathbf{Q}}^{\sigma}\}+\delta_{\mathbf{k},\mathbf{p}+\mathbf{Q}}\delta_{\sigma\sigma'}[(\overline{\Delta}\sigma+I)-F_{\mathbf{k}}^2B_{\mathbf{k}}^{-\sigma}]\}/D_{\mathbf{k}}^{\sigma}(\omega), \tag{55}$$

$$D_{\mathbf{k}}^{\sigma}(\omega)=[\omega-\varepsilon_{\mathbf{k}}+\nu_{\mathbf{k}}-F_{\mathbf{k}}^2A_{\mathbf{k}}^{\sigma}][\omega+\varepsilon_{\mathbf{k}}-\nu_{\mathbf{k}}-F_{\mathbf{k}}^2A_{\mathbf{k}+\mathbf{Q}}^2]-[(\overline{\Delta}\sigma+I)-F_{\mathbf{k}}^2B_{\mathbf{k}}^{-\sigma}]^2, \tag{56}$$

where

$$\nu_{\mathbf{k}}=-\frac{1}{8}\sum_{\alpha}\,\frac{\partial^2\varepsilon_{\mathbf{k}}}{\partial^2k_{\alpha}^2}\,l_{\alpha}^{-2}, \tag{57}$$

$$F_{\mathbf{k}}^2=\frac{1}{4}\sum_{\alpha}\,\left(\frac{\partial\varepsilon_{\mathbf{k}}}{\partial k_{\alpha}}\right)^2l_{\alpha}^{-2}, \tag{58}$$

and l_{α} is the transverse correlation length of the spin-density fluctuations in the αth direction. The poles of the averaged Green's function $\langle G\rangle$ determine the spectrum of single-particle excitations in a short-range-order antiferromagnet.

The zeroes of the denominator (56) of the function (55) are given by the dispersion relation

$$(\omega^2-E_{\mathbf{k}\sigma}^2)^3+(2\varepsilon_{\mathbf{k}}\nu_{\mathbf{k}}-2F_{\mathbf{k}}^2-\nu_{\mathbf{k}}^2)(\omega^2-E_{\mathbf{k}\sigma}^2)^2+F_{\mathbf{k}}^2(F_{\mathbf{k}}^2-4E_{\mathbf{k}\sigma}^2+2\varepsilon_{\mathbf{k}}\nu_{\mathbf{k}}+4I^2+4I\sigma\overline{\Delta})\times(\omega^2-E_{\mathbf{k}\sigma}^2)+4I\sigma\overline{\Delta}E_{\mathbf{k}\sigma}^4=0. \tag{59}$$

Thus, there are twelve branches of energy dispersion. To order $(a/l)^2$, we see that

$$(\omega_0^{\sigma})^2=E_{\mathbf{k}\sigma}^2+\frac{\sigma\overline{\Delta}IF_{\mathbf{k}}^2}{E_{\mathbf{k}\sigma}^2-I(I+\sigma\overline{\Delta})}, \tag{60}$$

$$(\omega_{1,2}^{\sigma})^2=E_{\mathbf{k}\sigma}^2\pm[4F_{\mathbf{k}}^2(E_{\mathbf{k}\sigma}^2-I(I+\sigma\overline{\Delta}))]^{1/2}+F_{\mathbf{k}}^2-\varepsilon_{\mathbf{k}}\nu_{\mathbf{k}}-\frac{\sigma\overline{\Delta}IF_{\mathbf{k}}^2}{2(E_{\mathbf{k}\sigma}^2-I(I+\sigma\overline{\Delta}))}. \tag{61}$$

Note that to first order in a/l, with $(\omega_0^{\sigma})^2=E_{\mathbf{k}\sigma}^2$, only eight new branches $\pm\omega_{1,2}^{\sigma}$ found. It is not hard to show that the parameter $F_{\mathbf{k}}$ associated with the Hamiltonian H_1 (19) corresponds to the scattering of a quasiparticle by a spin fluctuation with spin flip. Thus, in the short-range-order phase, the motion of an electron with ''right'' or ''wrong'' spin direction is possible in each of the two antiferromagnetic sublattices if they are in part randomly oriented. As a

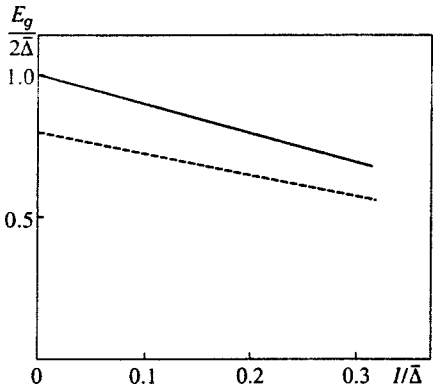

FIG. 2. $E_g(I)$ dependence for large (solid line) and short (dashed line) range order phases.

consequence, the dispersion branches are ''split'' and the gap for single-particle excitations is reduced. For example, (see Fig. 2),

$$E_g \approx 2\left(|I-\bar{\Delta}| - \frac{a}{l}\, t\left(\frac{\bar{\Delta}}{|I-\bar{\Delta}|} \right)^{1/2} \right) \tag{62}$$

for the one-dimensional system. Note that the reduction in E_g and the spin amplitude $\bar{\Delta}$, as compared with the long-range-order state $(a/l=0)$, is due exclusively to the spin-flip scattering Hamiltonian H_1. The scattering process without spin flip is described by the Hamiltonian H_2 and the corresponding parameter $\nu_{\mathbf{k}}$. This process leads to effective narrowing of the energetic band $\tilde{\varepsilon}_{\mathbf{k}}=\varepsilon_{\mathbf{k}}-\nu_{\mathbf{k}}$, where

$$\tilde{\varepsilon}_{\mathbf{k}} = -\sum_{\alpha} 2t_{\alpha}\left[1-\frac{1}{8}\left(\frac{a_{\alpha}}{l_{\alpha}} \right)^2 \right] \cos(k_{\alpha}a_{\alpha}). \tag{63}$$

We now address the influence of spin fluctuations on the mean SDW amplitude $\bar{\Delta}$. Since the order parameter $\bar{\Delta}(T)$ has very weak temperature dependence in the temperature range $T_N<T\ll T_N^0$ due to the Fermi statistics of the quasiparticles, we use the self-consistency equation (local Hartree-Fock approximation), in which the Fermi function is replaced by the step function:

$$\bar{\Delta} = \frac{U}{2} \int \frac{d\omega}{2\pi i} \frac{1}{N^2} \sum_{\mathbf{kp}} \sum_{\sigma\sigma'} \sigma^z_{\sigma\sigma'} \langle G^{\sigma\sigma'}_{\mathbf{k},\mathbf{p}+\mathbf{Q}} \rangle. \tag{64}$$

The amplitude is temperature-dependent only because of the transverse spin fluctuations with correlation length $\{l_{\alpha}(T)\}$. We substitute the average Green's function (55) into Eq. (64), and after some manipulations we obtain the following equation to second order in a/l:

$$\bar{\Delta} = \frac{U}{4N} \sum_{\mathbf{k}} \sum_{\sigma} \frac{\sigma}{E_{\mathbf{k}\sigma}} \left[(I+\sigma\bar{\Delta}) \left(1+\frac{\varepsilon_{\mathbf{k}}\nu_{\mathbf{k}}}{E^2_{\mathbf{k}\sigma}} \right) + \frac{IF^2_{\mathbf{k}}}{E^2_{\mathbf{k}\sigma}} \left(1-\frac{3(I+\sigma\bar{\Delta})^2}{2E^2_{\mathbf{k}\sigma}} \right) \right]. \tag{65}$$

Note that in the absence of a staggered potential $(I=0)$, renormalization of the SDW amplitude is due to the process without spin flip $(\nu_{\mathbf{k}})$, since the term proportional to $F^2_{\mathbf{k}}$ vanishes on the right-hand side of (65). As a consequence of the above mentioned narrowing of the band $\varepsilon_{\mathbf{k}}$ (63) the resulting change in the mean spin density per site $\bar{\Delta}$ is positive, i.e., it increases compared with Δ. For a one-dimensional chain one can take Eq. (65) in the form

$$\bar{\Delta} = \frac{U}{4\pi t} \sum_{\sigma} k_{\sigma} \left[\sigma K(k_{\sigma})(I+\sigma\bar{\Delta}) + \frac{\bar{\Delta}}{8}\left(\frac{a}{l} \right)^2 [K(k_{\sigma})-E(k_{\sigma})] \right], \quad k_{\sigma}=k_{\sigma}(\bar{\Delta}). \tag{66}$$

The temperature dependence of $\bar{\Delta}$ is due only to the dependence of $l(T)$. If, in agreement with our estimates in Sec. 3, we express the correlation length $l(T)$ in terms of the model parameters, it becomes clear that the contribution of transverse spin fluctuations to the renormalization of the SDW amplitude significantly exceeds the corresponding contribution arising from the temperature dependence of the Fermi function. It can be shown from (66) that $\bar{\Delta}(T)-\Delta \sim [a/l(T)]^2$, but to calculate the exchange integrals J_{ij} in Sec. 3 it was assumed for simplicity that $\bar{\Delta}(T)\approx\Delta$, where Δ is defined by Eq. (13). An explicit expression for $\bar{\Delta}$ can be obtained in the limit $I\ll\Delta$:

$$\bar{\Delta}(I) \approx \tilde{\Delta}_0(1-\tilde{I}^2/\tilde{\Delta}_0^2), \quad \tilde{\Delta}_0 = 8t\, \exp\left[-\frac{2\pi t}{U(1+\lambda)} \right],$$
$$\tilde{I}^2 = \frac{I^2}{1+\lambda}, \quad \lambda = \frac{1}{8}(a/l)^2. \tag{67}$$

Omitting the detailed analysis of Eq. (66) we note that in the phase diagram $\Delta(I)$, the metastable region $(I_c<I<I_0)$ is shifted to the right as compared with the case $a/l=0$. This tendency is shown qualitatively in Fig. 1 by the dashed line.

6. TRANSVERSE DYNAMICAL SUSCEPTIBILITY IN RANDOM-PHASE APPROXIMATION

A general theory of the dynamical susceptibility of itinerant magnets with short-range order was developed in Ref. 12. A calculation scheme (the socalled ''RPA with exchange'') was proposed to draw some qualitative conclusions about the nature of the response to an external magnetic field. In Ref. 13, a method for the calculation of the dynamical susceptibility that is somewhat different from Ref. 12 was proposed, based on the spin-fluctuation approach.[2,3] The two-particle Green's function for itinerant antiferromagnets without a staggered potential was calculated.[9,13] This method enables one to describe SDW antiferromagnets above the Néel point.

According to Ref. 12 one introduces a spin-density correlator for each fluctuation $\{\mathbf{e}_i\}$:

$$\chi^{\alpha\beta}_{\mathbf{qq}'}(t,\{\mathbf{e}_i\}) = \frac{i}{2N} (TS^{\alpha}_{\mathbf{q}}(t)S^{\beta}_{-\mathbf{q}'}(0))_{\{\mathbf{e}_i\}}. \tag{68}$$

Parentheses on the right-hand side of Eq. (68) denote averaging over quantum states for the fixed set of vectors $\{\mathbf{e}_i\}$ that specify the SDW configuration: the symbol T denotes time ordering in t. The complete response to an exter-

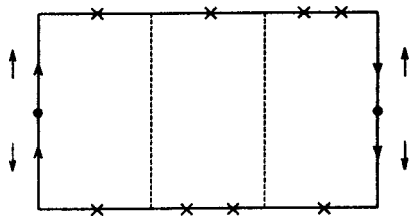

FIG. 3. Diagram for process with spin flip H_1.

nal variable magnetic field (the magnetic susceptibility at frequency Ω) is defined as the average over the SDW configurations:

$$\chi^{\alpha\beta}_{\mathbf{qq}'}(\Omega)=\int_0^\infty \exp(i\Omega t)\langle\chi^{\alpha\beta}_{\mathbf{qq}'}(t,\{\mathbf{e}_i\})\rangle dt. \tag{69}$$

We now discuss the calculation of the averages in (69). The two-particle Green's function of the system with fluctuating spin density can be usefully represented as an infinite series of ''ladder'' diagrams that take into account electron scattering by randomly oriented magnetic moments at sites. One of the simplest diagrams is shown in Fig. 3. The solid line denotes the unperturbed function G^0 given by Eq. (10), the dashed line denotes the Coulomb repulsion of particles with opposite spins U, the cross represents the matrix element a_{ij} for processes with spin flip H_1, and the arrows $\downarrow$ and $\uparrow$ correspond to the projections of electron spin.

Let us turn our attention to the alternation in the arrangement of matrix elements a_{ij}, a^*_{ij} and spin indices. In the diagrammatic language, the averaging procedure (69) amounts to taking into account all the possible ways of joining the crosses pairwise by wavy lines. A graphical element consisting of a pair of crosses joined by a wavy line corresponds to the spin correlator f_{ij}. Figure 4 displays some typical diagrams that contribute to the average two-particle Green's function to first order in U, and to fourth order in the disorder a_{ij}. Below we consider only diagrams of type a, assuming that diagrams b, c and d have already been taken into account in the renormalization of the interaction U. Note that U is not renormalized to second order in a_{ij}. In Fig. 3 and Fig. 4, processes without spin flip corresponding to H_2 are omitted, since their direct average $\langle|a^2_{ij}|\rangle$ in the single-particle channel is not needed in any additional explanations.

Thus, calculation of the susceptibility of an antiferromagnet in the short-range-order phase reduces to the modified RPA with quasiparticles ''dressed'' by disorder. In other words, it reduces to summation of the infinite sequence of ''ladder''-type diagrams shown in Fig. 5, where the double line represents the average single-particle Green's function.

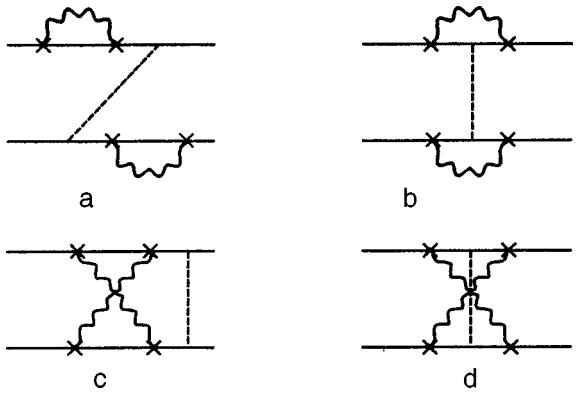

FIG. 4. Diagrams for average two-particle Green's function to fourth order in a_{ij}.

The first diagram in the series in Fig. 5, which does not explicitly contain the interaction U, corresponds to the following function of the two quasimomenta:

$$\overline{\chi}^{+-}_{\mathbf{qq}'}(\Omega)=\frac{1}{N^2}\int\sum_{\mathbf{kk}'}\langle G^{\downarrow\downarrow}_{\mathbf{k},\mathbf{k}-\mathbf{q}'}(\omega)\rangle\langle G^{\uparrow\uparrow}_{\mathbf{k},\mathbf{k}+\mathbf{q}}(\omega-\Omega)\rangle\frac{d\omega}{2\pi i}. \tag{70}$$

Apart from the function $\langle G\rangle$, the expression for the response $\overline{\chi}^{+-}$ contains boths diagonal and off-diagonal contributions in the quasimomentum space, each of which depends on just one quasimomentum:

$$\overline{\chi}^{+-}_{\mathbf{qq}'}(\Omega)=\delta_{\mathbf{qq}'}\overline{\chi}^{+-}_0(\mathbf{q},\Omega)+\delta_{\mathbf{q},\mathbf{q}'+\mathbf{Q}}\overline{\chi}^{+-}_{\mathbf{Q}}(\mathbf{q},\Omega). \tag{71}$$

Summing the geometric series in Fig. 5 (in fact, solving the Dyson equation), we obtain the response function

$$\overline{\chi}^{+-}_{\mathbf{qq}'}(\Omega)=\sum_{\mathbf{q}_1}\overline{\chi}^{+-}_{\mathbf{q},\mathbf{q}_1}(\Omega)[1-U\overline{\chi}^{+-}_{\mathbf{q}_1,\mathbf{q}'}(\Omega)]^{-1}. \tag{72}$$

where

$$[1-U\overline{\chi}^{+-}_{\mathbf{qq}'}(\Omega)]^{-1}=\frac{[1-U\overline{\chi}^{+-}_0(\mathbf{q}+\mathbf{Q},\Omega)]\delta_{\mathbf{qq}'}+U\overline{\chi}^{+-}_{\mathbf{Q}}(\mathbf{q},\Omega)\delta_{\mathbf{q},\mathbf{q}'-\mathbf{Q}}}{\mathrm{Det}(\mathbf{q},\Omega)}, \tag{73}$$

$$\mathrm{Det}(\mathbf{q},\omega)=[1-U\overline{\chi}^{+-}_0(\mathbf{q},\Omega)][1-U\overline{\chi}^{+-}_0(\mathbf{q}+\mathbf{Q},\Omega)]-U^2\overline{\chi}^{+-}_{\mathbf{Q}}(\mathbf{q},\Omega)\overline{\chi}^{+-}_{\mathbf{Q}}(\mathbf{q}+\mathbf{Q},\Omega). \tag{74}$$

The structure of the denominator of the transverse dynamical susceptibility (72) determines the spectrum of magnetic excitations of the system. We now proceed to some detailed calculations.

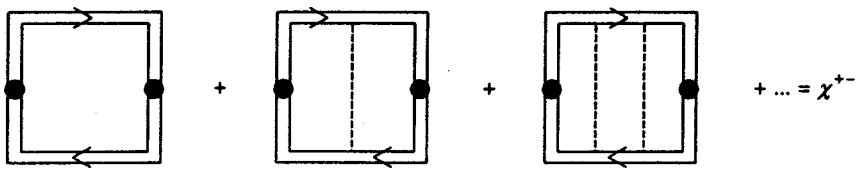

FIG. 5. Ladder diagrams to calculate the transverse susceptibility.

7. SPECTRUM OF LOW-FREQUENCY EXCITATIONS

We first calculate the irreducible susceptibility $\overline{\chi}_{\mathbf{qq}'}(\Omega)$. We substitute the function $\langle G\rangle$ (55) into Eq. (70) sum over the quasimomentum $\mathbf{k}'$, and integrate over frequency ω. We restrict ourselves to lowest-order terms in the fluctuations, i.e., terms of order of $(a/l)^2$.

To determine the spectrum of the long-wave length and low-frequency excitations, we expand function that enter into the expression for the susceptibility (72) in low frequencies $\Omega(|\Omega|\ll\overline{\Delta})$ and the quasimomenta $\boldsymbol{\delta}=\mathbf{q}-\mathbf{Q}(\delta a \ll 1, \delta a \ll \overline{\Delta}/t)$. We also drop corrections generated by terms of order $(a/l)^2(\Omega/\overline{\Delta})^2$ or $(a/l)^2(\delta a)^2$. Omitting some lengthy algebraic transformations, we write the series expansions for the irreducible components of the responses $\overline{\chi}_0^{+-}(\mathbf{q},\Omega)$ and $\overline{\chi}_{\mathbf{Q}}^{+-}(\mathbf{q},\Omega)$ for the one-dimensional case:

$$\overline{\chi}_0^{+-}(0,0)=\chi, \tag{75}$$

$$\overline{\chi}_0^{+-}(\mathbf{Q}+\delta,\Omega)=\frac{1}{U}+\alpha\left(\frac{\omega}{2\Delta}\right)^2-B\left(\frac{\delta a}{2\Delta}\right)^2+\left(\frac{a}{l}\right)^2\gamma, \tag{76}$$

$$\overline{\chi}_{\mathbf{Q}}^{+-}(0,\Omega)=\overline{\chi}_{\mathbf{Q}}^{+-}(\mathbf{Q},\Omega)=\frac{\Omega}{2\Delta}\,\alpha, \tag{77}$$

where

$$\chi=\frac{1}{2N}\sum_{\mathbf{k}}\ \frac{1}{E_{\mathbf{k}}^{\uparrow}+E_{\mathbf{k}}^{\downarrow}}\left(1-\frac{\varepsilon_{\mathbf{k}}^2+I^2-\Delta^2}{E_{\mathbf{k}}^{\uparrow}E_{\mathbf{k}}^{\downarrow}}\right), \tag{78}$$

$$\alpha=\frac{1}{2N}\sum_{\mathbf{k}}\ \frac{2\Delta^2}{E_{\mathbf{k}}^{\uparrow}E_{\mathbf{k}}^{\downarrow}(E_{\mathbf{k}}^{\uparrow}+E_{\mathbf{k}}^{\downarrow})}\left(1-\frac{4I^2}{(E_{\mathbf{k}}^{\uparrow}+E_{\mathbf{k}}^{\downarrow})^2}\right), \tag{79}$$

$$B=\frac{1}{2N}\sum_{\mathbf{k}}\ \frac{2(\varepsilon_{\mathbf{k}})^2a^{-2}}{E_{\mathbf{k}}^{\uparrow}E_{\mathbf{k}}^{\downarrow}(E_{\mathbf{k}}^{\uparrow}+E_{\mathbf{k}}^{\downarrow})}\times\left[1-\frac{4I^2\varepsilon_{\mathbf{k}}^2}{E_{\mathbf{k}}^{\uparrow}E_{\mathbf{k}}^{\downarrow}}\left(\frac{1}{E_{\mathbf{k}}^{\uparrow}E_{\mathbf{k}}^{\downarrow}}+\frac{1}{(E_{\mathbf{k}}^{\uparrow}+E_{\mathbf{k}}^{l})^2}\right)\right], \tag{80}$$

$$\gamma=\frac{1}{2N}\sum_{\mathbf{k}}\ \frac{(\varepsilon_{\mathbf{k}})^2a^{-2}}{16\Delta^2}\left[\frac{\varepsilon_{\mathbf{k}}^2}{I\Delta}\left(\frac{1}{E_{\mathbf{k}}^{\uparrow}}-\frac{1}{E_{\mathbf{k}}^{\downarrow}}\right)+\frac{1}{E_{\mathbf{k}}^{\uparrow}}+\frac{1}{E_{\mathbf{k}}^{\downarrow}}-I\left(\frac{I+\Delta}{E_{\mathbf{k}}^{\uparrow 3}}+\frac{I-\Delta}{E_{\mathbf{k}}^{\downarrow 3}}\right)\right]. \tag{81}$$

In all of these expressions we assume $\overline{\Delta}\approx\Delta$ since we need only the second order in the a/l expansion. Note that the self-consistency equation (64) is taken into account to obtain (75)–(81), but only after this we can write $\overline{\Delta}\approx\Delta$, where Δ is given by Eq. (12).

Inserting the expressions (75)–(81) into Eq. (72)–(74), we obtain the dynamical susceptibility in the low-frequency limit:

$$\overline{\chi}_{\mathbf{qq}'}^{+-}(\Omega)=\delta_{\mathbf{qq}'}/\mathrm{Det}(\delta,\Omega), \tag{82}$$

$$\mathrm{Det}(\delta,\Omega)=-U^2\left[\alpha\left(\frac{\Omega}{2\Delta}\right)^2\left(1+\frac{\alpha}{1/U-\chi}\right)-B\left(\frac{\delta a}{2}\right)^2+\frac{\gamma a^2}{l^2}\right]. \tag{83}$$

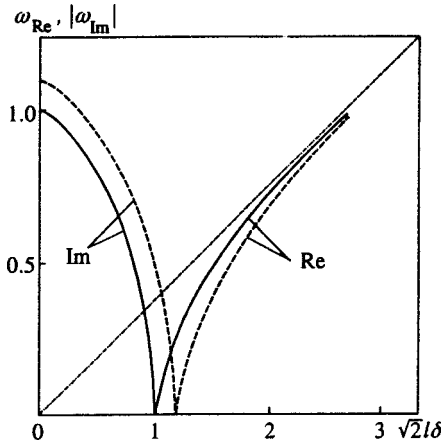

FIG. 6. Paramagnon spectrum $\omega_{\mathrm{Re}}=(\Omega_{\mathrm{Re}}/2t)(l/a)$, $\omega_{\mathrm{Im}}=(2\Omega_{\mathrm{Im}}/t)(l/a)$ for $I=0$ (solid line) and $I\neq 0$ (dashed line).

The poles of the response (82) determine the paramagnon spectrum $\Omega(\delta)$ in the short-range-order phase. The equation $\mathrm{Det}(\delta,\Omega)=0$ can only be solved numerically, so we consider the case of a weakly-dimerized chain $(I\ll\Delta)$ for qualitative analysis, where

$$\alpha=\chi=\Delta^2\varphi_3-\frac{5\Delta^2I^2}{2}\,(\varphi_5-\Delta^2\varphi_7), \tag{84}$$

$$B=-\varphi_1+[(2t)^2+\Delta^2]\varphi_3+\frac{I^2}{2}\{13\varphi_3-15\Delta^2\varphi_5+[(2t)^2+\Delta^2][15\Delta^2\varphi_7-13\varphi_5]\}, \tag{85}$$

$$\frac{1}{U}=\varphi_1-\frac{3I^2}{2}\,(\varphi_3-\Delta^2\varphi_5), \tag{86}$$

$$8\gamma=-\varphi_1+[(2t)^2+\Delta^2]\varphi_3+\frac{I^2}{2}[-\varphi_3-5\Delta^2\varphi_5+[(2t)^2+\Delta^2][\varphi_5+5\Delta^2\varphi_7]], \tag{87}$$

$$\varphi_n(\Delta)=\frac{1}{2N}\sum_{\mathbf{k}}\ \frac{1}{(\varepsilon_{\mathbf{k}}^2+\Delta^2)^{n+2}}. \tag{88}$$

Finally, after some calculations, we obtain the excitation spectrum of paramagnons, taking into account the dependence $\Delta(I)$ from Eq. (2). For $U\gg t$,

$$\Omega=\frac{(2t)^2}{U}\left[(a\delta)^2\left(1-2\left(\frac{2t}{U}\right)^2+\frac{8I^2}{U^2}\right)-\frac{1}{2}\left(\frac{a}{l}\right)^2\left(1-2\left(\frac{2t}{U}\right)^2+\frac{16I^2}{U^2}\right)\right]^{1/2}; \tag{89}$$

for $U\ll t$,

$$\Omega=2t\left[(a\delta)^2\left(1-\frac{U}{2\pi t}+\frac{1}{3}\left(\frac{I}{\Delta_0}\right)^2\frac{U}{2\pi t}\right)-\frac{1}{2}\left(\frac{a}{l}\right)^2\left(1-\frac{U}{2\pi t}+\frac{1}{3}\left(\frac{I}{\Delta_0}\right)^2\left(6-\frac{5U}{2\pi t}\right)\right)\right]^{1/2}. \tag{90}$$

Note that in the paramagnetic phase with finite correlation length, the collective excitations (89) and (90) have a very different character for $\delta<\delta_c$ and $\delta>\delta_c$, where δ_c is the critical quasimomentum (see Fig. 6). For $U\ll t$ (weak-coupling limit) the value of δ_c is determined qualitatively by the equation

$$\sqrt{2}l\delta_c\approx 1+\left(\frac{I}{\Delta_0}\right)^2, \quad I\ll\Delta_0. \tag{91}$$

For $t\ll I\ll U$ (strong-coupling limit), the equation for δ_c has the form

$$\sqrt{2}l\delta_c\approx 1+\left(\frac{I}{U}\right)^2. \tag{92}$$

When $\delta>\delta_c$, a weakly decaying spin-wave mode exists with a quasi-Goldstone spectrum $\Omega\approx v\delta$ for $(\delta l)^2>1$, where v is the spin-wave velocity. The decay of paramagnons results from thermal excitations of electrons through the antiferromagnetic gap ($\mathrm{Im}\,\Omega\sim\exp(-Eg/2T)$). Here we completely neglect this effect. In the region of strong magnetic disorder ($\delta<\delta_c$), a diffusive mode exists and the frequency Ω is imaginary. The effect of the staggered potential (at least in the case $I\ll\Delta$) on paramagnon dynamics amount to an increase in the critical quasimomentum δ_c, and in the spin-wave velocity v as compared with the case $I=0$ (see Ref. 9). The staggered potential also leads to an increase in diffusion frequency. In the case $\delta=0$, Eqs. (89) and (90) yield

$$i\Omega(\delta=0)=-\Omega_{\mathrm{Im}}\approx\frac{(2t)^2}{U}\frac{a}{l}\frac{1}{\sqrt{2}}\left(1+\frac{8I^2}{U^2}\right), \quad t\ll I\ll U, \tag{93}$$

$$i\Omega(\delta=0)=-\Omega_{\mathrm{Im}}\approx\frac{2t}{\sqrt{2}}\frac{a}{l}\left(1+\frac{I^2}{\Delta_0^2}\right), \quad U\ll t, \quad I\ll\Delta_0. \tag{94}$$

Note also that the decrease in $l(T)$ in Eq. (94) (see Sec. 4) is very small ($\sim I^2/U$) as compared with the term I^2/Δ_0^2, since $U\gg\Delta_0$ in the weak-coupling limit.

Our analysis confirms the tendency of Ω_{Im} and δ_c to increase for all I in the stability region of the phase diagram $\Delta(I)$.

8. CONCLUSION

In the SDW model with a staggered potential, we have shown that short-range antiferromagnetic order exists in this system far above the Néel point. We have calculated the dependence of the SDW amplitude on correlation length $l(T)$ and the staggered potential I. We have also obtained the renormalization of the Néel temperature T_N and effective exchange integral J on $l(T)$ and I, which both increase with $l(T)^{-1}$ and I in the weak-coupling regime.

In the short-range-order regime, the dielectric type of electron spectrum is preserved and the singularities in the density of states are smeared out. There exist twelve branches of this spectrum $\varepsilon(I,l(T))$, and the energy gap E_g narrows with increasing I and $l(T)^{-1}$. Finally, we have obtained parameters of the magnon spectrum that depend on I and $l(T)$.

It would be very interesting to extend our analysis to other models of itinerant magnets, for example the $t-J$ model in the weak-coupling regime.[14]

We gratefully acknowledge Dr. S. Caprara for valuable discussions and corrections to the manuscript.

This research was partially supported by the Russian Fund for Fundamental Research (Grant No. 96-02-19693).

*)E-mail: dunin@pliv.kial.su

[1] T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism*, Springer, Berlin (1985).
[2] V. Korenman and J. L. Murray, and R. E. Prange, Phys. Rev. B **16**, 4032 (1977); R. E. Prange and V. Korenman, Phys. Rev. B **19**, 4691 (1979).
[3] H. Capellmann and R. E. Prange, Phys. Rev. B **23**, 4709 (1981).
[4] V. Men'shov and V. Tugushev, Zh. Éksp. Teor. Fiz. **102**, 1637 (1992) [Sov. Phys. JETP **75**, 885 (1992)].
[5] V. Tugushev, in *Electronic Phase Transitions*, ed. by W. Hanke and Yu. Kopaev, Elsevier, North-Holland (1992).
[6] B. Halperin and M. Rice, Solid State Phys. **21**, 115 (1968); T. Rice, Phys. Rev. B **2**, 3619 (1970).
[7] V. Tugushev, M. Avignon, and S. Caprara, Phys. Rev. B **54**, 5466 (1996).
[8] M. Lannoo and G. Allen, Solid State Commun. **47**, 153 (1983).
[9] M. Avignon, V. Men'shov, and V. Tugushev, Europhys. Lett. **34**, 373 (1996).
[10] A. I. Akhieser, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin waves*, Wiley, New York (1968).
[11] V. G. Bar'yakhtar, V. N. Krivoruchko, and D. A. Yablonskii, *Green Functions in the Theory of Magnetism*, Naukova Dumka, Kiev (1984).
[12] H. Capellmann and V. Vieira, Phys. Rev. B **25**, 3333 (1982).
[13] V. Men'shov and V. Tugushev, Zh. Éksp. Teor. Fiz. **108**, 303 (1995) [JETP **81**, 163 (1995)].
[14] M. Avignon and V. Tugushev, Phys. Lett. A **209**, 198 (1995).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Elastic properties of ultrahard fullerites

V. D. Blank*)

Research Center for Superhard Materials, Ministry of Science and Technology of the Russian Federation, 142092 Troitsk, Moscow Region, Russia; Institute of Spectroscopy, Russian Academy of Sciences, 142092, Troitsk, Moscow Region, Russia

V. M. Levin

Center of Acoustic Microscopy, Institute of Biochemical Physics, Russian Academy of Sciences, 117334 Moscow, Russia

V. M. Prokhorov, S. G. Buga, and G. A. Dubitskiĭ

Research Center for Superhard Materials, Ministry of Science and Technology of the Russian Federation, 142092 Troitsk, Moscow Region, Russia

N. R. Serebryanaya

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 6 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1365–1374 (October 1998)

The elastic properties of C_{60} fullerite samples synthesized under pressure $P=13.0$ GPa at high temperatures were investigated using acoustic microscopy. The velocities of longitudinal ($c_L=17-26$ km/s) and transverse ($c_T=7.2-9.6$ km/s) elastic waves in the samples were measured. It was established that the longitudinal sound velocity of ultrahard fullerites is higher than that of any other known solid. The bulk modulus of these ultrahard samples is higher than that of diamond and reaches a value greater than 1 TPa. The high bulk modulus, the relatively large shear moduli, and the substantial Poisson ratio indicate that the structure of the ultrahard fullerites is fundamentally different from that of diamond. © *1998 American Institute of Physics.* [S1063-7761(98)01410-3]

It has been established during the last few years that under high pressures and temperatures C_{60} fullerene molecules form a large number of previously unknown phases, extending to super- and ultrahard fullerites, obtained from the initial fcc fullerite under high pressures $P\geqslant 18$ GPa at room temperature[1–3] and $P\geqslant 8$ GPa at high temperatures $T\geqslant 700$ K.[4–12] The phases obtained in this manner remain metastable under normal conditions, forming new allotropic forms of carbon. We were the first to establish that the three-dimensionally polymerized fullerites form a new class of superhard materials.[6] The actual structure of these materials is distinguished by great diversity as a function of the synthesis conditions—from crystalline phases of various symmetry to disordered, x-ray amorphous modifications. Figure 1 shows the $P-T$ diagram of the synthesis of super and ultrahard structures from the initial fcc C_{60}. This diagram was refined on the basis of the latest data on the state of carbon phases.[8–10] Ultrahard fullerites as fundamentally new carbon materials are studied in Refs. 11 and 12 .

The anomalous hardness of fullerites (50–300 GPa[4–9,13,14]) makes it of great interest to investigate their elastic characteristics. The elastic properties of the initial crystalline C_{60} have been investigated quite well, including measurements of the complete matrix of the elastic moduli by the ultrasonic method.[15] Just as all molecular crystals, crystalline C_{60} is characterized by low sound velocities: $c_L=3.0-3.4$ km/s for longitudinal waves and $c_T=1.6-2.0$ km/s for transverse waves. In the present work the velocities of longitudinal and transverse acoustic waves were measured and the elastic moduli of samples of super- and ultrahard fullerites were determined for the first time. The measurements show that these materials are characterized by uniquely high velocities of longitudinal elastic waves and a wide range of these values from 11 km/s to 26 km/s, depending on the structure of the materials, which is determined by the synthesis conditions. The value $c_L=26.0$ km/s measured for one of the fullerite phases is a record; it is almost 20% higher than the velocity of longitudinal waves in graphite along the atomic layers ($c_L=21.6$ km/s— until this work the highest value among known substances[16]) and 40% higher than the corresponding velocity in diamond ($c_{L\max}=18.6$ km/s (Ref. 17)). The velocities of transverse waves in solid fullerite phases are also high, with the values of c_T ranging from 7 km/s to 9.7 km/s. Nonetheless, they are lower than in diamond ($c_T=11.6-12.8$ km/s (Ref. 17)), which remain the highest among known substances as before. Data on the densities and sound velocities have made it possible to determine the complete set of elastic characteristics of ultrahard fullerites—the bulk modulus K, the shear modulus G, Young's modulus E, and the Poisson ratio σ—to determine the characteristic features of their elastic properties, and to

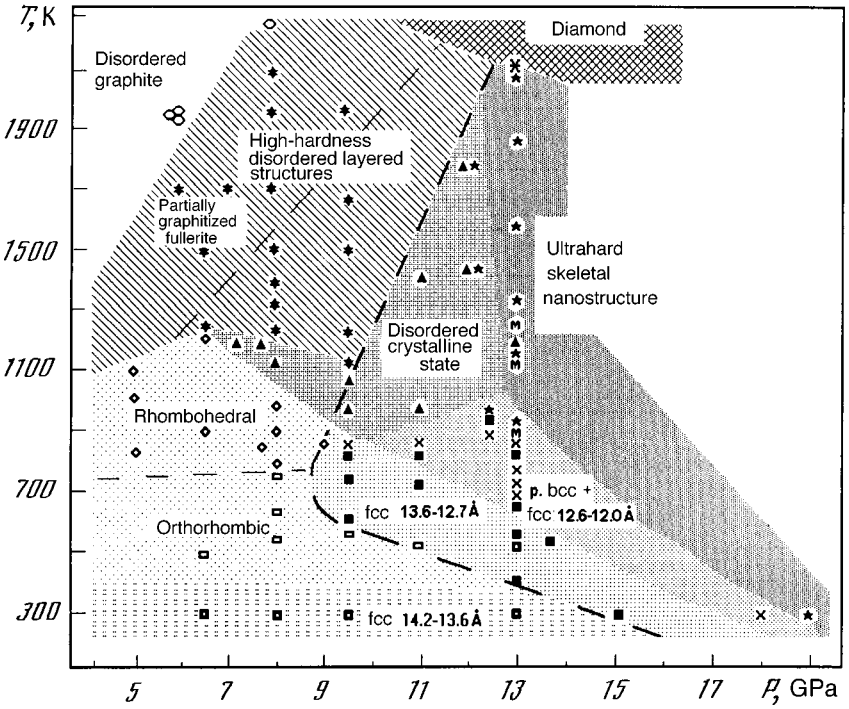

FIG. 1. $P-T$ diagram of the synthesis of metastable phases of carbon from C_{60} fullerite. The thick dashed line separates the hard and ''soft'' states (for $T<1000$ K) and semimetallic and semiconductor states (for $T>1000$ K).

make an attempt to relate the observed features with the structure of the materials.

1. FORMULATION OF THE PROBLEM AND DESCRIPTION OF THE METHOD

The standard methods for measuring sound velocities are difficult to use for solid fullerite phases because of the small sizes and considerable heterogeneity of the samples. For this reason, we employed a modification of the echo-pulse method, using the focusing system of an acoustic microscope. The sample was placed in the region of the waist of the focused ultrasonic beam, produced in an immersion liquid by an acoustic lens (Fig. 2). We employed ultrashort probe pulses, which made it possible to detect individual signals due to the reflection of a pulse from the sample walls.[18,19] Measurements of the time intervals between the echo pulses made it possible to calculate the velocities of the sound waves in a sample from the known sample thickness d.

The method requires long-focus systems with small apertures $2\theta_m$. In this case the probe beam can be treated as a narrow column of radiation with an almost planar wave front. Echo signals from the sample are formed by the paraxial rays propagating in the sample at small angles with respect to the beam axis. The path length traversed by such rays between two successive interactions with the boundaries of the sample equals, to adequate accuracy, the sample thickness d. Despite the small angular apertures, beams focused in solid samples with high sound velocities excite not only longitudinal but also transverse waves on account of mode conversion in the case of oblique incidence of a wave on an interface. As a result, the output signal of the acoustic focusing system includes signals corresponding to reflection of both longitudinal and transverse waves.[19]

The measurements are performed on the reliable identification of the echo signals. Here not only the distance between the signals but also their polarity are important. The base signal B, due to reflection of the incident beam from the front boundary of the sample, is always present in the oscillograms. When the focal point of the lens is shifted into the interior of the sample, an L pulse due to single passage of a pulse of longitudinal waves through the sample and reverse passage back through the sample after reflection from the back wall appears in the echogram. In our measurements the

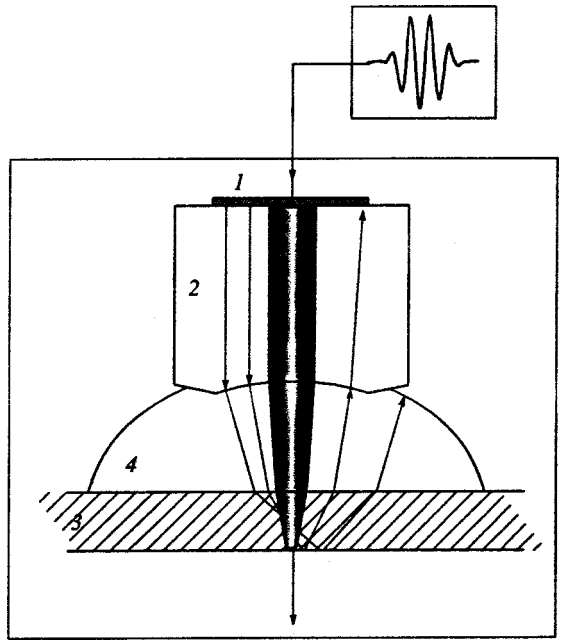

FIG. 2. Diagram of the acoustic microscope: *1* — transducer; *2* — wave guide with a spherical refracting surface; *3* — sample; *4* — immersion liquid.

polarity of the L pulse must be opposite to that of the B pulse. The time interval τ_L between the echo pulses is determined by the velocity c_L of the longitudinal ultrasonic wave in the sample and by the sample thickness. The measurement of this interval makes it possible to calculate the velocity of longitudinal sound as $c_L = 2d/\tau_L$.

For certain positions of the lens an LT signal due to successive mode conversion on the front and back boundaries of the sample is observed. A longitudinal wave in the liquid is converted on the front surface of the sample into a transverse mode T, reaches the back surface of the sample, and as a result of conversion there a longitudinal wave forms once again and reaches the front face of the sample, passing through it into the liquid. The LT signal can be formed by an alternative path: A longitudinal wave enters the sample and is converted on the back surface into a transverse wave, which reaches the front surface of the sample and once again is converted here into a longitudinal wave in the liquid. The delay of the echo signal is the same in both cases, and the LT pulse itself is inverted with respect to the main B pulses. The LT signal is used to determine the velocity of transverse sound from the relation $c_T = d/(\tau_{LT} - 0.5\tau_L)$, where τ_{LT} is the time interval between the LT pulse and the main B signal.

In principle, it is possible to observe the weaker T signal that arises as a result of the double conversion of waves on the front boundary of the sample: The incident signal enters the sample in the form of a T wave, which is reflected by the back face of the sample and is converted into a longitudinal wave in the liquid on the front boundary of the sample. The time delay of the T signal with respect to the B signal is determined by twice the propagation time of a pulse of transverse waves through the sample: $\tau_T = 2d/c_T$. The T signal, if observed, has the same sign as the main B signal.

2. SAMPLES

Ultrahard fullerite samples were synthesized under pressure $P = 13$ GPa at different temperatures ranging from 1500 K to 1870 K. The procedure for synthesizing super- and ultrahard fullerites is described in Ref. 6 . Pressure calibration of the synthesis chamber was done with respect to the phase transition in Pb ($P = 13$ GPa). Sample No. 1 was synthesized at $P = 12.5$ GPa and $T = 1000$ K. The size of the synthesized samples reached $\oslash 3 \times 4$ mm^3 and their mass reached 80 mg. The density ρ of the samples was measured by the method of submersion in liquids of different density obtained by mixing diiodomethane ($\rho = 3.32$ g/cm^3) with acetone. The structure of the samples was investigated by x-ray diffraction and Raman scattering spectroscopy.[7,8] Sample No. 1 possesses a disordered structure with residues of 3D-polymerized crystal phases of C_{60}. No traces of crystal phases were observed in the structure of samples Nos. 2–4. The x-ray diffraction patterns and Raman scattering spectra of the samples are different from those of other known superhard amorphous carbon materials.

For the acoustic measurements, the samples were glued inside openings in plane-parallel steel rings and ground with diamond pastes in a manner so that the working surfaces of the samples also became plane-parallel and the thickness of the ground samples was the same as that of the steel rings ($1 < d < 2$ mm). The surfaces of ultrahard samples can be ground with diamond pastes but cannot be polished, because they are harder than diamond. After grinding the surfaces of all samples possessed a sparse network of shallow (< 1 μm) and narrow (< 25 μm) grooves; their dimensions are small compared with the size of the ultrasonic spot on the sample surface. The presence of such defects had virtually no effect on the output signal of the acoustic microscope. Defects in the volume of the samples could scatter a portion of the energy of the ultrasonic beam, but they could not produce the ''false sample bottom'' effect.

3. MEASUREMENTS

The measurements were performed in a WFPAM pulsed scanning acoustic microscope (working frequency $f = 50$ MHz, frequency band $\Delta f = 20$ MHz). The ultrasonic beam was introduced into the sample volume through an immersion liquid (water or mercury) from a small-aperture acoustic lens ($\theta_m = 11°$). The duration of the ultrasonic probe pulses was ≈ 30 ns. The beam diameter on the irradiated sample surface was ≈ 100 μm. Together with recording oscillograms of the echo signals, the acoustic microscope was used in the B and C scanning modes. A B scan consists of a display of an oscillogram produced by scanning the acoustic lens along one direction (Figs. 3 and 4). The signal level at the scan points is conveyed by different degrees of brightness. B scans display both the topological features of the samples (geometry of the front and back surfaces, the presence of voids, cracks, and other defects) and the nonuniformity of the distribution of elastic properties (sound velocities) through the sample. We employed B scans to monitor the geometry of the sample and the uniformity of its elastic properties. Ordinary acoustic images (C scans) were used to visualize the sample surface and to select measurement points on it.

4. RESULTS

The main results of the measurements and calculations of the elastic characteristics of ultrahard fullerites along with the standard errors are presented in Table I. For comparison, the data on synthetic polycrystalline ''carbonado'' diamond (sample No. 5), which we obtained in the same measurements, as well as published data on single-crystal diamond, graphite, and the initial C_{60} fullerite are also displayed.

We shall discuss in greater detail the results obtained for specific samples.

Sample No. 1. The sample possessed only one ground surface because of the small sample size. For this reason, the sound velocities in the sample were measured by the method of $V(z)$ curves, which is used in acoustic microscopy to measure the velocity of surface acoustic waves.[20] The Rayleigh wave velocity c_R was measured using water as the immersion liquid; mercury was used as the immersion liquid to measure the velocity of longitudinal sound (skimming wave). The results are $c_L = 17.0 \pm 0.9$ km/s and $c_R = 8.6 \pm 0.5$ km/s. It should be emphasized that because of the

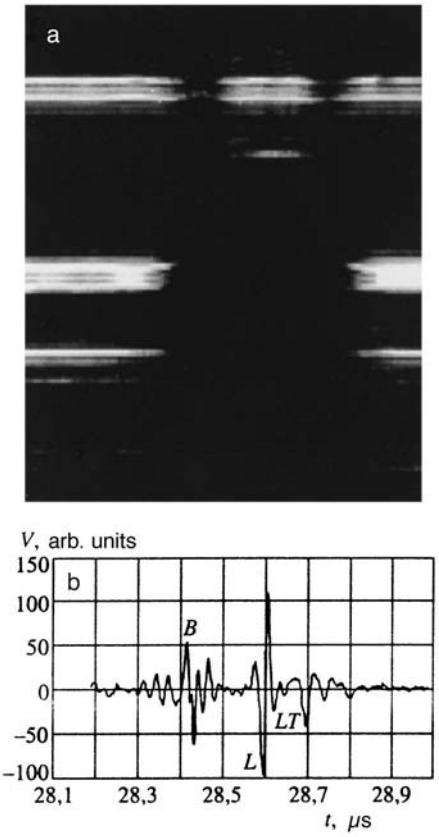

FIG. 3. *B* scan (a) and oscillogram of the echo-signals of the acoustic microscope (b) for sample No. 3. In the *B* scan on each side of the image there are echoes reflected from the steel holder; upper *B* is caused by reflection at the face, two lower ones are due to reflection from the bottom with mode conversion (*LT* signal) and without it (*L* signal). The echoses caused by reflection from the specimen are seen in the central region of the *B* scan (only the basic signal *B* and the signal *L* due to the round-trip of the longitudinal wave pulse through the speciman).

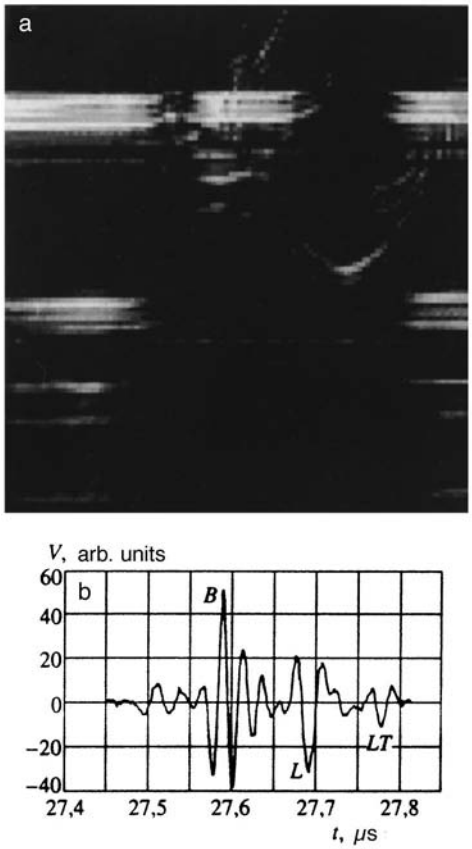

FIG. 4. *B* scan (a) and oscillogram of the measuring signal (b) of an acoustic microscope for sample No. 4.

small sample sizes and surface defectiveness, the corresponding $V(z)$ curves contained only several damped oscillations. For this reason, the results for sample No. 1 are approximate.

Sample No. 2. A series of measurements was performed with the ultrasonic beam focused on different points of the sample surface. At all points values of τ_L and τ_{LT} close to one another were obtained: $\tau_L=(130\pm 7)$ ns and $\tau_{LT}=(219\pm 6)$ ns. This attests to the fact that the elastic properties of the sample are of adequate uniformity. For sample thickness $d=1.11\pm 0.01$ mm the measured time intervals correspond to velocities $c_L=17.0\pm 0.9$ km/s for longitudinal waves and $c_T=7.2\pm 0.4$ km/s for transverse waves.

Sample No. 3. The sample was quite uniform with respect to all elastic properties. This is demonstrated by the *B* scan presented in Fig. 3a: The *L* signal reflected from the sample bottom consists of an even line parallel to the main *B* signal. To determine c_L a series of measurements was performed at different points, giving $\tau_L=(177\pm 7)$ ns. An acoustic transformer (a steel hemisphere, which decreased the refraction and reflection losses but did not change the structure of the ultrasonic beam entering the sample) was used to obtain an oscillogram containing three peaks, *B*, *L*, and *LT* (Fig. 3b), separated by intervals $\tau_L=(180\pm 10)$ ns and $\tau_{LT}=(280\pm 10)$ ns. For sample thickness $d=1.66\pm 0.01$ mm the velocities are $c_L=18.4\pm 1.0$ km/s and $c_T=8.7\pm 0.7$ km/s.

Sample No. 4. The *B* scan shown in Fig. 4a shows that the sample is divided into at least two parts with different values of the longitudinal ultrasonic wave velocity—the signal due to the pulse reflected from the sample bottom is located at a different distance from the main signal in the left- and right-hand parts of the sample. This circumstance could be due only to heterogeneity of the elastic properties of the sample, it cannot be due to defects of the sample shape or nonparallelism of the faces. The latter was checked visually in an optical microscope and using an acoustic microscope with the sample inverted. Accordingly, the oscillograms measured in different parts of the sample give different values for the velocity of longitudinal waves. Some oscillograms give $\tau_L=(130\pm 10)$ ns, which corresponds to $c_L=22.3\pm 1.7$ km/s. Figure 4b shows an oscillogram taken in a different, stiffer part of the sample. This oscillogram gives the time intervals $\tau_L=(103\pm 8)$ ns and $\tau_{LT}=(189\pm 10)$ ns. For sample thickness $d=1.335\pm 0.005$ mm they correspond to $c_L=26\pm 2$ km/s and $c_T=9.7\pm 1.0$ km/s. Inclusions of a third phase were also observed in the sample. This phase corresponds to the time interval $\tau_L=(190\pm 10)$ ns and longitudinal wave velocity $c_L=15.2\pm 0.8$ km/s.

Our data on the sound velocities and the measured values of the density (average) made it possible to calculate for each sample a complete set of elastic characteristics: the bulk

TABLE I.

No.	P, GPa/T, K	ρ, g/cm^3	c_L, km/s	c_T, km/s	K, GPa	G, GPa	E, GPa	σ
1	12.5/1000	3.10±0.05	17.0±0.9	9.4±0.5	540±60	270±30	700±170	0.28±0.04
2	13/1670	3.10	17.0±0.9	7.2±0.4	690±70	160±20	450±100	0.39±0.06
3	13/1770	3.30	18.4±1.0	8.7±0.7	790±70	250±40	680±180	0.36±0.04
4	13/1870	3.15	26±2	9.7±1.0	1700±250	300±60	850±300	0.42±0.08
5	Synt. diamond ''carbonado''	3.74	16±0.5	9.6±0.3	490±30	340±20	850±120	0.22±0.04
	Diamond[17]	3.51	17.5–18.6	11.6–12.8	417–586	354–535	884–1144	0.1
	Graphite[16]	2.27	4.0–21.6	0.3–14.0				
	Cryst. C_{60}[15]	1.68	3.0–3.4	1.6–2.0	10.8*	4.85*	12.6*	0.31*

*Calculation for a polycrystal.[15]

modulus K, the shear modulus G, Young's modulus E, and the Poisson ratio σ.

5. DISCUSSION

Ultrahard fullerites possess very high longitudinal sound velocities. Until the present work the highest values of c_L were observed in diamond, 17.5–18.6 km/s depending on direction,[17] and in crystalline graphite in a direction along the atomic planes: $c_L=21.6$ km/s from direct measurements of the elastic moduli by the ultrasonic method[16] and $c_L=22.6$ km/s, according to data on low-temperature specific heat and thermal expansion of natural graphite.[21] In the case of the fullerites investigated in the present work, longitudinal wave velocities close to the values in diamond (samples Nos. 1–3) or much higher values $c_L=22.3$ km/s and 26.0 km/s (sample No. 4) are observed. The latter value is 20% higher than the value of c_L measured in graphite and is the current record. The large range of variation of c_L both from sample to sample and in different parts of the same sample is interesting. Sample No. 4 was found to consist of regions with considerably different values of c_L; the variations reached 40%.

There is essentially no correlation between the density (averaged over the sample) and the observed values of c_L. For media with a local isotropy, this fact is quite interesting and possibly indicates partial ordering and pronounced elastic anisotropy at the atomic level, similarly to quasicrystalline graphite (HOPG)—parallel alignment of the atomic planes results in the highest known value of c_L in the direction of these planes. A variant with a nonuniform distribution of the local density in the sample is also possible.

The observed values of the transverse wave velocities are also found to be very high (7.2–9.6 km/s) in ultrahard fullerites, but their values are still 25–40% lower than in single-crystal diamond.

The elastic moduli K and G of ultrahard fullerites, calculated from the measured values of c_L, c_T, and ρ, and their relative values differ substantially from the same values in diamond. The behavior of the bulk modulus for samples synthesized at high temperatures is especially interesting; the value of the bulk modulus increases to teh highest known high values. This result requires further investigations, both theoretical and experimental. Elastic anisotropy or texturing of the sample could serve as a possible explanation. In this case, the anomalously high value of c_L will determined by one component (or a set of components) of the elastic constants tensor, similarly to the component C_{11} in crystalline graphite,[16] while the bulk modulus K of a temerpature-and-pressure treated sample (ceramic) should be calculated by averaging the components of this tensor according to existing models. On the other hand, the existing data on the pressure dependence of the compressibility of diamond and graphite show[22] that at high pressures the bulk modulus in graphite at first becomes equal to K in diamond (at $P\approx 10$ GPa) and then exceeds it by a factor of 1.5–2 at $P\approx 20$ GPa.

As a result of the large difference between the longitudinal and transverse sound velocities, for fullerites a Poisson ratio ($\sigma=0.36-0.42$) close to the maximum value $\sigma=0.5$ is typical, while the values of Young's modulus ($E=450-850$ GPa) are appreciably less than the value of the bulk modulus K. Sample No. 1 ($\sigma=0.28$; $E=660$ GPa) is an exception. The structure of this sample is different from that of the samples synthesized at higher pressures. The synthesis conditions for this sample are the boundary conditions for obtaining super- and ultrahard materials (Fig. 1).

The unusually high value of the bulk modulus, its value relative to Young's modulus, and the high value of σ all indicate that the structure of ultrahard fullerites is fundamentally different from that of diamond. The bulk modulus K for samples synthesized in the temperature range 1000–1670 K corresponds to the value 600–650 GPa calculated in Ref. 23 theoretically for the dense fcc structure of C_{60}. The existence of higher values of K has not been discussed in the literature, but further growth of K could correspond to a transition from a material consisting of 3D-polymerized C_{60} molecules to fullerite based on clusters of much smaller size. Specifically, as the 2D-polymerized {111} planes of fcc fullerite come closer to one another, C_{22-28} clusters (where all carbon atoms are in the sp^3 state, whereas some atoms of the C_{60} molecules retain sp^2 hybridization[13]) can form at the location of the tetrahedral voids in the fcc C_{60} lattice. As temperature increases further (for $P>12.5$ GPa) fragments of C_{60} molecules probably form a three-dimensional network with conjugate bonds between carbon atoms with different types of hybridization.

As follows from the relation for estimating the bulk modulus of spherical carbon molecules $K\approx 2h/3(S_{11}+S_{12})R$ (h is the interlayer spacing in graphite, R os the radius of the molecule, and S_{11} and S_{12} are components of

the compressibility tensor),[23] the value of K increases as the radius of the molecules (clusters) decreases. The type of interatomic bonds (sp^2 or sp^3) has virtually no effect on this quantity, since the corresponding components of the compressibility tensor, which appear in the denominator of this expression, are almost equal for graphite and diamond. Possible triple bonds (sp hybridization), which have a much higher energy compared with other types of bonds, could also strongly influence the elastic properties of fullerites. Additional investigations are required to establish the existence of such bonds.

Theoretical calculations have been performed of the elastic properties of hypothetical carbon polymers and composite materials based on nanotubes with Young's modulus of the order of 1300–1800 GPa and density 3.0–3.3 g/cm^3.[24,25] Calculations of the bulk modulus were not performed in these studies.

In Refs. 26 and 27 the possibility of obtaining materials based on C_{60} fullerite with hardness exceeding that of diamond is disputed. The results of investigations of the structure and some properties of samples of polymerized fullerites, obtained by heating C_{60} fullerite powder under a pressure of 12.5 GPa at temperatures from 20 to 900 °C, are presented as experimental proof. In Ref. 26 values of Young's modulus are presented for three samples. The maximum value $E=700$ GPa was obtained for a sample with density 3.15 g/cm^3 (the measurement method and the values of the ultrasonic velocities and the other moduli are not presented in this work). The maximum hardness $H_v=87$ GPa was obtained for a sample synthesized at temperature 900 °C ($\rho=3.0$ g/cm^3). The Raman scattering spectra of the samples presented in these works correspond to the Raman scattering spectra which we published earlier for fullerite samples synthesized under a pressure of 9.5 GPa.[6,7] The x-ray data of these works likewise correspond better to the data which we obtained on samples synthesized at $P=9.5-11$ GPa.[5,7,9] Comparison of the experimental data in Refs. 26 and 27 and the results of the present work (Fig. 1) shows that ultrahard fullerites were not obtained in Refs. 26 and 27.

6. CONCLUSIONS

The acoustic measurements performed in this work made it possible to determine the elastic moduli of ultrahard fullerites and to compare them with the corresponding values for single-crystal and polycrystal diamonds. It was shown that ultrahard fullerites are characterized by record high values of the bulk modulus, quite high shear moduli, and a substantial Poisson ratio. The elastic characteristics of ultrahard fullerites indicate that the structure of these materials is fundamentally different from that of diamond. The method developed for synthesizing the samples allows materials with uniquely high hardness and bulk modulus to be made from C_{60}.

This work was supported by the Russian Fund for Fundamental Research (Project 96-02-18015) and the Fund for Intellectual Collaboration (Project 95076).

*)E-mail: buga@ntcstm.msk.ru

[1] V. D. Blank, S. G. Buga, M. Yu. Popov *et al.*, Zh. Tekh. Fiz. **64**, 153 (1994) [Tech. Phys. **39**, 828 (1994)].
[2] V. D. Blank, M. Yu. Popov, S. G. Buga *et al.*, Phys. Lett. A **188**, 281 (1994).
[3] V. D. Blank, S. G. Buga, M. Yu. Popov *et al.*, New J. Chem. **19**, 253 (1995).
[4] V. D. Blank *et al.*, in *Abstracts of Invited Lectures and Contributed Papers of the 2nd Int. Workshop on ''Fullerenes and Atomic Clusters''*, St. Petersburg, Russia, June 19–24, 1995, pp. 31, 109, 132.
[5] V. D. Blank, B. A. Kulnitskiy, and Ye. V. Tatyanin, Phys. Lett. A **204**, 151 (1995).
[6] V. D. Blank, S. G. Buga, N. R. Serebryanaya *et al.*, Phys. Lett. A **205**, 208 (1995).
[7] V. D. Blank, S. G. Buga, G. A. Dubitsky *et al.*, Mol. Mater. **7**, 251 (1996).
[8] V. D. Blank, S. G. Buga, N. R. Serebryanaya *et al.*, Phys. Lett. A **220**, 149 (1996).
[9] V. D. Blank, S. G. Buga, N. R. Serebryanaya *et al.*, Appl. Phys. A **64**, 247 (1997).
[10] V. D. Blank, Ye. V. Tatyanin, and B. A. Kulnitskiy, Phys. Lett. A **225**, 121 (1997).
[11] M. Takano and A. Onodera, Curr. Opin. Solid State Mater. Sci. **2**, 166 (1997).
[12] H. Kuzmany, J. Winter, and B. Burger, Synth. Met. **85**, 1173 (1997).
[13] V. D. Blank, S. G. Buga, G. A. Dubitsky *et al.*, in *Molecular Nanostructures*, edited by H. Kuzmany *et al.*, World Scientific, 1997, p. 506.
[14] V. Blank, M. Popov, N. Lvova, K. Gogolinsky, and V. Reshetov, J. Mater. Res. **12**, 3109 (1997).
[15] N. P. Kobelev, R. K. Nikolaev, Ya. M. Soifer *et al.*, Chem. Phys. Lett. **276**, 263 (1997).
[16] O. Blackslee, J. Appl. Phys. **41**, 3373 (1970).
[17] I. S. Grigor'ev and E. Z. Melikhov [Eds.], *Physical Constants* [in Russian], Énergoatomizdat, Moscow, 1991, p. 149.
[18] R. S. Gilmore, K. G. Tam *et al.*, Philos. Trans. R. Soc. London, Ser. A **320**, 215 (1986).
[19] K. Maslov, Acoust. Imaging **20**, 245 (1993).
[20] A. Briggs, *Acoustic Microscopy*, Clarendon Press, Oxford, 1992.
[21] G. B. Spence, in *Proc. 5th Conf. on Carbon*, Pergamon, N. Y., 1961, Vol. 2, p. 531.
[22] R. W. Lynch and H. G. Drickamer, J. Chem. Phys. **44**, 181 (1966).
[23] R. S. Ruoff and A. L. Ruoff, Nature (London) **350**, 663 (1991).
[24] K. Tashiro, M. Kobayashi, and K. Yabuki, Synth. Met. **71**, 2101 (1995).
[25] S. B. Sinnott, O. A. Shenderova, C. T. White *et al.*, Carbon **36**, 1 (1998).
[26] V. V. Brazhkin, A. G. Lyapin, Yu. A. Antonov *et al.*, JETP Lett. **62**, 350 (1995).
[27] V. V. Brazhkin and A. G. Lyapin, Usp. Fiz. Nauk **166**, 893 (1996).

Translated by M. E. Alferieff

Semiclassical energy levels of electrons in metals with band degeneracy lines

G. P. Mikitik*) and Yu. V. Sharlaĭ

B. I. Verkin Institute for Low-Temperature Physics and Technology, National Academy of Sciences of Ukraine, 310164 Khar'kov, Ukraine

(Submitted 7 August 1997; resubmitted 27 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 1375–1392 (October 1998)

We show that in calculating the semiclassical energy levels of electrons in metals located in a magnetic field, one must determine whether or not the corresponding electron paths in the space of wave vectors **k** are attached to a band degeneracy line. Calculations in the two possible cases, i.e., with and without such attachment, differ by $|e|\hbar/2m^*c$, where e is the electron charge and m^* is the cyclotron mass of the electron. This shift in the energy levels is of a topological nature, and its existence depends neither on the specific form of the electron dispersion relation $\varepsilon(\mathbf{k})$ near the electron path nor on the shape or size of this path. The reason for this shift lies in the fact that the electron orbit is attached to the band degeneracy line, which is the line of singular points of the Bloch wave functions. In many respects this effect is similar to the Aharonov–Bohm effect if the band degeneracy line is considered an infinitely thin ''solenoid.'' This shift in energy levels should become apparent in studies of oscillation phenomena in metals. We give examples of metals in which the conditions for observing the shift is probably the most favorable. © *1998 American Institute of Physics.* [S1063-7761(98)01510-8]

1. INTRODUCTION

The degeneracy of electron energy states along the lines of the Brillouin zone is a fairly widespread phenomenon in crystals. In addition to the natural degeneracy, very often along the symmetry axes of the crystal there is an accidental contact of bands either along certain closed lines in the Brillouin zone or on curves that terminate at the zone's boundaries.[1] The possibility of such accidental degeneracy of bands for crystals with a center of inversion (most metals belong to this class of crystals, and only these are considered in this paper) follows from the assertion that if there is a point of accidental contact of bands on a symmetry axis of a crystal, and the band-to-band matrix element of the velocity operator is nonvanishing at that point, the line of degeneracy of the bands is sure to pass through that point.[1] This also leads to a statement about the abundance of such lines, since a frequent outcome of numerical calculations of the band structure of metals is the discovery of points of accidental degeneracy on a symmetry axis of the Brillouin zone (see, e.g., Ref. 2). Here, knowing the representations of the wave-vector group by which the wave functions of the corresponding bands are transformed, we can immediately establish whether or not the band-to-band matrix element of the velocity operator vanishes. Examples of crystals with lines of accidental band contact are aluminum, beryllium, magnesium, zinc, cadmium, and some others. Strictly speaking, accidental degeneracy of the bands exists only if spin–orbit coupling is neglected, i.e., allowing for this coupling lifts the degeneracy. However, if the coupling is weak, there is still strong attraction of the bands near the line of band contact in the absence of spin–orbit coupling. Thus, physically the idea of a band degeneracy line is justified even if spin–orbit coupling is taken into account, provided that the typical coupling-induced energy splitting Δ of the degenerate electron states is much less than E_0, the band gap in a crystal ($E_0 \sim 1-10$ eV).

The semiclassical quantization of the electron energy in a metal in the presence of an external magnetic field H has been examined by many researchers.[3–12] The semiclassical electron path in the space of wave vectors **k** is the curve along which the constant-energy surface $\varepsilon = \text{const}$ intersects a plane perpendicular to the magnetic field, whose direction we select as the z axis. The quantization rule for closed electron orbits with no self-intersection is

$$S(\varepsilon, k_z) = 2\pi \frac{|e|H}{\hbar c}\left(n + \gamma \pm \frac{gm^*}{4m}\right), \tag{1}$$

where S the area of the constant-energy surface in the plane $k_z = \text{const}$; n is a large positive integer; γ is a constant ($0 \leqslant \gamma < 1$); g is the electron g factor, equal to 2 if we neglect spin–orbit coupling (and Fermi-liquid effects); e and m are the electron charge and mass; and m^* is the electron's cyclotron mass. In this case, when in analyzing a physical effect we can neglect not only spin–orbit coupling but also electron spin, the last term on the right-hand side of Eq. (1) can be dropped, i.e., we set $g = 0$ (by definition, the g factor characterizes the energy splitting of the states of a particle in a magnetic field due to the particle's spin). If the semiclassical path in **k** space passes far from the singular points of the function $\varepsilon(\mathbf{k})$—the dispersion relation of an electron in the crystal—and nowhere near the path is the semiclassicality condition violated (e.g., the x and y projections of the electron velocity do not vanish), the constant γ has the universal value[5,13]

$$\gamma = 1/2, \tag{2}$$

which is commonly used in formulas that describe oscillatory effects in metals (the de Haas–van Alphen and Shubnikov–de Haas effects, among others).[14]

The foregoing constraints on a semiclassical path, under which Eq. (2) holds, can be given an instructive geometrical form. In addition to the natural requirement on the size of the cross-sectional area,

$$\frac{|e|H}{\hbar c S} \approx \frac{1}{n} \ll 1, \tag{3a}$$

the foregoing conditions amount to the requirement that all semiclassical orbits that pass too close to other paths in **k** space belonging to either the same or different bands be excluded. The condition that the paths not be too close to one another means that the probability of magnetic breakdown (both intraband and band-to-band) is low, i.e.,

$$\frac{|e|H}{\hbar c\, \delta S} \ll 1, \tag{3b}$$

where $\delta S \sim (\delta k)^2$, with δk the minimum distance between paths (or between segments of a single orbit) in the Brillouin zone, and δS the characteristic area of the ''neck'' separating the paths. At energies or values of k_z where (3b) breaks down, the set of semiclassical orbits in the space of wave vectors is close to altering its topology, and γ will depend heavily on ε and k_z (Refs. 6, 7, and 10). We do not consider such cases here and always assume that conditions (3a) and (3b) are met.

In the present paper we show that when electron spin can be completely neglected ($g=0$), we have, as an alternative to (2), another condition:

$$\gamma = 0, \tag{4}$$

provided that the electron path encloses either a line of accidental contact of two bands or a threefold symmetry axis on which there is natural degeneracy of the band under consideration and some other band. This is a topological result that depends neither on the specific form of the electron dispersion relation $\varepsilon(\mathbf{k})$ near the electron's semiclassical path nor the shape of the path or on how far it is from the line of band contact. It is related to the fact that this path is attached[15] (Fig. 1) to the band degeneracy line, which is the line of singular points of the Bloch electron wave functions (and the function $\varepsilon(\mathbf{k})$). Without such an attachment we have (2). When weak spin–orbit coupling is turned on and we allow for electron spin on those paths for which (4) holds, condition (2) is restored, but now the g factor differs substantially from 2:

$$g \approx 2 \pm 2\frac{m}{m^*}, \tag{5}$$

where the sign on the right-hand side is determined by the electron states on the line of band contact. Note that if we plug the second (purely orbital) term in this expression into (1), we obtain the same energy spectrum as in the simplest, spinless, case (bearing in mind that the values $\gamma=0$ and $\gamma=1$ are equivalent). Thus, although allowance for spin–orbit coupling generally lifts the degeneracy of bands on a line, the result (4) is stable when that coupling is turned on, as long as it is not too strong, $\Delta \ll E_0$ (note that the requisite weakness of spin–orbit coupling is equivalent to the condition under which the very notion of the band degeneracy line has meaning).

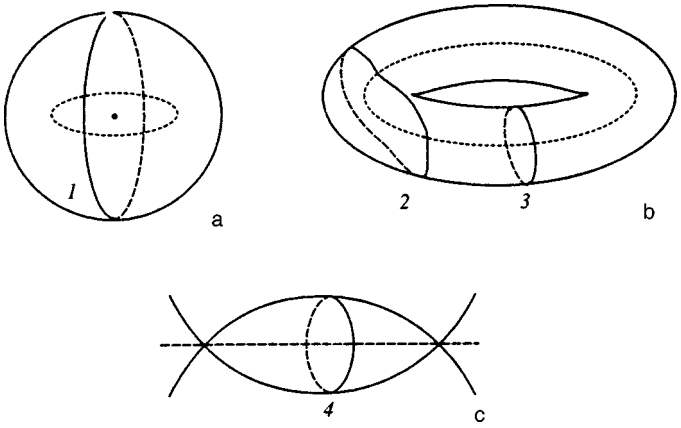

FIG. 1. Relative disposition of semiclassical paths (curves *1–4*) and band degeneracy lines (dashed curves) for a) simply connected, b) multiply connected, and c) self-intersecting Fermi surfaces. Paths *3* and *4* are attached to a band degeneracy line, while paths *1* and *2* are not.

In Sec. 2 we present equations that can be used to find γ and the g factor for any closed path without self-intersections that meets condition (3). These equations are used in Secs. 3 and 4 to analyze the case in which a metal has a band degeneracy line. To get to the heart of the problem, in Sec. 3 we examine the simplest case (spin–orbit coupling and electron spin are neglected). In Sec. 4 we take spin–orbit coupling and electron spin into account. The Conclusion is devoted to a discussion of possible experimental evidence for the results.

2. THE PARAMETER γ AND THE g FACTOR

We employ the following method to derive the energy quantization rule (1) and the equations for γ and the g factor. Using the appropriate Hamiltonian, we can find the electron wave function to second order in the small parameter (3a). We then impose the constraint that the wave function be single-valued (the phase change of the wave function over a closed semiclassical path can only be $2\pi n$), which yields the required results. The equations that make it possible to calculate γ and the g factor for closed orbits with and without allowance for spin–orbit coupling were derived in Ref. 9[1)] on the basis of the effective one-band Hamiltonian of a Bloch electron in a magnetic field.[16,17] What is important is that if condition (3) is met, this Hamiltonian can be used in the case of band degeneracy.[17] Below we repeat the results of Roth[9] and Blount[17] that will be needed in our analysis.

We expand the effective one-band Hamiltonian of a Bloch electron in a magnetic field in powers of H (see Ref. 17). To find γ and the g factor we need only keep the two first terms in the expansion. Then, in the **K** representation, the Hamiltonian is

$$\mathscr{H}=\varepsilon_0^{(s)}(\hat{\mathbf{k}})+\frac{e}{c}Hm_0^{(s)}(\hat{\mathbf{k}}), \tag{6}$$

where $\varepsilon_0(\mathbf{k})$ is the electron dispersion relation in the band being investigated, which hereafter we denote by the subscript 0. The ''S'' on the functions $\varepsilon_0(\hat{\mathbf{k}})$ and $m_0(\hat{\mathbf{k}})$ means that these functions are completely symmetrized in the components of the operator $\hat{\mathbf{k}}=\mathbf{K}-(e/\hbar c)\hat{\mathbf{A}}\cdot(i\partial/\partial\mathbf{K})$, where $\mathbf{A}(\mathbf{r})$ is the vector potential of the magnetic field. The function $m_0(\mathbf{k})$ can be written

$$m_0(\mathbf{k})=\mathbf{i}_z\cdot(\mathbf{v}\times\mathbf{\Omega})+\frac{\hbar}{2i}\sum_{l\neq 0}\frac{(v_x)_{0l}(v_y)_{l0}-(v_y)_{0l}(v_x)_{l0}}{\varepsilon_l(\mathbf{k})-\varepsilon_0(\mathbf{k})}, \tag{7}$$

where $\mathbf{i}_z$ is the unit vector in the direction of the magnetic field, $\mathbf{v}=(1/\hbar)(\partial\varepsilon_0/\partial\mathbf{k})$, $(\mathbf{v})_{0l}$ is the band-to-band matrix element of the velocity operator calculated at $\mathbf{k}$, and $\mathbf{\Omega}$ is the part of the position operator that is periodic in $\mathbf{k}$:

$$\mathbf{\Omega}(\mathbf{k})=i\int_v u^*_{\mathbf{k}0}(\mathbf{r})\frac{\partial}{\partial\mathbf{k}}u_{\mathbf{k}0}(\mathbf{r})\,d\mathbf{r}. \tag{8}$$

Here $u_{\mathbf{k}l}(\mathbf{r})$ is a periodic factor that is present in the Bloch wave function of an electron belonging to the lth band, and integration is over the volume v of a cell of the crystal lattice.

These equations can be used immediately if electron spin is neglected. If it is not, all electron states in a crystal with a center of inversion are doubly degenerate, and the band index now consists of the band index l and the spinor indices ρ ($\rho=1,2$). Accordingly, the periodic Bloch factor $u_{\mathbf{k}l}(\mathbf{r})$ becomes a periodic spinor $u_{\mathbf{k}l\rho}(\mathbf{r})$, and the quantities in (7) and (8) are two-dimensional matrices in ρ. Moreover, in (7) we must add a purely spin contribution to the Hamiltonian:

$$\frac{\hbar}{m}s_{\rho\rho'}=-\frac{\hbar}{2m}\int_v u^*_{\mathbf{k}0\rho}(\mathbf{r})\sigma_3 u_{\mathbf{k}0\rho'}(\mathbf{r})\,d\mathbf{r}, \tag{9}$$

where σ_3 is a Pauli matrix.

As already noted, $g=0$ in the simplest case, i.e., when electron spin can be completely neglected. Here, if (3a) and (3b) hold, semiclassical energy quantization leads to[9]

$$\gamma-\frac{1}{2}=-\frac{1}{2\pi}\oint_\Gamma\frac{m_0(\mathbf{k})}{v_\perp(\mathbf{k})}d\kappa, \tag{10}$$

where the integral is calculated along a closed semiclassical path Γ in $\mathbf{k}$ space, i.e., along the line of intersection of the surface $\varepsilon_0=\text{const}$ and the plane $k_z=\text{const}$ ($d\kappa$ is the element of arc length on this path), and $v_\perp$ is the absolute value of the projection of $\mathbf{v}$ on the (k_x,k_y) plane. Usually it is assumed that $m=0$ in a crystal with a center of inversion, so that $\gamma=1/2$. However, in Sec. 3 we show that generally, when a crystal has a band degeneracy line, the second term on the right-hand side of Eq. (6) is nonvanishing.

When electron spin is taken into account, m_0 is a two-dimensional matrix, which can be conveniently written as a linear combination of the Pauli spin matrices σ_i (see Ref. 18):

$$(m_0)_{\rho\rho'}=-\frac{\hbar}{4m}\sum_{i=1}^{3}G_{zi}\sigma_i, \tag{11}$$

where, generally speaking, the coefficients G_{zi} depend on $\mathbf{k}$ and are given by

$$\frac{\hbar}{4m}G_{zx}=-\operatorname{Re}(m_0)_{12},\quad \frac{\hbar}{4m}G_{zy}=\operatorname{Im}(m_0)_{12},$$

$$\frac{\hbar}{4m}G_{zz}=-(m_0)_{11}.$$

The fact that (11) contains no identity matrix follows from the behavior of the matrix m_0 under the transformation $U=(i\sigma_2)KI$ (see Ref. 18), where I, K, and $i\sigma_2K$ are the spatial inversion, complex conjugation, and time reversal operators. Here we have also taken $u_{\mathbf{k}l2}=Uu_{\mathbf{k}l1}$. The vanishing trace of matrix (11) yields $\gamma=1/2$, and m_0 contributes only to the g factor. The latter then becomes (see Eq. (64) of Ref. 9)

$$g=\frac{\hbar}{2\pi m^*}\oint_\Gamma\frac{d\kappa}{v_\perp}[G_{zz}+G_{zy}\operatorname{Im}\tau+G_{zx}\operatorname{Re}\tau], \tag{12}$$

where the cyclotron mass m^* is given by the well-known formula

$$m^*=\frac{\hbar^2}{2\pi}\frac{\partial S(\varepsilon,k_z)}{\partial\varepsilon}=\frac{\hbar}{2\pi}\oint_\Gamma\frac{d\kappa}{v_\perp}.$$

The complex function $\tau(\kappa)$ is defined on the semiclassical path Γ, and specifies the direction of the unit vector $\mathbf{e}$ in the spinor space,

$$\mathbf{e}=\frac{1}{\sqrt{1+|\tau|^2}}\begin{pmatrix}1\\ \tau\end{pmatrix},$$

which is a factor in the electron wave function ($\psi_{sc}\propto\times\exp(-i\hbar cS/eH)\,\mathbf{e}$). The function $\tau(\kappa)$ can be found by solving the equation

$$i\frac{4m}{\hbar}v_\perp\frac{d\tau}{d\kappa}+[G_-\tau^2+2G_{zz}\tau-G_+]=0 \tag{13}$$

(which can be derived from Eq. (52) of Ref. 9), where $G_\pm\equiv G_{zx}\pm iG_{zy}$. The boundary condition for Eq. (13) is

$$\tau(0)=\tau(\kappa_0), \tag{14}$$

where κ_0 is the length of a semiclassical path in $\mathbf{k}$ space.

It can be shown (although we will not do so here) that the complex Riccati equation (13) with boundary condition (14) has a solution for any κ_0. However, for arbitrary $G_{zi}(\mathbf{k})$ it is impossible to reduce the solution to quadratures, which means we cannot find an explicit expression for the g factor. In the special case in which all G_{zi} can be assumed constant,[18] the solutions of Eq. (13) with boundary condition (14) are

$$\tau=-\frac{G_{zz}\pm\sqrt{G_{zz}^2+G_+G_-}}{G_-}.$$

Inserting this expression into (12) yields $g^2=G_{zz}^2+G_+G_-$, which coincides with the result obtained in Ref. 18.

3. THE SIMPLEST CASE

In this section we use the results of Sec. 2 to determine the admissible values of γ in the simplest case in which spin–orbit coupling and electron spin are completely neglected. General considerations would seem to imply that crystals with an inversion center have $m_0=0$. Indeed, in the absence of spin, the operator U introduced in Sec. 2 is simply the product of the spatial inversion and complex conjugation operators.[19] This operator transforms an electron state with wave vector $\mathbf{k}$ into itself, so that a suitable choice of wave-function phases ensures that

$$U u_{\mathbf{k}l}(\mathbf{r}) = u_{\mathbf{k}l}(\mathbf{r}). \tag{15}$$

With such phases, all matrix elements of the velocity vector are real quantities and $\mathbf{\Omega}=0$.[19,20] Then (7) immediately yields $m_0=0$.

We now see how m_0 depends on the choice of phases of the Bloch functions. If these phases are changed in such a way that

$$u_{\mathbf{k}l} \to u'_{\mathbf{k}l} = u_{\mathbf{k}l}\ \exp[i\varphi_l(\mathbf{k})], \tag{16}$$

where $\varphi_l(\mathbf{k})$ are arbitrary smooth functions specified over the entire Brillouin zone, we have

$$\mathbf{v}_{ll'} \to \mathbf{v}'_{ll'} = \mathbf{v}_{ll'}\ \exp\{i[\varphi_{l'}(\mathbf{k}) - \varphi_l(\mathbf{k})]\},$$

and the sum over l in (7) remains equal to zero. The first term on the right-hand side of Eq. (7), which is the contribution the electron orbital angular momentum that is completely diagonal in the band numbers, is not invariant under (16). Indeed, according to (8), for such a transformation we have

$$\mathbf{\Omega} \to \mathbf{\Omega}' = \mathbf{\Omega} - \frac{\partial \varphi_0}{\partial \mathbf{k}}. \tag{17}$$

Hence, the new value of $\mathbf{\Omega}$ and the corresponding value of m_0 will generally be nonzero. However, such noninvariance of m_0 has no effect on measurable variables. To show that this is so for γ, we write (10) with allowance for the fact that the sum over l in (7) is zero. As a result we have

$$\gamma - \frac{1}{2} = -\frac{1}{2\pi} \oint_\Gamma \mathbf{\Omega}\cdot d\mathbf{k}, \tag{18}$$

where $d\mathbf{k} \equiv d\kappa\,(\mathbf{i}_z \times \mathbf{v})/v_\perp$, and integration is along a semiclassical path Γ (since the electron velocity in ordinary space is orthogonal to that in $\mathbf{k}$ space, the vector $d\mathbf{k}$ is directed along the orbit element in $\mathbf{k}$ space). The invariance of γ under (16) follows directly from (17) and (18).

Actually, the foregoing argument that m_0 can be made to vanish by properly selecting the phases of the wave functions is true only for a nondegenerate band. We now examine the case in which the crystal has a line along which the band with $l=0$ is in contact with another band, for example the one with $l=\lambda$. It turns out (see Ref. 20) that in this case it is impossible to choose the phases of the Bloch factors in such a way that (15) holds on the one hand, while on the other these factors are continuous in $\mathbf{k}$ in the Brillouin zone.

This claim can be made more specific. If (15) holds, there will be a surface in $\mathbf{k}$ space with an edge on which $u_{\mathbf{k}\lambda}$ and $u_{\mathbf{k}0}$ experience a discontinuity and $\mathbf{\Omega}$ is singular, and the edge coincides with the band degeneracy line. To show this, we select a point O on the band degeneracy line as the origin of $\mathbf{k}$ and write the Hamiltonian of an electron in the absence of a magnetic field in the Luttinger–Kohn representation,[21] i.e., we use the

$$e^{i\mathbf{k}\cdot\mathbf{r}} u_{0l}(\mathbf{r}),$$

as basis functions. These functions differ from the Bloch functions only in the factors $u_{\mathbf{k}l}$, which are evaluated at $\mathbf{k}=0$ (at which point the two sets of functions coincide). In this representation the Hamiltonian is not diagonal in l. So as not to overburden our discussion with technical details, only two bands will be taken into account by the Hamiltonian: $l=0$ and $l=\lambda$. The Hamiltonian then becomes

$$\hat{\mathscr{H}} = \begin{pmatrix} E_{00} & E_{0\lambda} \\ E^*_{0\lambda} & E_{\lambda\lambda} \end{pmatrix}, \tag{19}$$

where

$$E_{ij} = \hbar\mathbf{k}\cdot\mathbf{v}_{ij}(0) + \delta_{ij}\hbar^2\mathbf{k}^2/2m, \quad i,j = 0,\lambda,$$

where δ_{ij} is the Kronecker delta, and the $\mathbf{v}_{ij}(0)$ are the matrix elements of the velocity vector calculated at $\mathbf{k}=0$. We reckon the energy from the energy of band degeneracy at $\mathbf{k}=0$. The transformation matrix from the Luttinger–Kohn representation to the Bloch representation is

$$S_{ij}(\mathbf{k}) = \langle u_{0i} | u_{\mathbf{k}j} \rangle = \int_v u^*_{0i} u_{\mathbf{k}j}(\mathbf{r})\ d\mathbf{r}. \tag{20}$$

This unitary matrix diagonalizes the Hamiltonian (19), whose eigenvalues determine the electron dispersion relations in the Bloch representation for bands 0 and λ:

$$\varepsilon_{0,\lambda} = \tfrac{1}{2}(E_{00}+E_{\lambda\lambda}) \pm \sqrt{\tfrac{1}{2}(E_{00}-E_{\lambda\lambda})^2 + |E_{0\lambda}|^2}. \tag{21}$$

If (15) holds, then both the E_{ij} and $S_{ij}(\mathbf{k})$ will be real, which means that S will be an orthogonal matrix. For the Hamiltonian (19), the orthogonal matrix S can be calculated directly (the extent to which we are free to select S is discussed below):

$$S = \begin{pmatrix} \cos(\phi/2) & -\sin(\phi/2) \\ \sin(\phi/2) & \cos(\phi/2) \end{pmatrix}, \tag{22}$$

where

$$\sin\phi = E_{0\lambda}/\delta\varepsilon, \quad \cos\phi = E_{00} - E_{\lambda\lambda}/2\,\delta\varepsilon, \tag{23}$$

$$\delta\varepsilon = \sqrt{\tfrac{1}{4}(E_{00}-E_{\lambda\lambda})^2 + E^2_{0\lambda}},$$

with $-\pi \leqslant \phi \leqslant \pi$. Using this matrix, we can easily find the dependence of the Bloch factors $u_{\mathbf{k}l}$ on $\mathbf{k}$ since, according to (20)

$$u_{\mathbf{k}i} = (S')_{i0} u_{00} + (S')_{i\lambda} u_{0\lambda}, \tag{24}$$

where S' is the transpose of S.

We can now easily justify the foregoing statement. Through O we pass a plane that intersects the line of band

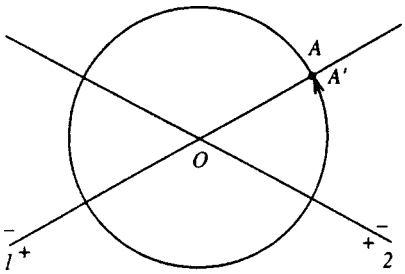

FIG. 2. Lines of intersection of the surfaces $E_{0\lambda}=0$ and $E_{00}-E_{\lambda\lambda}=0$ with a plane passing through O (straight lines *1* and *2*, respectively). The plus and minus sign on line *1* represent the sign of the product $\mathbf{k}\cdot\mathbf{v}_{0\lambda}(0)$ on either side of *1*, while the plus and minus on line *2* represent the sign of the product $\mathbf{k}\cdot[\mathbf{v}_{00}(0)-\mathbf{v}_{\lambda\lambda}(0)]$ on either side of *2*.

contact. In this plane there are two straight lines that pass through O, on one of which $E_{0\lambda}$ vanishes, while $E_{00}-E_{\lambda\lambda}$ vanishes on the other (Fig. 2). In the plane we consider a closed contour AA' that encircles O, and study the behavior of the Bloch factors as a point travels along the contour. From (23) we see that in this process ϕ varies from $-\pi$ to π. Here all the elements of S and hence all the $u_{\mathbf{k}i}$ are smooth functions of $\mathbf{k}$. However, according to (22) and (24), as a result of this process we obtain $u_{\mathbf{k}i}|_{A'}=-u_{\mathbf{k}i}|_{A}$.

Thus, on the ray OA the function $u_{\mathbf{k}i}(\mathbf{k})$ experiences a discontinuity. Clearly, there exists a surface of discontinuities of the functions $u_{\mathbf{k}0}$ and $u_{\mathbf{k}\lambda}$, while the ray is the line along which this surface intersects the plane we have selected.

Note that there still is a certain arbitrariness in calculating the matrix S, since for any $\mathbf{k}$ the signs of the eigenvectors of $\hat{\mathscr{H}}$ are undetermined. This corresponds to the possibility of replacing $u_{\mathbf{k}i}$ by $-u_{\mathbf{k}i}$ in a certain region of the Brillouin zone, which obviously does not violate (15). Such a transformation can be used to deform the discontinuity surface (for Eqs. (20) and (23) this means specifying a new range of admissible values of ϕ: $\phi_0\leqslant\phi\leqslant\phi_0+2\pi$, where ϕ_0 is a fixed angle). However, the very fact that the functions $u_{\mathbf{k}i}$ experience discontinuities when (15) holds is independent of the gauge of the wave functions (the edge of the discontinuity surface always coincides with the band degeneracy line). Everywhere that the functions $u_{\mathbf{k}i}$ are continuous we can apply the line of reasoning developed at the beginning of this section, and $\mathbf{\Omega}=0$. However, $\mathbf{\Omega}$ is singular on the discontinuity surface, as the definition (8) implies, and the integral in (18) is nonzero.

Strictly speaking, the two-band Hamiltonian (19) adequately describes electron states only near the band degeneracy line, as long as $|\varepsilon_0(\mathbf{k})-\varepsilon_\lambda(\mathbf{k})|\ll E_0$, where E_0 is the energy scale characterizing energy gaps in the electron spectrum of the crystal. To extend the foregoing claims to all values of $\mathbf{k}$, we must allow for other bands in the Luttinger–Kohn Hamiltonian. The matrix S can then be written as a product of two orthogonal matrices, S_1 and S_2. The first forces the E_{0l} and $E_{\lambda l}$ with $l\neq 0$, λ to zero, i.e., reduces the Hamiltonian to block form. Here the Hamiltonian block of interest will have the same form as (19), but now the E_{ij} $(i,j=0,\lambda)$ in it are complicated functions of $\mathbf{k}$. What is important, however, is that the E_{ij} and the elements of S_1 are smooth single-valued functions of the wave vector (on the band degeneracy line in particular),[19] and the $\mathbf{k}\cdot\mathbf{v}_{ij}(0)$ are simply the first terms in the series expansions of E_{ij}. The line of band contact is now the line of intersection of two surfaces,

$$E_{0\lambda}(\mathbf{k})=0 \tag{25}$$

and

$$E_{00}(\mathbf{k})-E_{\lambda\lambda}(\mathbf{k})=0. \tag{26}$$

S_1 is also a block matrix, and the part of interest can be described by the same equations (22) and (23). The foregoing conclusion about the existence of a surface of discontinuities of the functions $u_{\mathbf{k}i}$ is based on the fact that one circuit around the band degeneracy line changes ϕ by 2π. This property still holds in the general case, since it is not related to the explicit form of the E_{ij} but is determined only by the fact that the surfaces (25) and (26) intersect. This means that another assertion still holds, i.e., in the presence of a band degeneracy line, $\mathbf{\Omega}$ can vanish in the neighborhood of any point not belonging to this line, but not simultaneously at all the points on a closed curve that encircles the line. We also note that when there is no band degeneracy (the surfaces (25) and (26) do not intersect), following a closed contour changes ϕ by less than 2π, and it ultimately returns to its original value. In such a situation it is possible to make the functions $u_{\mathbf{k}i}$ continuous in $\mathbf{k}$ (provided that (15) holds), and $\mathbf{\Omega}=0$ over the entire Brillouin zone.

We now examine the case of natural degeneracy of two bands on a symmetry axis of the Brillouin zone. If it is a fourfold or sixfold symmetry axis, the series expansions of the left-hand sides of Eqs. (25) and (26) in powers of $\mathbf{k}$ contain no linear terms; one of the series then begins with k_1k_2, and the other with $k_1^2-k_2^2$ (k_3 along the symmetry axis).[19] Now each of the equations, (25) or (26), determines two surfaces that intersect along the straight line $k_1=k_2=0$. An analysis similar to the one carried out above shows that ϕ changes by 4π around a closed contour that encircles such an axis, which means that if (15) holds, we can choose the $u_{\mathbf{k}i}$ to be continuous functions of $\mathbf{k}$. Note that Eq. (21) imply that in this special case, as the wave vector moves away from the axis, the band energies split quadratically in $k_\perp=\sqrt{k_1^2+k_2^2}$. But if natural degeneracy occurs on a threefold symmetry axis, then, as in the case of accidental band contact, the series expansions of the left-hand sides of Eqs. (25) and (26) start with terms linear in $\mathbf{k}$ (Ref. 19), and all of the foregoing results still hold. One typical feature of this case is band splitting that is linear in $k_\perp$.

Since $\mathbf{\Omega}$ can vanish in the neighborhood of any point not located on the band degeneracy line by a suitable choice of the phases of the wave functions,

$$\operatorname{curl}\mathbf{\Omega}=0 \tag{27}$$

at such a point (the curl is calculated in $\mathbf{k}$ space). Here, due to the invariance of curl$\mathbf{\Omega}$ under the transformation (17), Eq. (27) actually holds for any selection of the phases of the wave functions. On the other hand, on the band degeneracy line the value of curl$\mathbf{\Omega}$ is undefined. This follows from the

fact that such a line is a line of the singular points of the Bloch factor $u_{\mathbf{k}0}$ if the factor is considered a function of $\mathbf{k}$ (see Ref. 19 and Eqs. (21)–(24)).

The integral of $\mathbf{\Omega}$ around a closed contour encircling the line of band contact is invariant under transformation (16), and furthermore, according to (27), it is also independent of the size and shape of the contour. To evaluate the integral, we remove the constraint (15) and choose the phases of the Bloch factors only on the basis of the requirement that $u_{\mathbf{k}l}$ be continuous in $\mathbf{k}$ space (as was taken for granted in deriving the energy quantization rule). To do so, we need only set $\varphi_l = q\phi/2$ at $l=0,\ \lambda$ and transform $u_{\mathbf{k}l}$ in a manner similar to (16). Generally, q is any odd number here, but allowance for spin–orbit coupling shows (see below) that we must limit ourselves to $q=\pm 1$. Then the new quantity $\mathbf{\Omega} = \pm(1/2)\partial\phi/\partial\mathbf{k}$, and

$$\oint \mathbf{\Omega}\cdot d\mathbf{k} = \pm\pi, \tag{28}$$

where integration is along any contour that encircles the band degeneracy line, and the sign of the right-hand side is determined by the direction in which the contour is traversed (and the sign of $\mathbf{q}$). When there is natural degeneracy on a fourfold or sixfold symmetry axis, the integral vanishes.

These results suffice to find γ for any relative position of the path and the line of band contact. If the semiclassical path is attached to the line of accidental band degeneracy (or to a threefold symmetry axis, on which there is natural degeneracy), then, according to what was said earlier, $\gamma - 1/2 = \pm 1/2$. Bearing in mind that a simple change of notation in (1) from n to $n-1$ transforms $\gamma=1$ into $\gamma=0$, in the case at hand we obtain (4). But if there is no attachment, then according to a well-known theorem[15] there is sure to be a surface whose edge coincides with the path, i.e., a film stretched over the path and does not intersect the line of band contact (in Fig. 1, for such a film for the paths *1* and *2* we have taken a section of the depicted constant-energy surface). Using Stokes' theorem to transform the integral in (18) into an integral over the specified surface and then using (27), we obtain (2).

To conclude this section we note a certain analogy between our result and the Aharonov–Bohm effect.[22] In 1962, Blount[20] pointed out a similarity between $\mathbf{\Omega}$ and the vector potential of a magnetic field (see Eqs. (16) and (17)). Then Eqs. (27) and (28) make it possible to think of the band degeneracy line as an infinitely thin ''solenoid'' carrying a certain flux that is generated by the ''field'' curl$\mathbf{\Omega}$. As soon as this is done, the analogy becomes clear. Indeed, although an electron moving along a semiclassical path enclosing the band degeneracy line does not enter the region with the ''field,'' it experiences effects of the ''vector potential'' $\mathbf{\Omega}$, which cannot be made equal to zero over the entire path. The semiclassical steady state of an electron with energy given by (1) is a standing wave produced by two waves traveling in opposite directions. The presence of a band degeneracy line, i.e., a ''solenoid,'' inside the path shifts the interference pattern relative to its position when there is no such line. This shift shows up as a change in γ. We also note that this shift is the same as that induced by introducing a real solenoid, with a flux equal to the magnetic-flux quantum $\pi\hbar c/|e|$, into the orbit of an electron in real coordinate space.

4. ALLOWANCE FOR SPIN AND SPIN–ORBIT COUPLING

We start by allowing only for electron spin, i.e., we still neglect electron–orbit coupling. Now all the bands are doubly degenerate in the projection of the spin on an axis, and the Bloch factors $u_{\mathbf{k}l\rho}$ are products of the previously derived functions $u_{\mathbf{k}l}$ for a spinless particle and constant spinors s_ρ, i.e., $u_{\mathbf{k}l1} = u_{\mathbf{k}l}s_1$ and $u_{\mathbf{k}l2} = Uu_{\mathbf{k}l1} = (KIu_{\mathbf{k}l})s_2$, where $s_2 = i\sigma_2 s_1$ (it is in this manner that the specified quantity $u_{\mathbf{k}l2}$ is obtained in the weak spin–orbit coupling limit). The orthogonality of s_1 and s_2 implies that the orbital parts of G_{zx} and G_{zy} vanish. If we put

$$s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

the same is true of their spin parts, so that on the whole $G_{zx} = G_{zy} = 0$. Then, if we allow for the fact that the sum in (7) vanishes, formula (12) becomes

$$g = -\frac{2m}{\pi m^*}\oint_\Gamma \frac{d\kappa}{v_\perp}(m_0)_{11} = 2 - \frac{2m}{\pi m^*}\oint_\Gamma \mathbf{\Omega}\cdot d\mathbf{k}.$$

Here the first term on the right-hand side comes from the pure spin contribution (9) to m_0, and the integral in the second term coincides with the integral in (18). Taking condition (28) into account, we obtain (5). The sign in (5) remains undefined, since without spin–orbit coupling, electron spin and electron orbital motion are uncorrelated.

Allowance for spin–orbit coupling leads to a situation in which the Bloch factors $u_{\mathbf{k}l\rho}$ cannot be reduced to the products $u_{\mathbf{k}l}s_\rho$, but are instead linear combinations of the functions $u_{\mathbf{k}l'}s_{\rho'}$ with different l' and ρ'. The size of the ''admixture'' of these functions in the $u_{\mathbf{k}l}s_\rho$ is estimated to be $\Delta/\delta\varepsilon(\mathbf{k})$, where $\delta\varepsilon(\mathbf{k})$ is of the order of the (energy) distance of the nearest neighboring band from the band in question. The corrections to the g factor due to spin–orbit coupling are of the same order. When $\Delta \ll E_0$, these corrections are small ($\sim\Delta/E_0$) for semiclassical orbits far from the band degeneracy line. However, these corrections can be large for paths that are close to that line. Hence, to estimate the size of the corrections to (5), we must calculate the g factor for the most ''dangerous'' orbits. Here we use the two-band model of the electron energy spectrum (19), in which the strongest mixing of the functions $u_{\mathbf{k}0}s_\rho$ and $u_{\mathbf{k}\lambda}s_{\rho'}$ (related to spin–orbit coupling) is accounted for exactly, while the ''admixture'' of the wave functions of the other bands is neglected. This ensures an accuracy of calculations of the g factor of order Δ/E_0, which is quite sufficient for analyzing large corrections.

After spin–orbit coupling has been taken into account, we obtain a new Hamiltonian, which in the Luttinger–Kohn representation replaces the Hamiltonian (19):

$$\hat{\mathscr{H}} = \begin{pmatrix} \tilde{E}_{00} & \tilde{E}_{0\lambda} \\ \tilde{E}^+_{0\lambda} & \tilde{E}_{\lambda\lambda} \end{pmatrix}, \tag{29}$$

where $\tilde{E}_{ij}$ are the two-dimensional matrices

$$\tilde{E}_{00}=[\Delta+\tfrac{1}{2}(E_{00}+E_{\lambda\lambda})]\hat{1},$$
$$\tilde{E}_{\lambda\lambda}=[-\Delta+\tfrac{1}{2}(E_{00}+E_{\lambda\lambda})]\hat{1}, \tag{30}$$
$$\tilde{E}_{0\lambda}=E_{0\lambda}\hat{1}+\frac{i}{2}(E_{00}-E_{\lambda\lambda})(\boldsymbol{\beta}\cdot\boldsymbol{\sigma}).$$

Here 2Δ is the spin–orbit splitting of the bands at $\mathbf{k}=0$ (for the sake of definiteness we assume that Δ is positive, i.e., the specified band with $l=0$ is higher on the energy scale than the band with $l=\lambda$), $\hat{1}$ is the identity matrix, σ_i are the Pauli spin matrices, and $\boldsymbol{\beta}$ is a constant real vector of unit length characterizing the spin–orbit mixing of the wave functions of the specified bands at $\mathbf{k}=0$. We assume that a semiclassical path is attached to the band degeneracy line (determined without allowance for spin–orbit coupling), and for the origin of $\mathbf{k}$ we take the point at which the line intersects the orbital plane (the plane $k_z=0$). We also assume that at $\mathbf{k}=0$, (15) holds and the element $E_{0\lambda}$ is real. Moreover, in (30) we have neglected the term $\hbar^2\mathbf{k}^2/2m$, since this term becomes important only in the range of wave vectors at which the two-band model provides a poor description of the energy spectrum. Finally, we introduce the notation

$$2\mathbf{a}=\hbar[\mathbf{v}_{00}(0)+\mathbf{v}_{\lambda\lambda}(0)],$$
$$2\mathbf{a}'=\hbar[\mathbf{v}_{00}(0)-\mathbf{v}_{\lambda\lambda}(0)],$$
$$\mathbf{t}=\hbar\mathbf{v}_{0\lambda}(0).$$

Then Eqs. (29) and (30) imply that the electron dispersion relation $\varepsilon_0(\mathbf{k})$ is

$$\varepsilon_0(\mathbf{k})=\mathbf{a}\cdot\mathbf{k}+\sqrt{\Delta^2+(\mathbf{a}'\cdot\mathbf{k})^2+(\mathbf{t}\cdot\mathbf{k})^2}, \tag{31}$$

and the corresponding formula for $\varepsilon_\lambda(\mathbf{k})$ merely has the opposite sign for the radical.

We now discuss a way to calculate m_0. Knowing m_0, we can calculate the g factor with the equations of Sec. 2. Bearing in mind the definitions (8) for $\boldsymbol{\Omega}$ and (20) for S and the well-known relationship between the band-to-band matrix elements of the velocity and position operators,[13] we obtain the relationship

$$\left(S^+\frac{\partial S}{\partial\mathbf{k}}\right)_{0\rho,l\rho'}=-i\delta_{0l}\boldsymbol{\Omega}_{\rho\rho'}(\mathbf{k}) -(1-\delta_{0l})\hbar\frac{\mathbf{v}_{0\rho,l\rho'}(\mathbf{k})}{\varepsilon_0(\mathbf{k})-\varepsilon_l(\mathbf{k})}, \tag{32}$$

which makes it possible, knowing the matrix $S(\mathbf{k})$, to find $\boldsymbol{\Omega}_{\rho\rho'}$ and $\mathbf{v}_{0\rho,l\rho'}$ at any point of the Brillouin zone, and thus the orbital part of $m_0(\mathbf{k})$. We can calculate the spin contribution (9) to m_0 using

$$s_{\rho\rho'}=(S^+s(0)S)_{0\rho,0\rho'},$$

where $s(0)$ is the matrix describing this contribution in the Luttinger–Kohn representation, with $s(0)=-(1/2)\times\langle u_{0l\rho}\sigma_3 u_{0l'\rho'}\rangle$. This matrix can be found together with the Hamiltonian (29), and in the given case is

$$s(0)=-\frac{1}{2}\begin{pmatrix}\beta_z(\boldsymbol{\beta}\cdot\boldsymbol{\sigma}) & (\boldsymbol{\sigma}\times\boldsymbol{\beta})_z\\ (\boldsymbol{\sigma}\times\boldsymbol{\beta})_z & \beta_z(\boldsymbol{\beta}\cdot\boldsymbol{\sigma})\end{pmatrix}.$$

Thus, calculating $m_0(\mathbf{k})$ amounts to determining the transformation S that diagonalizes the Hamiltonian written in the Luttinger–Kohn representation. Below we list the final results of calculating the g factor in the two-band model of the energy spectrum (29), (30).

Ignoring the spin contribution (9) to m_0, we first analyze the orbital part of the g factor. In this case we can find an S such that m_0 becomes a diagonal matrix in the spin index, and accordingly $G_{zx}=F_{zy}=0$. For G_{zz} we obtain

$$G_{zz}=\frac{2m}{\hbar^2}\frac{(\mathbf{i}_z\cdot(\mathbf{a}'\times\mathbf{t}))(\varepsilon+\Delta)}{\delta\varepsilon(\delta\varepsilon+\Delta)}, \tag{33}$$

where

$$\delta\varepsilon\equiv\sqrt{\Delta^2+(\mathbf{a}'\cdot\mathbf{k})^2+(\mathbf{t}\cdot\mathbf{k})^2}.$$

The result for the g factor can be represented by two terms:

$$g=g_1+g_2,$$

where g_1 is the intraband contribution to g due to the first term on the right-hand side of Eq. (7), and the band-to-band contribution g_2 is given by the sum over l in the expression for m_0. For g_1 and g_2 we have

$$g_1=2\,\mathrm{sign}(\mathbf{i}_z\cdot\boldsymbol{\nu})\frac{m}{m^*}\left(1-\frac{\Delta}{\sqrt{\varepsilon^2-\varepsilon_{\min}^2+\Delta^2}}\right), \tag{34}$$
$$g_2=2\,\mathrm{sign}(\mathbf{i}_z\cdot\boldsymbol{\nu})\frac{m}{m^*}\frac{\Delta}{\sqrt{\varepsilon^2-\varepsilon_{\min}^2+\Delta^2}},$$

where $\mathrm{sign}\,x=1$ if $x>0$, $\mathrm{sign}\,x=-1$ if $x<0$, $\boldsymbol{\nu}\equiv(\mathbf{a}'\times\mathbf{t})$, and $\varepsilon_{\min}$ is the minimum value of (31) at $k_z=0$. Such a minimum value must exist for a closed orbit ($\varepsilon_{\min}\propto\Delta$ and $0<\varepsilon_{\min}/\Delta\leqslant 1$). The equations in (34) suggest that although the contributions to g_1 and g_2 related to spin–orbit coupling are of order $g(\Delta/\varepsilon)$ and may be large, they balance each other perfectly. Thus, not only for orbits that are far from the band degeneracy line but also for semiclassical orbits passing close to this line, spin–orbit coupling provides only a relatively small correction overall (of order Δ/E_0) to the orbital part of the g factor considered here. We also note in connection with Eqs. (34) that the result $|g|=2m/m^*$ of Cohen and Blount[18], who used the two-band model of the electron energy spectrum, has been known for rather a long time. However, the semiclassical paths examined in Ref. 18 were those that pass near the extrema of the electron dispersion relation (i.e., as $\varepsilon\to\varepsilon_{\min}$), and only the band-to-band contribution to the g factor was taken into account, since $\mathbf{v}$ and hence g_1 are zero at $\varepsilon=\varepsilon_{\min}$. Equations (34) generalize the Cohen–Blount result to orbits with energies within the range $\varepsilon_{\min}\leqslant\varepsilon\ll E_0$, with the intraband contribution to the g factor being the principal one when $\varepsilon\gg\Delta$. We also note that the exact electron spectrum for the Hamiltonian (29), (30) obtained without taking into account the direct interaction of electron spin and a magnetic field $\mathbf{H}$ was obtained in Ref. 23. Equations (34) together with (1) and (2) determine the electron energy levels, and the result coincides with the exact result for all n, including $n=0$.

If in addition to the orbital part we allow for the spin contribution (9) to m_0, then

$$\delta G_{zz}=2\beta_z, \tag{35}$$

is added to the right-hand side of Eq. (33), and furthermore G_{zx} and G_{zy} are nonzero:

$$G_{zx}-iG_{zy}=2(\beta_y+i\beta_x)[(\mathbf{t}\cdot\mathbf{k})+i(\mathbf{a}'\cdot\mathbf{k})]/\delta\varepsilon\,. \tag{36}$$

In the g factor there appears a spin part g_s (in addition to the orbital contribution (34)), which can be calculated using (12), where G_{zz} must be replaced by δG_{zz}. Direct calculations show that at $\Delta=0$ the function

$$\tau_0=\frac{1-\beta_z}{\beta_y+i\beta_x}\,\frac{(\mathbf{t}\cdot\mathbf{k})+i(\mathbf{a}'\cdot\mathbf{k})}{\delta\varepsilon},$$

specified on the semiclassical path $k_z=0$, $\varepsilon_0(\mathbf{k})=\varepsilon$, is an exact solution of Eq. (13) with the boundary condition (14). Substituting this function into the formula for g_s yields $g_s=2$, as expected. From (33), (35), and (36) it follows that for $\Delta\neq 0$ the solution of Eq. (13) differs from τ_0 by a quantity of order $\tau_0(\Delta/\varepsilon)$. Accordingly, the correction to g_s is $\delta g_s\sim\Delta/\varepsilon$. Bearing in mind that $g\sim m/m^*\sim E_0/\varepsilon$, we obtain the estimate $\delta g_s/g\sim\Delta/E_0$, which suggests that in the adopted approximation spin–orbit coupling has no effect on the spin contribution to the g factor.

If we now take (34) into account, we obtain the final expression

$$g=2+2\ \mathrm{sign}(\mathbf{i}_z\cdot\boldsymbol{\nu})\,\frac{m}{m^*}. \tag{37}$$

Note that by allowing for spin–orbit coupling we were able to determine the sign with which the orbital part enters into the g factor (and also justify our choice of q in deriving (28)). Here, although Eq. (37) was derived for $\varepsilon\ll E_0$, the continuity of the function $g(\varepsilon)$ suggests that the sign of the orbital part is the same outside that range. Hence (37) holds for semiclassical paths far from the line of band contact.

The properties of the vector $\boldsymbol{\nu}$, on which the aforementioned sign depends, can be established by analyzing the two-band Hamiltonian (19) with allowance for Herring's result.[1] These properties reduce to the following statements. The vector $\boldsymbol{\nu}$ is directed along the tangent to the band degeneracy line, determined without allowance for spin–orbit coupling. It is independent of the choice of basis functions u_{00} and $u_{0\lambda}$ at $\mathbf{k}=0$, and in that sense is an invariant characteristic of degenerate electron states. The vector $\boldsymbol{\nu}$ vanishes only at the points of intersection of the given line of band contact with a different line of this type (if such points exist), so that in the intervals between the points of intersection its orientation does not change. The dot product $\mathbf{i}_z\cdot\boldsymbol{\nu}$ is nonzero for the paths considered here, which are attached to the line of band contact (the product vanishes only for orbits of type *2* in Fig. 1).

In closing, we again dwell on the relationship between the results obtained in this section and in Sec. 4. When spin–orbit coupling is neglected, the wave functions of an electron in a magnetic field can be represented as products of constant spinors s_ρ and purely orbital wave functions of a spinless particle. Naturally, such states correspond to $\gamma=0$ and $g=2$. In essence, this is simply a different classification of the electron states in a magnetic field that emerge in the limit $\Delta\to 0$. Although such an approach is approximate when spin–orbit coupling is taken into account, for $\varepsilon\gg\Delta$ it provides a good description not only of the energy spectrum but also of the matrix elements of operators. Only when $\varepsilon\sim\Delta$ must the matrix elements be calculated by the rigorous approach used in this section.

5. CONCLUSION

The substantial difference between the real value of the g factor and 2 (or a change in γ) predicted in the present paper should become manifest in oscillatory effects. For the sake of definiteness we discuss magnetization oscillations in the de Haas–van Alphen effect. The amplitude of the pth harmonic of this quantity is proportional to the factor[14]

$$\cos\left(\pi p\,\frac{g\,m^*}{2\,m}\right), \tag{38}$$

which depends on the g factor. Substituting (5) into (38), we find that for paths attached to the band degeneracy line, the sign of this factor for the fundamental harmonic ($p=1$) is opposite that expected for the case with $g=2$. Methods of measuring the absolute value and sign of the cosine (38) have been described in a monograph by Shoenberg.[14] Shoenberg's analysis shows that finding the sign of (38) is simplest when small extremal cross sections of the Fermi surface are involved. In this connection we point to several metals in which the semiclassical paths corresponding to such cross sections are attached to the line of band contact and in which spin–orbit coupling is weak, namely beryllium,[24] magnesium,[25] and graphite[26,27] (the latter is a semimetal). In beryllium and magnesium, which have a hexagonal close-packed crystal structure, the line of accidental degeneracy of the second and third bands lies in the basal plane of the crystal and passes through the multiply connected hole surface of the second band, the so-called monster (Mg) or crown (Be). If the magnetic field lies in the same plane, then (5) holds for orbits that correspond to the minimum cross sections of this surface (on its ''necks''). Neglecting a small amount of ''openness'' of the monster in certain directions, since this feature plays no role in our discussion, we can say that qualitatively the situation is similar to that depicted in Fig. 1b. Note that in zinc and cadmium, which are isovalent with beryllium and magnesium and have the same crystal structure, the line of accidental degeneracy of the second and third bands land inside the electron ''lens'' of the third band[28] and is not attached to semiclassical paths (see Fig. 1a). In beryllium, magnesium, and graphite there is natural band degeneracy on the threefold symmetry axis (the vertical edge HKH of their Brillouin zone, which is a hexagonal prism). When there is a magnetic field directed along this axis, (5) holds for an extremal orbit that encloses K. In beryllium and magnesium this orbit is located on the so-called ''cigar' of the third band. Note, however, that in magnesium with such an orientation of the magnetic field, even a weak

field leads to magnetic breakdown between the cigar and the monster, i.e., to a violation of (3b). There is no such difficulty with graphite, whose Fermi surface has a self-intersecting shape similar to that depicted in Fig. 1c (the path *4* corresponds to the orbit of interest in the vicinity of K). Hence this semimetal is probably the most convenient object for experimental studies of this effect in the case of natural band degeneracy. Note finally that $m^* \ll m$ on all the paths we have mentioned here, and that at $p=1$ the cosine in (38) is close to -1.

One could scarcely hope to distinguish between cases in which there are orbits in the metal that are attached to the band degeneracy line and those in which there are no such paths by studying the g factor via the EPR method. Indeed, the results of Sec. 4 show that if we neglect spin–orbit coupling, excitation of the spin degrees of freedom by a variable magnetic field has no effect on the orbital state of an electron. In other words, resonance will be observed at a frequency corresponding to $g=2$. Only if spin–orbit coupling is strong enough will resonance emerge at frequencies corresponding to values given by (37), which are extrema in k_z. However, resonance in such metals is possible even without a band degeneracy line.[29]

In conclusion, we discuss our reasons for arguing that Eqs. (4) and (5), which were obtained in the one-electron approximation, probably do not change when we allow for the Fermi-liquid interaction of electrons (not counting the well-known renormalizations of the spin part of the g factor and the cyclotron mass m^*). Since $\delta\gamma$, which we define to be the difference in γ for paths attached and unattached to the band degeneracy line, changes sign when H is replaced by $-H$ (see (18)), the invariance of the semiclassical spectrum under such a transformation of the field immediately implies that $2\delta\gamma$ must be an integer. In other words, $|\delta\gamma|$ is either 0 or 1/2. At the same time, $\delta\gamma$ depends only on the electron states on the Fermi surface, and in the spirit of Landau's Fermi-liquid theory we would expect that the value of $\delta\gamma$ can change only gradually when the interaction between electrons is ''turned on.'' Bearing in mind that (2) still holds even when this interaction is taken into account,[13] we ultimately obtain the foregoing assertion.

*)E-mail: mikitik@ilt.kharkov.ua

1)The method used in by Roth[9] to derive the appropriate equations differs somewhat from the one described here, but of course the results are the same.

[1] C. Herring, Phys. Rev. **52**, 365 (1937).
[2] D. A. Papaconstantopoulos, *Handbook of the Band Structure of Elemental Solids*, Plenum Press, New York (1986).
[3] L. Onsager, Philos. Mag. **43**, 1006 (1952).
[4] I. M. Lifshits and A. M. Kosevich, Zh. Éksp. Teor. Fiz. **29**, 730 (1955) [Sov. Phys. JETP **2**, 636 (1956)].
[5] G. E. Zil'berman, Zh. Éksp. Teor. Fiz. **32**, 296 (1957) [Sov. Phys. JETP **5**, 208 (1957)];**33**, 387 (1957) [Sov. Phys. JETP **6**, 299 (1958)].
[6] G. E. Zil'berman, Zh. Éksp. Teor. Fiz. **34**, 748 (1958) [Sov. Phys. JETP **7**, 513 (1958].
[7] M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **39**, 1276 (1960) [Sov. Phys. JETP **12**, 891 (1961)]; **46**, 929 (1964) [Sov. Phys. JETP **19**, 634 (1964)].
[8] L. A. Fal'kovskiĭ, Zh. Éksp. Teor. Fiz. **49**, 609 (1965) [Sov. Phys. JETP **22**, 423 (1966)].
[9] L. M. Roth, Phys. Rev. **145**, 434 (1966).
[10] A. A. Slutskin, Zh. Éksp. Teor. Fiz. **53**, 767 (1967) [Sov. Phys. JETP **26**, 474 (1968)].
[11] S. S. Nedorezov, Fiz. Nizk. Temp. **2**, 1047 (1976) [Sov. J. Low Temp. Phys. **2**, 515 (1976)].
[12] B. M. Gorbovitskiĭ and V. I. Perel', Zh. Éksp. Teor. Fiz. **85**, 1812 (1983) [Sov. Phys. JETP **58**, 1054 (1983)].
[13] E. M. Lifshitz and L. P. Pitaevskiĭ, *Statistical Physics*, 3rd ed., Part 2: *Theory of the Condensed State*, Pergamon Press, Oxford (1979).
[14] D. Shoenberg, *Magnetic Oscillations in Metals*, Cambridge Univ. Press, Cambridge (1984).
[15] B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern Geometry*, 3 parts, Springer-Verlag, Berlin (1991, 1985, 1990).
[16] L. M. Roth, J. Phys. Chem. Solids **23**, 433 (1962).
[17] E. I. Blount, Phys. Rev. **126**, 1636 (1962).
[18] M. H. Cohen and E. I. Blount, Philos. Mag. **5**, 115 (1960).
[19] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, Wiley, New York (1974).
[20] E. I. Blount, in *Solid State Physics: Advances in Research and Applications*, Vol. 13, H. Ehrenreich, F. Seitz, and D. Turnbull (eds.), Academic Press, New York (1962), p. 305.
[21] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[22] Y. Aharonov and D. Bohm, Phys. Rep. **115**, 485 (1959).
[23] G. P. Mikitik and Yu. V. Sharlaĭ, Fiz. Nizk. Temp. **22**, 762 (1996) [Low Temp. Phys. **22**, 585 (1996)].
[24] J. H. Tripp, P. M. Everett, W. L. Cordon, and R. W. Stark, Phys. Rev. **180**, 669 (1969).
[25] J. C. Kimball, R. W. Stark, and F. M. Mueller, Phys. Rev. **162**, 600 (1967).
[26] J. C. Slonczewski and P. R. Weiss, Phys. Rev. **109**, 272 (1958).
[27] R. O. Dillon, I. L. Spain, and J. W. McClure, J. Phys. Chem. Solids **38**, 635 (1977).
[28] R. W. Stark and L. M. Falikov, Phys. Rev. Lett. **19**, 795 (1967).
[29] M. Ya. Azbel', Fiz. Tverd. Tela (Leningrad) **4**, 568 (1962) [Sov. Phys. Solid State **4**, 415 (1962)].

Translated by Eugene Yankovsky

Limits of applicability of the Kramers–Kronig relations in the presence of an additional light wave

M. I. Strashnikova and E. V. Mozdor*)

Institute of Physics, National Academy of Sciences of the Ukraine, 252650 Kiev, Ukraine; Institute of Semiconductor Physics, National Academy of Sciences of the Ukraine, 252650 Kiev, Ukraine

(Submitted 11 November 1997)
Zh. Éksp. Teor. Fiz. **114**, 1393–1406 (October 1998)

We analyze the possibility of applying the Kramers–Kronig relations to media with substantial spatial diffusion of the permittivity, in which case additional light waves of the Pekar type emerge. The calculations are done within a broad range of values of the exciton damping constant Γ with and without a ''dead layer'' at the crystal surface. We establish the condition in which the use of the Kramers–Kronig relations in calculating the optical constants of the substance yields correct results. © *1998 American Institute of Physics.* [S1063-7761(98)01610-2]

1. INTRODUCTION

In his pioneering work,[1] Pekar examined the interaction of light and media that have substantial spatial dispersion of the permittivity $\varepsilon(\omega,\mathbf{K})$ in the excitonic region of the spectrum. He was the first to predict the existence of additional light waves and suggested new ideas concerning the excitonic absorption of light. According to Refs. 1 and 2, ''absorption is due to transitions of the system from the exciton states generated by the light to other states (i.e., with the exception of the initial state). If these transitions are accompanied by luminescence, we have Raman scattering of the initial light. If these transition are thermal and there is excitation of thermal vibrations, we have ordinary light absorption.'' Thus, absorption is due to the finiteness of the exciton lifetime in relation to nonradiative transitions. If there are no such transitions, there is no absorption, no matter how high the oscillator strength.

Thus, one prediction of the theory was that there can be large dispersion of the refractive index if exciton–photon coupling is strong (when oscillator strength of the transition is high), and yet the absorption of energy may be very low if exciton–photon coupling is weak (for weak nonradiative transitions). Hence Pekar repeatedly stated that if the spatial dispersion of ε is substantial, the oscillator strength of the excitonic transition should be determined not from the area under the absorption curve but from the curvature of the dispersion curve of the refractive index. This actually means that the classical Kramers–Kronig relations break down.

These relations establish the integral relationship that exists between the real and imaginary parts of the permittivity, $\varepsilon=\varepsilon'+i\varepsilon''$ (Ref. 3):

$$\varepsilon'(\omega)-1=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\varepsilon''(x)\,dx}{x-\omega}, \tag{1}$$

$$\varepsilon''(\omega)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\varepsilon'(x)-1}{x-\omega}\,dx. \tag{2}$$

Here the integral from $-\infty$ to $+\infty$ is understood in the principal-value sense. The Kramers–Kronig relations were derived on the basis of the causality principle for a local relationship between the exciting electromagnetic wave and the response of the crystal. This means that the medium's polarization **P** at a given point in space is determined by the electric field **E** at a given moment in time and at all preceding moments in time at the same point in space. As shown by Landau and Lifshitz,[3] in the entire upper half of the complex plane of the variable $\tilde{\omega}$, the function $\varepsilon(\tilde{\omega})$ is single-valued and does not become infinite, i.e., has no singular points. It is for such a function that Kramers[4] and Kronig[5] obtained their relations (1) and (2).

If we allow for the fact that $\varepsilon'(\omega)$ is an even function of ω and $\varepsilon''(\omega)$, an odd function, formulas (1) and (2) can be reduced to a form convenient for comparison of experimental and theoretical data, i.e., for real positive ω:

$$\varepsilon'(\omega)=1+\frac{2}{\pi}\int_{0}^{\infty}\frac{x\varepsilon''(x)\,dx}{x^2-\omega^2}, \tag{3}$$

$$\varepsilon''(\omega)=-\frac{2\omega}{\pi}\int_{0}^{\infty}\frac{\varepsilon'(x)}{x^2-\omega^2}\,dx. \tag{4}$$

As is known, permittivity can be expressed in terms of the square of the complex-valued refractive index $\tilde{n}=n+i\kappa$ as follows:

$$\varepsilon=\tilde{n}^2,\quad \varepsilon'=n^2-\kappa^2,\quad \varepsilon''=2n\kappa.$$

The real and imaginary parts of $\tilde{n}$, i.e., the refractive index n and the absorption coefficient κ, are also linked by integral relationships:[6–8]

$$n(\omega)=1+\frac{2}{\pi}\int_{0}^{\infty}\frac{x\kappa(x)\,dx}{x^2-\omega^2}, \tag{5}$$

$$\kappa(\omega)=-\frac{2\omega}{\pi}\int_{0}^{\infty}\frac{n(x)}{x^2-\omega^2}\,dx. \tag{6}$$

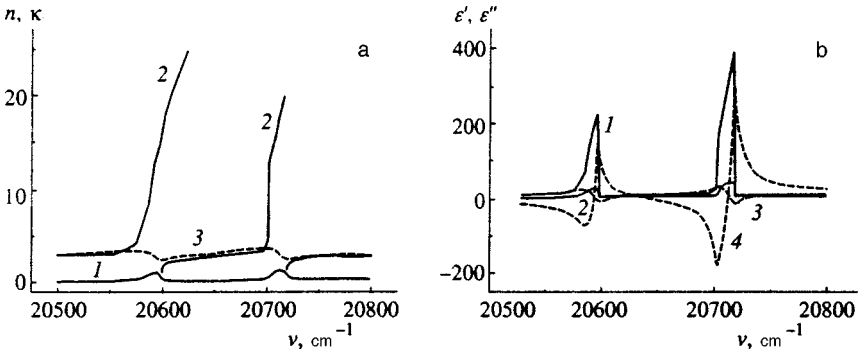

FIG. 1. a) The experimentally measured dispersion of the refractive index $n(\omega)$ (curve *2*) and the curve obtained through calculations involving the absorption spectrum $\kappa(\omega)$ (curve *1*) by (5) (curve *3*). b) The functions $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ obtained from experimentally measured $n(\omega)$ and $\kappa(\omega)$ (curves *1* and *2*) and calculated by (3) and (4) (curves *3* and *4*). The abscissa is the wave number $\nu = 1/\lambda = \omega/2\pi c$.

By measuring the absorption spectrum of a substance we can, in principle calculate the dispersion of its refractive index, and vice versa.

There are also integral formulas, widely used in spectroscopy, that relate the reflection coefficient $R(\omega)$ and the variation of the phase of the light wave in scattering, $\theta(\omega)$ (see Ref. 6):

$$\theta(\omega) = -\frac{2\omega}{\pi} \int_0^\infty \frac{\ln\sqrt{R(x)}}{x^2-\omega^2}\,dx, \tag{7}$$

$$\ln\sqrt{R(\omega)} = 1 + \frac{2}{\pi} \int_0^\infty \frac{x\,\theta(x)\,dx}{x^2-\omega^2}, \tag{8}$$

where $R = \tilde{r}\tilde{r}^*$, with $\tilde{r} = |\tilde{r}|\exp(i\theta)$ the amplitude reflection coefficient. Using these formulas and the Fresnel formulas, which express R and θ in terms of the optical constants n and κ, and knowing the reflection spectrum from measurements, one can calculate the optical constants of the substance when measuring these constants directly is difficult. In the simplest case of normal incidence, the Fresnel formulas are

$$R = \frac{(n-1)^2+\kappa^2}{(n+1)^2+\kappa^2}, \tag{9}$$

$$\tan\theta = \frac{2\kappa}{n^2+\kappa^2-1}. \tag{10}$$

When spatial dispersion of the permittivity is significant, the relationship between excitation and response is nonlocal, i.e., the medium's polarization **P** at a given point in space is determined by the electric field vector **E** not only at that point but also in the entire infinite space. Strictly speaking, in this case we can use only the notion of a ''polarizability core,'' which is determined by the integral relationship between the polarization of the medium and the electric field. Here the permittivity (or dielectric constant) ε proves to be dependent not only on the frequency but also on the wave vector $\mathbf{K} = \tilde{\mathbf{n}}\,\omega/c$ of the light wave in the crystal. As shown by Pekar,[1,2] in this case additional light waves, or Pekar waves,[9] emerge, and $\varepsilon(\omega,\mathbf{K}(\omega))$ becomes a multivalued function of frequency, since for each value of ω there are two values of $\varepsilon(\omega,\mathbf{K}(\omega))$ corresponding to two waves propagating in the crystal. Here it is impossible to use the relations (1) and (2) even formally, since it is unclear what values of ε' and ε'' or their combination should be plugged into these relations.

The first to generalize the Kramers–Kronig relations to the case of media with spatial dispersion and with independent ω and **K** was Leontovich.[10] Later the problem was studied by Davydov,[11] Mead,[12], and Ginzburg, Meĭman, and Agranovich.[13,14] An important stage in these studies was the publication of a series of papers by Solov'ev *et al.*,[15–19] who derived additional dispersion relations (see below).

The goal of the present investigation is to establish the limits of applicability of the Kramers–Kronig relations (3)–(8) to media in which spatial dispersion of ε is significant. In our investigation we show to what extent the values of the optical constants calculated by (3)–(8) for a crystal in which Pekar's additional waves play a significant role can be erroneous. We also establish the characteristics of the medium that correspond to the quantities calculated by the Kramers–Kronig relations. Finally, we determine the values of the parameters of the theory and the experimental conditions under which these relations become valid.

2. COMPARISON OF THE EXPERIMENTALLY MEASURED n VS. ω AND κ VS. ω CURVES WITH THOSE CALCULATED BY THE KRAMERS–KRONIG RELATIONS

The fact that the Kramers–Kronig relations do not work in the excitonic absorption region at low temperatures was first discovered for molecular crystals[20] and for the CdS crystal.[21,22] These papers compare the span of the dispersion curve for the refractive index $n(\omega)$ and the maximum value of the absorption coefficient, $\kappa^{\max}$, which according to the Kramers–Kronig relations are approximately equal. However, in experiments the span of $n(\omega)$ is always larger than $\kappa^{\max}$ (in Ref. 22, by a factor of almost ten). Besides (as shown in Ref. 23), according to the Kramers–Kronig relations, the absorption band always contains an anomalous dispersion section. At the same time, direct measurements of dispersion in the $1A$-exciton region of the CdS crystal at 4.2 K, first done by Brodin *et al.*,[24] showed that there is no such section. In the present paper we do a more detailed and thorough comparison of the experimental curves with the curves calculated by the Kramers–Kronig relations in a spectral range most suitable for observations.

The experimentally measured spectrum of the absorption coefficient $\kappa(\omega)$ in the regions of $1A$-, $1B$-, and $2A$-exciton

states of the CdS crystal at 4.2 K is depicted by curve *1* in Fig. 1a. (The measurements of this spectrum for a CdS single crystal 0.18-μm thick have been reported in Ref. 25.) Using this spectrum and the relation (5), we calculated the dispersion of the refractive index $n(\omega)$ in the same spectral range. In these calculations we replace the 1 in (5) by the background refractive index n_0, which allows for the contributions of all other resonances, i.e., with the exception of the one considered here, to n. The resulting curve is curve *3* in Fig. 1a. In the same figure, curve *2* represents the dispersion of the refractive index $n(\omega)$ measured in the experiment. Here we draw the average curve, i.e., a curve obtained by different researchers in different experiments: (1) via a Jamin interferometer crossed with a spectrograph and birefringence;[24,26] (2) via a very thin wedge-shaped single crystal;[27–29] and (3) via a Fabry–Perot interferometer.[30]

Figure 1a clearly shows that the calculated functions contain a section of anomalous variation of $n(\omega)$ in the half-width interval of the absorption bands and that the span of this section is approximately $\kappa^{\max}$. At the same time, the curves measured in the experiments have no such section and their span is much larger than $\kappa^{\max}$. Moreover, the experiment established that there are two waves transmitted through the crystal simultaneously and that the waves have different values of n, while (5) implies that there can be only one wave. Thus, Fig. 1a proves that the dispersion of the refractive index $n(\omega)$ calculated by the Kramers–Kronig relations differs from the dispersion measured experimentally in the case where the spatial dispersion $\varepsilon(\omega,\mathbf{K})$ is substantial.

The measured values of $n(\omega)$ and $\kappa(\omega)$ can be used to calculate the real and imaginary parts of permittivity. The frequency dependence of these quantities is depicted in Fig. 1b by curves *1* and *2*, respectively (the solid curves). In our calculations, for each frequency we took only one measured value of $n(\omega)$: the value of n of the wave whose intensity was the highest. Hence, starting with the longitudinal frequency $\omega_L = 20\,600\ \mathrm{cm}^{-1}$, the values of n of the second wave were discarded. At the same time, knowing $\varepsilon''(\omega)$ we can calculate $\varepsilon'(\omega)$ by (3), while knowing $\varepsilon'(\omega)$ we can calculate $\varepsilon''(\omega)$ by (4). The results of these calculations are depicted in Fig. 1b by curves *3* and *4*. We see that curves *1* and *3* (and *2* and *4*) differ significantly, which illustrates the fact that the relations (3) and (4) must not be used when the spatial dispersion $\varepsilon(\omega,\mathbf{K})$ is substantial.

Note that the conditions formulated in Ref. 14 under which the Kramers–Kronig relations are valid have not been violated. According to Ref. 14, the tensor $\varepsilon_{ij}(\omega,\mathbf{K})$ ''probably acquires a singularity in the upper half-plane of the complex variable $\tilde{\omega}$ only if the condition $Ka \ll 1$, where a is the lattice constant or some other characteristic size, is violated.'' (At $n=10$ we have $aK=0.063$, while at $n=20$ we have $aK=0.126$.) Hence it is the presence of an additional light wave (and not the above conditions) that is the reason the Kramers–Kronig relations break down.

3. COMPARISON OF THE DISPERSION AND ABSORPTION CURVES CALCULATED IN PEKAR'S THEORY WITH THOSE CALCULATED BY THE KRAMERS–KRONIG RELATIONS

We have verified the applicability of the relations (5) and (6) (and hence (3) and (4)) to dispersion and absorption curves calculated by Pekar's theory in the $1A$-exciton region of the CdS crystal for different values of the damping constant Γ. The other parameters of the theory were selected in such a way so as to ensure the best agreement between the calculated and experimentally measured curves $n(\omega)$ and $\kappa(\omega)$ in the $1A$-exciton region at 4.2 K. To achieve the best approximation we were forced to allow explicitly for the effect of the nearest $1B$- and $2A$-excitons on the frequency dependence of the background permittivity $\varepsilon_0(\omega)$ in the $1A$-exciton region. The spatial dispersion effects in the $1B$- and $2A$-exciton regions were not taken into account and $\varepsilon(\omega)$ was written as follows:

$$\varepsilon(\omega)=\varepsilon_0(\omega)+\frac{\Delta_{LT}^{1A}\varepsilon_0}{\omega_0^{1A}+\hbar K^2/2M^{1A}-\omega-i\Gamma^{1A}}, \tag{11}$$

where

$$\varepsilon_0(\omega)=\varepsilon_0\left(1+\frac{\Delta_{LT}^{1B}}{\omega_0^{1B}-\omega-i\Gamma^{1B}}+\frac{\Delta_{LT}^{2A}}{\omega_0^{2A}-\omega-i\Gamma^{2A}}\right). \tag{12}$$

Here $\varepsilon_0=7.74$, the longitudinal–transverse splitting is $\Delta_{LT}^{1A}=16\ \mathrm{cm}^{-1}$, the resonant frequency is $\omega_0^{1A}=20\,584\ \mathrm{cm}^{-1}$, $\Gamma^{1A}=2\ \mathrm{cm}^{-1}$, $M^{1A}=0.8m_e$, $\Delta_{LT}^{1B}=12\ \mathrm{cm}^{-1}$, $\omega_0^{1B}=20\,707\ \mathrm{cm}^{-1}$, $\Gamma^{1B}=6\ \mathrm{cm}^{-1}$, $\Delta_{LT}^{2A}=2\ \mathrm{cm}^{-1}$, $\omega_0^{2A}=20\,770\ \mathrm{cm}^{-1}$, and $\Gamma^{2A}=8\ \mathrm{cm}^{-1}$.

The solution of the dispersion equation with the permittivity ε specified by (11) is given by Pekar's formula[1]

$$\tilde{n}_{\pm}^2=\frac{1}{2}[\mu+\varepsilon_0(\omega)]\pm\sqrt{\frac{1}{4}[\mu-\varepsilon_0(\omega)]^2+b}\,, \tag{13}$$

where

$$\mu=\alpha(\omega-\omega_0^{1A}+i\Gamma^{1A}),\quad b=\alpha\Delta_{LT}^{1A}\varepsilon_0,\quad \alpha=\frac{2M^{1A}c^2}{\hbar(\omega_0^{1A})^2}.$$

Pekar's double-wave crystal optics becomes the classical single-wave optics starting from a certain critical value of the damping constant, Γ_{cr} (Refs. 31 and 32). Note that the function $\tilde{n}_{\pm}^2$ has a branch point in the complex $\tilde{\omega}=\omega'+i\omega''$ plane, whose position can be found if we zero out the radicand in (13). Below we present an analysis of the movement of the branch point in the complex $\tilde{\omega}=\omega'+i\omega''$ plane as Γ increases, an analysis done by V. N. Piskavoĭ (in Ref. 33 the position of the branch point in the complex $\tilde{\omega}=\omega'+i\omega''$ plane remains fixed):

$$\omega'=\omega_0+\varepsilon_0/\alpha,\quad \omega''=\sqrt{\Delta_{LT}\varepsilon_0/\alpha}-\Gamma. \tag{14}$$

At $\Gamma=\Gamma_{\mathrm{cr}}=\sqrt{\Delta_{LT}\varepsilon_0/\alpha}$ we have $\omega''=0$, and the branch point is on the real frequency axis at $\omega'=\omega_{\mathrm{cr}}=\omega_0+\varepsilon_0/\alpha$. Here $\tilde{n}_+(\omega_{\mathrm{cr}},\Gamma_{\mathrm{cr}})=\tilde{n}_-(\omega_{\mathrm{cr}},\Gamma_{\mathrm{cr}})$, and the dispersion curves

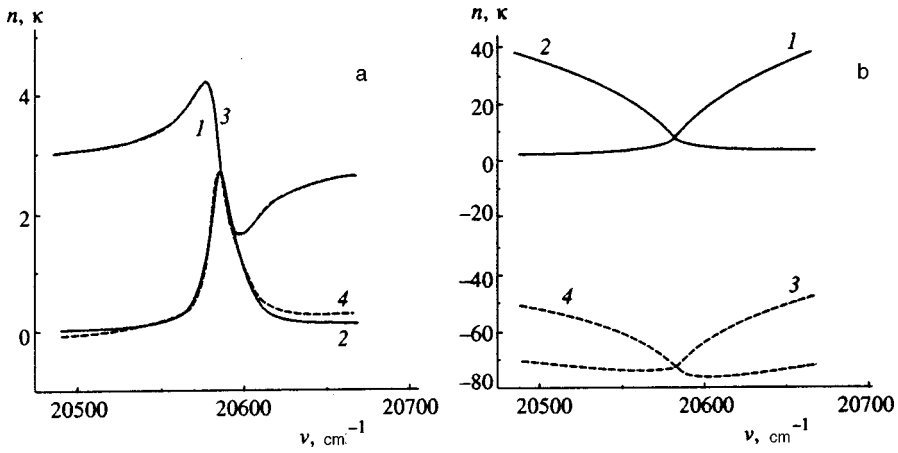

FIG. 2. The Pekar functions $n(\omega)$ (curve *1*) and $\kappa(\omega)$ (curve *2*) and similar functions calculated by (5) and (6) (curves *3* and *4*, respectively) for a) the ''plus–minus'' wave, and b) the ''minus–plus'' wave; $\Gamma > \Gamma_{cr}$.

intersect at real values of $\tilde{\omega}$ (for the sake of simplicity we will write ω instead of ω'). When $\Gamma < \Gamma_{cr}$, the branch point is in the upper half-plane of the complex-valued frequency $\tilde{\omega}$, and when $\Gamma > \Gamma_{cr}$, the branch point is in the lower half-plane. Hence for $\Gamma < \Gamma_{cr}$ the integration contour C, which passes along the real axes and is closed at infinity, encloses the singular point, which means that the ordinary Kramers–Kronig relations (3) and (4) break down. For $\Gamma > \Gamma_{cr}$ the branch point lies outside the integration contour and the Kramers–Kronig relation may be valid. However, only the wave that has classical limits and the classical shape satisfies these relations, while the other wave, in which n tends to infinity as $\omega \to \infty$, cannot satisfy the finiteness condition for the function, and so it does not satisfy the Kramers–Kronig relations.

These facts are illustrated by the results of calculations depicted in Figs. 2a and 2b, where for the sake of comparison we have depicted the Pekar curves $n(\omega)$ and $\kappa(\omega)$ and similar curves calculated by (5) and (6) at $\Gamma = 8\ \mathrm{cm}^{-1}$, which is larger than $\Gamma_{cr} = 5.64\ \mathrm{cm}^{-1}$. Figure 2a depicts the results of calculations done for the classical wave. For $\omega < \omega_{cr}$ this is a Pekar ''plus'' wave, which corresponds to the positive values of the root in (13), while for $\omega > \omega_{cr}$ it is a ''minus'' wave, which corresponds to the negative values of the root. Below we call such a classical wave a ''plus–minus'' wave. In Fig. 2b a similar comparison can be done for the nonclassical wave, which for $\omega < \omega_{cr}$ is a ''minus'' wave and for $\omega > \omega_{cr}$, a ''plus'' wave. Below we call such a nonclassical wave a ''minus–plus'' wave. The comparison shows that the functions $n(\omega)$ and $\kappa(\omega)$ calculated by (5) and (6), which are depicted by dashed curves, are in good agreement with the corresponding initial curves (solid curves) in the case of a classical wave and differ entirely from the corresponding curves in the case of a nonclassical wave. When the crystal is excited by light, the amplitude of the nonclassical wave tends to zero.[32]

As noted in Ref. 11, if two types of wave differing in the values of $n(\omega)$ and $\kappa(\omega)$ propagate in a medium, then ''to each wave we can assign a value of $\varepsilon(\omega)$.'' Bearing this in mind, we compared the curves $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ for each of these waves, with their frequency dependence calculated by the relations (3) and (4). We found that for the classical wave these relations are satisfied and for the nonclassical wave they are not.

In this connection we believe it is appropriate to quote a passage from Davydov's paper:[11]

> ''When there is spatial dispersion, the refractive index and the absorption coefficient for ordinary transverse waves satisfy the usual Kramers–Kronig relations. For additional transverse waves, in some frequency ranges the quantities n, κ, and ε assume infinite values, i.e., lose their physical meaning as macroscopic characteristics of the medium. Hence the function $\varepsilon(\omega)$ is not analytic in the upper half-plane of the complex-valued frequency $\tilde{\omega}$, including the real axis, and no integral relation of the Kramers–Kronig type can be written for it.''

Figure 2 shows that this is indeed the case when $\Gamma > \Gamma_{cr}$, i.e., when we can speak of principal and additional waves as existing in the entire frequency range. We note once more that here the amplitude of the additional wave tends to zero. The situation is much more complicated when $\Gamma < \Gamma_{cr}$, since, depending on the frequency range, each of the two waves may be a principal wave (with a higher amplitude) or an additional wave (with a lower amplitude), or the amplitudes of the two waves are the same and there is no way to tell which wave is the principal one and which is the additional.

We did a series of calculations for the case where $\Gamma = 1\ \mathrm{cm}^{-1} < \Gamma_{cr}$. First we calculated the Pekar curves $n(\omega)$ and $\kappa(\omega)$. Then we applied the Kramers–Kronig relations (5) and (6) to these curves. Finally, we compared the result with the initial functions. Such comparison was done for a ''plus'' wave, a ''minus'' wave, a ''minus–plus'' wave, and a ''plus–minus'' wave. We also considered the case of a compound wave, where the transition from the ''plus'' branch to the ''minus'' branch takes place at the frequency ω_l. The last variant of calculations was selected on the basis of the results of Akhmediev's paper,[33] in which such a wave was used to achieve a good approximation for the integral absorption contour for the $1A$-exciton of the CdS crystal.

We found that there is not a single case where the initial and calculated functions coincide. Figure 3 illustrates the re-

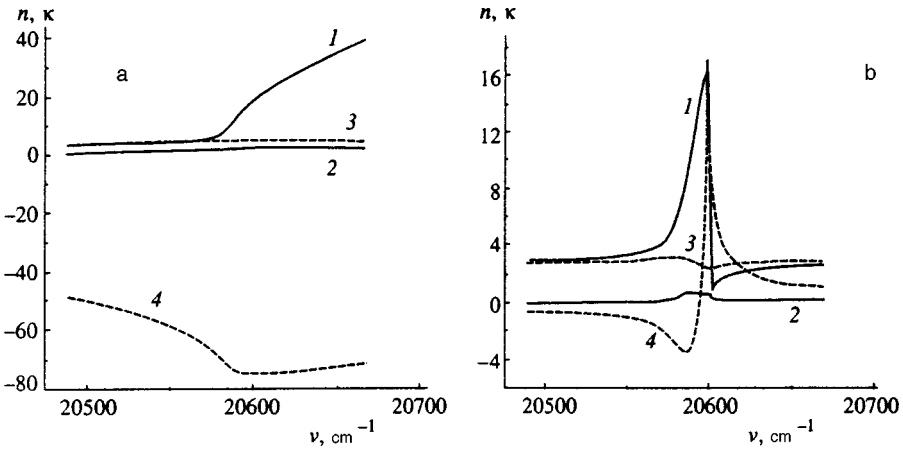

FIG. 3. The same as in Fig. 2 at $\Gamma<\Gamma_{cr}$ for (a) the ''plus'' wave, and ''plus–minus'' wave with the solutions matched at ω_l (b).

sults of two variants of calculation. Similar results were obtained for $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$.

Thus, the Kramers–Kronig relations work well if the additional wave becomes insignificant. If, however, the amplitudes of the two waves are roughly the same, the relations operate neither for each wave separately nor for a combination of the waves, including the classical one.

4. ANALYSIS OF THE REFLECTION SPECTRA CALCULATED IN PEKAR'S THEORY

According to Pekar's theory,[1,2,34–36] the reflection spectrum $R(\omega)$ and the spectrum of variations in the phase of the light wave in reflection, $\theta(\omega)$, are described by the classical Fresnel formulas (9) and (10) if instead of n and κ we plug in the ''effective'' values n^{eff} and κ^{eff}, where

$$\tilde{n}^{\text{eff}}=n^{\text{eff}}+i\kappa^{\text{eff}}=\frac{\tilde{n}_+}{1-\tilde{q}}+\frac{\tilde{n}_-}{1-1/\tilde{q}}\equiv\frac{\varepsilon_0+\tilde{n}_+\tilde{n}_-}{\tilde{n}_++\tilde{n}_-}, \qquad (15)$$

$$\tilde{q}=|q|e^{i\Phi}=-\frac{E_-}{E_+}=\frac{\varepsilon_0-\tilde{n}_+^2}{\varepsilon_0-\tilde{n}_-^2}.$$

Figure 4a depicts the spectral curves $n^{\text{eff}}(\omega)$ and $\kappa^{\text{eff}}(\omega)$ in the $1A$-exciton region of the CdS crystal (curves *1* and *2*). These curves were calculated with the same values of the parameters of the theory as in Sec. 3. After that we used (6) and (5) to calculate the conjugate quantities (in the sense of the Kramers–Kronig relations) (curves *3* and *4*). Clearly, the curves *1* and *3* agree very well, as do the curves *2* and *4*. Thus, we can assume that Eqs. (5) and (6) are valid if they are applied to the ''effective'' quantities n^{eff} and κ^{eff}.

Note that both n^{eff} and κ^{eff} differ dramatically from the refractive indices for the ''plus'' and ''minus'' waves, which play an important role in the passage of light through a crystal plate, a result corroborated by experimental measurements (cf. Figs. 1 and 4a). Hence using these values in calculations of such fundamental optical characteristics as ε' and ε'' results in an error. If formally, i.e., by analogy with the one-wave theory, we write $(\varepsilon'')^{\text{eff}}=2n^{\text{eff}}\kappa^{\text{eff}}$, there is no way in which this quantity can describe energy losses in the crystal. For instance, in the $\Gamma=0$ limit, there is no light absorption, while the value of $(\varepsilon'')^{\text{eff}}$ is finite and large. Nevertheless, the relationship between $(\varepsilon')^{\text{eff}}=(n^{\text{eff}})^2-(\kappa^{\text{eff}})^2$ and $(\varepsilon'')^{\text{eff}}$ is still described by (3) and (4).

Next we used Pekar's theory and the ''effective'' quantities n^{eff} and κ^{eff} to calculate the reflection spectrum $R^P(\omega)$ and the spectrum of variations of the reflected-wave phase, $\theta^P(\omega)$ (curves *1* and *2* in Fig. 4b). We applied the Kramers–Kronig relations (8) and (7) to these spectra and calculated $R^K(\omega)$ and $\theta^K(\omega)$, respectively (curves *4* and *3*). The good agreement of both pairs of curves in Fig. 4b is an indication that Eqs. (7) and (8) are valid if they are applied to the spectra R and θ calculated with n^{eff} and κ^{eff}.

However, as discovered in Refs. 37–39, the shape of

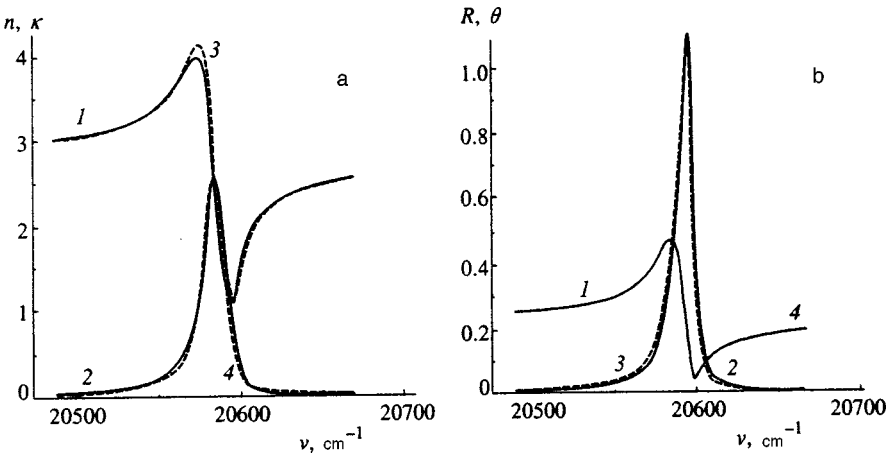

FIG. 4. a) The Pekar functions $n^{\text{eff}}(\omega)$ (curve *1*) and $\kappa^{\text{eff}}(\omega)$ (curve *2*) and similar functions calculated by (5) and (6) (curves *3* and *4*). b) The Pekar functions $R(\omega)$ (curve *1*) and $\theta(\omega)$ (curve *2*) and the same functions calculated by (7) and (8) (curves *3* and *4*). The crystal surface has no dead layer.

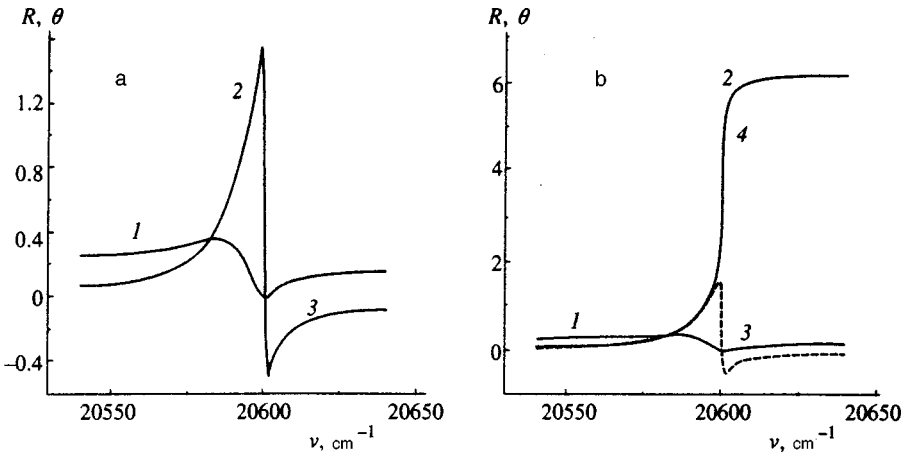

FIG. 5. Curves representing the functions $\theta^P(\omega)$ (curves *2*) and the phase spectra $\theta^K(\omega)$ calculated on the basis of $R^P(\omega)$ (curves *1*) by (7) (curves *3*) or by the modified dispersion relation (16) (curve *4*): a) $\Gamma > \Gamma^{\lim}$, and b) $\Gamma < \Gamma^{\lim}$. The crystal surface contains a dead layer.

experimental phase curves of reflection, $\theta(\omega)$, differs from that in Fig. 4b. The reason for this is the presence of a excitonless ''dead layer'' at the surface of the crystal. In this case the curve representing $\theta(\omega)$ is either *S*-shaped or *N*-shaped, depending on Γ (cf. the curves in Figs. 5b and 5a). A detailed analysis of the applicability of the Kramers–Kronig relations (7) and (8) to reflection spectra can be found in Refs. 15–19. There it was shown that for small values of Γ (low temperatures) one should use not the ordinary dispersion relations but dispersion relations with additional terms:

$$\theta(\omega)=\frac{\omega}{\pi}\int_0^{\infty}\frac{\ln[R_0/R(x)]}{x^2-\omega^2}dx+2\arctan\frac{\Gamma^{\lim}-\Gamma}{\omega^{\lim}-\omega}+a, \tag{16}$$

$$\ln\frac{R(\omega)}{R_0}=\frac{4}{\pi}\int_0^{\infty}\frac{x[\theta(x)-a]}{x^2-\omega^2}\,dx+2\ln\left[1+\left(\frac{\Gamma^{\lim}-\Gamma}{\omega^{\lim}-\omega}\right)^2\right]. \tag{17}$$

Here R_0 is the reflection coefficient far from resonance; $\Gamma^{\lim}$ is what is known as the damping-constant limit, at which *S*-shaped phase curves become, with increasing Γ, *N*-shaped; $\omega^{\lim}$ is the frequency at which the additional term in (16) becomes equal to $\pm\pi$; and $a=0$ for $\omega<\omega^{\lim}$ and $a=2\pi$ for $\omega>\omega^{\lim}$. When $\Gamma>\Gamma^{\lim}$, the additional terms in (16) and (17) should be discarded, which automatically leads to (7) and (8).

As in the previous case (see Fig. 4b), we carried out comparative calculations of the spectra in the 1*A*-exciton region of the CdS crystal with allowance for the dead layer (Fig. 5). The thickness of the dead layer was set to 70 Å, which is close to the value expected from theoretical calculations and the values obtained in Refs. 37–39. By varying Γ we found the values of $\Gamma^{\lim}$ and $\omega^{\lim}$ from the transition from *S*- shaped curves to *N*-shaped. The other parameters in the theory are the same as in Fig. 4. First we calculated the spectra $R^P(\omega)$ and $\theta^P(\omega)$ in Pekar's theory with a dead layer. Then, on the basis of the reflection spectrum $R^P(\omega)$ we calculated the corresponding phase curve $\theta^K(\omega)$ either by (7) or by (16) and compared the results with Pekar's phase curve $\theta^P(\omega)$.

The results of such comparison for two values of Γ, one larger than $\Gamma^{\lim}=1.21\,\mathrm{cm}^{-1}$ and the other smaller than $\Gamma^{\lim}=1.21\,\mathrm{cm}^{-1}$, are depicted in Figs. 5a and 5b, respectively. We see that when there is a dead layer and $\Gamma>\Gamma^{\lim}$, the relationship between the phase and amplitude curves are described fairly well by the ordinary Kramers–Kronig relation (7). However, here the *N*-shaped dependence has a negative section due to the interference of the waves reflected from the crystal surface and the boundary of the dead layer, so that the optical constants n and κ defined by the ordinary formulas (9) and (10) prove to be incorrect. First, the κ vs. ω curve in this case automatically also acquires a negative section, which has no physical meaning. Second, the absolute values obtained are incorrect.

When $\Gamma<\Gamma^{\lim}$, the dispersion relations derived by Moskovskiĭ and Solov'ev are satisfied fairly well, since the *S*-shaped phase curves *2* and *4* coincide. But determining the optical constants and the spectrum of the phase of the reflected wave may lead to even greater errors. More precisely, if formula (7) is applied to the reflection spectrum of an unknown substance in order to calculate the optical constants of this substance, a method widely used in physics, not only the values of n, κ, ε', and ε'' prove to be incorrect, but also the phase curve *3* differs substantially from the true curve *2*.

A remark is in order at this point. The phase curve calculated from the reflection spectrum by (7) (curve *3* in Fig. 5b) is *N*-shaped and has a negative section. On the other hand, sometimes in experiments involving the calculation of $\theta(\omega)$ from $R(\omega)$, sections with negative values of the phase curve also emerge. We believe that this is an indication that there is either a dead layer at the crystal surface or simple an ordinary layer with modified optical constants. The phase curve should not be ''lifted up'' arbitrarily to extinguish the region with negative values, as is sometimes done.

5. ANALYSIS OF THE REFLECTION SPECTRUM IN THE OSCILLATORY MODEL

Since the frequency curves of the effective optical characteristics of Pekar's theory, used to calculate the spectra $R(\omega)$ and $\theta(\omega)$, are similar to those of a classical oscillator, we did the necessary calculations for this case, too. The calculation parameters were the same as in Secs. 3 and 4, but the exciton mass was assumed infinite. It was found that $\Gamma^{\lim}=3.21\,\mathrm{cm}^{-1}$ in this case. We calculated the energy and

phase reflection spectra with and without a dead layer at the surface of the hypothetical crystal. It appeared that *S*- and *N*-shaped phase curves are also obtained for an ordinary oscillator in the case where there is a dead layer. Thus, it is the dead layer (rather than spatial-dispersion effects) that is responsible for the shape of the θ vs. ω curves.

As shown in Ref. 15, the reason *N*-shaped phase curves become *S*-shaped is the presence of a zero point in the $R(\omega)$ spectrum. In the complex $\tilde{\omega}=\omega'+i\omega''$ plane this is a singular point, and in integrating along the contour C with $\Gamma<\Gamma^{\lim}$ the contour must encircle this point, which leads to an additional term in (7) and the need to add 2π at frequencies $\omega>\omega^{\lim}$. Thus, the Kramers–Kronig relation (7) becomes the additional dispersion relation (16). At the frequency $\omega^{\lim}$ with $\Gamma=\Gamma^{\lim}$ the R vs. ω curve passes through zero.

When there is absorption in the medium, only an interference effect may be the reason for the R vs. ω curve passing through zero. Such an effect may occur when there is spatial dispersion in the medium or when there is no spatial dispersion, as in the case of an oscillator. Hence *N*-shaped phase curves with a negative section are possible even without spatial dispersion, but in the presence of a dead layer, i.e., at high temperatures and in media where there are sure to be no spatial-dispersion effects. *S*-shaped curves are possible only at low temperatures and small $\Gamma<\Gamma^{\lim}$, while the presence of spatial dispersion in ε is not obligatory.

6. ANALYSIS OF THE EXPERIMENTALLY MEASURED REFLECTION SPECTRUM

To conclude our investigation, we use the above analysis in applying the Kramers–Kronig relations to the experimentally measured spectrum $R^e(\omega)$ in the $1A$-excitonregion of the CdS crystal at 4.2 K. The literature devoted to such spectra is vast, and the results have been generalized by Kiselev *et al.*[40] Different crystal samples of the same substance in similar conditions may have diverse reflection spectra, which differ in additional structures (spikes) and in the span of the R vs. ω curve. We selected a typical structureless spectrum, which was used in Ref. 23 to compare with the reflection spectrum calculated by the Fresnel formula. In Ref. 23 this spectrum was approximated fairly well by a dependence calculated in Pekar's theory without a dead layer at the crystal surface. In the present paper we used Eq. (7) for a better approximation and found that the phase curve has a negative section. This is proof that there is a dead layer at the surface of this crystal. By setting the thickness of this layer to 70 Å and by varying the damping constant we achieved the best possible agreement between the calculated and experimental R vs. ω curves (Fig. 6).

In our approximation we found that it is highly important to allow for the $1B$-exciton, since it changes the short-wavelength limit behavior of the dispersion $n(\omega)$. One stage in the fitting was the search for the best agreement between the phase curves calculated by (7) from $R^e(\omega)$ and from $R^P(\omega)$. Our goal was to match the spans of the negative sections in the phase curves, and this was achieved at $\Gamma=0.7\,\mathrm{cm}^{-1}$. It is important to note that the two theoretically calculated curves correspond to no real variation of phase in scattering, since at this value of Γ, which is much smaller than $\Gamma^{\lim}$, the phase curve must be calculated by the modified relation (16) and be *S*-shaped, a case actually observed with samples used by Solov'ev and Babinskiĭ[37] and Pevtsov *et al.*[39]

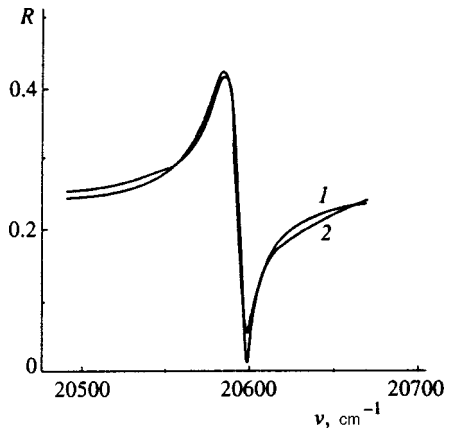

FIG. 6. The experimentally measured reflection spectrum (curve *1*) and the reflection spectrum calculated in Pekar's theory with a dead layer at the crystal surface (curve *2*).

Thus, the crucial information that can be obtained when (7) is applied to the reflection spectrum is the indication that there is a dead layer at the crystal surface. However, the true phase curve cannot be obtained without additional study. The reader will recall that finding the parameters of the theory only from the spectrum $R(\omega)$ is an ambiguous procedure, since very close curves can be obtained for different combinations of the parameters (see also Ref. 40). Only by finding Γ from other data and comparing it with $\Gamma^{\lim}$ and Γ_{cr} can we be sure that the phase curve and the optical constants of the substance are really the true ones.

7. CONCLUSION

1. We have compared the experimentally measured curves $n(\omega)$ and $\kappa(\omega)$ for a CdS crystal within the range of the ground exciton states at 4.2 K with the respective conjugate quantities calculated on the basis of these curves by the Kramers–Kronig relations. The two groups of curves are shown to be very different.

2. We have found that the relations (5), (6) and (3), (4) are true only when $\Gamma>\Gamma_{\mathrm{cr}}$ and only for a classical wave (a ''plus–minus'' wave). In all other cases these relations do not hold: when $\Gamma>\Gamma_{\mathrm{cr}}$, the relations do not hold for a nonclassical wave (a ''minus–plus'' wave), and when $\Gamma<\Gamma_{\mathrm{cr}}$, the relations do not hold for a ''plus'' or ''minus'' wave separately or for a combination of such waves, including the classical one.

3. All the Kramers–Kronig relations, (3)–(8), hold if they are applied to the effective quantities n^{eff} and κ^{eff}, which are parameters in Pekar's theory and can be expressed in a complicated way in terms of the quantities $n_{\pm}(\omega)$ and $\kappa_{\pm}(\omega)$ of true waves propagating in the crystal. Here the formally written quantity $(\varepsilon'')^{\mathrm{eff}}=2n^{\mathrm{eff}}\kappa^{\mathrm{eff}}$ does not reflect the real energy losses in the medium.

4. When there is a dead layer at the crystal surface, the relationship between the reflection phase and energy curves are described well either by the ordinary relation (7) when the phase curve is N-shaped($\Gamma > \Gamma^{\lim}$) or by the additional dispersion relation (16) when the phase curve is S-shaped ($\Gamma < \Gamma^{\lim}$). However, the optical constants of a substance obtained from the spectra $R(\omega)$ and $\theta(\omega)$ by ordinary calculation methods (without allowing for a dead layer or spatial dispersion) will have incorrect values.

5. We have shown that S- and N-shaped phase curves can be obtained even if there is no spatial dispersion of ε, i.e., for a classical oscillator, if there is a dead layer at the crystal surface.

6. Finally, on the basis of Pekar's theory via the Kramers–Kronig relation (7) we have approximated the experimentally measured reflection spectrum $R(\omega)$ in the $1A$-exciton region of the CdS crystal at 4.2 K.

ACKNOWLEDGMENTS

We would like to express our gratitude to V. N. Piskavoĭ and V. I. Sheka for numerous and productive discussions of the results.

*)E-mail: class@class.semicond.kiev.ua

[1] S. I. Pekar, Zh. Éksp. Teor. Fiz. **33**, 1022 (1957) [Sov. Phys. JETP **6**, 785 (1958)].
[2] S. I. Pekar, Zh. Éksp. Teor. Fiz. **34**, 1176 (1958) [Sov. Phys. JETP **7**, 813 (1958)].
[3] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).
[4] H. A. Kramers, Atti Congr. Int. Fis. Como **2**, 545 (1927).
[5] E. de L. Kronig, J. Opt. Soc. Am. Rev. Sci. Instrum. **12**, 547 (1926).
[6] T. Moss, G. Burrell, and B. Ellis, *Semiconductor Optoelectronics*, Butterworths, London (1973).
[7] H. Nussenzweig, *Causality and Dispersion Relations*, Academic Press, New York (1972).
[8] V. V. Sobolev and V. V. Nemoshkalenko, *Methods of Computational Physics in Solid-State Theory*, Naukova Dumka, Kiev (1988).
[9] Diploma for Discovery No. 323: ''The phenomenon of propagation of additional light waves (Pekar waves) in crystals,'' S. I. Pekar, OT-11003; application of September 27, 1984; published August 30, 1987; *Otkrytiya. Izobreteniya* [in Russian], No. 32, 2 (1987).
[10] M. A. Leontovich, Zh. Éksp. Teor. Fiz. **40**, 907 (1961) [Sov. Phys. JETP **13**, 634 (1961)].
[11] A. S. Davydov, Zh. Éksp. Teor. Fiz. **43**, 1832 (1962) [Sov. Phys. JETP **16**, 1293 (1963)].
[12] C. A. Mead, Radiat. Res. **20**, 101 (1963).
[13] V. L. Ginzburg and H. H. Meĭman, Zh. Éksp. Teor. Fiz. **46**, 243 (1964) [Sov. Phys. JETP **19**, 169 (1964)].
[14] V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion and Excitons*, 2nd ed., Springer-Verlag, Berlin (1984).
[15] S. B. Moskovskiĭ and L. E. Solov'ev, Zh. Éksp. Teor. Fiz. **86**, 1419 (1984) [Sov. Phys. JETP **59**, 831 (1984)].
[16] S. B. Moskovskiĭ, A. B. Novikov, and L. E. Solov'ev, Fiz. Tverd. Tela (Leningrad) **30**, 1431 (1988) [Sov. Phys. Solid State **30**, 826 (1988)].
[17] T. Musienko, V. Rudakov, and L. Solov'ev, J. Phys.: Condens. Matter **1**, 6745 (1989).
[18] S. B. Moskovskiĭ, A. B. Novikov, O. S. Omegov, and L. E. Solov'ev, Fiz. Tverd. Tela (Leningrad) **33**, 657 (1991) [Sov. Phys. Solid State **33**, 375 (1991)].
[19] S. B. Moskovskiĭ, A. B. Novikov, and L. E. Solov'ev, Zh. Éksp. Teor. Fiz. **105**, 994 (1994) [JETP **78**, 533 (1994)].
[20] M. S. Brodin, A. F. Prikhot'ko, and M. S. Soskin, Opt. Spektrosk. **6**, 28 (1959).
[21] M. S. Brodin and M. I. Strashnikova, Fiz. Tverd. Tela (Leningrad) **4**, 2454 (1962) [Sov. Phys. Solid State **4**, 1798 (1963)].
[22] M. I. Strashnikova, Fiz. Tverd. Tela (Leningrad) **17**, 729 (1975) [Sov. Phys. Solid State **17**, 467 (1975)].
[23] S. I. Pekar and M. I. Strashnikova, Zh. Éksp. Teor. Fiz. **68**, 2047 (1975) [Sov. Phys. JETP **41**, 1024 (1975)].
[24] M. S. Brodin, N. A. Davydova, and M. I. Strashnikova, JETP Lett. **19**, 297 (1974).
[25] M. I. Strashnikova and A. T. Rudchik, Fiz. Tverd. Tela (Leningrad) **14**, 984 (1972) [Sov. Phys. Solid State **14**, 845 (1972)].
[26] V. Ya. Reznichenko, M. I. Strashnikova, and V. V. Cherny, Phys. Status Solidi B **152**, 675 (1989).
[27] M. V. Lebedev, M. I. Strashnikova, V. B. Timofeev, and V. V. Chernyĭ, JETP Lett. **39**, 440 (1984).
[28] A. A. Demidenko, M. V. Lebedev, S. I. Pekar, M. I. Strashnikova, V. B. Timofeev, and B. E. Tsekvava, Zh. Éksp. Teor. Fiz. **89**, 330 (1985) [Sov. Phys. JETP **62**, 185 (1985)].
[29] M. I. Strashnikova and V. V. Chernyĭ, Fiz. Tverd. Tela (Leningrad) **32**, 1090 (1990) [Sov. Phys. Solid State **32**, 640 (1990)].
[30] I. V. Makarenko, I. N. Uraltsev, and V. A. Kiselev, Phys. Status Solidi B **72**, 161 (1975).
[31] A. S. Davydov and E. N. Myasnikov, Phys. Status Solidi B **68**, 325 (1974).
[32] M. I. Strashnikova and E. V. Bessonov, Zh. Éksp. Teor. Fiz. **74**, 2206 (1978) [Sov. Phys. JETP **47**, 1148 (1978)].
[33] N. N. Akhmediev, Zh. Éksp. Teor. Fiz. **79**, 1534 (1980) [Sov. Phys. JETP **52**, 773 (1980)].
[34] S. I. Pekar, Zh. Éksp. Teor. Fiz. **36**, 451 (1959) [Sov. Phys. JETP **9**, 314 (1959)].
[35] S. I. Pekar, Zh. Éksp. Teor. Fiz. **38**, 1786 (1960) [Sov. Phys. JETP **11**, 1286 (1960)].
[36] S. I. Pekar, *Crystal Optics and Additional Light Waves*, W. A. Benjamin, Reading, MA (1983).
[37] L. E. Solov'ev and A. V. Babinskiĭ, JETP Lett. **23**, 263 (1976).
[38] A. V. Komarov, S. M. Ryabchenko, and M. I. Strashnikova, Zh. Éksp. Teor. Fiz. **74**, 251 (1978) [Sov. Phys. JETP **47**, 128 (1978)].
[39] A. B. Pevtsov, S. A. Permogorov, Sh. R. Saĭfullaev, and A. V. Sel'kin, Fiz. Tverd. Tela (Leningrad) **22**, 2400 (1980) [Sov. Phys. Solid State **22**, 1396 (1980)].
[40] V. A. Kiselev, B. V. Novikov, and A. E. Cherednichenko, *Excitonic Spectroscopy of the Surface Region of Semiconductors*, Leningrad Univ. Press, Leningrad (1987).

Translated by Eugene Yankovsky

Photoinduced phase transitions in a Peierls system

A. L. Semenov*)

Ul'yanovsk State University, 432700 Ul'yanovsk, Russia
(Submitted 13 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1407–1420 (October 1998)

A theory is constructed that explains photoinduced phase transitions in a Peierls system being irradiated by light with a finite width of the optical spectrum and a central frequency close to the upper van Hove singularity of the first kind in the combined density of electron states. The electron spectrum and the matrix elements of the dipole-moment operator are calculated by Bogolyubov's method of canonical transformations. The interaction with the light is described by the Liouville equation for the density matrix of the electron subsystem in the dipole approximation. The light field is considered a quasimonochromatic time-independent random process with a Lorentzian spectrum. The derived equations are analyzed for two limits: (1) when the width of the optical spectrum tends to zero (a monochromatic light field), and (2) when the width of the optical spectrum is close to the upper limit (a bifurcation point) at which a photoinduced phase transition can still be observed. An existence criterion for such a transition is obtained, and the main parameters of the transition (the critical points and the size of the hysteresis loop) are calculated. The broadening of the optical spectrum of the incident light is shown to narrow the range of values of the central frequency of the light field and to reduce the size of the hysteresis loop. Finally, near the phase transition point, cavityless optical bistability sets in in the system, with light absorption increasing in the process.
 [S1063-7761(98)01710-7]

1. INTRODUCTION

The Peierls system examined in this paper is a one-dimensional chain of atoms each of which contains one outer electron. In the high-temperature metallic phase, the atoms in the chain are positioned equidistantly and the conduction band is half-filled. When the system is cooled and the temperature reaches a critical value T_0, the metallic phase becomes unstable and a metal–semiconductor phase transition occurs. As a result the chain atoms move closer together in a pairwise fashion.[1]

The theoretical results obtained in the Peierls model are used to describe the experimentally observed properties of many quasi-one-dimensional materials: complex platinum compounds,[1] TaS_3 compounds,[2] organic conductors,[3] oxide vanadium bronzes,[4] $NbSe_3$ and materials.[5–7] In particular, this model makes it possible to explain the effect on the metal–semiconductor transition in vanadium dioxide of such factors as uniaxial pressure,[8] alloying,[9,10] phonon–phonon coupling,[8,11] cohesion of a vanadium dioxide film and the substrate,[12] adsorption, and other factors.[13]

As is known, in a constant electric field with a certain threshold field strength E_t the low-temperature phase of the Peierls system undergoes a thermodynamically nonequilibrium phase transition to a state with a traveling charge density wave.[5] Depending on the conditions, the transition is either first-order or second-order. For instance, the current–voltage characteristic of $NbSe_3$ at $T=26.5$ K has a hysteresis dependence characteristic of first-order phase transitions, while at $T=34$ K there is no hysteresis dependence.[5]

The behavior of the low-temperature phase of the Peierls system in a light field that excites nonequilibrium electron–hole pairs in the system has been studied both theoretically[14] and experimentally (in vanadium dioxide).[15,16] It was found that because of strong electron–phonon coupling, as the electron–hole pair concentration increases the band gap narrows, and at high excitation levels there occurs a semiconductor–metal phase transition.[6]

Note that the narrowing of the band gap in the electron spectrum and the decrease of the frequency of one or several phonon modes down to zero as the concentration of nonequilibrium light-excited electrons grows is a fairly general property of a broad class of material media of various dimensionalities.[17–23] Here, near the point of the photoinduced semiconductor–phase transition accompanied by changes in the crystal lattice stricture, one can observe spatial[17–19] and temporal[20–22] periodic modulation of the parameters of the system.

This paper studies the behavior of the low-temperature phase of the Peierls system in a quasimonochromatic light field whose frequency is somewhat lower than the upper van Hove singularity of the first kind,[24] at which the combined density of electron states in direct optical band-to-band transitions becomes infinite (see Fig. 1 below).

Thanks to the strong electron–phonon coupling, the frequency that corresponds to the van Hove singularity decreases as the electron–hole pair concentration increases.[9] In such conditions, when the irradiation intensity increases, there may be a sudden narrowing of the band gap in the system's spectrum (a photoinduced semiconductor–semiconductor phase transition). The reason for this is positive feedback. The light field drives the electron–hole pair con-

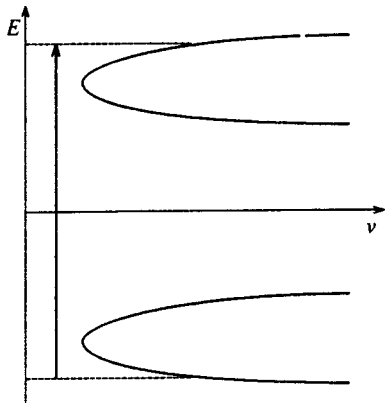

FIG. 1. Dependence of the density ν of electron states of the Peierls system on the energy E (Eq. (27)). The vertical arrow designates the optical band-to-band transitions (considered in this paper) with a frequency somewhat lower than that corresponding to the upper van Hove singularity of the first kind, where the combined density of electron states for direct optical transitions becomes infinite.

centration up and, as a result, the frequency corresponding to the van Hove singularity of the first kind decreases, approaching the light frequency. Thus, the electron–hole pair generation rate increases dramatically, which leads to a further decrease of the frequency of the van Hove singularity. This positive feedback reduces the stability of the system. If the positive feedback becomes strong enough, stability can be lost, with a sudden transition to a new state of equilibrium (a phase transition).

Thus, thanks to the strong coupling between the electron subsystem and the static phonon mode in the Peierls model,[8] one of the mechanisms causing a photoinduced phase transition and cavityless optical bistability with increasing absorption comes into play.[25] Generally, the given type of bistability is due to the nonlinear dependence of the absorption coefficient on the incident radiation intensity and may be brought on by different causes:[25] atomic correlations within a small volume, the temperature dependence of the band gap, renormalization of the band gap due to electron–phonon coupling,[25,26] and variations in the frequency of transitions in an ensemble of two-level atoms due to local-field effects or electron–phonon coupling.[27] However, the author knows of no mechanism, similar to that considered in this paper, of a photoinduced phase transition and cavityless optical bistability in the Peierls system irradiated by light field whose photons have an energy somewhat lower than the width of the total Peierls band.

2. ELECTRON SPECTRUM AND DIPOLE MOMENT OF THE SYSTEM

Consider a chain of atoms each of which has one outer electron. The Hamiltonian of the electron subsystem in the tight-binding approximation can be written as follows:[1]

$$H=\sum_n B_{n,n+1}(a_n^\dagger a_{n+1}+a_{n+1}^\dagger a_n), \tag{1}$$

where n is the number of the atom in the chain, $B_{n,n+}$-- is the overlap integral of the wave functions of neighboring electrons, and $a_n^\dagger$ and a_n are the operators of electron creation and annihilation at the nth atom.

For narrow-gap systems, e.g., for the Peierls model, the separation of adjacent atoms, $r_{n,n+1}$, exceeds the effective radius R of the atomic wave function of an electron several-fold. In this case $B_{n,n+1}\propto\exp(-r_{n,n+1}/R)$ (Ref. 28). The separation $r_{n,n+1}$ can be written

$$r_{n,n+1}=r_0+(-1)^{n+1}R\xi, \tag{2}$$

with r_0 the atomic separation in the metallic phase, and ξ the period-doubling parameter for a one-dimensional crystal. If we allow for (2), the overlap integral $B_{n,n+1}$ assumes the form

$$B_{n,n+1}=b\ \exp[(-1)^n\xi], \tag{3}$$

where b is the overlap integral in the metallic phase ($\xi=0$). With the Hamiltonian (1), the phases of the atomic wave functions are selected in such a way that b in (3) is a real quantity.

By employing Bogolyubov's method of canonical transformations[29] we reduce the Hamiltonian (1) to the diagonal form

$$H=\sum_k \varepsilon_k\alpha_k^\dagger\alpha_k, \tag{4}$$

where

$$\varepsilon_k=2b\ \mathrm{sign}(\cos k)\sqrt{\cos^2 k+\sinh^2\xi}, \tag{5}$$

α_k and $\alpha_k^\dagger$ are new collective second-quantization Fermi operators, and $k=0,\pm 2\pi/N,\ \ldots,\pm\pi$, $\alpha_{k+2\pi}=\alpha_k$, with N the number of atoms in the chain.

Equation (5) shows that for $\xi\neq 0$ the spectrum of ε_k has two bands, with the lower band in the ground state completely occupied and the upper band vacant (the semiconducting phase). At $\xi=0$ the spectrum (5) consists of one half-filled band (the metallic phase).

In the tight-binding approximation, the dipole moment operator of the Peierls system is

$$\mathbf{d}=\sum_n (\mathbf{d}_{n,n+1}a_n^\dagger a_{n+1}+\mathbf{d}_{n,n+1}^* a_{n+1}^\dagger a_n), \tag{6}$$

where the dependence of $\mathbf{d}_{n,n+1}$ on ξ is similar to (3):

$$\mathbf{d}_{n,n+1}=(\mathbf{d}_1+i\mathbf{d}_2)\exp[(-1)^n\xi]$$
$$=-e\int_v \psi_n^*(\mathbf{r})\mathbf{r}\psi_{n+1}(\mathbf{r})\,dv. \tag{7}$$

Here $\psi_n(\mathbf{r})$ is the atomic wave function of the electron at the nth site, and e is the electron charge. By selecting the phases of the wave functions $\psi_n(\mathbf{r})$ in such a way that the overlap integral (3) is real we ensure that both $\mathbf{d}_1$ and $\mathbf{d}_2$ in (7) are uniquely defined.

By introducing the Fermi operators α_k and α_k^+ into (6) we obtain

$$\mathbf{d}=\sum_k \left\{ \left[\frac{\mathbf{d}_1\varepsilon_k}{b} - \frac{2\mathbf{d}_2}{1+\varphi_k^2}[(1-\varphi_k^2)\cosh\,\xi\,\sin\,k + 2\varphi_k\,\sinh\,\xi\,\cos\,k] \right] \alpha_k^\dagger\alpha_k + i\frac{2\mathbf{d}_2}{1+\varphi_k^2}[(1-\varphi_k^2) \times \sinh\,\xi\,\cos\,k - 2\varphi_k\,\cosh\,\xi\,\sin\,k]\alpha_k^\dagger\alpha_{k-\pi} \right\}, \quad (8)$$

where

$$\varphi_k=\frac{\cosh\,\xi\,\cos\,k-\mathrm{sign}(\cos\,k)\,\sqrt{\cos^2 k+\sinh^2\xi}}{\sinh\,\xi\,\sin\,k}. \quad (9)$$

Note the formal similarity of the operators (1) and (6) at $\mathbf{d}_2=0$. The operators (4) and (8) are also formally similar.

Suppose that in the absence of an external electric field the total dipole moment of the system is zero. Then Eqs. (8) and (5) imply that $\mathbf{d}_1=0$. Thus, the fact that we have chosen the phases of the wave functions $\psi_n(\mathbf{r})$ in such a way that the integral $B_{n,n+1}$ in (1) is real ensures, in the present case, that the interstitial dipole-moment matrix element $\mathbf{d}_{n,n+1}$ in (6) is imaginary. The case of $\mathbf{d}_1\neq 0$ can probably be observed in systems exhibiting ferroelectric properties, but we will not consider such systems here. As $\xi\to 0$, as Eq. (9) indicates, $\varphi_k\to 0$ for all $k\neq\pm\pi/2$, with the result that $\mathbf{d}_{k,k-\pi}\to 0$ in (8) and all dipole transitions are forbidden. If $\xi\neq 0$, then $\mathbf{d}_{k,k-\pi}\neq 0$ in (8), and the corresponding dipole transitions are allowed. Since in this case the interval $k\in(-\pi/2,\pi/2)$ is the first Brillouin zone, the given transitions in the spectrum (5) are vertical band-to-band transitions.

3. EQUILIBRIUM EQUATION OF A THERMODYNAMICALLY NONEQUILIBRIUM PEIERLS SYSTEM

In describing the behavior of the Peierls system in a light field we assume that the electron intraband relaxation time is much shorter than the band-to-band relaxation time τ (Ref. 30). In this case, following the approximation adopted in Refs. 14 and 24, we can assume that the electron statistics in each band is described by a Fermi distribution with a specific Fermi quasilevel. Then, allowing for the symmetry of (5), we arrive at an expression for the concentration n of electron–hole pairs in a light field:

$$n=\frac{N}{2}-\sum_{|k|\leqslant\pi/2}\tanh\frac{\varepsilon_k-\mu}{2k_BT}, \quad (10)$$

where μ is the Fermi quasilevel of the conduction band (in view of the symmetry of the dispersion law (5) the Fermi quasilevel of the valence band has the opposite value), and k_B is Boltzmann's constant.

Treating the parameter ξ as a generalized coordinate, we can write the equilibrium equation for the Peierls system (within each band the electron gas is in thermodynamic equilibrium, but thermodynamic equilibrium between the bands is violated):

$$\left(\frac{\partial F_1}{\partial\xi}\right)_{T,N_1}+\left(\frac{\partial F_2}{\partial\xi}\right)_{T,N_2}+\left(\frac{\partial F_{\mathrm{ph}}}{\partial\xi}\right)_T=0, \quad (11)$$

where

$$F_j=\mu_jN_j-k_BT\sum_k\ln\left[1+\exp\left(\frac{\mu_j-\varepsilon_k}{k_BT}\right)\right], \quad (12)$$

$$F_{\mathrm{ph}}=\frac{A}{2}\xi^2,$$

with F_j, μ_j, and N_j the electron free energy, the Fermi quasilevel, and the number of electrons belonging to the jth band ($j=1,2$), and A the coefficient in the expansion of the free energy of the phonon subsystem, F_{ph}, in powers of ξ (see Ref. 1). Summing over k in (12) must be done within the jth band of the spectrum (5).

Plugging (12) into (11) and performing the necessary transformations, we finally get

$$-2\sum_{|k|\leqslant\pi/2}\frac{\partial\varepsilon_k}{\partial\xi}\tanh\frac{\varepsilon_k-\mu}{2k_BT}+A\xi=0. \quad (13)$$

The expansion coefficient A can be expressed in terms of the critical temperature T_0 of a thermodynamically equilibrium metal–semiconductor transition (in the absence of a light field) and other characteristics of the system:[8,11]

$$A=\frac{4bN}{\pi}\left[\ln\frac{\pi b}{2k_BT_0}+1\right]. \quad (14)$$

Thus we have obtained an equation (Eq. (10)) expressing the dependence of the concentration n of electron–hole pairs on the Fermi quasilevel μ and a parameter ξ, and an equilibrium equation (Eq. (13)) describing the relationship between the parameter ξ of the Peierls system and μ. The Fermi quasilevel μ is determined in turn by the degree to which light affects the system. An equation describing this effect must depend on the specific mechanism of interaction of the radiation and the electron subsystem.

4. INTERACTION WITH RADIATION

The interaction between the system and the light field is described by an operator V, which in the dipole approximation can be written as

$$V=-\mathbf{d}\cdot\mathbf{E}(t)=-\mathbf{d}\cdot\int\mathbf{E}_\omega\,\exp(-i\omega t)\,d\omega, \quad (15)$$

where $\mathbf{E}_\omega$ and ω are the amplitude and frequency of a spectral component of the light field. We assume the incident radiation $\mathbf{E}(t)$ to be a quasimonochromatic time-independent random process linearly polarized along the crystal's axis.[31,32]

Using Liouville's equation[33]

$$i\hbar\frac{\partial\rho}{\partial t}=[H+V,\rho], \quad (16)$$

and allowing for Eqs. (5) and (15), we arrive at a kinetic equation for the concentration $n=2\Sigma_{|k|\leqslant\pi/2}\rho_{kk}$ of the electron–hole pairs in second-order perturbation theory:

$$\frac{\partial n}{\partial t}=\frac{4\pi}{\hbar^2}\sum_{|k|\leqslant\pi/2}d_k^2\,\tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right)G\left(\frac{2\varepsilon_k}{\hbar}\right)-\frac{n-n_0}{\tau}, \quad (17)$$

where $n_0=n(\mu=0)$ is the concentration of electron–hole pairs in the absence of a light field,

$$d_k=\frac{2d_2}{1+\varphi_k^2}|(1-\varphi_k^2)\sinh\,\xi\,\cos\,k-2\varphi_k\,\cosh\,\xi\,\sin\,k|, \tag{18}$$

and $G(\omega)$ is the spectral density of the light field, which for a quasimonochromatic signal can be written as follows:[31]

$$G(\omega)=Ig(|\omega|-\omega_0), \tag{19}$$

with ω_0 the carrier frequency, and $g(x)$ a bell-shaped function, with its maximum at $x=0$, satisfying the normalization condition $\int g(x)\,dx=1$. The width $\Delta\omega$ of the spectrum $G(\omega)$ satisfies the inequality $\Delta\omega\ll\omega_0$. The quantity $I=\frac{1}{2}\int G(\omega)\,d\omega$ is the intensity of the light field (in the Gaussian system of units, to within a factor $cn/2\pi$, where c is the speed of light, and n is the medium's refractive index).

The last term on the right-hand side of Eq. (17), which has been introduced phenomenologically, allows for band-to-band relaxation with a relaxation time τ. Equation (17) shows that the variation in the electron–hole pair concentration is due to the interaction of the electron subsystem and the spectral components of the light field with frequencies $\omega_k=2\varepsilon_k/\hbar$.

Combining (19) with Eq. (17), in the steady-state mode ($\partial n/\partial t=0$) we finally arrive at an expression for the intensity:

$$I=\frac{(n-n_0)\hbar^2}{4\pi\tau}\left[\sum_{|k|\leqslant\pi/2} d_k^2\,\tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right)g\left(\frac{2\varepsilon_k}{\hbar}-\omega_0\right)\right]^{-1}. \tag{20}$$

Equations (10), (13), and (20) constitute a complete set of equations with respect to the internal parameters μ, n, and ξ of the Peierls system with given external parameters T, I, ω_0, etc.

5. DEPENDENCE OF THE CONCENTRATION OF ELECTRON–HOLE PAIRS ON THE BAND GAP

Let us analyze Eqs. (10) and (13) for the case where the Peierls system is a nondegenerate semiconductor:

$$\frac{\varepsilon/2-\mu}{k_BT}>2, \tag{21}$$

where ε is the band gap. Physically, the condition (21) means that (a) $4k_BT<\varepsilon$, and (b) the concentration of excited electrons is so low that the Fermi quasilevel of the conduction band is in the band gap. Then, taking $\exp[-(\varepsilon/2-\mu)/k_BT]$ as a small parameter, we can approximately transform Eq. (10) for the concentration n of electron–hole pairs into

$$n=\frac{N}{2}\sqrt{\frac{\varepsilon k_BT}{\pi b^2}}\exp\left(-\frac{\varepsilon/2-\mu}{k_BT}\right). \tag{22}$$

We write the equilibrium equation (13) in the following form:

$$\exp\left(-\frac{\varepsilon/2-\mu}{k_BT}\right)=\sqrt{\frac{\varepsilon}{4\pi k_BT}}\left[\arcsin\frac{\pi}{4}+\ln\frac{2\pi b}{\varepsilon}-\frac{\pi A}{4bN}\right]. \tag{23}$$

Here in addition to assuming that condition (21) is met we assume that $\xi\leqslant 0.5$. This is justified because such values of ξ are realized in most materials described by the Peierls model.[1,6,8]

Plugging (23) into (22) and allowing for (14), we arrive at a relationship linking the concentration n of electron–hole pairs to the band gap ε:

$$n=\frac{N\varepsilon}{4\pi b}\,\ln\frac{\varepsilon_0}{\varepsilon}, \tag{24}$$

where ε_0 is the band gap in the absence of light ($\mu=0$) for $k_BT\ll\varepsilon$,

$$\varepsilon_0=4k_BT_0\,\exp[\arcsin(\pi/4)-1]. \tag{25}$$

Equation (24) is valid when $\Delta\varepsilon\equiv\varepsilon_0-\varepsilon$ is small, and the reason for this lies in the adopted approximation (21), which with allowance for (23) and (25) can be written in the more convenient form

$$\Delta\varepsilon\equiv\varepsilon_0-\varepsilon<\sqrt{\varepsilon_0k_BT}/2\,, \tag{26}$$

where, as in (25), it is assumed that $T<T_0/2$.

An analysis of the equilibrium equations (11)–(13)shows that the solution $n(\varepsilon)$ given by (24) is stable if $\partial n/\partial\varepsilon$ is negative, i.e., if $\varepsilon\in(\varepsilon_0/e,\varepsilon_0]$.

6. A PHOTOINDUCED PHASE TRANSITION IN A MONOCHROMATIC LIGHT FIELD

We will begin our analysis of Eq. (20) and (24) with the case of a monochromatic light field, with the form factor $g(x)=\delta(x)$. In (20) we replace the sum by an integral with respect to E, bearing in mind that the density of electron states $\nu(E)$ corresponding to the spectrum (5) has the form (Fig. 1)

$$\nu(E)=\frac{2NE}{\pi\sqrt{(4b^2\cosh^2\xi-E^2)(E^2-4b^2\sinh^2\xi)}}, \tag{27}$$

and that, according to (18), (5), and (9), the matrix element of the dipole-moment operator, d_k, at $\varepsilon_k=E$ is given by the formula

$$d(E)\equiv d_k(\varepsilon_k=E)=\frac{4bd_2\,\cosh\,\xi\,\sinh\,\xi}{E}. \tag{28}$$

Then at $\tanh[(\hbar\omega_0/2-\mu)/2k_BT]=1$ and $n_0=0$ we have the following I vs. ε dependence

$$I=\frac{\hbar^2\omega_0 b\,\ln(\varepsilon_0/\varepsilon)\,\sqrt{[16b^2+\varepsilon^2-(\hbar\omega_0)^2][(\hbar\omega_0)^2-\varepsilon^2]}}{8\tau d_2^2\varepsilon(16b^2+\varepsilon^2)}. \tag{29}$$

We will assume that the initial (at $n=0$) offset ΔE of the energy $\sqrt{16b^2+\varepsilon_0^2}$ of the van Hove singularity of the first kind from the incident-photon energy $\hbar\omega_0$ is small:

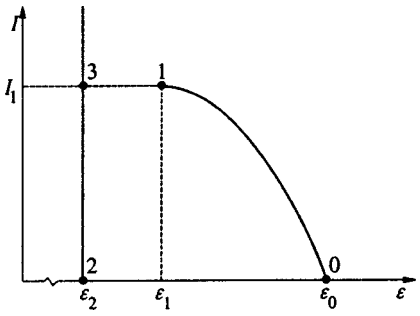

FIG. 2. Inverse dependence of the band gap ε on the intensity I of a monochromatic light field; $\varepsilon_1=(\varepsilon_0+2\varepsilon_2)/3$.

$$0<\Delta E\equiv\sqrt{16b^2+\varepsilon_0^2}-\hbar\omega_0\ll\hbar\omega_0. \tag{30}$$

In this case Eq. (29) can be approximately written as follows:

$$I=\frac{b^2\ \ln(\varepsilon_0/\varepsilon)\ \sqrt{\varepsilon-\varepsilon_2}}{\sqrt{2\varepsilon_0}\ \tau\omega_0 d_2^2}, \tag{31}$$

where

$$\varepsilon_2=\sqrt{(\hbar\omega_0)^2-16b^2}. \tag{32}$$

An analysis of the kinetic equation (17) shows that for the solution (31) to be stable $\partial I/\partial\varepsilon$ must be negative.

Figure 2 schematically represents I vs. ε for physically realizable systems. As the light intensity increases, a point on the graph moves along the curve $0\to1\to3$, while as the intensity decreases the point moves along the curve $3\to2\to0$. The jumps in the value of ε in the $1\to3$ and $2\to0$ sections corresponds to direct and inverse photoinduced phase transitions. The interval $I<I_1$ constitutes the bistability region.

The curve $0\to1$ is described by (31). At the point

$$\varepsilon_1=(\varepsilon_0+2\varepsilon_2)/3 \tag{33}$$

the solution (31) becomes unstable (since $\partial I/\partial\varepsilon=0$) and there is a transition to the states $\varepsilon\approx\varepsilon_2$ in which the frequency ω_0 exceeds the frequency corresponding to the van Hove singularity of the first kind by a small quantity, comparable to the spectral linewidth of the light field; it is assumed that for a monochromatic field the spectral linewidth is an extremely small (but finite) quantity compared to the other characteristic parameters of the problem.

This state is stable for all $I>0$, since $\nu(E)=\infty$ at the van Hove singularity. Indeed, under a fluctuation increase in ε the frequency corresponding to the van Hove singularity of the first kind also increases, approaching the frequency ω_0 of the light field. As a result the electron–hole pair concentration grows, which leads to a decrease in ε. Thus, negative feedback sets in in this case, and this ensures the stability of the system.

Plugging (33) into (31) and allowing for (32) and (30), we find the value I_1 of the intensity of the light field at which the phase transition occurs (see Fig. 2):

$$I_1=I(\varepsilon_1)=\sqrt{\frac{2}{27}}\ \frac{b^2\hbar^{3/2}\omega_0^{1/2}(\Delta E)^{3/2}}{\tau d_2^2\varepsilon_0^{5/2}}. \tag{34}$$

Thus, the intensity I_1 is proportional to $(\Delta E)^{3/2}$, which means that in experiments the intensity can be fairly low, so that there is no overheating of the sample.

7. PHOTOINDUCED PHASE TRANSITION IN A LIGHT FIELD WITH A FINITE WIDTH OF THE OPTICAL SPECTRUM

The possibility of observing a photoinduced phase transition at low light intensity means that we can use nonlaser sources of light, which do not emit monochromatic light. This means we must allow for the finiteness of the width of the optical spectrum of the incident radiation. Note that a separate investigation of the effect of nonmonochromaticity on the way in which a phase transition proceeds would also be of interest.

To solve Eq. (20) we must know the function $g(\omega)$ characterizing the spectral density of the light field. To this end we assume the $g(\omega)$ has the Lorentzian form[32]

$$g(\omega)=\frac{\tau_0}{\pi[1+(\omega\tau_0)^2]}, \tag{35}$$

where τ_0 is the reciprocal of the spectral halfwidth.

In (20) we replace the sum with an integral with respect to E, allowing for (24), (27), (28), and (35) and the approximate relationships $\tanh[(\hbar\omega_0/2-\mu)/2k_BT]=1$ and $n_0=0$. Here, as in Sec. 6, we limit ourselves to the case specified in (30), where the carrier frequency ω_0 of the light field is close to the initial frequency, which corresponds to the van Hove singularity of the first kind. Doing the necessary transformations, we arrive at an approximate I vs. ε dependence:

$$I=\frac{\hbar^{5/2}b^{3/2}|x_0|z}{\pi\tau_0^{3/2}\tau d_2^2\varepsilon_0^2}, \tag{36}$$

where

$$z(x)=\left(1-\frac{x}{x_0}\right)\sqrt{(1+x^2)(x+\sqrt{1+x^2}\,)}, \tag{37}$$

$$x=x(\varepsilon)=\tau_0\left(\omega_0-\frac{\sqrt{16b^2+\varepsilon^2}}{\hbar}\right), \tag{38}$$

with $x_0=x(\varepsilon_0)<0$ the initial dimensionless offset of the light-field frequency ω_0 from the frequency corresponding to the van Hove singularity of the first kind.

Equations (36)–(38) describe the relationship between the light-field intensity I and the band gap ε of the Peierls system in a stationary state of equilibrium. An analysis of the kinetic equation (17) shows that for the state to be stable $\partial I/\partial\varepsilon$ must be negative, which in a form more convenient for analysis can be expressed as

$$\frac{\partial z}{\partial x}<0. \tag{39}$$

Let us study the z vs. x dependence (37) in the range $|x|\leqslant1$. To this end we expand (37) in a Taylor series up to terms cubic in x:

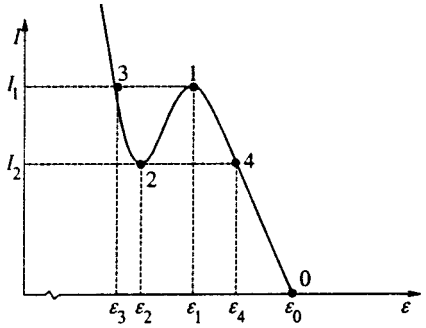

FIG. 3. Inverse dependence of the band gap ε on the intensity I of a light field with a finite width of the optical spectrum.

$$z(x)=1+a_1x+a_2x^2+a_3x^3, \tag{40}$$

where

$$a_1=\frac{x_0-2}{2x_0}, \quad a_2=\frac{5x_0-4}{8x_0}, \quad a_3=\frac{3x_0-10}{16x_0}. \tag{41}$$

The points x_1 and x_2 corresponding to the loss of stability in the system can be found by solving the equation $\partial z/\partial x=0$. Combining this with (40) yields

$$x_{1,2}=\frac{-a_2\mp\sqrt{a_2^2-3a_1a_3}}{3a_3}. \tag{42}$$

We see that the behavior of the system is critical if $a_2^2-3a_1a_3>0$. Combining this fact with (41) and allowing for the inequality $x_0<0$, we arrive at the necessary condition for a phase transition, a condition that imposes a constraint on the initial offset of the light-field carrier frequency from the frequency corresponding to the van Hove singularity of the first kind, $x_0<-9.55$, or, in the original notation,

$$\sqrt{16b^2+\varepsilon_0^2}-\hbar\omega_0>9.55\hbar/\tau_0. \tag{43}$$

If condition (43) is met, an analysis of Eqs. (36)–(38) shows that schematically the I vs. ε graph has the form depicted in Fig. 3. As the light intensity increases, a point on the graph moves along the curve $0\to4\to1\to3$, while as the intensity decreases the point moves along the curve $3\to2\to4\to0$. The jumps in the value of ε in the $1\to3$ and $2\to4$ sections correspond to direct and inverse photoinduced phase transitions. The range $I_2<I<I_1$ constitutes the bistability region.

Analyzing Eq. (40), we can find the points x_3 and x_4 corresponding to the new positions of equilibrium for the direct and inverse photoinduced phase transitions,

$$x_{3,4}=\frac{-a_2\pm2\sqrt{a_2^2-3a_1a_3}}{3a_3}, \tag{44}$$

and the dimensionless value Δx of the jump in the band gap, in a phase transition:

$$\Delta x=\frac{\sqrt{a_2^2-3a_1a_3}}{a_3}. \tag{45}$$

It then turns out that Δx is the same for direct and inverse photoinduced phase transitions. The jump in the band gap, $\Delta\varepsilon$, correspond to Δx of (45) can be approximately written as follows:

$$\Delta\varepsilon=-\frac{\hbar\Delta x\sqrt{16b^2+\varepsilon^2}}{\varepsilon\tau_0}, \tag{46}$$

where we have allowed for (38) and the fact that $\Delta\varepsilon\ll\varepsilon$.

Using (40), we can also derive an expression for the width of the hysteresis loop in a photoinduced phase transition (the size of the region of cavityless optical bistability in terms of the dimensionless light-field intensity):

$$\Delta z=z(x_1)-z(x_2)=\frac{8(a_2^2-3a_1a_3)^{3/2}}{27a_3^2}. \tag{47}$$

At the bifurcation point, where (43) becomes an equality, $x_0=-9.55$. In this case, in accordance with (41), (42), (44), and (47), $x_1=x_2=x_3=x_4=x_b\cong-0.89$ and $\Delta x=\Delta z=0$. Thus, within the present approximation $|x|\leqslant1$ we are able to describe the behavior of the Peierls system near the threshold value $x_0=-9.55$, below which there are critical singularities.

The above analysis shows that to observe a photoinduced phase transition and cavityless optical bistability one must ensure that the light-field intensity I is higher than I_b, the light-field intensity at the bifurcation point ($x_0=-9.55$, $x=-0.89$). Here is a numerical estimate of I_b. If in (36) we put $d_2\approx10^{-18}$ esu (Refs. 27 and 32), $\tau\approx10^{-10}$ s (Ref. 30), $1/\tau_0\approx10^2\ \mathrm{s}^{-1}$ (Ref. 34), and $\varepsilon_0\approx2b\approx0.5$ eV (Refs. 6 and 8), we find that $I_b\sim10^{-12}$ esu, which corresponds to an intensity $I_bc/2\pi\sim10^{-3}$ esu$\sim10^{-10}\,\mathrm{W\,cm^{-2}}$. When thin-film or thin-filament samples with good heat-removal characteristics are irradiated by light of such or even higher intensity, overheating can be avoided.

8. CONCLUSION

Our theory thus suggests that when the Peierls system is irradiated with light whose frequency is somewhat lower than the frequency corresponding to the upper van Hove singularity of the first kind, a photoinduced phase transition and cavityless optical bistability can be observed in such a system, with light absorption increasing in the process. These phenomena occur when condition (43) is met, i.e., when the light-field frequency ω_0 and the halfwidth $1/\tau_0$ are limited from above. The fact that such a condition exists can be explained by competition of two feedback mechanisms. Positive feedback is reflected by the first term on the right-hand side of Eq. (17), and the mechanism has been described in detail in the Introduction. Negative feedback consists in the following: as the concentration n of the nonequilibrium electrons in the conduction band grows, the recombination rate increases, which drives n down. This process is represented by the second term on the right-hand side of Eq. (17). If negative feedback is stronger than positive feedback, the equilibrium state of the system is stable, and slow variations in the external parameters lead to smooth variations in the internal parameters. But if positive feedback is dominant,

then, in accordance with (17), there is a sharp increase in fluctuations, which leads to a sudden variation in the internal parameters of the system.

Broadening of the optical spectrum of the incident radiation (an increase in the halfwidth $1/\tau_0$) reduces the possibility of observing the critical singularities of the system: the condition (43) imposed on the frequency ω_0 becomes more stringent and dimensions of the hysteresis loop, Δx of (45) and Δz of (47) decrease, as Eqs. (41) imply.

The final expressions (22)–(47) obtained in this paper hold if the initial offsets of ω_0 from the frequency corresponding to the van Hove singularity of the first kind are small:

$$\sqrt{16b^2+\varepsilon_0^2}-\hbar\omega_0<\sqrt{\varepsilon_0 k_B T}/2\,. \tag{48}$$

The present theory is unable to examine the frequency range where the offsets are large, i.e., where the condition (48) is not met, since the approximation (26) of a weak optical excitation is invalid in this range. At temperatures $T\approx 100\,\mathrm{K}$ ($k_B T\approx 8.6\times 10^{-3}$ eV) and a band gap $\varepsilon_0\approx 0.5\,\mathrm{eV}$ this condition imposes a constraint on the range of possible values of ε:

$$0<\varepsilon_0-\varepsilon<0.033\,\mathrm{eV}.$$

In conclusion it must be noted that the phenomenon of cavityless optical bistability with increasing absorption similar to that studied in the present paper (but caused by other mechanisms) was observed in experiments involving zinc selenide,[35] the amorphous semiconductor GeS_2 (Ref. 36), amorphous $GeSe_2$ films,[25] CdS plates, and other materials.[25,26] However, the author knows of no such experiments involving materials whose electronic properties can be described by the Peierls model.

*)E-mail: semenov@quant.univ.simbirsk.su

[1] L. N. Bulaevskiĭ, Usp. Fiz. Nauk **115**, 263 (1975) [Sov. Phys. Usp. **18**, 131 (1975)].
[2] V. E. Minakova, Yu. I. Latyshev, and V. A. Volkov, JETP Lett. **62**, 455 (1995).
[3] L. P. Gor'kov, Usp. Fiz. Nauk **144**, 381 (1984) [Sov. Phys. Usp. **27**, 809 (1984)].
[4] A. A. Fotiev, V. L. Volkov, and V. K. Kapustkin, *Oxide Vanadium Bronzes*, Nauka, Moscow (1978), p. 151.
[5] G. Gruner, Rev. Mod. Phys. **60**, 1129 (1988).
[6] A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *The Metal–Insulator Phase Transition and Its Application*, Nauka, Leningrad (1979), p. 28.
[7] R. O. Zaĭtsev, E. V. Kuz'min, and S. G. Ovchinnikov, Usp. Fiz. Nauk **148**, 603 (1986) [Sov. Phys. Usp. **29**, 322 (1986)].
[8] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **30**, 52 (1989).
[9] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **31**, No. 10, 261 (1989) [Sov. Phys. Solid State **31**, 1803 (1989)].
[10] A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1974 (1994) [Phys. Solid State **36**, 1079 (1994)].
[11] A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **39**, 925 (1997) [Phys. Solid State **39**, 826 (1997)].
[12] V. I. Emel'yanov and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **32**, 3083 (1990) [Sov. Phys. Solid State **32**, 1790 (1990)].
[13] V. I. Emel'yanov, N. L. Levshin, S. Yu. Poroĭkov, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **32**, 63 (1991).
[14] K. F. Berggren and V. A. Huberman, Phys. Rev. B **18**, 3369 (1978).
[15] A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, JETP Lett. **33**, 629 (1981).
[16] A. A. Bugaev, V. V. Gudyalis, B. P. Zakharchenya, and F. A. Chudnovskiĭ, JETP Lett. **34**, 430 (1981).
[17] V. F. Elesin, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **71**, 714 (1976) [Sov. Phys. JETP **44**, 375 (1976)].
[18] V. V. Kapaev, Yu. V. Kopaev, and S. N. Molotkov, Mikroélektronika **12**, 499 (1983).
[19] V. I. Emel'yanov and I. F. Uvarova, Zh. Éksp. Teor. Fiz. **94**, No. 8, 255 (1988) [Sov. Phys. JETP **67**, 1662 (1988)].
[20] I. M. Suslov, JETP Lett. **39**, 670 (1984).
[21] Yu. V. Kopaev, V. V. Menyaĭlenko, and S. N. Molotkov, Zh. Éksp. Teor. Fiz. **89**, 1404 (1985) [Sov. Phys. JETP **62**, 813 (1985)].
[22] V. B. Stopachinskiĭ and I. M. Suslov, Zh. Éksp. Teor. Fiz. **91**, 314 (1986) [Sov. Phys. JETP **64**, 183 (1986)].
[23] V. I. Emel'yanov and D. V. Babak, Laser Phys. **7**, 514 (1997).
[24] V. L. Bonch-Bruevich and S. G. Kalashnikov, *Semiconductor Physics*, Nauka, Moscow (1977), p. 255.
[25] H. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, New York (1985).
[26] P. I. Khadzhi, G. D. Shibarshina, and A. Kh. Rotaru, *Optical Bistability in a System of Excitons and Biexcitons in Semiconductors*, Shtiintsa, Kishinev (1988), p. 20.
[27] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions*, Nauka, Moscow (1988).
[28] O. Madelung, *Introduction to Solid State Physics*, Springer-Verlag, Berlin (1978), p. 449.
[29] N. N. Bogolyubov and N. N. Bogolyubov, Jr., *Introduction to Quantum Statistical Mechanics*, Nauka, Moscow (1984), p. 282.
[30] N. R. Belashenkov, V. B. Karasev, A. A. Solunin, I. A. Khakhaev, K. Sh. Tsibadze, and F. A. Chudnovskiĭ, Fiz. Tverd. Tela (St. Petersburg) **36**, 2475 (1994) [Phys. Solid State **36**, 1347 (1994)].
[31] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics*, Nauka, Moscow (1981), p. 42.
[32] D. N. Klyshko, *Physical Bases of Quantum Electronics*, Nauka, Moscow (1986), p. 22.
[33] A. S. Davydov, *Solid-State Theory*, Nauka, Moscow (1976), p. 296.
[34] A. N. Matveev, *Optics*, Vysshaya Shkola, Moscow (1985), p. 317.
[35] V. A. Stadnik, Fiz. Tverd. Tela (Leningrad) **29**, 3594 (1987) [Sov. Phys. Solid State **29**, 2059 (1987)].
[36] V. M. Lyubin and V. K. Tikhomirov, JETP Lett. **55**, 23 (1992).

Translated by Eugene Yankovsky

Microstructure of pair centers of Cr^{3+}–Cr^{2+} ions in the $KZnF_3$ crystal

M. V. Eremin,*) S. I. Nikitin,†) N. I. Silkin, S. Yu. Prosvirin, and R. V. Yusupov

Kazan State University, 420008 Kazan, Russia
(Submitted 11 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 1421–1429 (October 1998)

Studies involving the piezospectroscopy method have shown that the symmetry of the pair centers of Cr^{3+}–Cr^{2+} ions in the $KZnF_3$ crystal is tetragonal. In this paper we develop a microscopic model of a pair center. We use the temperature dependence of the integrated intensity of the absorption line to find the effective hopping integral for an e_g electron, $t_{\sigma\sigma}=205 \pm 10$ cm^{-1}, and the polaron reduction factor, equal to 0.11. By analyzing the selection rules for exchange-induced electric dipole transitions under double-exchange conditions we identify all the absorption lines of Cr^{3+}–Cr^{2+} pairs. © *1998 American Institute of Physics.* [S1063-7761(98)01810-1]

1. INTRODUCTION

Thanks to the recent discovery in $La_{1-x}A_xMnO_3$ compounds, where A is a bivalent cation, of giant magnetoresistance[1] and a huge isotopic effect,[2] experimental and theoretical studies of mixed-valence perovskites have become highly important. The microscopic theories of these compounds are based on the notions of double exchange and strong electron–phonon coupling.[3,4] One of the main difficulties encountered in these theories is the problem of estimating the microscopic parameters, such as the electron hopping integral and the Jahn–Teller coupling energy. A possible way of finding the parameters of the interaction between Mn^{3+} and Mn^{4+} ions or their isoelectronic analogs is to study the pair centers of these ions in dielectric crystals with a low pair-center concentration. In this case a large number of methods is available, e.g., optical spectroscopy, whose use in the case of concentrated compounds with a large conductivity is highly problematic.

In Ref. 5 we reported that the optical absorption spectrum of $KZnF_3$:Cr^{3+}, Cr^{2+} crystals has three groups of relatively narrow lines in the 500-nm, 580-nm, and 600-nm ranges (Fig. 1), in addition to the characteristic lines for single Cr^{3+} and Cr^{2+} ions. The experimental values of the transition energies are listed in Table I. The intensity of these absorption lines is proportional to the product of concentrations of the Cr^{3+} and Cr^{2+} ions, which made it possible to interpret them as lines related to the pair centers of the Cr^{3+} and Cr^{2+} ions. It was assumed that Cr^{3+} and Cr^{2+} occupy positions in neighboring unit cells and are coupled by the double-exchange mechanism, which arises through the intermediate fluorine ion (a double 180° exchange). The goal of the present investigation is an experimental verification of this assumption, an estimate of the electron hopping integral, and an identification of the transitions corresponding to the observed absorption lines.

The main condition for dynamic electron hopping between ions that are in different valence states (in our case these are ions of bivalent and trivalent chromium ions) is the total equivalence of the crystallographic positions occupied by these ions. When $KZnF_3$ crystals are activated, Cr^{3+} ions replace Zn^{2+} ions. Here, because of heterovalent substitution, centers of Cr^{3+} ions of cubic, trigonal, tetragonal, and monoclinic symmetries are formed.[6,7] Hence we may assume that pair centers of the Cr^{3+} and Cr^{2+} ions form either without local balance of charge (pairs coupled by the double-exchange mechanism) or with local balance of the excess charge of the Cr^{3+} ion. In the latter case the mechanism of superexchange interaction with virtual electron transfer is realized. Thus, finding the symmetry of a pair center may lead to conclusions concerning the center's microstructure.

2. EXPERIMENTAL RESULTS

$KZnF_3$ crystals have a perovskite cubic structure (the space group Pm3m), so that in the presence of anisotropic centers there is orientational degeneracy. It is known that the most effective method of determining the symmetry of anisotropic centers in cubic crystals in experiments is the one that employs external fields. To solve the problem we chose the piezospectroscopy method.[8]

The samples for our investigations were grown by the Bridgman–Stockburger method. For the initial materials we took potassium fluoride, pre-dried to exclude the possibility of hydrolysis, and zinc fluoride, purified by recrystallization from the melt. For alloying the crystals we used chromium fluorides CrF_3 and CrF_2, whose concentration in the mixture was about 1 percent by weight.

For piezospectroscopy studies, the grown single crystals were cut into samples in the form of rectangular parallelepipeds in such a way that the normal to two opposite faces is parallel either to the C_4 axisor the C_2 axis of the crystal. The observation of the piezodichroism spectra was done along an axis parallel to the long face of the parallelepiped, with this direction corresponding to the tetragonal axis of the crystal. The samples were oriented along the perfect cleavage of $KZnF_3$ single crystals in (100) planes, and the accuracy of orientation was checked on the DRON-2 x-ray diffractometer and was found to be $\pm 2°$.

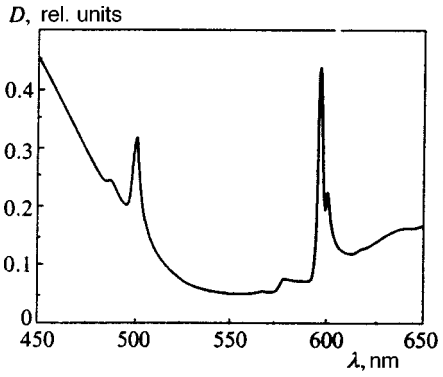

FIG. 1. The absorption spectrum of the $KZnF_3:Cr^{3+},Cr^{2+}$ crystal; $T=300$ K.

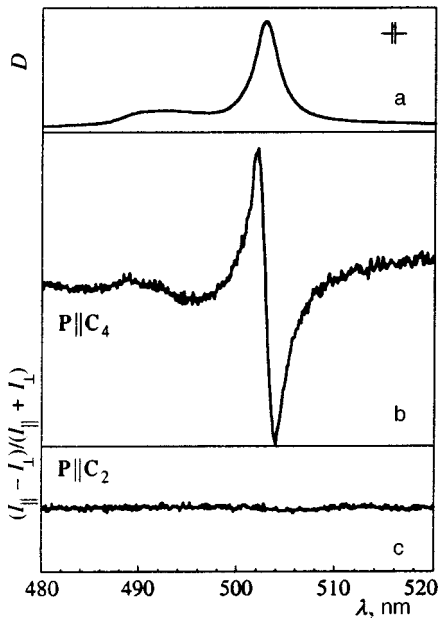

FIG. 2. Absorption (a) and piezodichroism (b,c) spectra of the $KZnF_3:Cr^{3+}$, Cr^{2+} crystal in the 500-nm range; $T=77$ K.

The splitting of the absorption lines due to the lifting of orientational degeneracy can be observed in the following cases: only under pressure along the C_4 axis if the symmetry of the centers is tetragonal, and only under pressure along the C_2 axis if the symmetry is trigonal. If the symmetry of the centers is lower, the lines will split under pressure along the C_2 axis and along the C_4 axis.

When nondestructive pressures are applied ($P \approx 20-25\,\mathrm{kgf\,mm^{-2}}$), the size of the splitting of the spectral lines may be much smaller than the width of these lines. Such splitting was detected by measuring the linear dichroism signal with a highly sensitive spectral polarimeter similar to the one used by Zaslavskiĭ and Natadze;[9] the measurements were done at $T=77$ K.

The absorption and linear dichroism spectra for groups of lines in the 500-nm and 600-nm ranges (where the absorption lines of a pair center are the strongest) are depicted in Figs. 2 and 3 for two directions of applying axial pressure: parallel to the C_4 axis and parallel to the C_2 axis, respectively. To make the picture more graphic, we have drawn only the absorption lines of pair centers isolated in the spectrum of the $KZnF_3:Cr^{3+},Cr^{2+}$ crystal. In both cases the axial pressure was about $15\,\mathrm{kgf\,mm^{-2}}$. Figures 2 and 3 show that in the uniaxial deformation field the linear dichroism signal, which indicates that both groups of lines are split, can be observed only when the pressure is applied parallel to the C_4 axis. These results clearly suggest that the pair centers are tetragonally symmetric and that the piezospectroscopy effect is due solely to the lifting of orientational degeneracy.

The results obtained in the experiments suggest that the chromium ions forming a pair center occupy neighboring positions of the Zn^{2+} ions on the tetragonal axis of the crystal. Of course, if the excess charge of the Cr^{3+} ion were to be locally balanced by a crystal lattice defect on the tetragonal axis, the symmetry of a center would also be tetragonal. However, it is unclear why this way of excess charge balance has preference to other possible variants of local balance of charge, e.g., as those for single Cr^{3+} ions.[6,7] On the basis of all this reasoning we concluded that an observed pair center of Cr^{3+}–Cr^{2+} ions in the $KZnF_3$ crystal is formed with nonlocal balance of charge similar to that for centers of single Cr^{3+} ions with cubic symmetry.

All the observed absorption lines of a pair center of Cr^{3+}–Cr^{2+} exhibit strong temperature dependence. When

TABLE I. Probabilities of electric dipole transitions and the positions of the levels of a Cr^{2+}–Cr^{3+} pair center in the $KZnF_3:Cr^{3+},Cr^{2+}$ crystal ($F(S)=(S+9/2)(7/2-S)$ and $\alpha^2 \gg \beta^2$).

Wavelength, nm (energy, cm^{-1})	Excited state	Transition probability	Calculated energy, cm^{-1}
598 (16 722)	$Cr^{2+}(^3E_u)Cr^{3+}(^4A_2)$	$\alpha^2 \frac{1}{16}(d^{ee}_{v,v}(z))^2 F(S)$	16 730
580 (17 241)	$Cr^{2+}(^3E_v)Cr^{3+}(^4A_2)$	$\beta^2 \frac{1}{16}(d^{ee}_{v,v}(z))^2 F(S)$	17 290
503 (19 881)	$Cr^{2+}(^3T_1(y,x))Cr^{3+}(^4T_2)$	$\frac{1}{8}(d^{tt}_{y,y}(z))^2 F(S)$	22 090
492 (20 325)	$Cr^{2+}(^3T_1(z))Cr^{3+}(^4T_2)$	$\frac{1}{8}(d^{tt}_{z,y}(z))^2 F(S)$	22 510

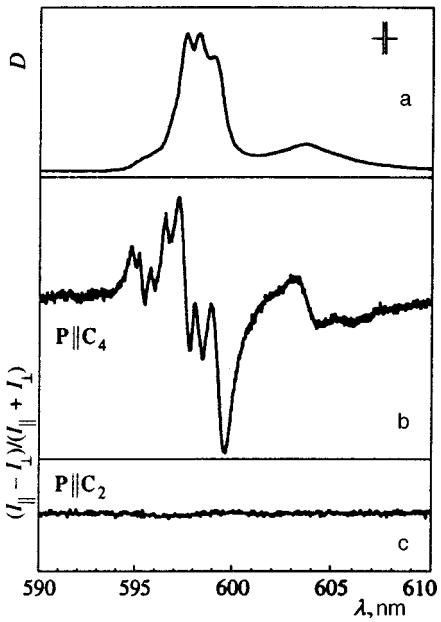

FIG. 3. Absorption (a) and piezodichroism (b,c) spectra of the $KZnF_3:Cr^{3+}$, Cr^{2+} crystal in the 600-nm range; $T=77$ K.

the temperature is lowered below 150 K, the intensity of these lines decreases, and at 10 K the absorption lines of the pair centers disappear completely. Since these lines are observed against the background of the broad absorption bands of single Cr^{3+} ions, we used the following method to analyze their behavior under temperature variations quantitatively. The absorption spectrum of the Cr^{3+} ions was determined from the spectrum of a crystal activated only by Cr^{3+} ions and was then subtracted from the absorption spectrum of the samples with a coefficient that depended on the ratio of the concentration of Cr^{3+} ions in the sample to that in the $KZnF_3:Cr^{3+}$ crystal. Since the shape of the electron-vibration absorption bands of the Cr^{3+} ions is also temperature-dependent, the spectra of the $KZnF_3:Cr^{3+}$ crystal were measured at the same temperatures as the absorption spectra of $KZnF_3:Cr^{3+},Cr^{2+}$ crystals. The temperature dependence of the integrated intensity of the absorption line with $\lambda_{max}=598$ nm is depicted in Fig. 4. An interpretation of the observed variations in the spectrum of pair centers is given below, in the discussion of the experimental results.

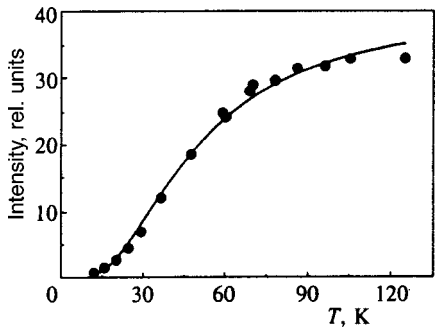

FIG. 4. Temperature dependence of the integrated intensity of the absorption line with $\lambda_{max}=598$ nm of a Cr^{2+}–Cr^{3+} pair center and its approximation by the Boltzmann distribution.

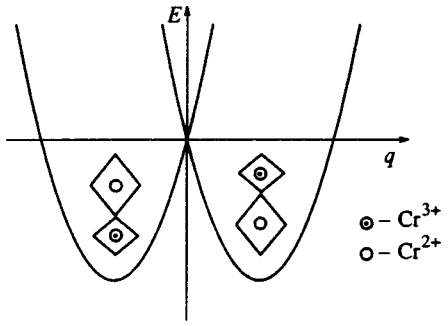

FIG. 5. Schematic of the adiabatic potential of the ground state of a $Cr^{2+}(^5E)$–$Cr^{3+}(^4A_2)$ pair center and the microstructure of distortions at the potential's minimum in the xz plane; q is the normal coordinate of the coordinated displacements of fluorine ions.

3. STRUCTURE OF THE ADIABATIC POTENTIALS OF MIXED-VALENCE PAIRS AND VIBRONIC REDUCTION OF THE HOPPING INTEGRAL

We examine the adiabatic potential in a pair in the ground state, $Cr^{2+}(t_2^3,e;{}^5E)$–$Cr^{3+}(t_2^3;{}^4A_2)$, where the Cr^{3+} ion is in position A and the Cr^{2+} ion, in position B. In the $KZnF_3$ crystal, the triply charged ion Cr^{3+} replaces the doubly charged Zn^{2+} ions. In view of this it is natural to expect that the nearest neighbors of the fluorine ions move closer to the Cr^{3+} ion. The corresponding term of the adiabatic potential for the $^4A_2$ state has the form

$$U_A=V_AQ_A+\frac{m\omega_A^2}{2}Q_A^2, \tag{1}$$

where Q_A is the normal component of the respiratory mode of an octahedron with a frequency ω_A. According to the existing theory of the Jahn–Teller effect, for the ground state ^5E of Cr^{2+} ions we have an adiabatic potential of the form[10]

$$U_B=V_BQ_B+\frac{m\omega_B^2}{2}Q_B^2+V_E(U_\theta Q_\theta+U_\varepsilon Q_\varepsilon)+\frac{m\omega_E^2}{2}\times(Q_\theta^2+Q_\varepsilon^2)+N_E\left[\frac{1}{\sqrt{2}}(Q_\varepsilon^2-Q_\theta^2)U_\theta+\sqrt{2}\,Q_\theta Q_\varepsilon U_\varepsilon\right]+V_3Q_\theta(Q_\theta^2-3Q_\varepsilon^2), \tag{2}$$

where Q_θ and Q_ε are the normal coordinates of the octahedron, which form the basis of the irreducible representation of type E. It is natural to assume that $V_B \ll V_A$. Pointing the z axes at the centers A and B along the axis of the pair and bearing in mind that in the Cr^{2+}–F^-–Cr^{3+} fragment the fluorine ion is common to the two chromium ions, we arrive at the adiabatic potential depicted in Fig. 5. In the same figure we give the diagram of the displacement of the fluorine ions from their equilibrium positions in the xz plane. In both minima, the pair centers of Cr^{2+}–Cr^{3+} have a tetragonal symmetry axis along the z axis. According to (1), the ambient structure of the fluorine ions surrounding Cr^{3+} is compressed along the z axis, while, according to (2), the ambient structure of the fluorine ions surrounding Cr^{2+} is stretched along the z axis.

An important feature of the microstructure of the Cr^{2+}–Cr^{3+} center is that the bridge fluorine ion is displaced from a crystal lattice site, which is significant if we want to explain the electric dipole transitions.

Near each minimum the adiabatic potential can be approximated by the expression

$$V_{\rm eff}(q-q_0)+\frac{m\omega_{\rm eff}^2}{2}(q-q_0)^2,$$

where $V_{\rm eff}$ and $\omega_{\rm eff}^2$ can even be larger than V_A and ω_A^2, respectively.

In the ground state of the pair the potential wells are equivalent (Fig. 5). The Hamiltonian of polaron hopping is

$$\sum_{\alpha,\beta} t_{\alpha\beta}(a_\alpha^\dagger b_\beta + b_\beta^\dagger a_\alpha), \tag{3}$$

where $t_{\alpha\beta}$ is the effective hopping integral. Representing the wave functions by products of electron and nuclear (oscillatory) functions and integrating with respect to the coordinate q, we easily find that

$$t_{\alpha\beta}=t_{\alpha\beta}^{(0)}\exp\left(-\frac{V_{\rm eff}^2}{\hbar\omega_{\rm eff}}\right), \tag{4}$$

where $t_{\alpha\beta}^{(0)}$ is the electron hopping integral in the absence of local deformations of the lattice. The value of this integral, $t_{\sigma\sigma}^{(0)}$, along the $(3z^2-r^2)$-orbitals for the Mn^{2+}–F^-–Cu^{2+} pairs in the $KMnF_3$ crystal is known[11] and amounts to roughly 1800–2000 cm^{-1}. For the case of pairs with different valencies, the operator (3) is known as the double-exchange operator. Anderson and Hasegawa[12] have calculated eigenvalues of this operator. In our case we have

$$E(S)=\pm\tfrac{1}{4}|t_{\sigma\sigma}|(S+1/2), \tag{5}$$

where S is the total spin of the pair, which in the ground state $Cr^{2+}(^5E)$–$Cr^{3+}(^4A_2)$takes the values 7/2, 5/2, 3/2, and 1/2; the state with $S=7/2$ is the ground state.

In the excited state of the pair $Cr^{2+}(^3E)$–$Cr^{3+}(^4A_2)$ corresponding to the absorption line with $\lambda_{\max}=598$ nm (the identification of the absorption line will be done in Sec. 4), the allowed values of the total spin S^* are 5/2, 3/2, and 1/2. To analyze the temperature dependence of the integrated intensity of this absorption line (Fig. 4), we note that the excited state of the pair does not contain the state with $S^*=7/2$. Hence the absorption line with $\lambda_{\max}=598$ nm must be assigned to the $S(5/2)\rightarrow S^*(5/2)$ transition, and the Boltzmann distribution should explain the temperature dependence of this line. A good approximation to the experimental temperature dependence below 150 K is the Boltzmann distribution (Fig. 4). The effective hopping integral determined from the temperature dependence of the line with $\lambda_{\max}=598$ nm is $t_{\sigma\sigma}=205\pm10$ cm^{-1}, which is nine to ten times smaller than the value of $t_{\sigma\sigma}^{(0)}$ for Mn^{2+}–F^-–Cu^{2+} pairs. Assuming that this difference is due primarily to the emergence of the vibronic reduction factor in (4), we arrive at a rough estimate,

$$\exp\left(-\frac{V_{\rm eff}^2}{\hbar\omega_{\rm eff}}\right)=0.10-0.11. \tag{6}$$

For the sake of comparison we note that to explain the huge isotopic effect in $La_{0.8}Ca_{0.2}MnO_3$ (Ref. 2) the polaron factor (6) should be set to 0.18.

4. SELECTION RULES FOR ELECTRIC DIPOLE TRANSITIONS AND IDENTIFICATION OF THE ABSORPTION LINE OF A PAIR CENTER OF Cr^{2+}–Cr^{3+}

In the second quantization representation, the operator of the effective dipole moment of a pair of similar ions with a mixed valence is

$$\mathbf{D}_{\rm eff}=\sum\langle\alpha|\mathbf{d}|\beta\rangle a_\alpha^\dagger b_\beta+\text{H.c.}, \tag{7}$$

where α and β are the sets of quantum numbers of one-electron states at the centers A and B, respectively, and $\langle\alpha|\mathbf{d}|\beta\rangle=d_{\alpha\beta}$ are the matrix elements of the electric dipole-moment operator calculated with one-electron functions.

The matrix elements corresponding to transitions of a pair center accompanied by e_g-electron transport, i.e., matrix elements of the type $d_{v,v}^{ee}(z)=\langle r_A^2-3z_A^2|d_z|r_B^2-3z_B^2\rangle$ and $\langle y^2-x^2|d_z|y^2-x^2\rangle$ (the π-polarization), are nonzero, since the bridge fluorine ion F^- (Fig. 5) is displaced from a crystal lattice site. Transitions in the σ-polarization are forbidden.

When there is t_{2g}-electron transport, the one-electron matrix elements that are finite are $d_{y,y}^{tt}(z)=\langle xz_A|d_z|xz_B\rangle=\langle yz_A|d_z|yz_B\rangle$ and $\langle xy|d_z|xy\rangle$ in the π-polarization and $d_{z,y}^{tt}(y)=\langle xy|d_y|xz_B\rangle=\langle xy|d_x|yz_B\rangle$ in the σ-polarization.

The matrix elements of the effective dipole moment were calculated by the formula

$$\begin{aligned}
&\langle S_A\Gamma_A Q_A,\overline{S}_B\overline{\Gamma}_B\overline{Q}_B;SM_S\\
&\quad\times\left|\sum d_{\alpha\beta}a_\alpha^\dagger b_\beta\right|\overline{S}_A\overline{\Gamma}_A\overline{Q}_A,S_B\Gamma_B Q_B;SM_S\rangle\\
&=\sum(-1)^{n_A+n_B}\left\langle\begin{matrix}\overline{\Gamma}_A & \gamma_A & \Gamma_A\\ \overline{Q}_A & q_A & Q_A\end{matrix}\right\rangle\left\langle\begin{matrix}\overline{\Gamma}_B & \gamma_B & \Gamma_B\\ \overline{Q}_B & q_B & Q_B\end{matrix}\right\rangle\\
&\quad\times d_{\gamma_\alpha q_\alpha,\gamma_\beta q_\beta}\sqrt{n_An_B[S_A][S_B][\Gamma_A][\Gamma_B]}\\
&\quad\times(l_A^{n_A}S_A\Gamma_A\{|l_A^{n_A-1}\overline{S}_A\overline{\Gamma}_A)(l_B^{n_B-1}S_A\Gamma_A)|\}l_B^{n_B}\overline{S}_B\overline{\Gamma}_B)\\
&\quad\times(-1)^{S_A+S_B+\overline{S}_B}\left\{\begin{matrix}S_A & \overline{S}_B & S\\ S_B & \overline{S}_A & 1/2\end{matrix}\right\},
\end{aligned} \tag{8}$$

where S is the total spin of the ion pair, and the other notation coincides with that used in Ref. 13. The $6j$-symbol in the above formula implies that in an electric dipole transition of a pair center, the spin quantum numbers of the initial and final states of the ions A and B can differ only by 1/2. The results of calculations of the probabilities of allowed electric dipole transitions from the ground state of the pair,

$$\begin{aligned}
\psi_u^{(0)}=\frac{1}{\sqrt{2}}[&{\rm Cr}^{2+}(t_2^3\,{}^4A_2,e;\,{}^5E_u){\rm Cr}^{3+}(t_2^3\,{}^4A_2)\\
&+{\rm Cr}^{3+}(t_2^3\,{}^4A_2){\rm Cr}^{2+}(t_2^3\,{}^4A_2,e;\,{}^5E_u)],
\end{aligned}$$

are listed in Table I. In calculating the transition probabilities we accounted for the splitting of levels in the tetragonal crystalline field and allowed for the orientational degeneracy of the pair centers in the crystal.

We associate the absorption line with $\lambda_{max}=598$ nm with the transition to the excited antisymmetric state of the pair,

$$\psi_u^{(1)}=\frac{1}{\sqrt{2}}[\mathrm{Cr}^{2+}(t_2^3\,{}^4A_2,e;\,{}^3E_u)\mathrm{Cr}^{3+}(t_2^3\,{}^4A_2)$$
$$-\mathrm{Cr}^{3+}(t_2^3\,{}^4A_2)\mathrm{Cr}^{2+}(t_2^3\,{}^4A_2,e;\,{}^3E_u)].$$

The transition to the state $\psi_v^{(1)}$ corresponds to the weak absorption line with $\lambda_{max}=580$ nm. The relative intensity of this transition is low since the transition is underresolved only because of the mixing of the states $\psi_v^{(1)}$ and $\psi_u^{(1)}$ due to spin–orbit coupling. The second-order effect leading to such mixing has been discussed by Fujiwara and Tanabe.[14] The absorption lines with $\lambda_{max}=503$ nm and $\lambda_{max}=492$ nm are associated with the transition to the excited state

$$\Psi^{(2)}=\frac{1}{\sqrt{2}}[\mathrm{Cr}^{2+}(t_2^4\,{}^3T_1)\mathrm{Cr}^{3+}(t_2^2\,{}^3T_1,e;\,{}^4T_2)$$
$$-\mathrm{Cr}^{3+}(t_2^2\,{}^3T_1,e;\,{}^4T_2)\mathrm{Cr}^{2+}(t_2^4\,{}^3T_1)],$$

which is split by the tetragonal crystalline field. The ratio of the values of the splitting of states $\psi^{(1)}$ and $\psi^{(2)}$ is 1.2 and is in good agreement with the calculated value if one assumes that the main source of the tetragonal field at the Cr^{2+} ion is the displacement of the bridge fluorine ion from the equilibrium position toward Cr^{3+}. The estimated values of the transition energies (see Table I) are in satisfactory agreement with the experimental data. The calculations were done with the following set of parameters: $Dq=1480\ \mathrm{cm}^{-1}$, $B=785\ \mathrm{cm}^{-1}$, and $C=3280\ \mathrm{cm}^{-1}$ for Cr^{3+} (these values were found from the spectra of single Cr^{3+} ions), and $Dq=1190\ \mathrm{cm}^{-1}$ (from the absorption line corresponding to the ${}^5E-{}^5T_2$ transition), $B=800\ \mathrm{cm}^{-1}$, and $C=3200\ \mathrm{cm}^{-1}$ for Cr^{2+}, in accordance with the data from the literature,[15] and the tetragonal field parameter C_2^{tetr} was set to $970\ \mathrm{cm}^{-1}$.

5. CONCLUSION

Thus, we have used the piezodichroism method to show that the absorption lines of the $KZnF_3:Cr^{3+},Cr^{2+}$ crystal in the 500-nm, 580-nm, and 600-nm ranges correspond to the pair centers of Cr^{2+}–Cr^{3+} ions positioned along the tetragonal axis of the crystal. We have developed a microscopic model of a pair center. The minima of the adiabatic potential of the pair correspond to compression of the ambient structure of the fluorine ions surrounding Cr^{3+} and coordinated Jahn–Teller stretching of the fluorine octahedron near Cr^{2+}. The strong vibronic interaction of the chromium ions and the crystal lattice substantially reduces the double-exchange parameter. The polaron reduction factor, determined from the temperature dependence of the total intensity of the absorption lines, proved to be about 0.11. We have also discussed the selection rules for exchange-induced electric dipole transitions under double-exchange conditions, and on the basis of this have identified all the observed absorption lines for Cr^{2+}–Cr^{3+} pairs.

This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant 97-02-185-98a).

*)E-mail: Mikhail.Eremin@ksu.ru
†)E-mail: Sergey.Nikitin@ksu.ru

[1] S. Jin, T. H. Tiefel, M. McCormack *et al.*, Science **264**, 413 (1994).
[2] Guo-meng Zhao, K. Conder, H. Keller, and K. A. Muller, Nature (London) **381**, 676 (1996).
[3] A. J. Millis, B. I. Shraiman, and R. Mueller, Phys. Rev. Lett **77**, 175 (1996).
[4] H. Röder, J. Zang, and A. R. Bishop, Phys. Rev. Lett. **76**, 1356, 4987 (1996).
[5] M. V. Eremin, S. I. Nikitin, N. I. Silkin, A. I. Tsvetkov, and R. V. Yusupov, JETP Lett. **61**, 612 (1995).
[6] M. Binois, A. Leble, and J. C. Fayer, J. de Phys. Coll. C9 **34**, 285 (1973).
[7] R. Yu. Abdulsabirov, L. D. Livanova, and V. G. Stepanov, Fiz. Tverd. Tela (Leningrad) **16**, 2135 (1974) [Sov. Phys. Solid State **16**, 1395 (1974)].
[8] A. A. Kaplyanskiĭ, Opt. Spektrosk. **16**, 602 (1964).
[9] V. S. Zaslavskiĭ and A. L. Natadze, Opt. Spektrosk. **32**, 1015 (1972).
[10] A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance in Transition Ions*, Clarendon Press, Oxford (1970).
[11] M. V. Eremin and Yu. V. Rakitin, J. Phys. C **14**, 247 (1981).
[12] P. W. Anderson and H. Hasegawa, Phys. Rev. **100**, 675 (1955).
[13] D. T. Smirnov and Yu. F. Sviridov, *Theory of Optical Spectra of Ions of Transition Metals*, Nauka, Moscow (1977).
[14] T. Fujiwara and Y. Tanabe, J. Phys. Soc. Jpn. **37**, 1512 (1974).
[15] N. W. Alcock, C. F. Putnik, and S. L. Holt, Inorg. Chem. **15**, 3175 (1976).

Translated by Eugene Yankovsky

Hysteresis of the characteristics of magnetostatic waves in ferrite films with stripe domains whose magnetization vectors are oriented close to the plane of the film

A. V. Vashkovskiĭ,*) É. G. Lokk, and V. I. Shcheglov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 141120 Fryazino, Moscow Province, Russia

(Submitted 7 April 1998)
Zh. Éksp. Teor. Fiz. **114**, 1430–1450 (October 1998)

The propagation of zero-exchange spin waves (magnetostatic waves) is investigated in yttrium iron garnet films having a regular stripe domain structure with almost in-plane orientation of the domain magnetization vectors. The characteristics of the waves are studied for magnetizations of the film parallel and perpendicular to projections of the [111] crystallographic axes onto the plane of the film. It is established, in contrast with films having the domain magnetization vectors oriented close to the normal to the plane of the film, that both the propagation of magnetostatic waves and the variation of the parameters of the domain structure exhibit a distinctly pronounced hysteretic character as the magnetizing field is varied. The hysteresis of the amplitude–frequency response, equiphase, and dispersion curves of the magnetostatic waves is investigated. The authors examine how the hysteresis of these parameters is related to the hysteresis of the domain structure. The spectrum of magnetostatic waves is found to have an interval of wavelengths (wave numbers) that are not excited in the unsaturated film when the applied field is close to the saturation value, and this phenomenon as well exhibits hysteresis. © *1998 American Institute of Physics.* [S1063-7761(98)01910-6]

1. INTRODUCTION

Investigations of the excitation spectra of ferromagnets with a domain structure have been reported in a great many papers,[1–18] including studies of spin-wave excitation spectra[1–16] and domain-wall excitation spectra,[17,18] the latter having disclosed both low-frequency (1–10 MHz, Ref. 17) and high-frequency (0.8–1.5 GHz, Ref. 18) resonances. Spin-wave excitations, or spin waves, in magnetic films and wafers have been investigated for the cases of regular[1–13] and irregular[14,15] stripe domain structures. In the majority of theoretical papers the propagation of spin waves has been treated on the assumption that the wavelength λ and the thickness s of the ferrite layer are much greater than the domain width d (Refs. 1 and 3–6), whereas experimental studies, usually involving yttrium iron garnet (YIG), have been concerned with the case $d \simeq s$, $d \ll \lambda$ (Refs. 5, 10–12, 14, and 15).

As a rule, the experiments have been carried out on rectangular samples of small dimensions,[5,14,15] resulting in strong edge distortions of the domain structure and the occurrence of irregularities in it.[14,15] The orientation of the crystallographic axes in the plane of a film is often not monitored or is chosen arbitrarily.[10,14,15] The films used in experimental work customarily have domains whose magnetization vectors are oriented in a direction almost perpendicular to the plane of the film, facilitating visual observation of the domain structure through the Faraday effect. As a rule, however, the cause of this orientation of the magnetization vector inside the domains has never been discerned, and the magnetic anisotropy has not been measured. These shortcomings make it difficult to compare experiment with theory or to compare results obtained by different authors, so that general functional relations cannot be discerned either in the behavior of the domain structure of the films or in the nature of the propagation of spin waves.

Previously,[11,12] using samples of large dimensions (greater than $30 \times 30\ \mathrm{mm}^2$), we have succeeded in eliminating the influence of the edges of the sample and obtaining a definite picture of the peculiarities of domain structure formation in YIG films and also of the propagation of long-wavelength ($\lambda > 50\ \mu\mathrm{m}$) spin waves in these structures. By monitoring the orientation of the crystallographic axes in the plane of the film we have been able to reveal sharp anisotropy in the behavior of magnetostatic waves (MSWs) due to significant differences in the modification of the domain structure and in the character of its phase transitions. It should be noted, however, that the results set forth in Refs. 11 and 12 are far from exhaustive. For example, in our experiments we have discovered a large class of YIG films in which the behavior of MSWs differs significantly from that described in Refs. 11 and 12.

The cause of the disparity is the equilibrium orientation of the magnetization vector of the domains close to the plane of the film, a situation that is strongly mirrored in all the MSW characteristics: Their field and orientation dependences acquire a previously unnoticed, very pronounced hysteresis, where the hysteretic properties are governed both by the crystallographic anisotropy and by the induced anisotropy of the films. The term ''hysteretic properties of magnetostatic waves'' has been used previously in Ref. 15, where the minimum static magnetic field at which MSWs begin to be excited in a film has been observed to vary by $\simeq 4$ Oe upon magnetization reversal of the film relative to zero field.

In this paper we interpret ''hysteresis'' in the conventional sense as a significant change in the parameters and characteristics of these waves in an identical static magnetic field H_0, depending on how the latter has varied in the past. Here we report an experimental study of the propagation of long spin waves through an ensemble of domains forming a regular structure over the entire area of the film.

2. EXPERIMENTAL ARRANGEMENT AND PARAMETERS OF THE MAGNETIC FILMS

The experiments were carried out on an apparatus similar to that described in Refs. 11 and 12. The investigated ferrite films were magnetized in-plane by a homogeneous field $\mathbf{H}_0$, which could be varied over the range 0–100 Oe. Magnetostatic waves in the frequency range 100–3000 MHz were generated and received by moving arrays of gold-plated tungsten wire transducers of length 3.5 mm and thickness 12 μm. The MSW wave number was measured by the floating probe method. Simultaneously with the measurement of the characteristics of these waves the domain structure of the samples was monitored by a magnetooptical technique utilizing the Faraday effect with the film illuminated by a light beam perpendicular to the plane of the film.

The investigated YIG films were grown by liquid-phase epitaxy on gadolinium gallium garnet substrates cut in the crystallographic (111) plane. The thickness of the films was 5–20 μm, and their diameter was 75 mm. Measurements of the film parameters by the method described in Ref. 19 revealed that the films had cubic as well as uniaxial anisotropy with the axis almost normal to the plane of the film. The following, most characteristic values of the film parameters were obtained: saturation magnetization $4\pi M_0 = 1750 \pm 50$ G; cubic anisotropy field $H_c = 80 \pm 5$ Oe; uniaxial anisotropy field less than or equal to 200 Oe. The angle θ of deviation of the uniaxial anisotropy axis from the normal to the plane of the film was less than 4°. The orientation of the cubic anisotropy axes in the plane of the film were determined within $\pm 0.5°$ error limits, and the orientation of the projection of the uniaxial anisotropy axis onto the plane of the film was determined within $\pm 3°$ limits (The latter error was greater on account of the extremely small angle θ). The values of the static magnetic fields sufficient for tangentially magnetizing a film to saturation along or perpendicular to the projection of a [111] axis onto the plane of the film were $H_{\text{sat}}^{\parallel} = 30$–40 Oe and $H_{\text{sat}}^{\perp} = 14$–25 Oe, respectively. The halfwidth of the ferromagnetic resonance line for all the films was less than 0.6 Oe.

Clearly, these values of the film parameters do not differ at all significantly from those used in previous work.[10–12] On the other hand, as will be described below, the behavior of MSWs in them differs considerably from anything known before and exhibits very distinct hysteresis. In the course of the investigations it has been established that the principal difference in the selected films is the orientation of the magnetization vector of the domains close to the plane of the film, whereas in previously investigated films, which have not exhibited hysteretic properties, the magnetization vector of the domains was oriented close to the normal to the plane of the film.[11,12] We have discovered that films of both groups possess uniaxial anisotropy with the axis perpendicular to the plane of the film, and the field is at most 150–200 Oe; in this case, as a rule, the magnitude of the anisotropy field in films having nearly in-plane magnetization is 20–70 Oe smaller than in films with the domains magnetized close to the normal. This does not mean that the uniaxial anisotropy field suddenly changes upon transition from films of one group to films of the other group. Only the orientation of the magnetization vector makes an abrupt change, as is entirely possible during a smooth variation of the anisotropy field and generally characteristic of orientational phase transitions in magnets.[20]

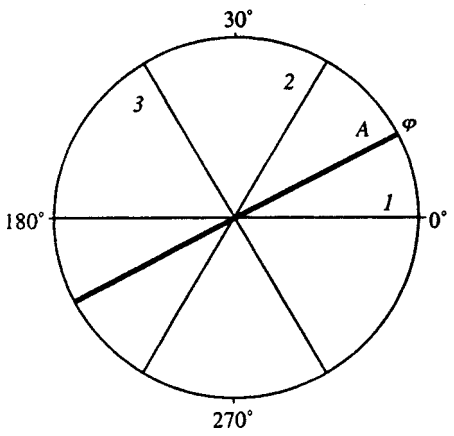

FIG. 1. Positions of special directions in the plane of an yttrium iron garnet film. *1–3*) Directions defined by the projections of [111] axes onto the plane of the film; *A*) direction defined by the projection of the uniaxial anisotropy axis.

The arrangement of the crystallographic axes in the plane of a YIG film and the corresponding choice of reference frame for the angles are shown in Fig. 1 (the plane of the film coincides with the plane of the figure). The projection of the uniaxial anisotropy axis onto the plane of the film is represented by a heavy line and is labeled with the letter *A*. It is obvious that all the in-plane projections of [111] axes shown in the figure are exactly equivalent in the crystallographic sense; hence, the number *1* labels the projection oriented closest to direction *A*, and the other two are numbered clockwise in increasing order. The angles between directions *1* and *A* is denoted by φ (we read all angles from axis *1*, positive in the counterclockwise direction).

The most general laws (observed in all the investigated films) characterizing the nature of MSW propagation and the domain-wall parameters are described below in the example of YIG film No. 1, which has the parameters $s = 17.9$ μm, $4\pi M_0 = 1780$ G, $\Delta H = 0.54$ Oe, $\Delta H = -80.5$ Oe, $\theta = 1.3°$, $\varphi = 29°$, $H_{\text{sat}}^{\parallel} = 33.5$ Oe, and $H_{\text{sat}}^{\perp} = 19.1$ Oe. In addition to this film, we have also investigated YIG films No. 2 of thickness 18.5 μm with $\varphi = -22°$, $H_{\text{sat}}^{\parallel} = 39$ Oe, and $H_{\text{sat}}^{\perp} = 23$ Oe and No. 3 of thickness 12.6 μm with $\varphi = -7°$, $H_{\text{sat}}^{\parallel} = 36$ Oe, and $H_{\text{sat}}^{\perp} = 14$ Oe. However, essentially all the laws of MSW propagation can be described in the example of film No. 1, as will be done below.

3. PROPAGATION OF MAGNETOSTATIC WAVES: METHODS OF INVESTIGATION AND GENERAL CHARACTERISTICS

The objective of the present study is to investigate the behavior of MSWs in the films described above with a domain structure, magnetized tangentially by a static magnetic field $\mathbf{H}_0$ parallel or perpendicular to one of the projections of a [111] axis onto the plane of the film. In our experiments we have investigated the amplitude-frequency and phase-frequency responses of the MSW transmission coefficient with a spacing of 15 mm between transducers and we have used the results to plot frequency–field curves in the regions where these waves exist, along with the equiphase curves and dispersion curves of MSWs in these regions. The method used to plot these curves is similar to the one described in Ref. 12.

The experiments show that with the transducers oriented along the direction of the field $\mathbf{H}_0$ magnetostatic surface waves propagating perpendicular to the direction of $\mathbf{H}_0$ are efficiently generated in films having a domain structure and any crystallographic orientation. The surface character of these waves is confirmed by the form of the dispersion curves and the substantial variation of the wave attenuation as the direction of wave propagation changes. Here the excitation frequency intervals, the nature of the dispersion, and the character of the hysteresis depend strongly on the crystallographic orientation of the films relative to the field. In fields H_0 strong enough to magnetize the film to saturation magnetostatic surface waves are excited in all the investigated films, in perfect agreement with theory,[21] and hysteresis phenomena are not observed.

We note that when the MSW transducers are oriented perpendicular to the direction of the field $\mathbf{H}_0$ (backward-traveling volume magnetostatic waves are excited in saturated films in this case), waves are not excited in any of the investigated films, regardless of their crystallographic orientation.

We now examine in detail the above-mentioned properties of magnetostatic surface waves for various crystallographic orientations of the films.

4. PROPAGATION OF MAGNETOSTATIC SURFACE WAVES IN YTTRIUM IRON GARNET FILMS MAGNETIZED ALONG THE PROJECTION OF A [111] AXIS ONTO THE PLANE OF THE FILM

When films are magnetized by a field parallel to one of the projections of [111] axes onto the plane of the film, three characteristic fields can be distinguished according to the behavior of the MSWs: $H^{\parallel}_{\min 1}$, $H^{\parallel}_{\min 2}$, and $H^{\parallel}_{\rm sat}$. In fields $H_0 > H^{\parallel}_{\rm sat}$ YIG films are magnetized to saturation, and MSW propagation in them complies with theory.[21] When the field is decreased from a value greater than $H^{\parallel}_{\rm sat}$ to $\simeq 0$ (from now on we briefly characterize such a variation of the field as a decrease of H_0 from $H^{\parallel}_{\rm sat}+\delta$ to $\simeq 0$, where δ is a small field increment of the order of 2–3 Oe), the excitation of MSWs is observed until H_0 decreases to the value $H^{\parallel}_{\min 1}$ (MSWs are not excited for $H_0 < H^{\parallel}_{\min 1}$). The value of $H^{\parallel}_{\min 1}$ falls between the limits 1–3 Oe for all the investigated films. When the field is increased from $\simeq 0$ to a value greater than $H^{\parallel}_{\rm sat}$ (from now on we briefly characterize such a variation of the field as an increase of H_0 from $\simeq 0$ to $H^{\parallel}_{\rm sat}+\delta$) the excitation of MSWs does not begin until H_0 exceeds the value $H^{\parallel}_{\min 2}$. The value of $H^{\parallel}_{\min 2}$ for all the investigated films falls within the limits 2–5 Oe, i.e., $H^{\parallel}_{\min 2} > H^{\parallel}_{\min 1}$ by 1–4 Oe, a fact that has been noted earlier[15] in an investigation of the magnetization reversal of YIG films. As the field H_0 varies from $\simeq 0$ to $H^{\parallel}_{\rm sat}+\delta$ and back again, the behavior of the MSWs exhibits a distinct hysteretic character, which we propose to describe below for the magnetization of the YIG film No. 1 along axis *1* in Fig. 1. For film No. 1 we have $H^{\parallel}_{\min 1}=1$ Oe, $H^{\parallel}_{\min 2}=2.5$ Oe, and $H^{\parallel}_{\rm sat}=33.5$ Oe.

We first consider the onset of hysteresis in the variation of the amplitude-frequency response of the MSW transmission coefficient from the exciting transducer to the receiving transducer. Figure 2 shows the response curves for H_0 increasing from $\simeq 0$ to $H^{\parallel}_{\rm sat}+\delta$ (graphs *1–4*) and for H_0 increasing between the same limits (graphs *5–8*). It is evident from the figure that when H_0 is equal to, for example, 4.5 Oe and the field is increasing, MSWs are excited in the frequency interval 1600 − 1900 MHz (Fig. 2, graph *1*), whereas in the case of a decreasing field MSWs are excited in a significantly different frequency interval, 850–1900 MHz, for the same value of H_0 (Fig. 2, graph *8*). A pairwise comparison of the amplitude-frequency response curves corresponding to higher values of H_0 (graphs *2* and *7*, *3* and *6*, *4* and *5* in Fig. 2) shows that in both cases these curves gradually broaden and at $H_0 = H^{\parallel}_{\rm sat}$ (when the film is saturated) become identical, occupying the frequency interval 700–2400 MHz, consistent with theory[21] (see graphs *4* and *5*). Consequently, like values of the field H_0 correspond to different frequency intervals occupied by the MSW amplitude-frequency response curves, depending on the direction in which H_0 varies, i.e., on the history of the process, this being the most characteristic general feature of hysteresis phenomena. Obviously, if an increase of the field H_0 from $H^{\parallel}_{\rm sat}+\delta$ to $\simeq 0$ is followed by H_0 once again beginning to increase from $\simeq 0$ to $H^{\parallel}_{\rm sat}+\delta$, MSWs having ceased to be excited at $H_0 < H^{\parallel}_{\min 1}$ begin to be excited again (at $H_0 > H^{\parallel}_{\min 2}$) in the interval 1600–1900 MHz (graph *1* in Fig. 2), i.e., all the peaks repeat.

We also note a special property observed on the amplitude-frequency response curves of the MSW spectrum: As the field H_0 decreases from $H^{\parallel}_{\rm sat}+\delta$ to $\simeq 0$, the amplitude-frequency response acquires an attenuation drop, which shifts from low frequencies to high frequencies as the field decreases, traversing the entire spectrum of magnetostatic waves. Inasmuch as the phase-frequency response and the dispersion of MSWs (described below) remain essentially continuous in the vicinity of the drop, the attenuation drop is not a place where two different spectra of these waves merge (even though the form of the amplitude-frequency response curve might suggest such an interpretation), but rather is attributable to resonance factors, which will be discussed below in Sec. 9. In Fig. 2 this drop is conspicuous on graph *6* near frequency 1600 MHz and on graph *7* near 1900 MHz. As the drop shifts from the low-

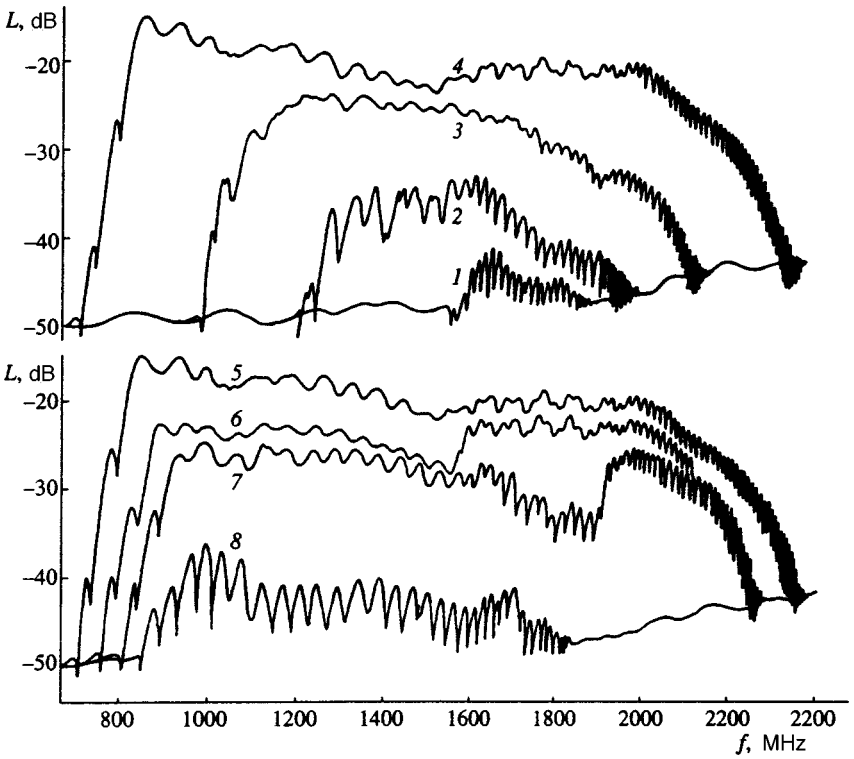

FIG. 2. Amplitude-frequency response curves of the MSW transmission coefficient for YIG film No. 1 magnetized along axis *1* in Fig. 1 with various applied fields: *1, 8*) $H_0=4.5$ Oe; *2, 7*) 17 Oe; *3, 6*) 24 Oe; *4, 5*) 33.5 Oe; curves *1–4* are plotted for a monotonically increasing field H_0; curves *5–8* are plotted for a monotonically decreasing field H_0 (the high-frequency part of the amplitude-frequency response described by curve *6* is not shown).

frequency part to the high-frequency part of the MSW spectrum, its magnitude gradually varies from $\simeq 1$ dB to $\simeq 9$ dB.

We now consider the onset of hysteresis phenomena in the example of the equiphase curves for magnetostatic waves.

The frequency–field regions of excitation of MSWs are labeled in Fig. 3a, which also shows equiphase curves obtained for film No. 1. The values of the wave number k (in cm^{-1}) corresponding to the equiphase curves are shown alongside each curve in the figure. The heavy lines distinguish equiphase curves corresponding to the initial frequency of the MSW spectrum with $k\simeq 0$ or to homogeneous ferromagnetic resonance. It is evident from the figure that as the field H_0 decreases from $H_{\mathrm{sat}}^{\parallel}+\delta$ to $\simeq 0$, two regions of wave excitation are observed, indicated by solid curves: a high-frequency region $S_{h1}^{\parallel}$ and a low-frequency region $S_{l1}^{\parallel}$. On the other hand, when $H_0\simeq 0$ increases from $\simeq 0$ to $H_{\mathrm{sat}}^{\parallel}+\delta$, two different regions of MSW generation (dashed curves) appear in the same field interval: a high-frequency region $S_{h2}^{\parallel}$ and a low-frequency region $S_{l2}^{\parallel}$. It is evident from Fig. 3a that, for example, in region $S_{h1}^{\parallel}$ (decreasing field) at $H_0\simeq 4.5$ Oe MSWs with $k=50\,\mathrm{cm}^{-1}$ are excited at a frequency of 1150 MHz, whereas in region $S_{h2}^{\parallel}$ (increasing field) at the

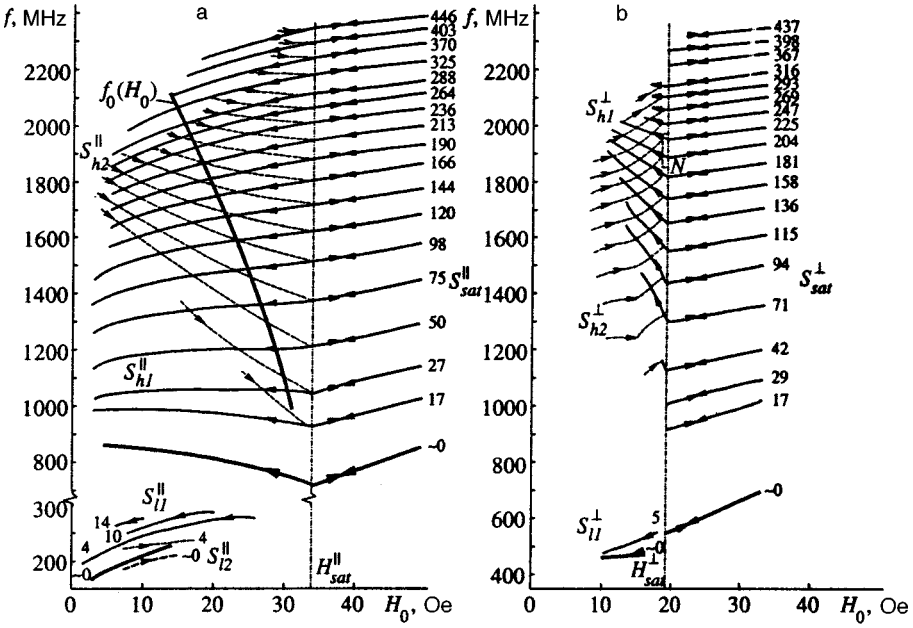

FIG. 3. Frequency–field regions in which magnetostatic surface waves exist and their equiphase curves for YIG film No. 1 (the values of k in cm^{-1} are given alongside each equiphase curve): a) field H_0 applied along one of the projections of [111] axes onto the plane of the film (along axis *1* in Fig. 1); b) field H_0 applied perpendicular to the same projection of a [111] axis. The equiphase curves corresponding to an increasing field H_0 are shown as dashed curves, and the curves corresponding to a decreasing field H_0 and the saturated state of the film are shown as solid curves. The low-frequency boundaries of the regions of existence of MSWs are represented by heavy lines.

same value of H_0 waves with $k=50\,\mathrm{cm}^{-1}$ are excited at a much higher frequency, 1680 MHz. It is evident from the equiphase curves that for any field H_0 MSWs with any wave number k are excited at significantly different frequencies in each of the regions $S^{\parallel}_{h1}$ and $S,^{\parallel}_{h2}$, where the higher the value of H_0, the smaller is the difference in the frequencies. At $H_0=H^{\parallel}_{\mathrm{sat}}$ (when the field is saturated) the equiphase curves of the two high-frequency regions $S^{\parallel}_{h1}$ and $S^{\parallel}_{h2}$ converge, pass through a kink, and make a continuous transition to equiphase curves of ordinary magnetostatic surface waves from the saturation region $S^{\parallel}_{\mathrm{sat}}$, in which the behavior of the MSWs is fully consistent with theory.[13,21] Consequently, conspicuous hysteresis of the MSW equiphase curves for the high-frequency regions is observed in the field interval $\simeq 0<H_0<H^{\parallel}_{\mathrm{sat}}$. As the field H_0 varies from $\simeq 0$ to $H^{\parallel}_{\mathrm{sat}}+\delta$ and back again, MSWs from $S^{\parallel}_{h1}$ described by the solid curves are generated instead of the waves from $S^{\parallel}_{h2}$ described by the dashed curves. A repeated variation of the field between the same limits is accompanied by repetition of the entire cycle, i.e., once again waves from $S^{\parallel}_{h2}$ and then from $S^{\parallel}_{h1}$ are generated (if the increasing field does not attain $H^{\parallel}_{\mathrm{sat}}$, MSWs from $S^{\parallel}_{h2}$ are generated, and if the decreasing field does not attain $H^{\parallel}_{\mathrm{min1}}$, waves from $S^{\parallel}_{h1}$ are generated). It is obvious that the equiphase curves for region $S^{\parallel}_{h1}$ correspond to the amplitude-frequency response curves represented by graphs *5–8* in Fig. 2, and the equiphase curves for region $S^{\parallel}_{h2}$ correspond to the response curves represented by graphs *1–4* in Fig. 2. Along with the equiphase curves, Fig. 3a also shows the H_0 dependence of the frequency f_p at which a drop in the attenuation of the signal is observed on the amplitude-frequency response curve (Fig. 2, graphs *6* and *7*). It is evident from Fig. 3a that the graph of $f_p(H_0)$ is close to a straight line and at $H_0=H^{\parallel}_{\mathrm{sat}}$ passes through the kink of the equiphase curve with $k\simeq 0$.

The variation of the MSW equiphase curves in the low-frequency regions $S^{\parallel}_{l1}$ and $S^{\parallel}_{l2}$ does not exhibit as pronounced a hysteretic behavior; in the behavior of these equiphase curves there is a noticeable trend toward convergence at $H_0\simeq 0$, and the dashed equiphase curves lie below the solid curves (the opposite of the situation in the high-frequency regions). Magnetostatic waves with small wave numbers ($k<20\,\mathrm{cm}^{-1}$) are always excited in regions $S^{\parallel}_{l1}$ and $S^{\parallel}_{l2}$, and the excitation efficiency is not very high (the attenuation is $\simeq -50$ dB). The excitation of waves in $S^{\parallel}_{l1}$ ceases for $H_0>17$ Oe, and in $S^{\parallel}_{l2}$ it ceases for $H_0>26$ Oe.

We now demonstrate the onset of hysteretic properties of magnetostatic waves in the example of the dispersion curves.

The MSW dispersion curves $f(k)$ measured at various fixed values of the field H_0 for regions $S^{\parallel}_{h2}$ (the field increasing from $\simeq 0$ to $H^{\parallel}_{\mathrm{sat}}+\delta$), $S^{\parallel}_{\mathrm{sat}}$ (at $H_0=H^{\parallel}_{\mathrm{sat}}$), and $S^{\parallel}_{h1}$ (the field decreasing from $H^{\parallel}_{\mathrm{sat}}+\delta$ to $\simeq 0$) are shown in Fig. 4 (curves *1–3*, *4*, and *5–8*, respectively; curves *1–3*, like their corresponding region, are dashed, and curves *4–8*, by analogy, are solid). It is evident from a pairwise comparison of curves *1* and *7*, *2* and *6*, *3* and *5* measured for the same fields that the $f(k)$ curves for region $S^{\parallel}_{h2}$ is situated much higher on the frequency scale than the analogous curves for $S^{\parallel}_{h1}$; the

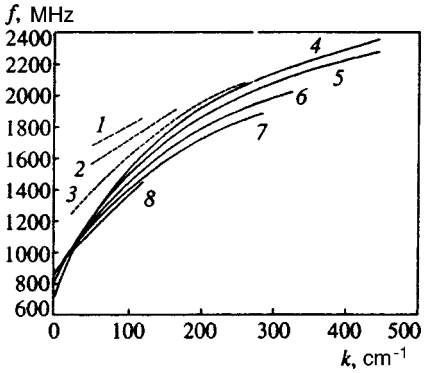

FIG. 4. Dispersion curves $f(k)$ of magnetostatic surface waves for YIG film No. 1 magnetized along one of the projections of [111] axes onto the plane of the film (along axis *1* in Fig. 1) for regions $S^{\parallel}_{h1}$ (curves *1–3*), $S^{\parallel}_{\mathrm{sat}}$ (curve *4*), and $S^{\parallel}_{h2}$ (curves *5–8*). The curves are the results of measurements for the following values of the magnetic field H_0: *1, 7*) 4.7 Oe; *2, 6*) 11.0 Oe; *3, 5*) 21.7 Oe; *4*) 33.5 Oe; *8*) 2.5 Oe.

dispersion curve $f(k)$ for $H_0=H^{\parallel}_{\mathrm{sat}}$ (curve *4*) occupies an intermediate position. All the dispersion curves are characteristic of surface-type magnetostatic waves. We also note an appreciable change in the slopes of the indicated curves corresponding to different fields and a change in the interval of excited wave numbers, which is especially noticeable for region $S^{\parallel}_{h2}$, in which waves with small k appear only when the field is close to $H^{\parallel}_{\mathrm{sat}}$. The hysteretic behavior of the dispersion of MSWs is manifested in the fact that as the field H_0 varies from $\simeq 0$ to $H^{\parallel}_{\mathrm{sat}}+\delta$ and back again, the MSW dispersion also makes a gradual transition from curve *1* to curve *8*, after which waves are no longer excited (at $H_0<H^{\parallel}_{\mathrm{min1}}$), and at the beginning of a new cycle ($H_0>H^{\parallel}_{\mathrm{min2}}$) their dispersion is again described by curve *1*.

We also discuss briefly the description of hysteresis phenomena for cases in which the field H_0 is applied along two other projections of [111] axes onto the plane of the film (axes *2* and *3* in Fig. 1). We note at once that the hysteretic behavior of the amplitude-frequency response, equiphase, and dispersion curves for these two cases do not differ qualitatively in any way from those described above for magnetization of the film along axis *1* (Fig. 1), but there are significant quantitative differences. The differences are best described by comparing the hysteresis of the equiphase curves for all three cases. We note at the outset that each of the fields $H^{\parallel}_{\mathrm{min1}}$, $H^{\parallel}_{\mathrm{min2}}$, and $H^{\parallel}_{\mathrm{sat}}$ is approximately identical in all three cases (differing at most by ± 1 Oe). All the MSW characteristics are also exactly the same in the part of the hysteresis cycle where H_0 decreases from $H^{\parallel}_{\mathrm{sat}}+\delta$ to $\simeq 0$ (this part of the cycle is described by the solid equiphase curves in Fig. 3a). But when the field decreases from $\simeq 0$ to $H^{\parallel}_{\mathrm{sat}}+\delta$, the slopes of the equiphase curves (represented by dashed lines in Fig. 3a) differ substantially in all three film magnetizations. For example, the slope of the dashed equiphase curve with $k=98\,\mathrm{cm}^{-1}$ in Fig. 3a is equal to $S^1_{98}=-9.7$ MHz/Oe (the dashed equiphase curves can be regarded approximately as straight lines), while the slopes of the analogous equiphase curves for magnetizations along axes *2* and *3* (Fig. 1) have the respective values

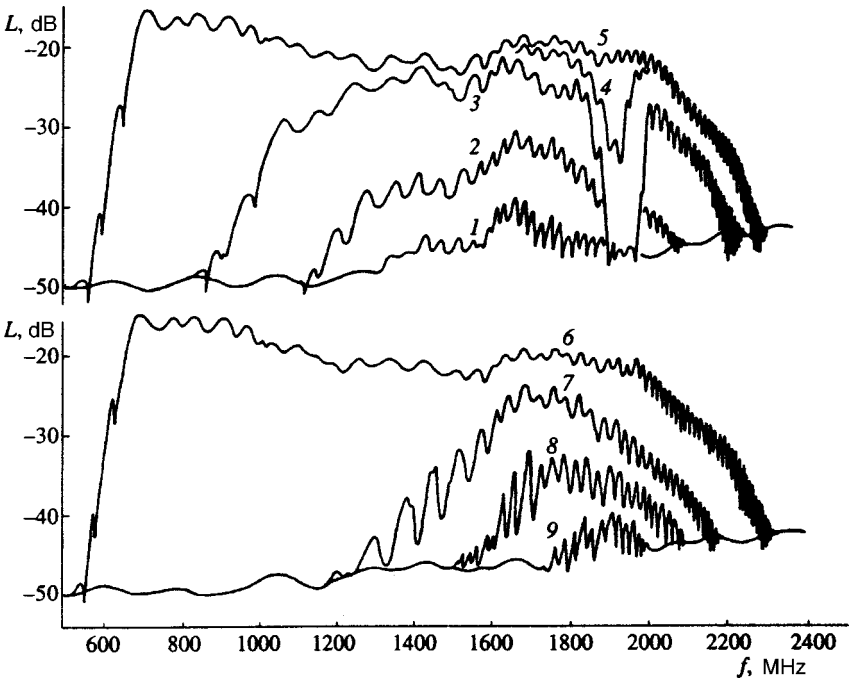

FIG. 5. Amplitude-frequency response curves of the MSW transmission coefficient for YIG film No. 1 magnetized perpendicular to axis *1* in Fig. 1 with various applied fields: *1*) $H_0=7.2$ Oe; *2*) 11.2 Oe; *3*) 18.4 Oe; *4*) 18.8 Oe; *5, 6*) 19.1 Oe; *7*) 17.8 Oe; *8*) 15.2 Oe; *9*) 11.2 Oe; curves *1–5* are plotted for a monotonically increasing field H_0; curves *6–9* are plotted for a monotonically decreasing field H_0 (to aid visualization of the figure, the amplitude-frequency responses described by curves *1, 2,* and *4* are not shown in their entirety).

$S_{98}^2=-13.1$ MHz/Oe and $S_{98}^3=+0.7$ MHz/Oe. Consequently, the smallest hysteresis loop of the equiphase curves is observed when the film is magnetized along axis *3*, the hysteresis loop of the equiphase curves is a maximum for magnetization along axis *2*, and for the case described in detail above, i.e., magnetization along axis *1*, the size of the hysteresis loop of the equiphase curves is intermediate between the other two. The other MSW characteristics for the three cases described here also differ appreciably when the field increases from $\simeq 0$ to $H_{\rm sat}^{\parallel}+\delta$. For example, a comparison of the amplitude-frequency responses has revealed that for each of the three film magnetizations (along axes *1*, *2*, and *3*) at $H_0=4.5$ Oe the frequency intervals in which MSWs are generated are 1600–1900 MHz, 1850–2020 MHz, and 1500–1800 MHz, respectively. Only the dashed MSW dispersion curves $f(k)$ exhibit any changes; for magnetization along axis *3* these curves shift closer in frequency to the dispersion curve for the saturated film (curve *4* in Fig. 4), and for magnetization along axis *2* they are situated farther along the frequency scale from the dispersion curve for the saturated film.

5. PROPAGATION OF MAGNETOSTATIC SURFACE WAVES IN YTTRIUM IRON GARNET FILMS MAGNETIZED PERPENDICULAR TO THE PROJECTION OF A [111] AXIS ONTO THE PLANE OF THE FILM

When the film is magnetized by a field H_0 perpendicular to one of the projections of [111] axes onto the plane of the film, the situation is analogous to the preceding case in that three characteristic fields can be distinguished according to the behavior of the MSWs: $H_{\rm min1}^{\perp}$, $H_{\rm min2}^{\perp}$, and $H_{\rm sat}^{\perp}$. In fields $H_0>H_{\rm sat}^{\perp}$ YIG films are magnetized to saturation, and the propagation of MSWs in them conforms to theory.[21] As the field decreases from $H_{\rm sat}^{\perp}$ to $\simeq 0$, the excitation of MSWs is observed until H_0 decreases to the value $H_{\rm min1}^{\perp}$ (waves are not excited for $H_0<H_{\rm min1}^{\perp}$). For all the investigated films the value of $H_{\rm min1}^{\perp}$ lies in the interval 7–16 Oe. As the field increases from $\simeq 0$ to $H_{\rm min2}^{\perp}+\delta$, MSWs are excited only for $H_0>H_{\rm min2}^{\perp}$. For all the investigated films the field $H_{\rm min2}^{\perp}$ lies in the interval 5–12 Oe, i.e., its position differs from the case of the film magnetization described in Sec. 4, $H_{\rm min2}^{\perp}<H_{\rm min1}^{\perp}$. As the field H_0 varies from $\simeq 0$ to $H_{\rm sat}^{\perp}+\delta$ and back again, the behavior of the MSW characteristics again exhibits hysteresis, which we describe below for the magnetization of film No. 1 perpendicular to axis *1* (Fig. 1). For film No. 1 we have $H_{\rm min1}^{\perp}=10.2$ Oe, $H_{\rm min2}^{\perp}=7$ Oe, and $H_{\rm sat}^{\perp}=19.1$ Oe.

As in Sec. 4, we begin our analysis by comparing the amplitude-frequency responses of the MSW transmission coefficient. Figure 5 shows the amplitude-frequency response curves corresponding to the cases of H_0 increasing from $\simeq 0$ to $H_{\rm sat}^{\perp}+\delta$ (graphs *1–5*) and of H_0 decreasing within the same limits (graphs *6–9*). It is evident from the figure that as the field H_0 increases, the frequency interval in which waves are excited broadens from 1400–1900 MHz at $H_0=7.2$ Oe (Fig. 5, graph *1*) to 1200–2100 MHz at $H_0=11.2$ Oe (Fig. 5, graph *2*). At $H_0=16.9$ Oe on the spectral curve a loss of MSW excitation efficiency takes place in the vicinity of 1100 MHz and the frequency $f_p'=2000$ MHz, but in the vicinity of 1100 MHz this loss of efficiency vanishes at $H_0=18.4$ Oe. In the vicinity of the frequency f_p' at $H_0=18.4$ Oe, conversely, a total lack of MSW generation is observed in the frequency band $\simeq 100$ MHz (Fig. 5, graph *3*), but excitation gradually resumes with a further increase in the field H_0 (Fig. 5, graphs *4* and *5*). We also note that in the interval $17.7\ {\rm Oe}<H_0<18.2$ Oe a degradation of MSW excitation is observed over the entire spectrum, appearing as a change $\simeq -12$ dB in the signal attenuation over the entire amplitude-frequency response curve (to preserve clarity in the figure, the corresponding graphs are not shown). As H_0

decreases from $H_{\text{sat}}^{\perp}+\delta$ to $\simeq 0$, the variation of the spectral curve is simpler than in the preceding case: The efficiency of wave excitation decreases, and the frequency interval in which MSWs are excited gradually shrinks from 550–2300 MHz at $H_0=19.1$ Oe (Fig. 5, curve *6*) to 1750–2000 MHz at $H_0=11.2$ Oe (Fig. 5, curve *9*). It is evident from Fig. 5 that the behavior of the MSW amplitude-frequency response for the cases of increasing and decreasing fields differs significantly, as is typical of hysteresis phenomena. These characteristics become identical for both cases at $H_0=H_{\text{sat}}^{\perp}=19.1$ Oe. We note, in addition, that if H_0 is decreasing from $H_{\text{sat}}^{\perp}+\delta$, but this decrease is halted at $H_0\simeq H_{\text{min1}}^{\perp}=10.5$ Oe, and H_0 begins to increase, the same degradation of MSW excitation will be observed in the interval of fields $16.9\text{ Oe}<H_0<19.1$ Oe as on graphs *3* and *4* in Fig. 5 for the case of the field increasing from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$.

We now turn our attention to the equiphase curves.

The frequency–field regions of MSW excitation are indicated in Fig. 3b, which also shows the curves obtained for film No. 1. It is evident from the figure that two regions of wave excitation, represented by solid curves, are observed as the field H_0 decreases from $H_{\text{sat}}^{\perp}+\delta$ to $\simeq 0$: a high-frequency region $S_{h1}^{\perp}$ and a low-frequency region $S_{l1}^{\perp}$. On the other hand, when H_0 increases from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$, a different high-frequency region of MSW excitation $S_{h2}^{\perp}$, represented by dashed curves, appears in the same field interval. It is evident from Fig. 3b that for any field H_0 a wave with a definite wave number k is excited at substantially different frequencies in each of the regions $S_{h1}^{\perp}$ and $S_{h2}^{\perp}$ (e.g., for $H_0=11$ Oe a magnetostatic wave with $k=158\text{ cm}^{-1}$ is excited at a frequency of 1880 MHz in region $S_{h1}^{\perp}$ and at 1650 MHz in region $S_{h2}^{\perp}$), where the higher the value of H_0, the smaller is the frequency difference. The equiphase curves in both regions $S_{h1}^{\perp}$ and $S_{h2}^{\perp}$ converge at $H_0=H_{cr}^{\perp}=18.1$ Oe, and the equiphase curves in $S_{h2}^{\perp}$ acquire a kink in this case. In the interval $18.1\text{ Oe}<H_0<19.1$ Oe the behavior of the equiphase curves is the same in $S_{h1}^{\perp}$ and $S_{h2}^{\perp}$, and for $H_0=H_{\text{sat}}^{\perp}=19.1$ Oe the equiphase curves of both regions, acquiring a kink, join with the equiphase curves of ordinary magnetostatic surface waves in the saturation region $S_{\text{sat}}^{\perp}$, in which the behavior of the MSWs is fully consistent with theory.[13,21] Consequently, as the field H_0 varies from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$ and back again, hysteresis is observed on the part of the equiphase curves for the high-frequency regions: Instead of the waves in $S_{h2}^{\perp}$ described by the dashed curves we now have waves in $S_{h1}^{\perp}$ described by the solid curves; in contrast with the film magnetization described in Sec. 4, here the dashed curves lie below the solid curves. A repeated variation of the field from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$ and back again leads to repetition of this cycle, i.e., once again waves described by the dashed curves and then waves described by the solid curves are excited. It is important to note a characteristic feature inherent in the dashed equiphase curves. These curves, shown in Fig. 3b, represent a certain average of the process actually observed, i.e., as H_0 is varied continuously, the phase-frequency response changes in small jumps, which occur after the field changes by 0.3–0.8 Oe, and to faithfully

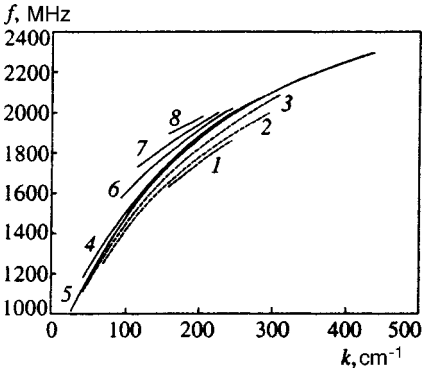

FIG. 6. Dispersion curves $f(k)$ of magnetostatic surface waves for YIG film No. 1 magnetized perpendicular to one of the projections of [111] axes onto the plane of the film (perpendicular to axis *1* in Fig. 1) for regions $S_{h1}^{\parallel}$ (curves *1–4*), $S_{\text{sat}}^{\parallel}$ (curve *5*), and $S_{h2}^{\parallel}$ (curves *4, 6–8*). The curves are the results of measurements for the following values of the magnetic field H_0: *1*) 8 Oe; *2, 8*) 10.5 Oe; *3, 6*) 15.1 Oe; *4*) 18.1 Oe; *5*) 19.1 Oe; *7*) 12.8 Oe [the initial part of $f(k)$ described by curve *5* in the frequency interval 500–1000 MHz is not shown.

reproduce the equiphase curves in the figure, it would be necessary to draw something like a stepladder with steps of width 0.3–0.8 Oe and a height that can vary from $\simeq 1$ MHz to 20 MHz. In the case of the solid equiphase curves, on the other hand, as H_0 varies continuously, the phase-frequency response also changes continuously. It is obvious that the dashed equiphase curves correspond to the MSW amplitude-frequency responses represented by graphs *1–5* in Fig. 5, and the solid equiphase curves corresponds to the MSW amplitude-frequency responses represented by graphs *6–9* in Fig. 5. The observable lack of wave excitation in the vicinity of the frequency f_p' on the MSW amplitude-frequency response curves (graph *3* in Fig. 5) corresponds to the region *N* in Fig. 3b (outlined by a dashed line) wherein the dashed equiphase curves with wave numbers $190\text{ cm}^{-1}<k<230\text{ cm}^{-1}$ (or wavelengths $270\ \mu\text{m}<\lambda<330\ \mu\text{m}$) suffer a discontinuity, i.e., MSWs with these values of k are not excited. We note that region *N* is observed in the field interval $18.1\text{ Oe}<H_0<19.1$ Oe, in which the dashed and solid equiphase curves coincide. Moreover, it is evident from a comparison of Figs. 3a and 3b that the frequency–field position of region *N* coincides with the frequency–field position of the $f_p(H_0)$ curve describing the behavior of the signal attenuation drop on the MSW amplitude-frequency response curve (region *N* begins just above this curve). The latter fact can most likely be interpreted as evidence that the two phenomena have a common cause. We also note that hysteresis of the MSW equiphase curves is not observed in the low-frequency region $S_{l1}^{\perp}$, owing to the strong attenuation of the MSW signal (more than 50 dB) as H_0 decreases from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$; this fact has made it impossible to measure the MSW parameters.

We now demonstrate the hysteretic behavior of the magnetostatic wave dispersion curves.

The MSW dispersion curves $f(k)$ measured for various fixed fields H_0 in the regions $S_{h2}^{\perp}$ (field increasing from $\simeq 0$ to $H_{\text{sat}}^{\perp}+\delta$) and $S_{h1}^{\perp}$ (field decreasing from $H_{\text{sat}}^{\perp}+\delta$ to $\simeq 0$) are shown in Fig. 6 (curves *1–8*). As in Fig. 3b, the dashed

curves correspond to region $S_{h2}^{\perp}$ (curves *1–3*), and the solid curves correspond to $S_{h1}^{\perp}$ (curve *6–8*). For $H_0 > 18.1$ Oe the $f(k)$ curves for both regions, like the equiphase curves in Fig. 3b, coincide, so that these $f(k)$ curves are also represented by solid lines in Fig. 6 (curves *4* and *5*). It is evident from Fig. 6, in contrast with Fig. 4, that the dashed dispersion curves corresponding to an increasing field are far lower on the frequency scale than the solid curves corresponding to a decreasing field. All the dispersion curves are characteristic of surface-type magnetostatic waves. We also note an appreciable change in the slopes of the indicated curves corresponding to different fields, along with a variation of the interval of excited wave numbers both in $S_{h1}^{\perp}$ and in $S_{h2}^{\perp}$ (low-k MSWs occur only for fields close to $H_{\text{sat}}^{\perp}$). The hysteretic behavior of the MSW dispersion is manifested in the fact that as the field H_0 varies $\simeq 0$ to $H_{\text{sat}}^{\perp} + \delta$) and back again, the dispersion also gradually changes in accordance with curves *1–8*, after which waves are no longer excited (for $H_0 < H_{\text{min1}}^{\perp}$), and at the beginning of a new cycle (for $H_0 > H_{\text{min2}}^{\perp}$) their dispersion is again described by curve *1*.

When the YIG film No. 1 is magnetized perpendicular to other projections of [111] axes onto the plane of the film (perpendicular to axes *2* or *3* in Fig. 1), we do not observe a significant difference, either qualitatively or quantitatively, in the hysteretic character of the amplitude-frequency response, equiphase, and dispersion curves from the case of film magnetization described above. Small differences are observed only for an increasing field H_0. They are attributed mainly to a variation of $H_{\text{min2}}^{\perp}$ (within ± 2-Oe limits) and a certain decrease in the average slope of the equiphase curves (by approximately 3 MHz/Oe) relative to the case described above.

6. DOMAIN STRUCTURES OF THE INVESTIGATED FILMS; GENERAL CHARACTERIZATION

It is evident from the preceding sections that hysteretic behavior of the MSWs is observed only in fields insufficient for the saturation of YIG films, i.e., when a domain structure exists in the sample. The significant influence of domains on MSW propagation has been noted previously.[1–15] To discern the nature of this influence in our case, we have measured the parameters of the domain structure of the films simultaneously with measurements of the MSW characteristics. We have performed parallel measurements of the domain structure and MSW parameters in an effort to exhibit an interrelationship between the hysteretic behavior of MSWs and the variation of the domain structure. We consider it appropriate to summarize the essence of this interaction right at the outset: As the field H_0 varies from 0 to a level above saturation and back again, hysteretic behavior of the MSW parameters is observed in parallel with hysteresis variation of the domain structure parameters. We emphasize that hysteresis of the domain structure has always been linked to a change in the orientation of the domain walls. The factors responsible for the change in orientation of the domain walls can be exposed by analyzing the behavior of the domain structure as the field H_0 varies from 0 to $H_{\text{sat}} + \delta$ and back again. Following is a brief description of this behavior.

We note at the outset that, in contrast with the films described in Refs. 11 and 12, a regular, symmetric stripe domain structure is established at $H_0 = 0$ in the majority of the investigated films, irrespective of the magnetization history, the walls of this structure being oriented somewhere between the two directions defined by those projections of [111] axes onto the plane of the film, between which lies the projection of the uniaxial anisotropy axis onto the plane of the film (axes *1* and *2* in Fig. 1). The pattern of the observed domain structure has high contrast, like the films described in Refs. 11 and 12, implying that the domain magnetizations are oriented close to the normal to the plane of the film. In YIG film No. 1 the period of this domain structure was found to be $T_0 = 23.1\ \mu$m, and the angle φ_0 at which the domain walls were tilted relative to axis *1* in Fig. 1 was equal to $(+6 \pm 3)°$.

When a field $H_0 = H_s = 0.5$ Oe is applied in any crystallographic direction of the tangent plane of a film, the contrast of the observed pattern of the domain structure decreases abruptly, i.e., the magnetizations of the domains are oriented close to the plane of the film. When the domain structure evolves into this low-contrast form, the domain walls probably also acquire the property of relative mobility (in comparison with the domain structure of the films described in Refs. 11 and 12, so that the walls can change their orientation as the energy advantage of the directions changes. Since the field H_0 has a strong influence on the depth of the energy minima corresponding to the four axes (*1–3* and *A* in Fig. 1), as H_0 increases in the interval $H_s < H_0 < H_{\text{min2}}$, the domain walls change their orientation from the direction specified by the angle φ_0 to some other direction, which depends on the crystallographic orientation of the film relative to $\mathbf{H}_0$. As a rule, the reorientation of the domain walls takes place through the emergence of a block domain structure in the field interval $H_s < H_0 < H_{\text{min2}}$, wherein the domains inside each block are in the form of straight stripes of identical width, and the domain walls are parallel to one of the energy-favorable directions. The value of H_{min2} for all the investigated films lies in the interval 2–12 Oe (in isolated cases, which will be described below, a block domain structure does not occur for $H_0 < H_{\text{min2}}$, and at $H_0 = H_{\text{min2}}$ the domain structure changes from one regular type to another regular type).

When the field is further increased from H_{min2} to the value H_{sat} at which the film is magnetized to saturation, a regular or quasiregular domain structure exists in the film. We apply the term ''quasiregular'' to a domain structure that is almost regular except for small distortions (e.g., the walls are not strictly straight lines, but are almost straight, or occasional dead-end domains might be encountered). We assume that this domain structure should be distinguished from other irregular domain structures, because when a quasiregular domain structure is present in the film, the efficiency of excitation of MSWs is essentially the same as when a true regular domain structure is present, whereas MSWs are not excited when an irregular domain structure exists in the film (i.e., in essence, a wave does not sense infrequent and very small — of a scale much smaller than the wavelength — distortions of the domain structure). A quasiregular domain

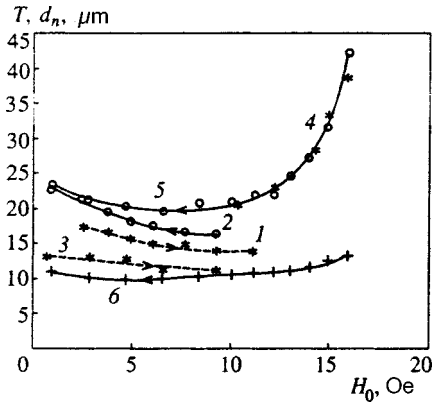

FIG. 7. Graphs of the periods T of the stripe domain structure and the narrow-domain widths d_n of the asymmetric domain structure versus the magnetic field H_0 for YIG film No. 1 magnetized along (curves *1* and *2*) and perpendicular to (curves *3*–*6*) one of the projections of [111] axes onto the plane of the film (axis *1* in Fig. 1): *1, 3, 4,* *) $T(H_0)$ for an increasing field H_0 (dashed curves; *4* coincides with *5* in the interval 10 Oe$<H_0<$16 Oe); *2, 5,* ○) $T(H_0)$ for a decreasing field H_0 (solid curves); *6,* +) narrow-domain width D_n versus field H_0 for a domain structure whose period is described by curve *5*.

structure usually occurs in films as the field H_0 increases from 0 to $H_0>H_{\min 2}$.

The process of H_0 decreasing from $H_{\rm sat}+\delta$ to values smaller than $H_{\rm sat}$ induces a new domain structure in the film. Since the external field $\mathbf{H}_0$ for the formation of a domain structure in this case is much higher than for the formation of a domain structure in the case of an increasing field, the influence of $\mathbf{H}_0$ on the orientation of the walls dominates the new domain structure: The walls are always straight and oriented either along the direction of $\mathbf{H}_0$ or deviate from this direction at most by $\pm 30°$. This domain structure exists in the film as long as H_0 is smaller than a certain value $H_{\min 1}$. For $H_0<H_{\min 1}$ a block domain structure similar to the one described for the case $H_0<H_{\min 2}$ is established in the films. The values of $H_{\min 1}$ lie in the interval 1–3 Oe for all the investigated films.

Consequently, as the field H_0 varies from 0 to $H_{\rm sat}+\delta$ and back again, hysteresis of the orientation of the domain walls occurs in the films.

It is important to note that a high-contrast domain structure does not occur at $H_0=0$ in some of the investigated films. A low-contrast block domain structure exists in such films at $H_0=0$, but in all other cases its variation corresponds exactly to that described above.

We also call attention to the following general property of the investigated domain structures, which will not be mentioned in the discussion that follows. In the field intervals $H_{\min 1}<H_0<H_{\rm sat}$ and $H_{\min 2}<H_0<H_{\rm sat}$ a drastic reduction in contrast of the pattern of the domain structure has always been observed for values of H_0 close to $H_{\rm sat}$, so that the domain structure could only be distinguished in the initial part of these field intervals. Moreover, the behavior of the structures for $H_0<H_{\min 1}$ and for $H_0<H_{\min 2}$ (MSWs are not excited in these cases) will also be ignored below, because by and large it has already been discussed above.

Now, when describing the behavior of the domain structure in general terms, we consider specifically how the orientation of the domain walls changes for the YIG film No. 1 is magnetized in different crystallographic directions.

7. DOMAIN STRUCTURES FOR MAGNETIZATION OF THE FILM ALONG PROJECTIONS OF [111] AXES ONTO THE PLANE OF THE FILM

As shown in Fig. 1, three projections of [111] axes (*1*, *2*, and *3*) lie in the plane of the film. When film No. 1 is magnetized by a field H_0 that increases from 0 to the saturation value $H_{\rm sat}^{\parallel}$ along one of these projections, a quasiregular domain structure appears in the film for $H_0>H_{\min 2}$ ($H_{\min 2}=2.5$ Oe and $H_{\rm sat}^{\parallel}=33.5$ Oe for film No. 1). The angle α between the domain walls and the direction of H_0 depends on the specific [111] projection (Fig. 1) along which the field is applied:

1) $\alpha=(+40\pm 3)°$ for magnetization of the film along axis *1*;

2) $\alpha=-60°$ or $\alpha=+60°$ for magnetization of the film along axis *2* (the MSW characteristics do not differ for different absolute values of α);

3) $\alpha=(+7\pm 3)°$ for magnetization of the film along axis *3*.

As the field decreases from $H_{\rm sat}^{\parallel}+\delta$, where $H_{\rm sat}^{\parallel}$ corresponds to the kink of the MSW equiphase curves (Fig. 3a), the domain walls in the field interval $H_{\min 1}^{\parallel}<H_0<H_{\rm sat}^{\parallel}$ are always oriented parallel to H_0 ($\alpha=0$), no matter which projection of a [111] axis the field is applied along ($H_{\min 1}^{\parallel}=1$ Oe for film No. 1).

In none of the above-described domain structures did the orientation of the walls (i.e., the values of α) change when H_0 was varied within the field intervals $H_{\min 1}^{\parallel}<H_0<H_{\rm sat}^{\parallel}$ and $H_{\min 2}^{\parallel}<H_0<H_{\rm sat}^{\parallel}$ (at least not for values of H_0 such that domain structure could be distinguished in the microscope).

The dependence of the periods of the domain structures on the field, $T(H_0)$, is shown in Fig. 7 for film No. 1 magnetized along axis *1* in Fig. 1 (the propagation of MSWs for this case is described in Sec. 3). Curve *1* corresponds to the field increasing from 0 to $H_{\rm sat}^{\parallel}+\delta$, and curve *2* corresponds to the field decreasing within the same limits. For magnetization of the film along axes *2* and *3*, on the whole, the $T(H_0)$ curves are similar to curves *1* and *2* in Fig. 7 (only the curves corresponding to an increasing field differ slightly).

Consequently, the hysteresis properties of the domain structure are seen in the way the field H_0 varies from 0 to $H_{\rm sat}^{\parallel}+\delta$ and back again, first of all, the orientation of the domain walls changes (e.g., the angle $\alpha=+40°$ changes to $\alpha=0$) and, second, the dependence of the period of the domain structure on H_0 changes (see Fig. 7, curves *1* and *2*). In a repeated variation of H_0 from 0 to $H_{\rm sat}^{\parallel}+\delta$ and back again the cycle of variation of the walls and the period of the domain structure is repeated.

8. DOMAIN STRUCTURES FOR MAGNETIZATION OF THE FILM PERPENDICULAR TO PROJECTIONS OF [111] AXES ONTO THE PLANE OF THE FILM

We first describe the modification of the domain structure when film No. 1 is magnetized along axis *1* in Fig. 1 (the propagation of MSWs for this case is described in detail in Sec. 4).

As the field H_0 increases from 0 to $H_{sat}^{\perp}+\delta$, the high-contrast, regular, symmetric stripe domain structure produced at $H_0=0$ (see Sec. 6) becomes a low-contrast structure at $H_0=H_s=0.5$ Oe, and this new regular, symmetric domain structure exists in the field interval $H_s<H_0<H_{min2}^{\perp}$, where $H_{min2}^{\perp}=10.2$ Oe. The angle α between the walls of the domain structure and the field H_0 remains constant and equal to $\alpha=-84°$ in the transition from the high-contrast to the low-contrast domain structure and in the interval $H_s<H_0<H_{min2}^{\perp}$. The field dependence of the period of the low-contrast domain structure in this field interval is shown in Fig. 7 (curve *3*). Magnetostatic waves are not excited in the film in the field interval $H_s<H_0<H_{min2}^{\perp}$. At $H_0=H_{min2}^{\perp}$ a phase transition takes place in the film in connection with a change in the orientation of the domain walls: The angle α changes abruptly from $-84°$ to $-30°$. As H_0 increases in the field interval $H_{min2}^{\perp}<H_0<H_{sat}^{\perp}$ ($H_{sat}^{\perp}=19.1$ Oe for film No. 1), the domain structure abruptly changes the orientation of the domain walls by a small angle after every increment of 0.2–0.5 Oe, so that the orientation of the walls gradually approaches the direction of the field H_0. At $H_0=15.9$ Oe, above which domain structure is undetectable, the angle α is equal to $-21.5°$. This discontinuous modification of the domain structure is attributable to the abrupt variation of the dashed equiphase curves in Fig. 3b, as mentioned in Sec. 4. The dependence of the period of the domain structure on H_0 for $H_0>H_{min2}^{\perp}$ is described by curve *4* in Fig. 7. We also note that for $H_0>15$ Oe the domain structure begins to transform from symmetric to asymmetric (the analogous transformation for the case of a decreasing field is described more in detail below).

As the field H_0 decreases from $H_{sat}^{\perp}+\delta$ to 0, a regular stripe domain structure is established in the film with an angle $\alpha=-30°$ (or $\alpha=+30°$) between the domain walls and H_0. The angle α remains constant (at least for fields H_0 such that domain structure is discernible) for variation of the field in the interval $H_{min2}^{\perp}<H_0<H_{sat}^{\perp}$. As H_0 decreases, the domain structure gradually transforms from its asymmetric pattern for fields close to $H_{sat}^{\perp}$ to a symmetric structure. The gradual nature of this transformation can be judged from Fig. 7, which shows the period of the domain structure and the width of the narrow domains as functions of the field H_0 (curves *5* and *6*, respectively).

When the film is magnetized perpendicular to one of the other two projections of [111] axes (axes *2* and *3* in Fig. 1), the behavior of the domain structure as H_0 increases in the interval $0<H_0<H_{min2}^{\perp}$ differs from the above-described behavior: In both cases the disappearance of the high-contrast domain structure is followed in this interval by the emergence of a low-contrast block structure, the values of $H_{min2}^{\perp}$ themselves differ only slightly (within 2-Oe limits). The behavior of the domain structures in the field interval $H_{min2}^{\perp}<H_0<H_{sat}^{\perp}$ is analogous, differing slightly as a result of a difference in the values of $H_{min2}^{\perp}$.

Consequently, the domain structure also manifests hysteresis properties in this situation: As the field H_0 varies from 0 to $H_{sat}^{\parallel}+\delta$ and back again, different types of domain structures (block or regular) appear in the film for the same values of H_0, and the orientation of the domain walls and the dependence of the period of the domain structure on H_0 vary altogether differently (see Fig. 7, curves *3–6*). If the variation of H_0 from 0 to $H_{sat}^{\parallel}+\delta$ and back again is repeated, the cycle of variation of the walls and the period of the domain structures is repeated.

9. ANALYSIS OF THE RESULTS

We now analyze the interrelationship between the hysterestic behavior of magnetostatic waves and the modification of the domain structure.

We first consider magnetization of the film along the projection of a [111] axis onto its plane. The analysis of the behavior of MSWs and domain structures in Secs. 3 and 7 leads to the conclusion that the hysteresis of the MSW characteristics (see Figs. 2, 3a, and 4) is associated with hysteretic variation of the orientation of the domain walls. Measurements of the MSW parameters for a film magnetized along three different projections of [111] axes onto its plane show that the more the orientations of the domain walls differ between an increasing field H_0 and a decreasing field (i.e., the greater the absolute value of the difference in the angles α corresponding to increasing and decreasing H_0), the greater will be the spread of the hysteresis loop observed on the MSW equiphase curves. Inasmuch as all three projections of [111] axes onto the plane of the film are crystallographically equivalent, the cause of such striking quantitative differences in the hysteresis of the equiphase curves (and other MSW characteristics) is, in the final analysis, uniaxial anisotropy (probably induced or of magnetoelastic origin), which introduces asymmetry into the plane of the film and whose axis is tilted, usually just a few degrees, relative to the plane of the film.

Consequently, when the film is magnetized along the projection of a [111] axis onto the plane of the film, a necessary condition for MSWs and the domain structure to exhibit hysteresis is the presence of a slight uniaxial anisotropy in the film, its axis deviated somewhat from the normal to the plane of the film.

We now consider magnetization of the film perpendicular to the projection of a [111] axis onto its plane. The analysis of the behavior of MSWs and domain structures in Sec. 4 and 8 leads to the conclusion that the hysteresis of the wave characteristics in the given situation (see Figs. 3b, 5, and 6) is also associated with hysteresis in the variation of the orientation of the domain walls. When the field decreases, the walls of the domain structure are oriented at a fixed angle α relative to the field H_0 (at least for values of H_0 such that a domain structure is distinguishable; $|\alpha|=30°$ holds as a rule, and only occasionally is $|\alpha|$ slightly smaller than 30°).

When H_0 increases, the walls of the domain structure change their orientation in small 0.5°–2° jumps every 0.3–0.8 Oe from $|\alpha| = 30°$, in all likelihood, down to 0°. Unfortunately, low contrast makes the domain structure indistinguishable in near-saturation fields, and in reality $|\alpha|$ has actually been observed to vary from 30° to $\simeq 20°$. This unobservability of the domain structure also makes it impossible to draw definite conclusions to the effect that MSWs with wavelengths $270\ \mu\text{m} < \lambda < 330\ \mu\text{m}$ are not excited in the frequency–field region *N* (Fig. 3b), although the following hypothesis is tenable: If the field dependence of the period of the domain structure (curve *4* in Fig. 7) is extrapolated to $H_0 = H_{\text{sat}}^{\perp} = 19.1$ Oe, it is evident at once that the widths of the broad domains in this case are fully capable of attaining $67-82\ \mu$m, i.e., are exactly equal to one quarter-wavelength for the MSWs not observed in the region *N*.

It appears, therefore, that for the first time we are witnessing the nongeneration of MSWs through a lattice of quarter-wave domain resonators. It is entirely conceivable that the attenuation drop on the MSW amplitude-frequency response curve (see Fig. 2, graphs *6* and *7*) is also of similar origin; the field dependence of the frequency $f_p(H_0)$ at which the drop occurs, shown in Fig. 3a, occurs in a frequency–field region that coincides with the frequency–field region *N* [this explanation is valid, of course, if we assume that the $T(H_0)$ curve (curve *2* in Fig. 7) in the field interval $9\ \text{Oe} < H_0 < 33.5$ Oe begins to rise like curve *4* or *5*).

We can conclude from the sum-total of these considerations that the hysteretic properties of MSWs are determined entirely by the hysteretic properties of regular and quasiregular domain structures. Magnetostatic waves are not excited when block domain structures are present in the film.

Apart from its purely physical appeal, the hysteretic behavior of MSWs as described in the present study can be a useful tool for the investigation of magnetization processes and oriented phase transitions and for measurements of the parameters of a material. Indeed, the majority of such investigations are carried out at the present time by magnetooptical methods involving the observation of domain structure through the Faraday effect.[20,22] However, in situations where the magnetization vector is oriented close to the plane of the film the accuracy of magnetooptical methods is reduced by the weakness of the Faraday effect in this case. On the other hand, the application of MSWs for determining certain parameters of a film and monitoring its domain structure is still reasonably effective. For example, in the present study the film saturation fields $H_{\text{sat}}^{\parallel}$ and $H_{\text{sat}}^{\perp}$ could not be determined by the magnetooptical method, but they have been determined quite accurately from the minimum of the MSW equiphase curves. In addition, the application of MSWs is the only way the hysteretic behavior of the magnetization vector can be assessed in the immediate vicinity of the saturation field, where domain structures are not visible.

Consequently, not only is the application of magnetostatic waves a useful adjunct to the magnetooptical method, in some cases it can even supersede the latter by virtue of its greater sensitivity.

10. CONCLUSIONS

We have investigated experimentally the propagation of zero-exchange spin waves (magnetostatic waves) in tangentially magnetized yttrium iron garnet films with domains in which the magnetization is oriented close to the plane of the film. Waves propagating perpendicular to the direction of the applied field, observed in the range 500–2500 MHz, are similar to magnetostatic surface waves in saturated films. As the applied field varies from 0 to a value above saturation and back again, hysteresis is observed in the variation of the characteristics of these waves, most likely due to the hysteresis observed in the variation of the orientation of the domain walls. The more pronounced the hysteresis of the domain-wall orientation, the greater is the difference in the characteristics and parameters of the waves in increasing and decreasing fields: Waves of identical wavelength can be excited at frequencies that differ by as much as 700 MHz as the field increases or decreases, or waves whose wavelengths differ severalfold can be excited at the same frequency. The difference in the wave parameters measured in increasing and decreasing fields decreases as the applied field approaches saturation. In the spectrum of observable MSWs we observe an interval of wavelengths $270\ \mu\text{m} < \lambda < 330\ \mu$m (or wave numbers $190\ \text{cm}^{-1} < k < 230\ \text{cm}^{-1}$) in which waves do not propagate for an applied field below, but close to, saturation. We have established the fact that waves are not excited along the direction of the applied field (when backward-traveling volume magnetostatic waves are excited in a saturated film).

In addition, we have observed a spectrum of low-frequency excitations (100–500 MHz), which are probably associated with collective resonance vibrations of the domain walls. For these excitations we have also observed hysteresis in the variation of the parameters, where the closer the applied field is to zero, the smaller is the difference between the parameters of these excitations in increasing and decreasing fields.

The authors are indebted to S. V. Gerus for consultation on Ref. 19 and to V. I. Zubkov for assistance in preparing the bibliographic references.

This work has received financial support from the Russian Fund for Fundamental Research (Project 96-02-17283a).

*)E-mail: svg318@ire216.msk.su

[1] I. A. Gilinskiĭ and R. G. Mints, Zh. Éksp. Teor. Fiz. **59**, 1230 (1970) [Sov. Phys. JETP **32**, 673 (1971)].
[2] I. A. Gilinskiĭ and K. A. Ryazantsev, Fiz. Tverd. Tela (Leningrad) **16**, 3008 (1974) [Sov. Phys. Solid State **16**, 1944 (1975)].
[3] D. D. Stancil, J. Appl. Phys. **56**, 1775 (1984).
[4] I. V. Zavislyak and V. V. Danilov, Pis'ma Zh. Tekh. Fiz. **8**, 72 (1982) [Sov. Tech. Phys. Lett. **8**, 31 (1982)].
[5] S. A. Vyzulin, S. A. Kirov, and N. E. Syr'ev, Vestn. Mosk. Univ. Fiz. Astron. **24**, 92 (1983).
[6] S. A. Vyzulin, S. A. Kirov, and N. E. Syr'ev, Vestn. Mosk. Univ. Fiz. Astron. **25**, 70 (1984).
[7] S. A. Kirov, A. I. Pil'shikov, and N. E. Syr'ev, Fiz. Tverd. Tela (Leningrad) **16**, 3051 (1974) [Sov. Phys. Solid State **16**, 1970 (1975)].
[8] L. V. Mikhaĭlovskaya and I. V. Bogomaz, Fiz. Tverd. Tela (Leningrad) **19**, 1245 (1977) [Sov. Phys. Solid State **19**, 725 (1977)].
[9] V. I. Kostenko and M. A. Sigal, Phys. Status Solidi B **170**, 569 (1992).
[10] G. T. Kazakov, A. G. Sukharev, and Yu. A. Filimonov, in *Abstracts of the*

Fifth All-Union School on Spin-Wave Microwave Electronics [in Russian] (Zvenigorod, 1991), p. 83.
[11] A. V. Vashkovskiĭ, E. G. Lokk, and V. I. Sheglov, JETP Lett. **63**, 572 (1996).
[12] A. V. Vashkovskiĭ, É. G. Lokk, and V. I. Sheglov, Zh. Éksp. Teor. Fiz. **111**, 1016 (1997) [JETP **84**, 560 (1997)].
[13] A. G. Gurevich and G. A. Melkov, *Magnetic Oscillations and Waves*, CRC Press, Boca Raton, FL, 1996 [Russ. original Nauka, Moscow, 1994).
[14] Yu. V. Gulyaev, P. E. Zil'berman, G. T. Kazakov *et al.*, Pis'ma Zh. Tekh. Fiz. **11**, 97 (1985) [Sov. Tech. Phys. Lett. **11**, 38 (1985)].
[15] P. E. Zil'berman, G. T. Kazakov, and V. V. Tikhonov, Radiotekh. Elektron. **32**, 710 (1987).
[16] M. M. Farztdinov, *Spin Waves in Ferromagnets and Antiferromagnets with Domain Structure* [in Russian] (Nauka, Moscow, 1988).
[17] V. T. Synogach, Fiz. Tverd. Tela (Leningrad) **32**, 3475 (1990) [Sov. Phys. Solid State **32**, 2015 (1990)].
[18] V. T. Synogach and H. Dötsch, Phys. Rev. B **54**, 15266 (1996).
[19] A. V. Voronenko, S. V. Gerus, and L. A. Krasnozhen, Mikroelektronika **18**, 61 (1989).
[20] K. P. Belov, A. K. Zvezdin, A. M. Kadomtseva, and R. Z. Levitin, *Orientational Transitions in Rare-Earth Magnets* (Nauka, Moscow, 1979).
[21] R. W. Damon and J. R. Eshbach, J. Phys. Chem. Solids **19**, 308 (1961).
[22] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials* (Academic Press, New York, 1979) [Russ. transl., Mir, Moscow, 1982].

Translated by James S. Wood

Magnetoexciton light absorption in inhomogeneous quasi-two-dimensional systems

Yu. E. Lozovik*)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

A. M. Ruvinskiĭ

Moscow State Institute of Steel and Alloys, 117936 Moscow, Russia
(Submitted 25 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1451–1465 (October 1998)

The problem of exciton light absorption in quasi-two-dimensional inhomogeneous systems in a strong transverse magnetic field H is analyzed. We assume that a random Gaussian field (''white noise'') acting separately on an electron and a hole is due to (1) fluctuations in the quantum well thickness or (2) fluctuations in the concentrations of the solid solution components. The problem of a magnetoexciton in a random Gaussian white noise field has been reduced to the problem of the motion in an H-dependent effective field of a single particle with the effective magnetic mass of the exciton, which is a function of the magnetic field and parameters of the quantum wells, in a field characterized by ''colored noise,'' whose correlation function is different from that of the white noise field. In this approximation, the problem of a magnetoexciton in isolated and coupled quantum dots is considered. In the coherent-potential approximation, the exciton absorption in random fields of the first and second type in single and coupled quantum wells has been calculated. The absorption decreases as H increases in the range of strong magnetic fields, which is in agreement with experimental data.
 [S1063-7761(98)02010-1]

1. INTRODUCTION

Recently electron–hole ($e-h$) systems in low-dimensional semiconducting structures have attracted a lot of attention[1–9] in connection with the substantial effects of electron–hole interactions on optical and transport properties of quantum dots, wires, wells, and superlattices, which exhibit very interesting collective phenomena. For example, the superfluidity of $e-h$ pairs in a system of quantum wells separated in space has been predicted, and this effect can be seen in persistent electric currents flowing in coupled wells.[10] Interesting effects involving the entrainment of quasiparticles in one layer by those in the other also take place in such systems (see Ref. 11 and references therein). The attraction between electrons and holes of different quantum wells gives rise to spatially indirect excitons. The probability of tunneling recombination of an indirect exciton can be very low owing to the small overlap between the electron and hole wave functions. An electric field applied normally to quantum layers and drawing an electron and a hole apart reduces the overlap between their wave functions, thereby further reducing the recombination probability. The lifetime of an indirect exciton depends sensitively on the magnetic field,[3] and in the region of low temperatures ($T<1$ K) and strong magnetic fields ($H>7$ T) it is determined by scattering of excitons on the terraces of quantum well interfaces.[12] In this region, change in the magnetic field leads to a considerable drop in the exciton photoluminescence peak intensity.[3] The phase diagram and superfluidity in systems with indirect excitons and magnetoexcitons were studied previously in Ref. 13 (see also references therein).

The present work is aimed at analyzing optical properties of a composite particle (an exciton) in random two-dimensional static fields in a strong magnetic field applied in the vertical direction. Such random fields are generated in quantum wells mostly near interfaces and can be caused by (1) fluctuations in the quantum well width or (2) fluctuations in the 2D concentration of solid solution components. The problem of exciton absorption in the absence of a magnetic field H was analyzed in Refs. 14–16, and localization of excitons in random fields in Ref. 16. The problem of absorption by magnetoexcitons, however, becomes considerably more complicated even in the absence of random fields because the equations do not separate in the center-of-mass and relative coordinates. We have bypassed this difficulty by constructing an effective Schrödinger equation for a magnetoexciton in a slowly changing external field.

The effective Schrödinger equation for a magnetoexciton moving as a whole in an external potential will be derived in Sec. 2. In Sec. 3 this approximation will be applied to a magnetoexciton in isolated and coupled quantum dots. In Sec. 4, we will calculate the optical absorption by magnetoexcitons in an single quantum well in the coherent potential approximation. Similar methods will be used in calculating magnetoabsorption in coupled quantum wells (Sec. 5). In Sec. 6 we will consider spectral properties of a magnetoexciton interacting with a random field due to local fluctuations in the concentrations of a III–V solid solution components in a single quantum well.

2. MOTION OF A MAGNETOEXCITON IN AN EXTERNAL MAGNETIC FIELD (EFFECTIVE MAGNETIC MASS APPROXIMATION)

The Schrödinger equation describing motion of a magnetoexciton in an external potential $V(\mathbf{r}_e,\mathbf{r}_h)=V_e(\mathbf{r}_e)+V_h(\mathbf{r}_H)$ has the form

$$\left[\frac{1}{2m_e}\left(\frac{\hbar}{i}\nabla_e+\frac{e}{c}\mathbf{A}_e\right)^2+\frac{1}{2m_h}\left(\frac{\hbar}{i}\nabla_h-\frac{e}{c}\mathbf{A}_h\right)^2+V_e(\mathbf{r}_e)+V_h(\mathbf{r}_h)-\frac{e^2}{\epsilon\sqrt{D^2+(\mathbf{r}_e-\mathbf{r}_h)^2}}\right]\times\Psi(\mathbf{r}_e,\mathbf{r}_h)=E\Psi(\mathbf{r}_e,\mathbf{r}_h), \tag{1}$$

where $\mathbf{r}_{e,h}$ are the two-dimensional vectors of the electron (e) and hole (h) positions (the electron and hole are located in a single quantum well or two spatially separated wells, corresponding to direct and indirect excitons respectively); $\mathbf{A}=\mathbf{H}\times\mathbf{r}/2$ is the vector potential of the magnetic field in the symmetric gauge; D is the distance between the quantum wells containing electrons and holes; and we have written $\epsilon=(\epsilon_1+\epsilon_2)/2$, where $\epsilon_{1,2}$ are the permittivities of the materials around the electron and hole quantum wells. We will look for the wave function Ψ expanding in a series eigenfunctions $\Phi_{nmp}(\mathbf{R},\mathbf{r})$ describing a magnetoexciton in an ordered system:[17–19]

$$\Psi(\mathbf{R},\mathbf{r})=\sum_{nmp}a_{nmp}\Phi_{nmp}(\mathbf{R},\mathbf{r}), \tag{2}$$

where $\mathbf{R}=(m_e\mathbf{r}_e+m_h\mathbf{r}_h)/M$, $M=m_e+m_h$, and $\mathbf{r}=\mathbf{r}_e-\mathbf{r}_h$; $\mathbf{P}$ is the exciton quasimomentum in the magnetic field, which is an integral of motion in homogeneous systems, and n and m are magnetoexciton quantum numbers. In the case of a strong magnetic field, we neglect in Eq. (2) transitions between different Landau levels in the magnetoexciton due to scattering by the slowly changing potential $V(\mathbf{r}_e,\mathbf{r}_h)$ and the nondiagonal matrix elements of the Coulomb interaction, multiply both sides of Eq. (1) by Φ_p^* (Φ_p is the magnetoexciton wave function on the lowest Landau level), and integrate it with respect to $\mathbf{R}$ and $\mathbf{r}$; thus we obtain

$$a_pE(P)+\sum_{p'}V_{pp'}a_{p'}=Ea_p, \tag{3}$$

where $E(P)$ is the magnetoexciton spectrum on the lowest Landau level.[17–19] The magnetoexciton dispersion $E(P)$ at sufficiently small magnetic momenta is quadratic, as follows from the effective Hamiltonian invariance with respect to rotations and the analytical property of $E(P)$ at $P=0$.[20] Moreover, the point $P=0$ is the minimum of $E(P)$. For $Pl/\hbar\ll 1$ we have

$$E(P)=\frac{1}{2}\hbar\omega_c-\mathscr{E}(0,D)+\frac{P^2}{2M_{\rm exc}}, \tag{4}$$

where $l=\sqrt{\hbar c/eH}$ is the magnetic length, $\omega_c=eH/\mu c$ is the cyclotron frequency, $\mu=m_em_h/(m_e+m_h)$ is the exciton reduced mass in the quantum well plane, $\mathscr{E}(P,D)$ is the magnetoexciton dispersion on the lowest Landau level, and $M_{\rm exc}$ is the magnetoexciton effective mass, which is a function of magnetic field H and distance D between the quantum wells: $M_{\rm exc}\sim 2^{3/2}\epsilon\hbar^2/e^2l\sqrt{\pi}$ for $D\ll l$ and $M_{\rm exc}\sim D^3\epsilon\hbar^2/e^2l^4$ for $D\gg l$.[17–19]

The matrix element of the potential $V(\mathbf{r}_{e,h})$ connecting the states $\langle n=m=0,\mathbf{P}|$ and $\langle n=m=0,\mathbf{P}'|$ has the form

$$\langle\mathbf{P}'|V_{e,h}(\mathbf{r})|\mathbf{P}\rangle=\frac{1}{S}\exp\left(-\frac{l^2}{4\hbar^2}(\mathbf{P}'-\mathbf{P})^2\right)V_{e,h}(\mathbf{P}'-\mathbf{P})\times\exp\left(\pm\frac{il^2}{2\hbar^2H}\mathbf{H}\cdot\mathbf{P}\times\mathbf{P}'\right)=\frac{1}{S}\widetilde{V}_{e,h}(\mathbf{P}'-\mathbf{P})\exp\left(\pm\frac{il^2}{2\hbar^2H}\mathbf{H}\cdot\mathbf{P}\times\mathbf{P}'\right), \tag{5}$$

where S is the area of the quantum well. Let us introduce the operator $E(-i\hbar\nabla)$ such that

$$E(-i\hbar\nabla)\exp\left(\frac{i}{\hbar}\mathbf{P}\cdot\mathbf{R}\right)=E(P)\exp\left(\frac{i}{\hbar}\mathbf{P}\cdot\mathbf{R}\right). \tag{6}$$

We multiply Eq. (3) by $\exp[(i/\hbar)\mathbf{P}\cdot\mathbf{R}]$ and sum with respect to $\mathbf{P}$. Equation (3) transforms to

$$E(-i\hbar\nabla)F(\mathbf{R})+\frac{1}{S}\sum_{P_1,P_2}a_{p_2}\exp\left(\frac{i}{\hbar}\mathbf{P}_1\cdot\mathbf{R}\right)\left[\widetilde{V}_e(\mathbf{P}_1-\mathbf{P}_2)\times\exp\left(\frac{il^2}{2\hbar^2H}\mathbf{H}\cdot\mathbf{P}_2\times\mathbf{P}_1\right)+\widetilde{V}_h(\mathbf{P}_1-\mathbf{P}_2)\times\exp\left(-\frac{il^2}{2\hbar^2H}\mathbf{H}\cdot\mathbf{P}_2\times\mathbf{P}_1\right)\right]=EF(\mathbf{R}), \tag{7}$$

where

$$F(\mathbf{R})=\sum_p a_p\exp\left(\frac{i}{\hbar}\mathbf{P}\cdot\mathbf{R}\right). \tag{8}$$

Using the ansatz $\mathbf{P}\to-i\hbar\nabla$, it is convenient to rewrite the second term on the left of Eq. (7) as

$$\sum_p a_p\left[\exp\left(\frac{l^2}{2\hbar H}\mathbf{H}\cdot\mathbf{P}\times\nabla\right)\widetilde{V}_e(\mathbf{R})\exp\left(\frac{i}{\hbar}\mathbf{P}\cdot\mathbf{R}\right)+\exp\left(-\frac{l^2}{2\hbar H}\mathbf{H}\cdot\mathbf{P}\times\nabla\right)\widetilde{V}_h(\mathbf{R})\exp\left(\frac{i}{\hbar}\mathbf{P}\cdot\mathbf{R}\right)\right]. \tag{9}$$

The summation over $\mathbf{P}$ transforms Eq. (9) to

$$\lim_{R'\to R}\left[\exp\left(-\frac{il^2}{2H}\mathbf{H}\cdot\nabla_{R'}\times\nabla_R\right)\widetilde{V}_e(\mathbf{R})+\exp\left(\frac{il^2}{2H}\mathbf{H}\cdot\nabla_{R'}\times\nabla_R\right)\widetilde{V}_h(\mathbf{R})\right]F(\mathbf{R}'). \tag{10}$$

For $P\ll\hbar/l$ the operator (6) can be expressed, with due account of Eq. (4), as a power series in $-i\hbar\nabla$:

$$E(-i\hbar\nabla)=\frac{1}{2}\hbar\omega_c+\mathscr{E}(0,D)-\frac{\hbar^2}{2M_{\rm exc}}\Delta. \tag{11}$$

By substituting Eqs. (10) and (11) in Eq. (7), we obtain the effective Schrödinger equation for the magnetoexciton:

$$-\frac{\hbar^2}{2M_{\rm exc}}\Delta F(\mathbf{R})+\lim_{R'\to R}\left[\exp\left(-\frac{il^2}{2H}\mathbf{H}\cdot\nabla_{R'}\times\nabla_R\right)\tilde V_e(\mathbf{R})+\exp\left(\frac{il^2}{2H}\mathbf{H}\cdot\nabla_{R'}\times\nabla_R\right)\tilde V_h(\mathbf{R})\right]F(\mathbf{R}')=\mathscr{E}F(\mathbf{R}),\quad(12)$$

where $\mathscr{E}=E-\hbar\omega_c/2-\mathscr{E}(0,D)$ and $F(\mathbf{R})$ is the magnetoexciton envelope function in the effective potential. The application region of Eq. (12) for the envelope function F is defined, as in the case of Eq. (3), by the inequalities

$$\hbar\omega_c\gg E_{\rm exc},\quad \hbar\omega_c\gg\sqrt{\langle V_{e,h}^2\rangle_{\rm av}},$$

where $E_{\rm exc}$ is the magnetoexciton binding energy in an ideal system as a function of magnetic field and thickness D of the barrier between electron and hole quantum wells: $E_{\rm exc}\sim e^2/l\epsilon\sqrt{\pi/2}$ for $D\ll l$ and $E_{\rm exc}\sim e^2/\epsilon D$ for $D\gg l$;[17–19] here $\langle\ldots\rangle_{\rm av}$ denotes averaging over random field fluctuations.

The characteristic wavelength of the envelope and effective potential $\tilde V_{e,h}$ is the correlation length L of potential $V(\mathbf{r}_e,\mathbf{r}_h)$. In this paper we consider the case $L\ll r_{\rm exc}$, where $r_{\rm exc}$ is the exciton mean size, i.e., the case of a smooth surface potential, which is realized, as was shown using a scanning tunneling microscope, on interfaces in AlGaAs—GaAs structures discussed here. A more rigorous condition of the smoothness of random potential can be expressed as[22]

$$r_{\rm exc}\sqrt{\langle\nabla V^2\rangle_{\rm av}}\ll E_{\rm exc}.\quad(13)$$

Given that the correlation length of the random potential is larger than the magnetoexciton mean size l, exponential functions in Eq. (10) can be expanded in series. In this case, expression (10) takes the form

$$(\tilde V_e(\mathbf{R})+\tilde V_h(\mathbf{R}))F(\mathbf{R})+\frac{il^2}{2H}\Big[\mathbf{H},\nabla_R(\tilde V_e(\mathbf{R})-\tilde V_h(\mathbf{R}))\Big]\times\nabla_R F(\mathbf{R})-\frac{l^4}{8}\left(\frac{\partial^2F(\mathbf{R})}{\partial X^2}\frac{\partial^2(\tilde V_e(\mathbf{R})+\tilde V_h(\mathbf{R}))}{\partial Y^2}+\frac{\partial^2F(\mathbf{R})}{\partial Y^2}\frac{\partial^2(\tilde V_e(\mathbf{R})+\tilde V_h(\mathbf{R}))}{\partial X^2}-2\frac{\partial^2F(\mathbf{R})}{\partial X\partial Y}\frac{\partial^2(\tilde V_e(\mathbf{R})+\tilde V_h(\mathbf{R}))}{\partial X\partial Y}\right)+\cdots\quad(14)$$

In the approximation of the exciton effective magnetic mass, the second term in Eq. (14) acts as a dipole interaction between a magnetoexciton and external electric field, which is a random function of coordinates in this case. Specifically, the exciton dipole moment on the lowest Landau level is determined by the expression

$$\langle\mathbf{d}\rangle=e\langle\mathbf{r}\rangle=e\sum_p|a_p|^2\frac{l^2}{\hbar H}\mathbf{H}\times\mathbf{P}=\frac{c}{SH^2}\sum_{p_1,p_2}a_{p_1}^*a_{p_2}\times\int\exp\left(-\frac{i}{\hbar}\mathbf{P}_1\cdot\mathbf{R}\right)\frac{\hbar}{i}(\mathbf{H}\times\boldsymbol{\nabla})\exp\left(\frac{i}{\hbar}\mathbf{P}_2\cdot\mathbf{R}\right)d\mathbf{R}=\int F^*(\mathbf{R})\frac{c\hbar}{iH^2}\mathbf{H}\times\boldsymbol{\nabla}F(\mathbf{R})d\mathbf{R},\quad(15)$$

i.e., $(c\hbar/iH^2)\mathbf{H}\times\nabla$ acts as an effective dipole moment in the space of envelope functions $F(\mathbf{R})$. Similarly, one can easily show that the diagonal matrix element of the second term in Eq. (14) for the envelope function $F(\mathbf{R})$ under the condition that $V(\mathbf{r}_r,\mathbf{r}_h)=e(\mathbf{r}_e-\mathbf{r}_h)\mathbf{E}$ is identical to the quantum-mechanical average of the dipole interaction between an exciton and quasi-homogeneous electric field $\mathbf{E}$ for the initial wave functions (2) [on the scale of the exciton size; see Eq. (13)]. The role of the second term is clearly illustrated by the example of an exciton in a quantum dot discussed in the next section. In a random field, the mean value of the exciton dipole moment $\mathbf{d}$ at zero magnetic quasimomentum is zero [for an exciton at $\mathbf{P}=0$, where $E(\mathbf{P})$ has a minimum].

By retaining terms of zero order in l/L, we obtain the Schrödinger equation in the form

$$\left(-\frac{\hbar^2}{2M_{\rm exc}}\Delta+V_{\rm eff}\,(\mathbf{R})\right)F(\mathbf{R})=\mathscr{E}F(\mathbf{R}),\quad(16)$$

where

$$V_{\rm eff}(\mathbf{R})=\frac{1}{\pi l^2}\int\exp\left(-\frac{(\mathbf{R}-\mathbf{r})^2}{l^2}\right)[V_e(\mathbf{r})+V_h(\mathbf{r})]d\mathbf{r}.\quad(17)$$

The effective Schrödinger equation (16) for the magnetoexciton is invariant under the substitution $t\to-t$, $F\to F^*$, unlike Eqs. (1) and (12) (for $m_h\neq m_e$). In fact, Eq. (16) treats the magnetoexciton as an electrically neutral composite particle. Since the particle is neutral, it does not directly interact with the magnetic field, which breaks the time-reversal symmetry in Eqs. (1) and (12). The interaction with the magnetic field can be seen indirectly through the renormalization of the exciton effective mass and modification of the correlation function of the random field. The inclusion of terms of the first and higher orders in l/L in strong magnetic fields, which describe the effect of scattering of the magnetoexciton as a whole on its internal degrees of freedom, breaks the symmetry with respect to the substitution $t\to-t$, $F\to F^*$ at $m_h\neq m_e$ (this circumstance must be taken into account when we study the weak localization of magnetoexcitons, cf. Ref. 16). In the next section, we will show that inclusion of the terms linear and quadratic in l/L leads to a renormalization of the exciton effective mass and generates in the Hamiltonian terms responsible for exciton polarization in an external field.

Thus, the Schrödinger equation (16) applies to magnetoexcitons in weak external fields $V_e(\mathbf{r})$ and $V_h(\mathbf{r})$ whose correlation lengths are considerably larger than the exciton size.

3. SPECTRA OF DIRECT AND INDIRECT MAGNETOEXCITONS IN QUANTUM DOTS

In order to clarify the nature of the approximation (16) used in this paper, let us consider a magnetoexciton containing an electron and a hole in one quantum dot and two spatially separated dots. This problem, of course, is interesting, irrespective of the issues discussed here, in connection with experimental studies of excitons in quantum dots in a magnetic field.[23]

Assume that the electron and hole are confined in quantum dots by parabolic potentials $V_{e,h}(r)=\alpha_{e,h}r^2$. In calculating the exciton spectrum we use Eq. (12). It holds under the condition that the interaction between the exciton and quantum dot potential does not induce transitions between exciton levels due to quantization of internal motion.

Using Eq. (12) and taking into account confining potentials, we obtain

$$\left(-\frac{\hbar^2}{2\tilde{M}}\Delta+(\alpha_e+\alpha_h)R^2+\frac{c(\alpha_h-\alpha_e)}{eH^2}\mathbf{H}\cdot\hat{\mathbf{L}}\right)F(\mathbf{R})=E'F(\mathbf{R}), \tag{18}$$

where $\hat{\mathbf{L}}=-i\hbar\mathbf{R}\times\nabla$ is the exciton angular momentum operator,

$$\frac{1}{\tilde{M}}=\frac{1}{M_{\rm exc}(D/l)}+\frac{l^4}{2\hbar^2}(\alpha_e+\alpha_h), \tag{19}$$

where $E'=\mathscr{E}-l^2(\alpha_e+\alpha_h)/2$. Note that the confining potentials reduce the exciton effective mass. The parameter $\tilde{M}$ monotonically increases with H. Note that Eqs. (18) and (19) have been derived from the ''exact'' equation (12) [but not from the approximate equation (16)] without expanding in powers of the ratio between the magnetic length and localization radius $L\sim a_0$ (see below). If we used Eq. (16), we would not obtain the third term on the left of Eq. (18), which describes the exciton polarization by external field, and the exciton effective mass would not be renormalized.

Thus, we obtain the following expression for $F(\mathbf{R})$:

$$F_{nm}(\mathbf{R})=\sqrt{\frac{n!}{2\pi(n+|m|!)}}\frac{e^{im\phi}}{a_0}\left(\frac{R}{\sqrt{2}a_0}\right)^{|m|}\times L_n^{|m|}\left(\frac{R^2}{2a_0^2}\right)\exp\left(-\frac{R^2}{4a_0^2}\right), \tag{20}$$

where

$$a_0=\left(\frac{\hbar^2}{2\tilde{M}(\alpha_e+\alpha_h)}\right)^{1/4} \tag{21}$$

is the exciton localization radius in quantum dots, and L_n^m are Laguerre polynomials. The exciton localization region becomes smaller with increasing D/l and magnetic field H.

The spectrum (18) is fully discrete:

$$E'_{nm}=ml^2(\alpha_h-\alpha_e)+4\hbar\sqrt{\frac{\alpha_e+\alpha_h}{2\tilde{M}}}\left(n+\frac{|m|+1}{2}\right). \tag{22}$$

Increasing H decreases the distance between the levels defined by Eq. (22), i.e., leads to finer structure for magnetoexciton absorption due to the internal motion of the exciton on the lowest Landau level. This narrowing of the spectrum is caused by the increase in the exciton effective mass (19) and the decrease in the magnetic length l with the increasing magnetic field (a similar effect is the narrowing of the exciton band in quantum wells in stronger fields[19]). Such properties of the exciton absorption were detected in absorption spectra of natural quantum dots.[23]

Up to this point we have discussed the exciton spectrum around the lowest Landau level. Similarly, the entire exciton spectrum in coupled quantum dots in a strong magnetic field consists of fine features around higher Landau levels. We merely note that in spectra around higher levels one should also take into account features corresponding to excitons formed at ''roton'' minima.[17–19] At $D=0$, our results are in good agreement with numerical calculations of magnetoexciton spectra in quantum dots.[24]

If we have $\alpha_e=\alpha_h$, the levels defined by Eq. (22) are degenerate in the quantum number $N=2n+|m|$. Each level except (0,0) is $N+1$-fold degenerate.

The approximation (18) is valid under the condition

$$\alpha_e+\alpha_h\ll\omega_c^2\mu. \tag{23}$$

As follows from this inequality, the applicability of our results is wider in stronger magnetic fields.

The probability of generating an exciton is known to be determined[25,26] by the parameter

$$\int|\Psi(\mathbf{R},\mathbf{r})|^2\delta(\mathbf{r})d\mathbf{R}d\mathbf{r}, \tag{24}$$

i.e., the probability of detecting an electron and a hole at the same point. Here we consider either an isolated quantum dot or generation of an exciton in coupled quantum dots. Using Eq. (2), we transform Eq. (24) to

$$\frac{1}{2\pi l^2}\sum_p|a_p|^2\exp\left(-\frac{P^2l^2}{2\hbar^2}\right). \tag{25}$$

The factor a_p is a Fourier transform of function (20), and at $n=m=0$ it is determined by the equation

$$a_p=\sqrt{\frac{2\pi}{S}}2a_0\exp\left(-\frac{P^2a_0^2}{\hbar^2}\right). \tag{26}$$

By substituting Eq. (26) in (25), we find that the magnetoexciton absorption in a quantum dot is proportional to

$$\frac{1}{2\pi^2}\left(\frac{a_0}{l}\right)^2\frac{1}{4a_0^2+l^2}, \tag{27}$$

which increases monotonically with H in the region of strong magnetic fields because the wave functions are compressed as the magnetic field intensity increases. The rate of this growth is lower at higher $\alpha_e+\alpha_h$.

4. MAGNETOEXCITON LIGHT ABSORPTION IN A SINGLE QUANTUM WELL

Fluctuations in the quantum well width generated in the process of its fabrication result in a random potential in the well. Provided that the amplitude of such fluctuations is smaller than average quantum well width and they are sufficiently smooth, the interaction between an exciton and this random field is described by the potential[12,27,28]

$$V(\mathbf{r}_e,\mathbf{r}_h)=\tilde{\alpha}_e[\xi_1(\mathbf{r}_e)-\xi_2(\mathbf{r}_e)]+\tilde{\alpha}_h[\xi_1(\mathbf{r}_h)-\xi_2(\mathbf{r}_h)], \quad (28)$$

where $\tilde{\alpha}_{e,h}=\partial E^{(0)}_{e,h}/\partial d$, d is the quantum well width, $E^{(0)}_{e,h}$ are the lowest levels of the electron and hole in the conduction and valence bands, and $\xi_{1,2}(\mathbf{r})$ are fluctuations in the coordinates of the upper and lower interface. Let us assume that fluctuations on different interfaces are statistically independent and characterized by a Gaussian correlation function:

$$\langle\xi_i(\mathbf{r}_1)\xi_j(\mathbf{r}_2)\rangle=g_i\delta_{ij}\delta(\mathbf{r}_2-\mathbf{r}_1), \quad (29)$$

where g_i is proportional to the squared amplitude of the ith interface fluctuation.[12,27,28]

The exciton absorption factor can be expressed as $\alpha=\alpha_0 A(E)$,[15,25] where α_0 is a factor weakly depending on the frequency $\omega=E+E_g+E_{\rm exc}$ (here E_g is the band gap width and $\hbar=1$), and $A(E)$ is given by the equation

$$A(E)=-\frac{1}{\pi}\mathrm{Im}\,G_1(0,E), \quad (30)$$

where $G_1(0,E)$ is the Fourier transform of the retarded Green's function at $k=0$. Using the replica technique, let us express the Green's function in the form of a functional integral over boson field:[16,29]

$$G_{1,2}(E,R_1,R_2)=\lim_{N\to 0}\int D\phi^p_a\phi^{1,2}_1(R_1)\phi^{1,2}_1(R_2)e^L,$$

$$L=\frac{is_p}{2}\int\phi^p_a(\mathbf{r})\left[E_p-\frac{\nabla^2_R}{2M_{\rm exc}}-V_{\rm eff}(\mathbf{R})\right]\phi^p_a(\mathbf{r})d\mathbf{r}. \quad (31)$$

Here $\phi^p_a(\mathbf{r})$ are real fields and summation over repeated indices is implied; the replica indices are $a=1,\ldots,N$, $p=1,2$, $s_1=-s_2=1$, $E_p=E+is_p\eta$, $\eta\to 0$. By performing Gaussian averaging in Eq. (31) over $\xi_i(\mathbf{r})$ we obtain

$$L=\frac{is_p}{2}\int\phi^p_a(\mathbf{R})\left(E_p-\frac{\nabla^2_R}{2M}\right)\phi^p_a(\mathbf{R})d^2\mathbf{r}-\frac{1}{8}\int d\mathbf{R}_1d\mathbf{R}_2\phi^p_a\times(\mathbf{R}_1)\phi^p_a(\mathbf{R}_1)s_pB(\mathbf{R}_1,\mathbf{R}_2)\phi^q_b(\mathbf{R}_2)\phi^q_b(\mathbf{R}_2)s_q, \quad (32)$$

where $B(\mathbf{R}_1-\mathbf{R}_2)=\langle V_{\rm eff}(\mathbf{R}_1)V_{\rm eff}(\mathbf{R}_2)\rangle_{\rm av}$, and the effective potential $V_{\rm eff}$ is given by Eq. (17). As a result the correlator of the random field, which determines properties of excitons, unlike that of the fields acting on the electron and hole separately, corresponds to a colored rather than white Gaussian noise:

$$B(\mathbf{R}_1-\mathbf{R}_2)=\frac{g_1+g_2}{2\pi l^2}(\tilde{\alpha}_e+\tilde{\alpha}_h)^2\exp\left(-\frac{(\mathbf{R}_1-\mathbf{R}_2)^2}{2l^2}\right). \quad (33)$$

Note that we have $l\to 0$ in the limit of strong magnetic fields, whence $B(\mathbf{R}-\mathbf{R}_2)\sim(g_1+g_2)(\tilde{\alpha}_e+\tilde{\alpha}_h)^2\delta(\mathbf{R}_1+\mathbf{R}_2)$, i.e., the random field acting on an exciton is again characterized by the white noise spectrum. The time-reversal symmetry in Schrödinger equation (16) and elimination of the long-range property of the random field rule out all causes that could lead to a difference between transport coefficients in the cooperon and diffuson.[30]

Owing to the translation symmetry and isotropy, the exciton correlation function depends only on $|\mathbf{R}|=|\mathbf{R}_1-\mathbf{R}_2|$.

To derive an effective potential describing long-wave fluctuations it is convenient to decouple the interaction part of the Lagrangian by introducing a bilocal field $Q(\mathbf{R}_1,\mathbf{R}_2)$:[16,29]

$$\exp(L_{\rm int})=\int D\hat{Q}\exp\left[\frac{i}{2}\int d\mathbf{R}_1d\mathbf{R}_2\sqrt{s_ps_q}Q^{pq}_{ab}\phi^p_a(\mathbf{R}_1)\times\phi^q_b(\mathbf{R}_2)-\frac{1}{2}\int d\mathbf{R}_1d\mathbf{R}_2\mathrm{Tr}\,\hat{Q}(\mathbf{R}_1,\mathbf{R}_2)\times B^{-1}(|\mathbf{R}_1-\mathbf{R}_2|)\hat{Q}(\mathbf{R}_1,\mathbf{R}_2)\right],$$

$$L_{\rm int}=-\frac{1}{8}\int d^2\mathbf{R}_1d\mathbf{R}_2\phi^p_a(\mathbf{R}_1)\phi^p_a(\mathbf{R}_1)s_pB(\mathbf{R}_1,\mathbf{R}_2)\times\phi^q_b(\mathbf{R}_2)\phi^q_b(\mathbf{R}_2)s_q. \quad (34)$$

Let us introduce a generating functional

$$Z[J]=\int D\hat{Q}D\phi^p_a\exp\left(L[\hat{Q},\phi]+\int J^{pq}_{ab}(\mathbf{R}_1,\mathbf{R}_2)\phi^p_a(\mathbf{R}_1)\phi^q_b(\mathbf{R}_2)d\mathbf{R}_1d\mathbf{R}_2\right), \quad (35)$$

where $J^{pq}_{ab}(\mathbf{R}_1,\mathbf{R}_2)=J^{qp}_{ba}(\mathbf{R}_2,\mathbf{R}_1)$. Then the expression for the Green's function is

$$\left.\frac{\delta Z[J]}{\delta J^{pp}_{11}}\right|_{J=0}=is_pG_p(\mathbf{R}_1,\mathbf{R}_2), \quad (36)$$

or[16]

$$\langle G_p(\mathbf{R}_1-\mathbf{R}_2)\rangle_{\rm av}=-2B^{-1}(R_1-R_2)\langle Q^{pp}_{11}(R_1,R_2)\rangle, \quad (37)$$

where $\langle\ldots\rangle$ denotes the functional integral over field Q.

The Lagrangian in Eq. (35) is quadratic in ϕ^p_a, so direct integration yields

$$L[\hat{Q}]=-\frac{1}{2}\mathrm{Tr}\,\ln\left[\hat{s}\left(E-\frac{\nabla^2_R}{2M_{\rm exc}}\right)\delta(\mathbf{R}_1-\mathbf{R}_2)+\tilde{Q}\right]-\frac{1}{2}\int d\mathbf{R}_1d\mathbf{R}_2\mathrm{Tr}\tilde{Q}(\mathbf{R}_1,\mathbf{R}_2)B^{-1}(\mathbf{R}_1,\mathbf{R}_2)\tilde{Q}(\mathbf{R}_1,\mathbf{R}_2), \quad (38)$$

where $\hat{s}^{pq}_{ab}=s_p\delta_{pq}\delta_{ab}$, $\tilde{Q}^{pq}_{ab}=\sqrt{s_ps_q}\,\hat{Q}^{pq}_{ab}$.

Let us calculate the stationary path of the Lagrangian. It should satisfy the equation

$$\frac{\delta L}{\delta Q}=0. \quad (39)$$

We seek a solution of this equation in the form

$$\tilde{Q}_{ab}^{pq}=\delta_{pq}\delta_{ab}Q_a^p. \tag{40}$$

The equation for Q_a^p in the saddle approximation is

$$G_p^0(\mathbf{R}_1,\mathbf{R}_2)=-2B^{-1}(|\mathbf{R}_1-\mathbf{R}_2|)Q_a^p(\mathbf{R}_1,\mathbf{R}_2), \tag{41}$$

where

$$G_p^0(\mathbf{R}_1,\mathbf{R}_2)=\left\langle \mathbf{R}_1\left|\frac{1}{E-\nabla_R^2/2M_{\rm exc}+Q_a^p}\right|\mathbf{R}_2\right\rangle. \tag{42}$$

Equations (41) and (42) determine the Green's function in the saddle approximation. The saddle approximation is known to be equivalent to the coherent potential approximation. It allows one to determine the scattering potential $-Q$ using a self-consistent procedure described by Eqs. (41) and (42). The coherent potential approximation not only yields qualitatively good results, but also quantitative calculations of spectra.[31]

Owing to the translation symmetry and isotropy, the solution of the integral equation with respect to Q_p depends only on $|\mathbf{R}_1-\mathbf{R}_2|$. By taking the Fourier transform of Eq. (41), we obtain

$$Q_p(\mathbf{k})=-\frac{1}{2}\int G_p^0(\mathbf{q})B(|\mathbf{k}-\mathbf{q}|)\frac{d\mathbf{q}}{(2\pi)^2}. \tag{43}$$

In the limit $g_i\to 0$ the parameter $\mathrm{Im}\,Q_p$ is small; hence, as in Refs. 16 and 29, we can use the ansatz

$$\mathrm{Im}\,G_p^0(q)\to-\pi\delta\left(E-\frac{q^2}{2M_{\rm exc}}\right). \tag{44}$$

By substituting the expression (44) in Eq. (43) and expressing the imaginary part of the Green's function in the coherent potential approximation, we obtain

$$\mathrm{Im}\,G_p^0(k)=-\frac{\mathrm{Im}\,Q_p(k)}{(E-k^2/2M_{\rm exc})^2+[\mathrm{Im}\,Q_p(k)]^2}, \tag{45}$$

where

$$\mathrm{Im}\,Q_p(k)=\frac{\pi}{2}\int\delta\left(E-\frac{q^2}{2M_{\rm exc}}\right)B(|\mathbf{k}-\mathbf{q}|)\frac{d\mathbf{q}}{(2\pi)^2},$$

or

$$\mathrm{Im}\,Q_p(k)=\frac{1}{2}B_0M_{\rm exc}\pi l^2\exp\left(-\frac{1}{2}k^2l^2-M_{\rm exc}El^2\right)\times I_0(kl^2\sqrt{2M_{\rm exc}E}), \tag{46}$$

where B_0 is the pre-exponential factor (33).

After turning to Eq. (30), we obtain the expression for the exciton absorption in the coherent potential approximation:

$$A(E)=\frac{1}{2\pi}\frac{M_{\rm exc}B_0\pi l^2\exp(-M_{\rm exc}El^2)}{E^2+(M_{\rm exc}B_0\pi l^2/2)^2\exp(-2M_{\rm exc}El^2)}. \tag{47}$$

The absorption peaks in the range of low energies $M_{\rm exc}El^2\ll 1$. Under the condition $d^2\ll a_{e,h}l$, where $a_{e,h}$ are effective Bohr radii of the electron and hole in the quantum well plane, the coefficient in the size quantization energy expansion in terms of the quantum well width fluctuations, obviously, has the form $\tilde{\alpha}_{e,h}=-\pi^2/m_{e,h}d^3$, where d is the mean quantum well width.[27] Therefore the parameter B in Eq. (47) [the amplitude of the random field correlator; see Eq. (33)] is proportional to $1/d^6$. As a result, the absorption coefficient at $E=0$ is

$$A(0)=\frac{4}{\pi^5}\frac{d^6\mu_z^2}{(g_1+g_2)M_{\rm exc}}, \tag{48}$$

where $\mu_z=m_{ze}m_{zh}/(m_{ze}+m_{zh})$. The absorption coefficient falls off with increasing H as $1/\sqrt{H}$ since the exciton effective mass $M_{\rm exc}$ increases with H as $\sqrt{H}$. The FWHM of the absorption line increases with the random field amplitude g.

Let us calculate $A(E)$ in the region of large negative energies E. In this band, the density of states is nonzero only due to relatively rarely occurring configurations of the random field. By virtue of the macroscopic homogeneity condition, regions of large negative potential fluctuations should be separated by distances considerably larger than their dimensions, and, moreover, $\int V_{\rm eff}d\mathbf{r}=0$. Thus, typical realizations (optimal fluctuations) responsible for the spectrum in the range of large negative energies E should have the shapes of relatively deep wells separated by regions with typical values of the potential.

$$\langle V_{\rm eff}\rangle_{\rm av}\pm\sqrt{\langle V_{\rm eff}^2\rangle_{\rm av}-\langle V_{\rm eff}\rangle_{\rm av}^2}=\pm\sqrt{\langle V_{\rm eff}^2\rangle_{\rm av}}.$$

The density of states in the limit $E\to-\infty$ can be calculated by the method of optimal fluctuations[32] or the saddle point method[33] applied to the functional integral (32). In the saddle point method the main contribution depending on the parameter $1/M_{\rm exc}El^2$ to the functional (32), which determines the Green's function, is due to instantons. The instanton contribution to the density of states for large E is asymptotically exact, and in one-dimensional systems it is identical to the result obtained by the optimal fluctuation method. This contribution is determined by the behavior of function $B(\mathbf{R})$ in the region $R\ll l$. Thus, retaining only the lowest-order component in the expansion of $B(\mathbf{R})$ in powers of $\mathbf{R}$, we obtain

$$B(R)\approx B_0\left(1-\frac{R^2}{2l^2}\right). \tag{49}$$

Using the expression for the Green's function in the instanton approximation,[33] we derive the exciton absorption coefficient:

$$A(E)=\frac{E^2l}{(\tilde{\alpha}_e+\tilde{\alpha}_h)^3(g_1+g_2)^{3/2}}\exp\left(-\frac{2\pi l^2E^2}{(\tilde{\alpha}_e+\tilde{\alpha}_h)^2(g_1+g_2)}\right). \tag{50}$$

5. MAGNETOEXCITON LIGHT ABSORPTION IN COUPLED QUANTUM WELLS

The interaction between an exciton whose electron and hole are in spatially separated quantum wells and a random field due to fluctuations in widths of electron and hole quantum wells has the form[12]

$$V(\mathbf{r}_e,\mathbf{r}_h)=\tilde{\alpha}_e[\xi_1(\mathbf{r}_e)-\xi_2(\mathbf{r}_e)]+\tilde{\alpha}_h[\xi_3(\mathbf{r}_h)-\xi_4(\mathbf{r}_h)], \tag{51}$$

where $\tilde{\alpha}_{e,h}=\partial E_{e,h}^{(0)}/\partial d_{e,h}$, $d_{e,h}$ are average widths of the electron and hole quantum wells, ξ_1 and ξ_2 (ξ_3 and ξ_4) are fluctuations in the widths of the electron (hole) wells on the upper and lower interfaces, respectively [the applicability condition for Eq. (51) is similar to that of Eq. (28)].

We still assume that fluctuations of different interfaces are statistically independent, whereas fluctuations of a specific interface are described by a Gaussian correlation function (29). This is possible if the distance D between the electron and hole quantum wells is larger than the amplitude of fluctuations on the nearest interfaces (the opposite case occurs in double quantum wells).

The correlation function of the effective potential has the form

$$B(\mathbf{R}_1-\mathbf{R}_2)=\frac{\tilde{\alpha}_e^2(g_1+g_2)+\tilde{\alpha}_h^2(g_3+g_4)}{2\pi l^2}\exp\left(-\frac{(\mathbf{R}_1-\mathbf{R}_2)^2}{2l^2}\right). \quad (52)$$

The absorption coefficient at $E=0$ is

$$A(0)=\frac{4}{\pi^5 M(\mathscr{D})}\left(\frac{g_1+g_2}{m_{ze}^2 d_e^6}+\frac{g_3+g_4}{m_{zh}^2 d_h^6}\right)^{-1}, \quad (53)$$

where

$$M(\mathscr{D})=\frac{2^{3/2}\epsilon\hbar^2}{e^2 l\sqrt{\pi}}\times\left[(1+\mathscr{D}^2)\exp\left(\frac{\mathscr{D}^2}{2}\right)\mathrm{erfc}\left(\frac{\mathscr{D}}{\sqrt{2}}\right)-\mathscr{D}\sqrt{\frac{2}{\pi}}\right]^{-1} \quad (54)$$

is the effective mass of the indirect exciton in the state $n=m=0$ in a coupled quantum well[18,19] and $\mathscr{D}=D/l$. For the reasons discussed in Sec. 4, $A(0)$ drops with the magnetic field as $1/\sqrt{H}$ for $\mathscr{D}\ll 1$, and $1/H^2$ for $\mathscr{D}\gg 1$ in agreement with the experimental data.[3] The peak intensity drops with $\mathscr{D}$.

In the instanton approximation for $M(\mathscr{D})l^2|E|\gg 1$, the magnetoexciton absorption has the form

$$A(E)=\frac{E^2 l}{[\tilde{\alpha}_e^2(g_1+g_2)+\tilde{\alpha}_h^2(g_3+g_4)]^{3/2}}\times\exp\left(-\frac{2\pi l^2 E^2}{[\tilde{\alpha}_e^2(g_1+g_2)+\tilde{\alpha}_h^2(g_3+g_4)]}\right). \quad (55)$$

Note that the parameter $M(\mathscr{D})l^2$ is a nonmonotonic function of H and D since the effective mass (54) increases and the magnetic length l decreases with the increasing magnetic field. So, this parameter increases with H for $D>0.7a^*$ (a^* is the exciton Bohr radius) in the strong-field range and drops for $D<0.7a^*$. Therefore, an increase in H at $D<0.7a^*$ can lead to a transition from the absorption type (55) to (53), and the energy $|E|$ above which the instanton approximation holds decreases with increasing magnetic field for $D>0.7a^*$.

6. MAGNETOEXCITON LIGHT ABSORPTION IN TWO-COMPONENT SOLID SOLUTIONS

The potential of interaction between an exciton and random field $\xi(\mathbf{r})$ due to local fluctuations in the concentration of the solid solution components can be expressed as[14–16]

$$\hat{V}(\mathbf{r}_e,\mathbf{r}_h)=\beta_e\xi(\mathbf{r}_e)-\beta_h\xi(\mathbf{r}_h), \quad (56)$$

where

$$\beta_e=\frac{1}{N}\frac{\partial E_c}{\partial x}, \quad \beta_h=\frac{1}{N}\frac{\partial E_v}{\partial x},$$

x is the mean concentration of sites occupied by atoms A, E_c is the energy of the conduction band bottom, E_v is the energy of the valence band top, N is the density of lattice sites where atoms of species A or B can be located, and $\xi(\mathbf{r})$ is the excess concentration of one solid solution component. The random Gaussian function $\xi(\mathbf{r})$ satisfies the relationship[15]

$$\langle\xi(\mathbf{r}_1)\xi(\mathbf{r}_2)\rangle=Nx(1-x)\delta(\mathbf{r}_1-\mathbf{r}_2). \quad (57)$$

Using Eqs. (17) and (57), we obtain the correlation function of the effective potential

$$B(\mathbf{R}_1-\mathbf{R}_2)=\frac{(\beta_e-\beta_h)^2}{2\pi l^2}g\exp\left(-\frac{(\mathbf{R}_1-\mathbf{R}_2)^2}{2l^2}\right), \quad (58)$$

where $g=Nx(1-x)$. At $\beta_e=\beta_h$ the correlation function vanishes. Including deviations of the random field from Gaussian of the form[22]

$$\sum_{\alpha,\beta=x,y} b_{\alpha\beta}\frac{\partial^2\delta(\mathbf{r}_1-\mathbf{r}_2)}{\partial r_{1\alpha}\partial r_{2\beta}}$$

in Eq. (57) does not change this result [$B(\mathbf{R})=0$ holds at $\beta_e=\beta_h$]. A weak interaction occurs only when virtual transitions to other Landau levels are taken into account.[34]

The absorption coefficient at $E=0$ is

$$A(0)=\sqrt{\frac{2}{\pi}}\frac{e^2 l}{\epsilon\hbar^2 g}\frac{1}{(\beta_e-\beta_h)^2}. \quad (59)$$

This coefficient drops with increasing H as $1/\sqrt{H}$.

The magnetoabsorption coefficient in a III–V quantum well in the instanton approximation has the form

$$A(E)=\frac{E^2 l}{|\beta_e-\beta_h|^3 g^{3/2}}\exp\left(-\frac{2\pi l^2 E^2}{(\beta_e-\beta_h)^2 g}\right). \quad (60)$$

7. CONCLUSIONS

We have calculated the coefficient of magnetoexciton absorption in a single quantum well and in coupled quantum wells taking into account quasi-two-dimensional random fields due to fluctuations in the widths of quantum wells or fluctuations in the concentrations of substitutional solid solution components. An increase in the random field dispersion results in a lower amplitude and a larger width of the absorption peak. Similar changes in the absorption spectra occur when the distance between coupled quantum wells containing electrons and holes increases. In the region of high magnetic fields, the peak amplitude drops with the in-

creasing magnetic field, in agreement with the experimental data.[3] The spectra of direct and indirect excitons in single and coupled quantum dots have been calculated.

The work was supported by the Russian Fund for Fundamental Research, INTAS, and the ''Physics of Solid-state Nanostructures'' program. One of the authors (A.M.R.) was sponsored by the Soros Postgraduate program financed by the ISSEP fund of G. Soros.

*)E-mail: lozovik@isan.troitsk.ru

[1] T. Fukuzawa, E. E. Mendez, and J. M. Hong, Phys. Rev. Lett. **64**, 3066 (1990).
[2] L. V. Butov, V. D. Kulakovskii, G. E. W. Bauer, A. Forchel, and D. Grützmacher, Phys. Rev. B **46**, 12765 (1992).
[3] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).
[4] J.-P. Cheng, J. Kono, B. D. McCombe, I. Lo, W. C. Mitchel, and C. E. Stutz, Phys. Rev. Lett. **74**, 450 (1995).
[5] U. Sivan, P. M. Solomon, and H. Strikman, Phys. Rev. Lett. **68**, 1196 (1992).
[6] L. V. Butov, A. Zrenner, G. Abstreiter, A. V. Petinova, and K. Eberl, Phys. Rev. B **52**, 12153 (1995).
[7] M. Bayer, V. B. Timofeev, F. Faller, T. Gutbrod, and A. Forchel, Phys. Rev. B **54**, 8799 (1996).
[8] A. I. Filin, V. B. Timofeev, S. I. Gubarev, D. Birkedel, and J. M. Hwang, JETP Lett. **65**, 656 (1997).
[9] V. B. Timofeev, A. V. Larionov, P. S. Dorozhkin, M. Bayer, A. Forchel, and Zh. Straka, JETP Lett. **65**, 877 (1997).
[10] Yu. E. Lozovik and V. I. Yudson, JETP Lett. **22**, 274 (1975); Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)].
[11] Yu. E. Lozovik and M. V. Nikitkov, Zh. Éksp. Teor. Fiz. **111**, 1107 (1997) [JETP **84**, 612 (1997)].
[12] Yu. E. Lozovik and A. M. Ruvinskiĭ, Fiz. Tekh. Poluprovod. **32**, 596 (1998) [Semiconductors **32**, 533 (1998)]; Yu. E. Losovik and A. M. Ruvinsky, Phys. Scr. **58**, 90 (1998).
[13] Yu. E. Lozovik and O. L. Berman, Zh. Éksp. Teor. Fiz. **111**, 1879 (1997) [JETP **84**, 1027 (1997)]; Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, JETP Lett. **66**, 355 (1997).
[14] S. D. Baranovskiĭ and A. L. Éfros, Fiz. Tekh. Poluprovod. **12**, 2233 (1978) [Sov. Phys. Semicond. **12**, 1328 (1978)].
[15] N. N. Ablyazov, M. É. Raikh, and A. L. Éfros, Fiz. Tverd. Tela **25**, 353 (1983) [Sov. Phys. Solid State **25**, 199 (1983)].
[16] Zh. S. Gevorkyan and Yu. E. Lozovik, Fiz. Tverd. Tela **27**, 1800 (1985) [Sov. Phys. Solid State **27**, 1079 (1985)].
[17] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 1079 (1980)].
[18] Yu. E. Lozovik and A. M. Ruvinskiĭ, Fiz. Tverd. Tela **39**, 2220 (1997) [Phys. Solid State **39**, 1981 (1997)].
[19] Yu. E. Lozovik and A. M. Ruvinskiĭ, Zh. Éksp. Teor. Fiz. **112**, 1791 (1997) [JETP **85**, 979 (1997)]; Yu. E. Lozovik and A. M. Ruvinsky, Phys. Lett. A **227**, 271 (1997).
[20] Yu. E. Lozovik, submitted to Physica E.
[21] D. Bimberg, J. Christen, T. Fukunaga, H. Nakashima, D. E. Mars, and J. N. Miller, J. Vac. Sci. Technol. **5**, 1191 (1987); M. Tanaka and H. Sakaki, J. Cryst. Growth **81**, 153 (1987).
[22] V. L. Bonch-Bruevich, I. P. Zvyagin, R. Kaĭper, A. G. Mironov, R. Énderlaĭn, and B. Ésser, in *Electronic Theory of Disordered Semiconductors*, Nauka, Moscow (1981), p. 350.
[23] A. Zrenner, L. V. Butov, M. Hagn, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **72**, 3382 (1994).
[24] V. Halonen, T. Chakraborty, and P. Pietiläinen, Phys. Rev. B **45**, 5980 (1992).
[25] R. J. Elliott and R. Loudon, J. Phys. Chem. Solids **15**, 196 (1960).
[26] L. P. Gor'kov and I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **53**, 717 (1967) [Sov. Phys. JETP **26**, 449 (1968)].
[27] P. K. Basu and P. Ray, Phys. Rev. B **44**, 1844 (1991).
[28] A. B. Dzyubenko and G. E. W. Bauer, Phys. Rev. B **51**, 14524 (1995).
[29] S. John and M. J. Stephen, Phys. Rev. B **28**, 6358 (1983).
[30] P. I. Arseyev and A. B. Dzyubenko, Phys. Rev. B **52**, R2261 (1995).
[31] R. J. Elliott, J. A. Krumhansl, and P. L. Leath, Rev. Mod. Phys. **46**, 465 (1974).
[32] B. I. Halperin and M. Lax, Phys. Rev. **148**, 722 (1966); Phys. Rev. **153**, 802 (1967).
[33] S. John and M. J. Stephen, J. Phys. C **17**, L559 (1984).
[34] A. B. Dzybenko and Yu. E. Lozovik, J. Phys. A **24**, 415 (1991).

Translation provided by the Russian Editorial office.

Non-Fermi-liquid effects in tunneling

L. A. Manakova*)

Russian Research Center ''Kurchatov Institute,'' 123182 Moscow, Russia
(Submitted 9 April 1998)
Zh. Éksp. Teor. Fiz. **114**, 1466–1486 (October 1998)

Non-Fermi-liquid tunneling mechanisms in a quantum structure with its own two-dimensional continuum doped with transition metal impurities are considered. New physical realizations of the two-channel Kondo orbital model with mechanisms different from those previously described in literature occur in such quantum structures. The tunneling transparency is anomalously high owing to new channels generated by multiparticle Fermi-liquid resonances near the edge of the two-dimensional energy band in the process of tunneling. The widths of new edge resonances can be much smaller than the width of the ''bare'' non-Fermi-liquid resonance at the Fermi level in the banks. The additional scattering due to tunneling induces a transition from the non-Fermi-liquid to the Fermi-liquid state as the separation between the Fermi level in the banks and the two-dimensional band edge in the quantum well varies.
 [S1063-7761(98)02110-6]

1. INTRODUCTION

The importance of many-body effects in tunneling mechanisms in quantum structures is well known.[1–4] They not only lead to essential renormalization of tunneling parameters, but also generate new tunneling channels. Such effects were probably first reported in Refs. 1 and 2 for the case of Kondo scattering in the presence of strong Coulomb repulsion on a localized level.

Quantum structures with a negative differential resistance, which find important applications, often have an energy profile shaped like a double-barrier quantum well. This type includes GaAlAs/GaAs/GaAlAs heterostructures, whose inner GaAs layer forms a quantum well with its own two-dimensional continuum of space-quantized band states. Previously the role of the two-dimensional continuum was considered to be trivial, so the mechanisms of tunneling via a quantum well[5] and a localized level below a barrier[6] were deemed identical.

Tunneling mechanisms in which the two-dimensional continuum in a double-barrier quantum well is of fundamental importance were first investigated in our earlier work.[7–19] An exponential growth in the transparency due to an elementary tunneling event was predicted using the single-particle approach. The competition between multiparticle and single-particle resonances in the process of tunneling for the case of one-channel Kondo scattering of electrons from the banks via an orbitally nondegenerate impurity state was considered.[8,9] Conditions under which the main contribution to the tunneling transparency comes from single-particle resonances were determined.

However, the state of a real transition metal impurity is usually orbitally degenerate. In this case,[11] the multichannel exchange between conduction electrons and an impurity state for $S<2n$, where S is the impurity spin and n is the number of orbital channels, results in formation of a non-Fermi-liquid (NFL) spectrum of elementary excitations on the Fermi level (see also Refs. 12 and 13). The tunneling mechanisms in quantum structures with an NFL continuum in their banks have not been sufficiently studied heretofore. The temperature dependence of the conductance was estimated in solving the problem of resonant tunneling via a quasi-localized state.[14] These calculations were undertaken with a view to interpreting experimental data,[15] which probably indicated a cross-over between non-Fermi-liquid and Fermi-liquid states in an external magnetic field.

This paper suggests a new multiparticle tunneling mechanism in a double-barrier quantum well, containing a two-dimensional continuum of electron band states, and doped with transition metal impurities. The multichannel orbital or spin Kondo scattering of conduction electrons from the banks via impurity states in the well is taken into account. Although the multi-channel Kondo spin model was proposed almost two decades ago, the researchers took a new interest in non-Fermi-liquid states after the emergence of physical realizations of the two-channel orbital Kondo model in heavy-fermion systems and high-temperature superconductors.[17]

For this reason, the problem of tunneling through a double-barrier quantum well containing a two-dimensional continuum of band states and doped with transition metal impurities is of fundamental importance. First, it allows us to design new physical realizations of the two-channel orbital Kondo model with mechanisms different from those suggested previously,[17–19] and second, such realizations may allow investigators to directly observe non-Fermi-liquid effects.

In this paper, the mechanism generating a two-channel exchange (in terms of the impurity quadrupole moment) scattering is breaking of the axial symmetry of space-quantized states in the quantum well due to virtual transitions between states in the banks and in the well.

The key role in the tunneling mechanisms suggested in the paper is played by the two-dimensional continuum in the

quantum well. Multiparticle resonances with widths much smaller than those of the ''bare'' non-Fermi-liquid resonance at the Fermi level can be formed near the two-dimensional band edge in the process of tunneling through a doped quantum well. The width of edge resonances is a power function of the tunneling parameters. The edge resonances are due to additional scattering of quasi-two-dimensional electrons in the well by non-Fermi-liquid excitations from the Fermi level, which is caused by tunneling processes. These resonances are multiparticle Fermi-liquid states, since they correspond to simple poles in electron Green's functions in both the banks and the quantum well instead of a power-law singularity corresponding to non-Fermi-liquid excitations without tunneling. The edge resonance make the principal contribution to the transparency and determine the Fermi-liquid regime of tunneling till the separation between the Fermi level and the edge of the two-dimensional band is smaller than the width of the bare non-Fermi-liquid resonance in this region. If the gap between the Fermi level and two-dimensional band gap is larger than the width of the bare non-Fermi-liquid resonance, the latter make the main contribution to the transparency, so it determines the non-Fermi-liquid regime of tunneling.

To sum up, when the Fermi level in the banks approaches the two-dimensional band gap in the well, a crossover between the non-Fermi-liquid and Fermi-liquid tunneling modes takes place.

2. TUNNELING HAMILTONIAN AND TUNNELING PROBABILITY

1. We consider the situation when a transition metal impurity generates a deep level with energy E_d in the band gap of the inside layer of the double-barrier quantum well. The spectrum in this layer also contains a two-dimensional continuum with dispersion $\varepsilon_{\mathbf{k}\perp}$. In solving this specific problem, we focus on the case when the Fermi level in the banks is near the bottom of the conduction band of the inside layer. Given that the five-fold degenerate d level is split by the crystal field into the doubly degenerate e_g level and the three-fold degenerate t_{2g} level, the eigenfunctions of an electron on the d level are the cubic d functions proper, and the quantum numbers are the numbers μ of columns of the irreducible representation of the point group: $\mu_{e_g}=\pm 1$, $\mu_{t_{2g}}=0,\pm 1$, $E_d=E_{e_g,t_{2g}}$. In a bulk semiconductor, these states are not hybridized with the conduction band states.[20] In a quantum well, however, the situation is different. As was shown previously,[10] the e_g and t_{2g} levels can hybridize with the conductance band states owing to the lowering of their symmetry. The Hamiltonian of a tunneling system with the E_d level in the quantum well has the form

$$H=H_{00}+H_t+H_{\rm int}, \tag{1}$$

where $H_{00}=H_{00}^{\nu}+H_{00}^{d}+H_{00}^{c}$ is the Hamiltonian of the noninteracting banks and quantum well:

$$H_{00}=\sum_{\mathbf{k},\nu=L,R,\sigma}\varepsilon_{\mathbf{k}}^{\nu}a_{\mathbf{k}\sigma\nu}^{+}a_{\mathbf{k}\sigma\nu}+\sum_{\mu\sigma}E_d d_{\mu\sigma}^{+}d_{\mu\sigma}$$

$$+\sum_{\mathbf{k}_\perp\sigma}\varepsilon_{\mathbf{k}_\perp}c_{\mathbf{k}_\perp\sigma}^{+}c_{\mathbf{k}_\perp\sigma}, \tag{2}$$

H_t is the tunneling Hamiltonian:

$$H_t=H_{td}+H_{tc}=\sum_{\mathbf{k}\nu\mu\sigma}(T_{\mathbf{k}d}^{\nu\mu}a_{\mathbf{k}\nu\sigma}^{+}d_{\mu\sigma}+\text{H.c.})$$

$$+\sum_{\mathbf{k}\nu\sigma}\sum_{\mathbf{k}_\perp'}(T_{\mathbf{k}\mathbf{k}_\perp'}^{\nu}a_{\mathbf{k}\nu\sigma}^{+}c_{\mathbf{k}_\perp'\sigma}+\text{H.c.}). \tag{3}$$

The operators $a_{\mathbf{k}\nu\sigma}$ describe electron states in the left-hand ($\nu=L$) and right-hand ($\nu=R$) banks of the tunneling junction. The operators d_σ and $c_{\mathbf{k}\perp}$ correspond to the wave functions of hybridized localized $\psi_{d\mu}(\mathbf{r})$ and band $\Psi(\mathbf{k}_\perp,\mathbf{k})$ states.[10,20]

The tunneling matrix elements in Eq. (3) are[8]

$$T_{\mathbf{k}d}^{\nu\mu}=B_\mu(\mathbf{k}_\perp)T_d^{\nu}(k_l),\quad T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}=T_0^{\nu}(k_l)\delta_{\mathbf{k}_\perp\mathbf{k}_\perp'}+T_{\mathbf{k}c}^{\nu}B(\mathbf{k}_\perp'),$$

$$T_{\mathbf{k}c}^{\nu}=T_c^{\nu}(k_l)B(\mathbf{k}_\perp),\quad B(\mathbf{k}_\perp)=\sum_{\mu}B_\mu(\mathbf{k}_\perp). \tag{4}$$

Here $B_\mu(\mathbf{k}_\perp)=v_{\mathbf{k}_\perp d}/(E_{d\mu}-\varepsilon_{\mathbf{k}_\perp})$, and $V_{\mathbf{k}_\perp d}^{\mu}$ is the hybridization matrix element in the quantum well:[10]

$$V_{\mathbf{k}_\perp d}^{\mu}=\int d\mathbf{r}\varphi_{d\mu}(\mathbf{r})U(\mathbf{r})\Psi(\mathbf{k}_\perp,\mathbf{r})$$

$$=\int d\mathbf{r}\varphi_{d\mu}(\mathbf{r})U(\mathbf{r})\psi(\mathbf{k}_\perp,\rho)\varphi(z). \tag{5}$$

In Eq. (4) we have written $\mathbf{k}=(\mathbf{k}_\perp,k_l)$, and we assume that the electron motion in the vertical and lateral directions are decoupled: $\varepsilon_{\mathbf{k}}=\varepsilon_{\mathbf{k}_\perp}+\varepsilon_{k_l}$; hence the impurity contribution to $T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}$ is separable, which is essential in solving the tunneling problem. The difference between the matrix elements $T_0^{\nu}(k_l)$, $T_d^{\nu}(k_l)$, and $T_c^{\nu}(k_l)$ is that they contain different factors composed of normalization constants, but all three parameters are proportional to the matrix elements of the tunneling potential between vertical components of the wave functions in the banks ν and in the doped layer:

$$T_\nu(k_l)=\int\psi_\nu^{*}(k_l,z)V(z)\varphi(z)dz.$$

In addition to the standard term H_{td} (which exists, however, owing to the presence of the ''Bloch tail'' of the impurity wave function), the tunneling Hamiltonian contains the second term H_{tc}. This is the term responsible for the new resonance states near the edge of the 2D band.[8]

Before proceeding to the solution of the problem of tunneling via impurity states, we should take into account the restructuring of the band spectrum in the well due to the tunneling between the banks and well described by terms with $T_0^{\nu}(k_l)$ in $T_{\mathbf{k},\mathbf{k}_\perp'}^{\nu}$. It was shown previously[8] that, since the tunneling takes place between bands of different dimensionalities (the 3D bands in the banks and the 2D band in the

well), evanescent states with complex wave vectors $\mathbf{k}_\perp$ and complex energies $\tilde{\varepsilon}_{\mathbf{k}_\perp}$ such that $\mathrm{Re}\,\mathbf{k}_\perp \approx \mathrm{Im}\,\mathbf{k}_\perp$ are created near the bottom of the 2D band. These states exist in the region $\varepsilon - \varepsilon_c < \gamma \ll W_c$ and are described by the density of states

$$\rho_c(\varepsilon) = \frac{\rho_{0c}}{\pi}\left[\arctan\frac{\varepsilon-\varepsilon_c}{\gamma_0} - \arctan\frac{\varepsilon - W_c}{\gamma_0}\right]. \tag{6}$$

Here ρ_{0c} is the density of states at the unperturbed 2D band edge, ε_c and $W_c \sim \rho_{0c}^{-1}$ are the edge position and width of the 2D band, respectively; $\gamma_0 \sim \Sigma_\nu |T_0^\nu(\varepsilon_c)|^2 \rho_{0\nu}$, $\rho_{0\nu} \sim W_a^{-1}$ is the corresponding tunneling width, and W_a is the width of the conduction band in the banks.

Thus, the only impurity term in the tunneling Hamiltonian H_{tc} in the interesting range of energies is that proportional to $T_c^\nu(k_l)$ in Eq. (4), but the density of states in the 2D continuum is given by Eq. (6).

The tunneling Hamiltonian H_t can be transformed to the ''single-band'' form, which is convenient for analyzing interactions. To this end, let us introduce, instead of $a_{\mathbf{k}\nu\sigma}$, where $\nu = L,R$, the new operators $a_{\mathbf{k}\sigma}$ and $b_{\mathbf{k}\sigma}$ defined by the linear transformation:

$$a_{\mathbf{k}\sigma} = u_{\mathbf{k}} a_{\mathbf{k}L\sigma} + v_{\mathbf{k}} a_{\mathbf{k}R\sigma}, \quad b_{\mathbf{k}\sigma} = u_{\mathbf{k}} a_{\mathbf{k}R\sigma} - v_{\mathbf{k}} a_{\mathbf{k}L\sigma},$$

$$u_{\mathbf{k}} = \frac{T_{\mathbf{k}d}^{L\mu}}{[(T_{\mathbf{k}d}^{L\mu})^2 + (T_{\mathbf{k}d}^{R\mu})^2]^{1/2}}, \quad u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2 = 1. \tag{7}$$

One can check directly that, in the new representation, quasiparticles of only one sort described by the operators $a_{\mathbf{k}\sigma}$ are hybridized with both the localized states and those of the two-dimensional continuum. The Hamiltonians H_{00}, H_t, and H_{int} are replaced in accordance with the formulas

$$H_{00} \to H_{00}^a + H_{00}^b, \quad H_t \to H_t^{(a)}, \quad H_{\mathrm{int}} \to H_{\mathrm{int}}^{(a)}. \tag{8}$$

If the electron spectra in both banks are identical, $\epsilon_{\mathbf{k}}^L = \epsilon_{\mathbf{k}}^R = \epsilon_{\mathbf{k}a}$, we have

$$H_{00}^{(a)} + H_{00}^{(b)} = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} a_{\mathbf{k}\sigma}^+ a_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} b_{\mathbf{k}\sigma}^+ b_{\mathbf{k}\sigma}.$$

The Hamiltonian $H_t^{(a)}$ is derived from H_t in Eq. (3) using the substitutions

$$a_{\mathbf{k}\nu\sigma}^+ \to a_{\mathbf{k}\sigma}, \quad T_{\mathbf{k}d}^{\nu\mu} \to T_{\mathbf{k}d}^{a\mu} = [(T_{\mathbf{k}d}^{L\mu})^2 + (T_{\mathbf{k}d}^{R\mu})^2]^{1/2},$$

$$T_{\mathbf{k}\mathbf{k}'_\perp}^{\nu} \to T_{\mathbf{k}\mathbf{k}'_\perp}^{a} = T_{\mathbf{k}\mathbf{k}'_\perp}^{L} u_{\mathbf{k}} + T_{\mathbf{k}\mathbf{k}'_\perp}^{R} v_{\mathbf{k}}.$$

Since the tunneling Hamiltonian acts only on the states $a_{\mathbf{k}\sigma}$, the interaction Hamiltonian H_{int} is defined only for these states.

2. In calculating the elastic tunneling probability, we use a formula similar to the Landauer formula, which expresses the probability in terms of the scattering matrix:

$$W(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}'}^R) = 2\pi |\mathscr{T}(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}'}^R)|^2 \delta(\varepsilon_{\mathbf{k}}^L - \varepsilon_{\mathbf{k}'}^R), \tag{9}$$

where $\mathscr{T} = H_t G H_t$ and $G = \hat{I}(z - H)^{-1}$ is the Green's function ($\hat{I}$ is the unit matrix). The matrix $\mathscr{T}$ describes tunneling via both band and localized impurity states with due account of all elastic scattering processes within the quantum well, which determine the Green's function G. The general expression for $\mathscr{T}$ in a double-barrier quantum well has the form

$$\begin{aligned}\mathscr{T}(\mathbf{k}, \varepsilon_{\mathbf{k}}^L; \mathbf{k}', \varepsilon_{\mathbf{k}'}^R) &= \langle a_{\mathbf{k}L}|H_t|d\rangle\langle d|G|d\rangle\langle d|H_t|a_{\mathbf{k}'R}\rangle \\ &+ \sum_{\mathbf{p}\mathbf{p}'} \langle a_{\mathbf{k}L}|H_t|c_{\mathbf{p}}\rangle\langle c_{\mathbf{p}}|G|c_{\mathbf{p}'}\rangle \\ &\times \langle c_{\mathbf{p}'}|H_t|a_{\mathbf{k}'R}\rangle + \sum_{\mathbf{p}} \langle a_{\mathbf{k}L}|H_t|c_{\mathbf{p}}\rangle \\ &\times \langle c_{\mathbf{p}}|G|d\rangle\langle d|H_t|a_{\mathbf{k}'R}\rangle + \sum_{\mathbf{p}} \langle a_{\mathbf{k}L}|H_t|d\rangle \\ &\times \langle d|G|c_{\mathbf{p}}\rangle\langle c_{\mathbf{p}}|H_t|a_{\mathbf{k}'R}\rangle. \end{aligned} \tag{10}$$

Here the spin indices are omitted. Assuming that the main contribution to the tunneling amplitude comes from the matrix element containing the Green's function, $\langle c_{\mathbf{p}}|G|c_{\mathbf{p}'}\rangle \equiv G_{cc}(\mathbf{k}_\perp \mathbf{k}'_\perp; \varepsilon_{\mathbf{k}}^L)$, we obtain an expression for the tunneling probability via the band channel:

$$\begin{aligned} W(\mathbf{k}_\perp, \mathbf{k}'_\perp, \varepsilon_{\mathbf{k}}^L) = \sum_{\mathbf{p}_\perp, \mathbf{p}'_\perp} &\Gamma_L^c(\mathbf{k}_\perp, \mathbf{p}_\perp; \varepsilon_{\mathbf{k}}^L)\Gamma_R^c(\mathbf{k}'_\perp, \mathbf{p}'_\perp; \varepsilon_{\mathbf{k}'}^R) \\ &\times \left|\frac{d\varepsilon_{\mathbf{k}'}}{dk'_l}\right| |G_{cc}(\mathbf{p}_\perp, \mathbf{p}'_\perp; \varepsilon_{\mathbf{k}}^L)|^2 \delta(\varepsilon_{\mathbf{k}}^L - \varepsilon_{\mathbf{k}'}^R). \end{aligned} \tag{11}$$

For one-dimensional states near the Fermi level, which will be used in what follows, the tunneling widths are determined by the expression

$$\begin{aligned}\Gamma_\nu^c(\mathbf{p}_\perp; \varepsilon) = \Gamma_\nu^c(\mathbf{p}_\perp; \varepsilon_k^\nu) &= \sum_q \langle a_{q\nu}|H_{tc}|c_{\mathbf{p}_\perp}\rangle\langle c_{\mathbf{p}_\perp}|H_{tc}|a_{q\nu}\rangle \\ &= \sum_k |T_{k\mathbf{p}_\perp}^\nu|^2 \delta(\varepsilon - \varepsilon_k^\nu). \end{aligned} \tag{12}$$

By expressing $a_{kL,R}$ in Eq. (10) in terms of a_k and b_k through Eq. (7), we obtain a formula which will be useful hereafter:

$$\Gamma_{L,R}^c = \Gamma_a^c \frac{\Gamma_{L,R}^{c0}}{\Gamma_L^{c0} + \Gamma_R^{c0}}, \tag{13}$$

where $\Gamma_{L,R}^{c0}$ are tunneling widths without interaction:

$$\sum_\nu \Gamma_{0\nu}(\varepsilon_F) = \sum_\nu |T_0^\nu(\varepsilon_F)|^2 \rho_\nu(\varepsilon_F) \sim \gamma_0,$$

and the arguments of all the functions are the same as in the previous formula.

The Green's function G_{cc} in Eq. (11) describes propagation of a tunneling electron through a quantum well with due account of all resonant and potential scattering processes within the well which are caused by the Hamiltonian H_t. The character of elementary excitations in the banks, which scatter quasi-two-dimensional electrons in the well, is determined by interactions in H_{int}^a. The latter interactions, together with H_{tc}, generate singularities in the interesting range of energies near the two-dimensional band edge. Thus, in solving the tunneling problem, the main thing is to deter-

mine the Hamiltonian of the interaction in Eq. (8) coupling the states within the well and $a_{\mathbf{k}\sigma}$ states, which are symmetrical linear combinations of those in the right- and left-hand banks.

3. MECHANISMS OF TWO-CHANNEL ORBITAL KONDO SCATTERING

1. Physical realizations of the two-channel orbital Kondo model in heavy-fermion and HTSC systems,[17] along with metallic glasses doped with two-level impurities,[18,19] are well known.

This paper describes quantum structures in which new physical realizations of the two-channel orbital Kondo model take place.

The interactions between conduction electrons in the banks and an orbitally degenerate impurity state in the well is due to the Hubbard repulsion between electrons on the deep level. It is described by the following term in $H_{\rm int}$ of Eq. (1):

$$H_U=\sum_{\mu,\mu';\sigma,\sigma'} U_{\mu\mu'}n_{d\mu\sigma}n_{d\mu'\sigma'}(1-\delta_{\mu\mu'}\delta_{\sigma\sigma'}). \qquad (14)$$

By applying the Schrieffer–Wolff transformation to the generalized Anderson Hamiltonian

$$H_A=H_{00}^a+H_{td}^a+H_{00}^d+H_U, \quad H_{00}^d=\sum_{\mu\sigma} E_d d_{\mu\sigma}^+ d_{\mu\sigma},$$

we find that, in the general case, the effective interaction between an impurity state in the well and electrons in the banks has the form

$$H_{\rm int}=\sum_{\mathbf{k}\mathbf{k}'}\sum_{\mu\mu'\sigma_i} V_{\mathbf{k}\mathbf{k}'}^{\mu\mu'} a_{\mathbf{k}\sigma_1}^+ a_{\mathbf{k}'\sigma_2} d_{\mu\sigma_3}^\dagger d_{\mu'\sigma_4},$$

$$V_{\mathbf{k}\mathbf{k}'}^{\mu\mu'}=-T_{d\mathbf{k}}^{a\mu}T_{\mathbf{k}'d}^{a\mu'}\left(\frac{1}{E_d-\varepsilon_{\mathbf{k}}^a}-\frac{1}{E_d+U_{\mu\mu'}-\varepsilon_{\mathbf{k}'}^a}\right). \qquad (15)$$

The new physical mechanism leading to multichannel exchange scattering is breaking of the axial symmetry of space-quantized states in the quantum well. Specifically, if the dependence of the 2D wave continuum wave functions in Eq. (5) on the transverse space coordinates is taken into account, the matrix elements $V_{\mathbf{k}_\perp d}^\mu$ are nonzero for all components of the d state with $\mu_{e_g}=\pm 1$, $\mu_{t_{2g}}=0,\pm 1$, and the terms with $\mu_{e_g}\neq +1$ are nonzero only when the axial symmetry of quantized states in the well is broken. Accordingly, the tunneling matrix elements $T_{\mathbf{k}d}^{\alpha\mu}$ are nonzero for all $\mu_{e_g}=\pm 1$ and $\mu_{t_{2g}}=0,\pm 1$. As follows from the definition (4), the direct consequence of the axial symmetry breaking is the momentum $\mathbf{k}$ direction dependence of the tunneling matrix elements $T_{\mathbf{k}d}^\mu$ (the spatial nonlocality of $T_{\mathbf{k}d}^\mu$).

The matrix elements of interaction with $\mu\neq\mu'$ in Eq. (15) are nonzero owing to the mechanism of axial symmetry breaking in the quantum well described above.

Of all interactions in Eq. (15), we will consider only those which have the exchange form with respect to either the orbital or spin index.

The interaction

$$H_{\rm int}=\sum_{\mathbf{k}\mathbf{k}'}\sum_{\mu\mu'\sigma} V_{\mathbf{k}\mathbf{k}'}^{\mu\mu'} a_{\mathbf{k}\sigma}^+ a_{\mathbf{k}'\sigma} d_\mu^+ d_{\mu'}, \qquad (16)$$

which describes exchange with respect to the orbital index, can be dominant in the following cases: (1) the spin degeneracy is fully lifted owing to the combined action of crystal fields (they have been taken into account previously) and exchange interaction within the d shell (the Hund rule); (2) both spin degenerate levels are below the Fermi level; (3) the temperature under consideration is much smaller than all energy parameters of the problem, except the Kondo temperature in the problem with the spin exchange. The latter condition can be satisfied when the exchange constants are sufficiently small, since the ''spin'' Kondo temperature is an exponential function of the exchange parameters, whereas the orbital exchange energies are power functions of the exchange parameters (see Refs. 13 and 26, and also Eqs. (32) and (33) in this paper). Let us rely on the fact that, usually, the gap between the e_g and t_{2g} states is sufficiently large (in comparison with the energy scales of the problem), so their mixing in the interaction matrix elements can be neglected. Suppose also that the deep level nearest to the 2D band edge is the e_g doublet. We expand the operators $a_{\mathbf{k}\sigma}$ and matrix elements $V_{\mathbf{k}\mathbf{k}'}^{\mu\mu'}$ in Eq. (16) for the e_g doublet in terms of cubic harmonics $K_{\Gamma\gamma}(\Omega_{\mathbf{k}})$, where $\Gamma=e_g$, $\gamma=\pm 1$ is the number of the line in the point group irreducible representation:

$$a_{\mathbf{k}\sigma}=\sum_\gamma K_{\Gamma\gamma}(\Omega_{\mathbf{k}})a_{k\gamma\sigma},$$

$$V_{\mathbf{k}\mathbf{k}'}^{\mu\mu'}=\sum_{\gamma\gamma'} K_{\Gamma\gamma}^*(\Omega_{\mathbf{k}})K_{\Gamma\gamma'}(\Omega_{\mathbf{k}'})V_{\gamma\gamma'}^{\mu\mu'}(kk'), \qquad (17)$$

and $\Omega_{\mathbf{k}}$ is the solid angle. The interaction nondiagonal matrix elements with $\gamma\neq\gamma'$ are nonzero since the expansion of tunneling matrix elements contains both of the terms with $\gamma=\pm 1$ by virtue of their dependence on the momentum $\mathbf{k}$ direction:

$$T_{\mathbf{k}d}^\mu=\sum_{\gamma=\pm 1} K_{\Gamma\gamma}(\Omega_{\mathbf{k}})T_{d\gamma}^{a\mu}(k). \qquad (18)$$

Note that in the simplest case, when 2D continuum states are described by plane waves, we have $\gamma=\mu$ and $\gamma'=\mu'$ in Eqs. (17) and (18). By substituting Eq. (17) in (16) and using the normalization properties of the cubic harmonics, we obtain an expression for the orbital exchange Hamiltonian in the case of the e_g doublet:

$$H_{\rm ex}^\mu=\sum_{\mathbf{k}\mathbf{k}'\sigma}\sum_{\gamma\gamma'=\pm 1}\sum_{\mu\mu'=\pm 1} V_{\gamma\gamma'}^{\mu\mu'}(kk')a_{k\gamma\sigma}^+ a_{k'\gamma'\sigma}d_\mu^+ d_{\mu'}. \qquad (19)$$

The doubly degenerate e_g level containing one electron (or hole) is conveniently described in terms of the pseudospin variable $\hat{\tau}_d$, whose projection on the z-axis has two values:

$$\hat{\tau}_d^z=\frac{1}{12}[3L_z^2-L(L+1)]=\pm\frac{1}{2}, \tag{20}$$

corresponding to the occupation of the d_{z^2} orbital ($L_z=0$) and $d_{x^2-y^2}$ orbital ($|L_z|=2$), where $\hat{L}$ is the angular momentum operator. The operator $\hat{\tau}_d^x \sim L_x^2-L_y^2$ inverts the pseudospin. Note that the operators $\hat{\tau}_d^z$ and $\hat{\tau}_d^x$ are components of the quadrupole moment tensor. Thus, the two values of the quantum number $\mu=\pm 1$ in Hamiltonian (19) correspond to two projections (20) of the quadrupole moment on the z-axis, i.e., the interaction (19) is quadrupole exchange scattering. By defining the pseudospin operator in Eq. (19) as

$$\hat{\tau}_d^i=\sum_{\mu\mu'=\pm 1} d_\mu^+ \tau^i_{\mu\mu'} d_{\mu'}, \quad \sum_{\mu=\pm 1} d_\mu^+ d_\mu=1$$

and assuming that

$$V^{\mu\mu'}_{\mathbf{kk}'}=\sum_{i=x,y,z} V^i_{\mathbf{kk}'}(\hat{\tau}_d^i)_{\mu\mu'}, \quad \mu,\mu'=\pm 1,$$

we can rewrite $H_{\rm ex}^{(\mu)}$ as

$$H_{\rm ex}^{(\mu)}=\sum_{kk'\sigma}\sum_{i=x,y,z}\sum_{\gamma,\gamma'=\pm 1} V^i(kk') a^+_{k\gamma\sigma}\hat{\tau}^i_{\gamma\gamma'} a_{k'\gamma'\sigma}\hat{\tau}^i_d. \tag{21}$$

Equations (19) and (21) describe the Hamiltonian which corresponds to the two-channel quadrupole exchange, where the scattering channel number (''color'') is determined by one of the two projections of the conduction electron real spin.

The tunneling matrix elements in Eq. (15) are complex in the general case, so that Eq. (21) contains, along along with the term with τ_d^x, a term with τ_d^y. Below we will consider the case $V^x=V^y\neq V^z$. Interactions determined by the products $\tau_d^i\tau_d^j$, $i\neq j$ are beyond the scope of this paper.

The difference between the model with the quadrupole exchange described by Eq. (21) and those discussed earlier[17] is in the physical exchange mechanism. We stress once again that the physical cause of the ''quadrupole exchange'' is the axial symmetry breaking in space-quantized states in the quantum well owing to virtual transitions, hence the spatial nonlocality of the tunneling matrix elements.

2. It is noteworthy that the momentum dependence of the matrix elements is also a cause of the two-channel interaction exchanging the spins. Considering again only the e_g doublet and expanding the operators and matrix elements in Eq. (15) (those with $\sigma_1=\sigma$, $\sigma_2=\sigma'$ and $d^+_{\mu\sigma_3}\equiv d^+_\sigma$, $d_{\mu\sigma_4}\equiv d_{\sigma'}$) in terms of cubic harmonics, we obtain a two-channel Hamiltonian with spin exchange, in which the channel number is determined by the orbital quantum number:

$$H_{\rm ex}^{(s)}=\sum_{kk'\sigma\sigma'}\sum_{i=x,y,z}\sum_{\mu=\pm 1} J^i_\mu(kk') \times a^+_{k\mu\sigma}\hat{\tau}^i_{\sigma\sigma'} a_{k'\mu\sigma'} d^\dagger_\sigma \hat{\tau}^i_{\sigma\sigma'} d_{\sigma'}. \tag{22}$$

This is a particular case of the multichannel Kondo spin model suggested by Nozieres and Blandin[11] (see also Ref. 30).

When the initial interaction Hamiltonians (15) are transformed to the two-channel form, terms with the ''magnetic'' field turn up:

$$H_h=h\hat{\tau}_d^z. \tag{23}$$

In the case of spin exchange, h is the applied magnetic field, and in the case of the ''quadrupole exchange'' h is the doublet splitting due to the local symmetry breaking at an impurity site, first, in accordance with the mechanism discussed above, second, owing to Jahn–Teller distortions. Below, we consider the case when the magnitude of the Jahn–Teller effect is smaller than all energy parameters of the problem.

To conclude this section, we note that, in addition to the Hubbard repulsion (14), there is another mechanism of interaction between electrons in the banks and impurity states in this system. This is renormalization of the interactions within the doped layer due to tunneling between the banks and the quantum well described by the first term on the right of the second line of Eq. (4) containing the matrix element $T_0^a(k_l)$. The ''bare'' interactions within the well are given by

$$H_{\rm int}^W=\sum_{\mathbf{k}_\perp\mathbf{k}'_\perp\sigma_i}\sum_{\mu\mu'} W^{\mu\mu'}_{\mathbf{k}_\perp\mathbf{k}'_\perp} c^+_{\mathbf{k}_\perp\sigma_1} c_{\mathbf{k}'_\perp\sigma_2} d^+_{\mu\sigma_3} d_{\mu'\sigma_4}, \quad i=1,2,3,4. \tag{24}$$

Physically, these are interactions within a partially filled shell of an impurity atom[20], which includes a ''kernel'' and a ''Bloch tail.'' The tunneling $T_0^a(k_l)$ renormalizes the interactions in Eq. (24), as a result, we have an interaction between the banks and well of the form

$$H_{\rm int}=\sum_{\mathbf{kk}'\sigma_i}\tilde{W}^{\mu\mu'}_{\mathbf{kk}'} a^+_{\mathbf{k}\sigma_1} a_{\mathbf{k}'\sigma_2} d^+_{\mu\sigma_3} d_{\mu'\sigma_4}, \tag{25}$$

$$\tilde{W}^{\mu\mu'}_{\mathbf{kk}'}=W^{\mu\mu'}_{\mathbf{k}_\perp\mathbf{k}'_\perp} \times\frac{T_0^{a*}(p_l)T_0^a(p'_l)}{[\tilde{\varepsilon}_{\mathbf{k}_\perp}-\varepsilon^a(\mathbf{k}_\perp)-\varepsilon^a(p_l)][\tilde{\varepsilon}_{\mathbf{k}'_\perp}-\varepsilon^a(\mathbf{k}'_\perp)-\varepsilon^a(p'_l)]}.$$

All the interactions obtained in the previous calculations generate non-Fermi-liquid excitations at the Fermi level in the banks similar to boson-type excitations in the Luttinger liquid[16] in the case of Coulomb interaction[21,22] or bound states of n electrons (here n is the number of orbital channels) for the case of multichannel exchange scattering.[12,13,23]

It is convenient to first solve the problem of the interaction between electrons in the banks and localized states using one of the techniques developed for the problem with two-channel Kondo scattering[13,24] and then use this solution as a basis for solving the tunneling problem. In other words, the solution of the problem with Hamiltonian $H_0=H_{00}+H_{\rm int}$ should yield non-Fermi-liquid states at the Fermi level and the corresponding state at the impurity level. The inclusion of the tunneling term H_{tc} leads to additional scattering of 2D electrons in the doped layer by states in the banks and impurity levels, which is obtained when interaction is taken into account. This scattering fully determines the probability of elastic tunneling through the quantum well [see (11)].

4. GREEN'S FUNCTIONS AND DENSITY OF STATES IN THE BANKS

1. In the case of a level doubly degenerate in either orbital or spin variables, when the dominant effect of interaction is generation of a multiparticle resonance at the Fermi level, the Green's functions for conduction electrons can be calculated using the equation-of-motion technique:[24,25]

$$G^a_{0\beta}(\mathbf{k}\mathbf{k}';z)=\delta_{\mathbf{k}\mathbf{k}'}G^a_{00\beta}(\mathbf{k};z)+G^a_{00\beta}(\mathbf{k};z)T^{a\beta *}_{\mathbf{k}d}\times G_{d\beta}(z)T^{a\beta}_{\mathbf{k}'d}G^a_{00\beta}(\mathbf{k}';z). \tag{26}$$

Here $\beta=[\sigma,\mu]$ for the spin or orbital exchange, respectively, and $G^a_{00\beta}(\mathbf{k};z)$ is the Green's function for noninteracting electrons.

In the case of single-channel Kondo scattering,

$$G_{d\sigma}(z)=\frac{1}{z-\epsilon_{d\sigma}-i\gamma_d-\Sigma_K(z)},\quad \epsilon_{d\sigma}-i\gamma_d=\sum_{\mathbf{k}}\frac{|T^a_{\mathbf{k}d}|^2}{z-\epsilon^a_{\mathbf{k}}},$$

and for $|z|$ close to the Fermi level we have

$$\Sigma_K(z)=\sum_{\mathbf{k}}\frac{|T^a_{\mathbf{k}d}|^2 f(\varepsilon^a_{\mathbf{k}})}{z-\epsilon^a_{\mathbf{k}}}\sim\gamma_d\ln\frac{W_a}{z-\varepsilon_F},\quad G_{d\sigma}\approx\frac{\mathscr{Z}_K}{z-E_K}. \tag{27}$$

Here $E_K=\varepsilon_F+i\gamma_K$, γ_K is of order of the Kondo temperature T_K, $T_K\sim(W_a\gamma_d)^{1/2}\mathscr{Z}_K$. It is noteworthy that in the single-channel case the density of charged states has no singularity at the Fermi level. In the ''resonance-level'' formalism employed in Eq. (27), this is seen through the small residue $\mathscr{Z}_K$ which determines the pole contribution and is controlled by the small number of charged excitations at the Fermi level.

2. In order to determine $G_{d\beta}(z)$ in the case of two-channel exchange, let us use the resonance-level model suggested by Emery and Kivelson.[13] Rewrite the Hamiltonians

$$H^a_{00}=\sum_{k\mu\sigma}\epsilon^a_k a^+_{k\mu\sigma}a_{k\mu\sigma},\quad H^{(\mu)}_{\rm ex}+H_h,\quad \epsilon^a_k=(k-k_F)v_F,$$

[see Eqs. (21) and (23)] as

$$H^a_{00}=iv_F\sum_{\mu\sigma}\int_{-\infty}^{\infty}\psi^+_{\mu\sigma}(x)\partial_x\psi_{\mu\sigma}(x),$$

$$H^{(\mu)}_{\rm ex}=\frac{1}{2}\sum_{\sigma,\mu=1,2}\sum_{i=x,y,z}V^i\psi^+_{\mu\sigma}(0)\hat{\sigma}^i_{\mu\mu'}\psi_{\mu'\sigma}(0)\hat{\tau}^i_d+h\tau^z, \tag{28}$$

where

$$\psi_{\mu\sigma}(x)=\int_{-\infty}^{\infty}dk\ e^{ikx}a_{k\mu\sigma},$$

and $\hat{\tau}$ and $\hat{\sigma}$ are Pauli matrices. The one-dimensional kinetic energy

$$iv_F\sum_{\mu\sigma}\int_0^L dr[\psi^+_{1\mu\sigma}(r)\partial_r\psi_{1\mu\sigma}(r)-\psi^+_{2\mu\sigma}(r)\partial_r\psi_{2\mu\sigma}(r)],$$

where $\psi^+_{1\mu\sigma}(r)$ and $\psi^+_{2\mu\sigma}(r)$ are operators creating electrons moving to the right and left at the point r, respectively, transforms to the expression in Eq. (28) via the substitution $\psi_{2\mu\sigma}(r)\to\psi_{2\mu\sigma}(-r)$. This transformation makes sense for interactions in which operators ψ and ψ^+ are independent of x.

Emery and Kilverson[13] transformed the Hamiltonian (28) to that in the resonance-level model by the following operations:

(1) Introduction of a boson representation of the four fermion fields:

$$\psi_{\mu\sigma}(x)=\hat{\eta}\frac{\exp[-i\Phi_{\mu\sigma}(x)]}{\sqrt{2\pi a}},\quad \hat{\eta}^2=1,$$

$$\Phi_{\mu\sigma}(x)=\sqrt{\pi}\left[\int_{-\infty}^{x}dx'P_{\mu\sigma}(x')+\varphi_{\mu\sigma}(x)\right]. \tag{29}$$

The operator $\hat{\eta}$ is introduced to satisfy the anticommutation relations, a is the lattice constant, $\varphi_{\mu\sigma}(x)$ is the boson field, $P_{\mu\sigma}(x)$ is the canonically conjugate momentum: $[\phi_{\mu\sigma}(x),P_{\mu'\sigma'}(x')]=i\delta(x-x')\delta_{\mu\mu'}\delta_{\sigma\sigma'}$.

(2) Introduction of collective variables by means of the canonical transformation of the fields $\varphi_{\mu\sigma}(x)$ and $P_{\mu\sigma}(x)$:

$$\varphi_{c,f}=\frac{1}{2}[(\varphi_{11}+\varphi_{12})\pm(\varphi_{21}+\varphi_{22})],$$

$$\varphi_{s,(sf)}=\frac{1}{2}[(\varphi_{11}-\varphi_{12})\pm(\varphi_{21}-\varphi_{22})], \tag{30}$$

and likewise with the conjugate fields $P_{\mu\sigma}(x)$, $\mu,\sigma=1,2$. The collective variables described by Fourier components of boson fields $k^{1/2}\varphi_l(k)$ correspond to densities $\rho_l(k)$ of charge ($l=c$), spin ($l=f$), pseudospin ($l=s$), pseudospin–spin ($l=sf$).

(3) Transition to spinless fermion fields (''refermionization''):

$$\psi_l(x)=\frac{\exp[-i\Phi_l(x)]}{\sqrt{2\pi a}},\quad l=c,f,s,(sf). \tag{31}$$

As a result of these operations, the charge (c) and color (f) fields are separated and the Hamiltonian H_0 takes the form

$$H_0=H^a_{00}+H_{sf}+H_s,$$

where

$$H^a_{00}=iv_F\sum_{l=s,(sf)}\int_{-\infty}^{\infty}\psi^+_l(x)\partial_x\psi_l(x),$$

$$H_{sf}=\frac{V_x}{\sqrt{2\pi a}}[\psi^+_{sf}(0)+\psi_{sf}(0)](d^+-d)+h\left(d^+d-\frac{1}{2}\right),$$

$$H_s=2(V_z-\pi v_F)\psi^+_s(0)\psi_s(0)\left(d^+d-\frac{1}{2}\right). \tag{32}$$

Here we have used the Majorana representation of spin operators: $\hat{\tau}^+=d^+\hat{\eta}$, where d^+ is the fermion operator and $\hat{\eta}$ is the Majorana (real) fermion operator.

The Hamiltonian defined by Eq. (32) corresponds to the resonance-level model, which yields a multiparticle resonance at the Fermi level. A remarkable feature of the model is that the hybridization and interaction are performed via

different channels. This is its essential difference from the resonance-level model[27,28] for the single-channel Kondo scattering, in which both of them are in the same channel.

Since there is no conservation of the number of fermions in the model described by Eq. (32), the Green's function $\hat{G}_d(z)$ contains, besides the normal components $\sim\langle dd^+\rangle$, anomalous components $\sim\langle dd\rangle$ and $\sim\langle d^+d^+\rangle$. Without the interaction in the s channel ($V_z=\pi v_F$), the Green's function $\hat{G}_d^0(\varepsilon)$ obtained by Emery and Kilverson[13] at $h=0$ assumes the form

$$\hat{G}_d^0(\varepsilon)=\frac{1}{2}\,\frac{\hat{\tau}_0-\hat{\tau}_x}{\varepsilon+i\Gamma_K\ \mathrm{sign}\,\varepsilon}+\frac{1}{2}\,\frac{\hat{\tau}_0+\hat{\tau}_x}{\varepsilon+i\delta\ \mathrm{sign}\,\varepsilon},$$

$$\delta\to 0,\quad \Gamma_K\sim\frac{V_x^2}{\varepsilon_F} \tag{33}$$

The form of the Green's function corresponds to the situation when only half the impurity degrees of freedom are coupled to conduction electrons. The multiparticle resonance of width Γ_K at the Fermi level is generated by a mixed pseudospin–color (sf) model, which has a charge due to the pseudospin contribution, and for this reason the Green's function does not contain a small residue [cf. Eq. (27)].

For the tunneling problem discussed in this paper, however, the form of $G_d(z)$ for a finite interaction constant in the s channel, i.e., for $V_z-\pi v_F\neq 0$, is important. As will be shown below, there is a certain critical coupling constant at which the character of the scattering operator radically changes.

In order to obtain a solution for this case, let us use the technique that was previously applied to the well-known problem of x-ray absorption in metals.[21] First the Hamiltonian $H_{00}^s+H_s$ of Eq. (32) is diagonalized. To this end, we introduce boson operators

$$b_{sk}=k^{-1/2}\rho_s(k),\quad b_{sk}^+=k^{-1/2}\rho_s(-k),$$

where $\rho_s(k)$ are density operators:

$$\rho_s(k)=\frac{1}{N^{1/2}}\sum_{q=0}^{k_D-k}\psi_s^+(q)\psi_s(q+k),$$

$$\rho_s(-k)=\frac{1}{N^{1/2}}\sum_{q=k}^{k_D}\psi_s^+(q)\psi_s(q-k),\quad k\geqslant 0,$$

$\psi_s(k)$ are Fourier components of fields $\psi_s(x)$, and the cut-off takes place at $k_D\sim a^{-1}$. Using operators b_{sk} and b_{sk}^+, we write the Hamiltonian $H_{00}^s+H_s$ as

$$H_{00}^s+H_s=v_F\sum_{k>0}kb_{sk}^+b_{sk}+\lambda_z\left(d^+d-\frac{1}{2}\right)\sum_{k>0}k^{1/2}(b_{sk}^++b_{sk}). \tag{34}$$

Here $\lambda_z\equiv 2(V_z-\pi\varepsilon_F)/N^{-1/2}$. This Hamiltonian is diagonalized to become $v_F\Sigma_{k>0}kb_{sk}^+b_{sk}$ by the canonical transformation

$$U=\exp\left\{\lambda_z\rho_{0a}\left(d^+d-\frac{1}{2}\right)\sum_{k>0}k^{-1/2}(b_{sk}-b_{sk}^+)\right\},$$

$$\rho_{0a}\sim v_F^{-1}.$$

In this operation, the Hamiltonian H_{sf} is transformed to

$$\tilde{H}_{sf}=\frac{V_x}{\sqrt{2\pi a}}[\psi_{sf}^+(0)+\psi_{sf}(0)](\tilde{d}^+-\tilde{d})$$

$$+(h-\varepsilon_U)\left(\tilde{d}^+\tilde{d}-\frac{1}{2}\right),$$

$$\tilde{d}^+=Ud^+U^{-1}$$

$$=\exp\left\{\lambda_z\rho_{0a}\sum_{k>0}k^{-1/2}(b_{sk}-b_{sk}^+)\right\}d^+\equiv U_0d^+, \tag{35}$$

$\varepsilon_U=\lambda_z^2\rho_{0a}$ is the ''polaron shift.'' With due account of Eq. (35), the Green's function $\hat{G}_d(t)$ of the resonance level is

$$\hat{G}_d(t)=\hat{G}_d^0(t)\langle U_0^+(t)U_0(0)\rangle_D, \tag{36}$$

$U_0(t)$ is derived from $U_0(0)$ using the substitution $b_{sk}\to b_{sk}\exp(i\varepsilon_kt)$. Here $\langle\ldots\rangle_D$ denotes averaging over the states of the diagonalized Hamiltonian $H_{00}^s+H_s$, $\hat{G}_d^0(t)$ is the Fourier transform of the function given by Eq. (33). The averaging is performed in the conventional manner[21] using the relationships

$$e^{\hat{A}}e^{\hat{B}}=e^{\hat{A}+\hat{B}+(1/2)[\hat{A},\hat{B}]},\quad \langle e^{[F(b^+,b)]}\rangle=e^{(1/2)\langle F^2(b^+,b)\rangle},$$

where F is an arbitrary linear combination of boson operators. As a result, we find that, at large times $\varepsilon_Ft\gg 1$, the function in Eq. (36)

$$\hat{G}_d(t)\sim\hat{G}_d^0(t)t^{-\alpha_d}.$$

Consequently, we obtain in the energy representation the expression

$$\hat{G}_d(z)=A\left[\frac{\hat{\tau}_0-\hat{\tau}_x}{z-E_{2K}}\left(\frac{z-E_{2K}}{\varepsilon_F}\right)^{\alpha_d}+\frac{\hat{\tau}_0+\hat{\tau}_x}{z}\left(\frac{z}{\varepsilon_F}\right)^{\alpha_d}\right], \tag{37}$$

$$A=\exp(i\pi(1/2-\alpha_d))\Gamma(1-\alpha_d),$$

$\alpha_d=(\delta/\pi)^2$, δ is the phase shift, $E_{2K}=\varepsilon_F+i\Gamma_K$, $\Gamma(x)$ is the gamma-function. The cut-off parameter in the first term on the right $\sim\varepsilon_F$ since the velocity of excitations in the collective channels is v_F, as follows from the Hamiltonian (32). The interaction in the pseudospin channel has the screening nature and leads to an effective broadening of the resonance level.

The logarithmic behavior of thermodynamic quantities in Refs. 13 and 26 takes place, first, in the region where the perturbation theory with respect to α_d applies and, second, at very large times, $\Gamma_Kt\gg 1$. In this study, we will discuss only the region of power-law energy dependences.

3. By expressing Eq. (26) in terms of partial states $a_{k\gamma\sigma}$ from Eq. (17) and taking into account that only components diagonal in β in the second term of $G_{0\beta}^{\alpha}(\mathbf{kk}';z)$ contribute to the density of states, we obtain

$$\rho(\varepsilon)=\rho_0(\varepsilon)+\frac{1}{\pi}\sum_{\beta}\ \mathrm{Im\,Tr}\ \hat{G}_{d\beta}(\varepsilon)\sum_{k\gamma}\ |f_{\gamma\beta}(k;\varepsilon)|^2, \tag{38}$$

where

$$f_{\mathbf{k}\beta}(z)\equiv T^a_{\mathbf{k}d}G^a_{00\beta}(\mathbf{k};z)=\sum_\gamma f_{\gamma\beta}(k;z)K_\gamma(\Omega_\mathbf{k}).$$

The sum over k in the second term on the right can be estimated as

$$\sum_{k\gamma}|f_{\gamma\beta}(k;\varepsilon_F)|^2\sim\gamma_d\rho_{0a}.$$

With the help of Eq. (37), we derive from Eq. (38) the density of states at the Fermi level:

$$\rho_a(\varepsilon)=\rho_{0a}+A(\gamma_d\rho_{0a})\sum_{i=1,2}\frac{\sin[(1-\alpha_d)\tan^{-1}(\Gamma_i/\varepsilon)]}{\varepsilon_F^{\alpha_d}(\varepsilon^2+\Gamma_i^2)^{(1-\alpha_d)/2}},$$

$$\varepsilon>0,\quad \gamma_d\sim\gamma_0|B|^2. \tag{39}$$

The widths $\Gamma_1\equiv\delta\to 0$ and $\Gamma_2\equiv\Gamma_K$ corresponds to the ''free'' and ''bound'' contributions to the spectral function [see Eq. (33)], $A\sim 1$.

5. SCATTERING MATRIX AND ITS POLES

1. Let us reconsider the tunneling problem described by the Hamiltonian in Eq. (1). Let us take into account the additional scattering of quasi-two-dimensional electrons within the quantum well by excitations at the Fermi level in the banks due to the tunneling term H^a_{tc}. In this case, the electron states at the Fermi level and impurities are described by Green's functions (26) and (37), respectively. The scattering matrix $\mathcal{T}^{cc}_\sigma(\mathbf{k}_\perp,\mathbf{k}'_\perp;z)$ for the well electrons derives from the Green's function $G^{cc}_\sigma(\mathbf{k}_\perp,\mathbf{k}'_\perp;z)$:

$$G^{cc}_\sigma(\mathbf{k}_\perp,\mathbf{k}'_\perp;z)=\langle c_{\mathbf{k}_\perp}|\hat{I}(z-\hat{H})^{-1}|c_{\mathbf{k}'_\perp}\rangle=\delta_{\mathbf{k}_\perp,\mathbf{k}'_\perp}G_{0\mathbf{k}_\perp}(z)+G_{0\mathbf{k}_\perp}(z)\mathcal{T}^{cc}_\sigma(\mathbf{k}_\perp,\mathbf{k}'_\perp;z)G_{0\mathbf{k}'_\perp}(z),$$

$$G_{0\mathbf{k}_\perp}(z)=[z-\tilde{\varepsilon}_{\mathbf{k}_\perp}]^{-1}, \tag{40}$$

$$\mathcal{T}^{cc}_\sigma(\mathbf{k}_\perp,\mathbf{k}'_\perp;z)=\frac{T_0(z)}{1-T_0(z)J_c(z)}B(\mathbf{k}_\perp)B^*(\mathbf{k}'_\perp), \tag{41}$$

$$T_0(z)=|\Sigma_{dc}(z)|^2G_{d\sigma}(z)+\Sigma_{cc}(z), \tag{42}$$

$$\Sigma_{cc}(z)=\sum_\mathbf{k}\frac{|T^a_{\mathbf{k}c}|^2}{z-\varepsilon_{\mathbf{k}a}},\quad \Sigma_{dc}(z)=\sum_{\mathbf{k}\mu}\frac{T^{a*}_{\mathbf{k}c}T^{a\mu}_{\mathbf{k}d}}{z-\varepsilon_{\mathbf{k}a}},$$

$$J_c(z)=\sum_{\mathbf{k}_\perp}\frac{|B(\mathbf{k}_\perp)|^2}{z-\tilde{\varepsilon}_{\mathbf{k}_\perp}}.$$

For a deep level

$$B(\mathbf{k}_\perp)\approx B(\varepsilon_c)\equiv\sum_\mu B_\mu(\varepsilon_c).$$

As follows from definition (41) of the scattering matrix, the function $T_0(z)$ plays the role of an effective scattering potential for 2D electrons. The first term in $T_0(z)$ describes resonant scattering due to virtual transitions with amplitude $\Sigma_{dc}(z)$ between the 2D continuum and impurity state in the well via electron states in the banks. The second term in $T_0(z)$ describes potential scattering of quasi-two-dimensional electrons.

The integral $J_c(z)$ is a Hilbert transform of the quasi-two-dimensional density of states $\rho_c(\varepsilon)$ given by Eq. (6). For $|z-\varepsilon_c|/\gamma_0\ll 1$, this integral has a logarithmic singularity:

$$J_c(z)=\int d\varepsilon\frac{\rho_c(\varepsilon)|B(\varepsilon)|^2}{z-\varepsilon}=-\frac{1}{2}\tilde{\rho}_{0c}\ln\frac{z-\varepsilon_c}{\gamma_0},$$

$$\tilde{\rho}_{0c}=\rho_{0c}(\varepsilon_c)|B(\varepsilon_c)|^2. \tag{43}$$

Functions $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$ in the absence of interaction are Hilbert transforms of the three-dimensional density of states in the banks ''weighted'' with the tunneling integrals. In the interesting spectral range near the gap edge, these are smooth functions of energy, as compared with $J_c(z)$. In this case, the poles of the scattering matrix are determined by the logarithmic singularity in $J_c(z)$ and have the shape of exponentially narrow resonances[8] near the 2D band edge with the width

$$\gamma^{(0)}_r=\gamma_0\exp\left(-\frac{\varepsilon_c-\epsilon_{d\mu}}{\Lambda'}\right),\quad \epsilon_d=\mathrm{Re}\,\overline{E}_d,$$

$$\overline{E}_d=E_d+\Sigma_d\equiv\epsilon_d+i\gamma_d, \tag{44}$$

$$\Lambda'=\mathrm{Re}\,\Lambda=\mathrm{Re}\left[\frac{|\Sigma_{dc}|^2-(\overline{E}_d-\varepsilon_c)\Sigma_{cc}}{2(\varepsilon_c-\overline{E}_d)}\tilde{\rho}_{0c}\right],$$

$$\Sigma_d=\sum_{\mathbf{k}\mu}\frac{|T^{a\mu}_{\mathbf{k}d}|^2}{z-\varepsilon^\nu_\mathbf{k}}.$$

The values of all self-energy functions are taken at $z=\varepsilon_c$.

The functions $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$ of an interacting system in the interesting energy region near the 2D band edge can be conveniently expressed as spectral expansions of conduction electron Green's functions:

$$\Sigma_{cc}(z)=\sum_{k\gamma\sigma}\frac{|T^a_{kc\gamma}|^2f(\varepsilon_{ka})}{(z-\varepsilon_c)-(\varepsilon_{ka}-\varepsilon_c)}=\sum_\gamma|T^a_{k_Fc\gamma}|^2\int_{-\infty}^0 d\varepsilon\frac{\rho_a(\varepsilon)}{(z-\varepsilon_c)-\varepsilon}, \tag{45}$$

$\rho_a(\varepsilon)$ is given by Eq. (39), and $f(\varepsilon)$ is Fermi's distribution function. The energies are measured with respect to $\varepsilon_F\to\varepsilon_c$. This is the region where one-particle [like those described by Eq. (44)] and multiparticle resonances have strong effect on each other.

It follows from Eq. (45) that $\Sigma_{cc}(z)$ and $\Sigma_{dc}(z)$ in an interacting system are Hilbert transforms of the multiparticle density of states, consequently, they have at the Fermi level features corresponding to non-Fermi-liquid peaks in the density of states.

By substituting Eq. (39) in (45), we obtain the contribution of the resonance level E_{2K} to the self-energy functions $\Sigma_{cc,dc}(z)$:

$$\Sigma^{\rm int}_{cc,dc}(z)=A_{1,2}\gamma_d^2\frac{(z+i\Gamma_K)^{1-\alpha_d}-(z-i\Gamma_K)^{1-\alpha_d}}{\varepsilon_F^{\alpha_d}(z^2+\Gamma_K^2)^{(1-\alpha_d)}} \quad (46)$$

($|A_{1,2}|\sim 1$). In this calculation, we have used the formula

$$\tan^{-1}x=(1/2i)\ln[(1+ix)(1-ix)].$$

This formula holds only for those α_d at which the function in Eq. (37) is a power function of energy.

The logarithmic behavior of the self-energy part $J_c(z)$ means that it generates one-particle resonances in the energy band of the multiparticle resonance, which determines properties of the effective scattering potential $T_0(z)$. For this reason, low-energy poles of the scattering matrix are determined by the self-consistent equation

$$1-T_0(z)J_c(z)=0. \quad (47)$$

2. By analyzing Eqs. (43) and (46), one can easily check out that Eq. (47) has a solution of the type of an edge resonance with energy $z_r=\varepsilon_c+i\gamma_r$. This resonance is largely due to the resonance component $T_0(z)$, which is much larger than the potential one:

$$|\Sigma^{\rm int}_{cc}(z_r)|\ll|\Sigma^{\rm int}_{dc}(z_r)|^2|G_d(z_r)|.$$

This relationship is easily derived, taking into account the following estimates of the quantities in Eq. (45):

$$|\Sigma^0_{cc,dc}(z_r)|\ll|\Sigma^{\rm int}_{cc,dc}(z_r)|, \quad |\Sigma^0_{cc,dc}|\sim\gamma_d.$$

For $\gamma_r\ll\Gamma_K$

$$|\Sigma^{\rm int}_{cc,dc}(z_r)|\sim\frac{\gamma_d^2}{\Gamma_K}\left(\frac{\Gamma_K}{\varepsilon_F}\right)^{\alpha_d}\sim\varepsilon_F\left(\frac{\gamma_d}{\varepsilon_F}\right)^{2\alpha_d}, \quad \Gamma_K\sim\frac{\gamma_d^2}{\varepsilon_F}. \quad (48)$$

It follows from Eq. (46) that for $|\Gamma_K-\gamma_r|\ll\Gamma_K$ the self-energy functions have power-law singularities in the interesting range of energies.

Taking into account the foregoing, we derive from Eq. (47) the following expression for the width γ_r of the edge resonance:

$$\gamma_r=A_{r1}\varepsilon_F|B|^{2/(1-\alpha_d)}\left(\frac{\rho_{0c}}{\rho_{0a}}\right)^{1/(1-\alpha_d)}\left(\frac{\gamma_d}{\varepsilon_F}\right)^{4/(1-\alpha_d)},$$

$$|\gamma_r|\ll\Gamma_K, \quad (49)$$

$$\gamma_r\approx\Gamma_K-A_{r2}\varepsilon_F|B|^{2/3(1-\alpha_d)}\left(\frac{\rho_{0c}}{\rho_{0a}}\right)^{1/3(1-\alpha_d)}\times\left(\frac{\gamma_d}{\varepsilon_F}\right)^{4/3(1-\alpha_d)}, \quad |\Gamma_K-\gamma_r|\ll\Gamma_K, \quad (50)$$

$A_{r1},A_{r2}\sim 1$. An edge resonance exists when

$$|B|^{2(6\alpha_d-1)}<\left(\frac{W_c}{W_a}\right)\left(\frac{\gamma_0}{\varepsilon_F}\right)^{2(1-3\alpha_d)}, \quad \alpha_d>\frac{1}{6}, \quad (51)$$

i.e., only for a finite interaction constant λ_z and sufficient depth of the d level.

At all other values of the parameters, including $\lambda_z=0$, the scattering matrix has no poles near the 2D band edge.

3. Tunneling widths $\Gamma^c_a(\mathbf{p}_\perp;\varepsilon)$ in Eqs. (12) and (13) can be expressed as

$$\Gamma^c_a(\mathbf{p}_\perp;\varepsilon_F)=\frac{1}{\pi}|B(\mathbf{p}_\perp)|^2{\rm Im}\,\Sigma_{cc}(z-\varepsilon_c), \quad {\rm Re}\,z=\varepsilon_F. \quad (52)$$

In the absence of an edge resonance, one can find, using Eq. (48), that the tunneling widths are

$$\Gamma^c_a(\varepsilon_F)=A_\Gamma|B|^2\frac{\gamma_d^2}{\Gamma_K}\left(\frac{\Gamma_K}{\varepsilon_F}\right)^{\alpha_d}\sim\gamma_0\left(\frac{\varepsilon_F}{\gamma_0}\right)^{1-2\alpha_d}|B|^{2(1+2\alpha_d)},$$

$$A_\Gamma\sim 1. \quad (53)$$

In the presence of an edge resonance, the tunneling widths in Eq. (52) are cut off at γ_r and determined by the expressions

$$\Gamma^c_a(\varepsilon_F)=A_{1\Gamma}\gamma_0\left(\frac{\varepsilon_F}{\gamma_0}\right)^{1-2\alpha_d}|B|^{2(1+2\alpha_d)}, \quad \gamma_r\ll\Gamma_K, \quad (54)$$

$$\Gamma^c_a(\varepsilon_F)=A_{2\Gamma}\gamma_0\left(\frac{\varepsilon_F}{\gamma_0}\right)^{1/3}|B|^{2/3}), \quad |\gamma_r-\Gamma_K|\ll\Gamma_K, \quad (55)$$

$A_{1\Gamma},A_{2\Gamma}\sim 1$. Note that for all widths in Eqs. (53)–(55) the condition $\Gamma^c_a\ll\gamma_0$ holds.

All the cases discussed above satisfy the condition $\Gamma^a_c(z_r)\ll|T_0(z_r)|$. This condition means that the characteristic tunneling times $\tau^{\rm int}_t\sim\Gamma^{a-1}_c(z_r)$ in an interacting system are much shorter than the characteristic scattering times $\tau^{\rm int}_{sc}\sim|T_0(z_r)|^{-1}$, so that the lifetime of an electron in the quantum well is large enough to form resonances due to scattering caused by the Hamiltonian H_t. Thus, the existence of multiparticle edge resonances is provided by the condition

$$\tau^{\rm int}_t\gg\tau^{\rm int}_{sc}.$$

Note also that, if condition (51) holds, edge resonances exist as long as $|\varepsilon_F-\varepsilon_c|<\Gamma_K$. The dominant contribution to the tunneling probability (11) is due to the resonance term in the Green's function $G_{cc}(z)$ [see Eq. (40)]. This corresponds to the Fermi-liquid (or, which is the same, resonance) tunneling mode.

There are no edge resonances in the case of shallow impurities, when condition (51) fails (although $|\varepsilon_F-\varepsilon_c|<\Gamma_K$), or at $|\varepsilon_F-\varepsilon_c|>\Gamma_K$ and arbitrary energies of the impurity level. In this case, the dominant contribution to the tunneling probability comes from the non-Fermi-liquid resonance at the Fermi level. This resonance controls the tunneling widths in Eq. (11) [as follows from their definition (52)]. The corresponding tunneling mode will be dubbed non-Fermi-liquid.

One can see that the additional scattering of electrons from the neighborhood of the two-dimensional band edge by non-Fermi-liquid excitations about the Fermi level due to the tunneling Hamiltonian $H^a_{\rm t}$ generates a Fermi-liquid resonance at the 2D band edge since it corresponds to a simple pole of the Green's function. If the scattering due to tunneling is not taken into account, the non-Fermi-liquid state of electrons in the banks corresponds to power-law singularities in Green's functions (26) and (37), density of states $\rho_a(\varepsilon)$, and, hence, in the tunneling widths.

Thus, when the Fermi level in the banks approaches the 2D band edge, a cross-over between the non-Fermi-liquid

and Fermi-liquid tunneling regimes is possible. The crossover conditions are identical to the conditions when solutions of Eq. (47) exist.

6. TUNNELING TRANSPARENCY

1. The tunneling transparency is given by the expression

$$\sigma(\varepsilon_F)=2e^2\int dE\,\delta(E-\varepsilon_F)\sum_{\mathbf{k}_\perp,\mathbf{k}'_\perp} W(\mathbf{k}_\perp,\mathbf{k}'_\perp;E).$$

It contains the non-Fermi-liquid and resonance contributions:

$$\sigma(\varepsilon_F)=\sigma_0(\varepsilon_F)+\sigma_r(\varepsilon_F).$$

The non-Fermi-liquid contribution $\sigma_0(\mu)$ is determined by the equation

$$\sigma_0(\varepsilon_F)=\frac{e^2}{\pi}\frac{\Gamma_L(\varepsilon_F)\Gamma_R(\varepsilon_F)}{\Gamma_L(\varepsilon_F)+\Gamma_R(\varepsilon_F)}\rho_c(\varepsilon_F).$$

By expressing $\Gamma^c_{L,R}$ in terms of Γ^a_c using Eq. (13), we have in the case of equal widths $\Gamma^c_{0L,0R}$

$$\sigma_0(\varepsilon_F)=\frac{e^2}{4\pi}\Gamma^a_c(\varepsilon_F)\rho_c(\varepsilon_F). \tag{56}$$

The tunneling widths in Eq. (53) are larger than these parameters in the absence of interaction, which are approximately equal to γ_d. This means that the non-Fermi-liquid contribution to the transparency is larger than the nonresonant contribution in the absence of interaction.[8]

By substituting Eq. (40) in (11), we find that the general expression for the resonance contribution to transparency has the form (at $\Gamma^c_{0L}=\Gamma^c_{0R}$)

$$\sigma_r(\varepsilon_F)=\frac{e^2}{4\pi}\Gamma^{c2}_a(z_r)\frac{|\tilde{T}_0(z_r)|^2}{|D'(z_r)|^2[(\epsilon_r-\varepsilon_F)^2+\gamma_r^2]}I^2(z_r). \tag{57}$$

Here

$$D(z)=1-T_0(z)J_c(z),\quad D'(z)\equiv(d/dz)D(z),$$

$$I(z_r)=\sum_{\mathbf{k}_\perp}|B(\mathbf{k}_\perp)|^2|G_{0\mathbf{k}_\perp}(z_r)|^2\approx\tilde{\rho}_{0c}\frac{1}{2\gamma_r}$$

when $\gamma_0\gg\gamma_r\gg|\varepsilon_F-\varepsilon_c|$.

One can easily show that, under these conditions,

$$D'(z_r)\approx\frac{T_0(\gamma_r)\tilde{\rho}_{0c}}{\gamma_r}F_r,$$

where F_r is the function of the parameters of order of unity. The factor $I^2(z_r)$ cancels out the small factor γ_r^2 in the numerator on the right of Eq. (57) due to the residue at the scattering matrix pole at $z=z_r$.

2. By substituting the expressions for the tunneling Γ^a_c and resonance γ_r widths, we obtain the maximum contribution of edge resonances to the transparency at $\varepsilon_r=\varepsilon_F$:

$$\sigma_r^{\max}(\varepsilon_F)=F_{1r}\frac{e^2}{4\pi}\left(\frac{\Gamma^c_a}{\gamma_r}\right)^2\equiv\frac{e^2}{4\pi}S(\varepsilon_F), \tag{58}$$

where $S(\varepsilon_F)$, which is the enhancement factor due to tunneling, can be written as

$$S(\varepsilon_F)=F_{2r}\left(\frac{\gamma_0}{\gamma_r}\right)^2\left(\frac{\varepsilon_F}{\gamma_0}\right)^{2(1-2\alpha_d)}|B|^{4(1+2\alpha_d)}\gg1,\quad \gamma_r\ll\Gamma_K, \tag{59}$$

$$S(\varepsilon_F)=F_{3r}\left(\frac{\gamma_0}{\gamma_r}\right)^2\left(\frac{\varepsilon_F}{\gamma_0}\right)^{2/3}|B|^{4/3}\gg1,\quad |\gamma_r-\Gamma_K|\ll\Gamma_K, \tag{60}$$

$F_{2r},F_{3r}\sim1$. The width γ_r is determined by Eqs. (49) and (50). In both limiting cases, the following inequality holds:

$$\sigma_r^{\max}(\varepsilon_F)\gg\sigma_0^{\max}(\varepsilon_F).$$

In the absence of interaction, the enhancement factor due to exponentially narrow resonances at the 2D band edge is[8]

$$S_0(\varepsilon_F)\sim\left(\frac{\gamma_0}{\gamma_r^{(0)}}\right)^2$$

where the width $\gamma_r^{(0)}$ is determined by Eq. (44). It turns out that this is the largest enhancement factor due to an elementary tunneling event among all results that have been known thus far. In fact, the enhancement factor due to tunneling via a quasi-localized level in single-channel Kondo scattering[1] is about unity. When the Coulomb interaction between electrons at impurity levels and in the banks is taken into account,[4] the enhancement factor is of order of $(\varepsilon_F/\gamma_0)^\alpha$.

Equations (56), (58)–(60) determine the transparency enhancement due to tunneling via the two-dimensional continuum with due account of interactions generating non-Fermi-liquid excitations at the Fermi level. It is clear that, in all the cases discussed above, the enhancement factor satisfies the conditions

$$S_0(\varepsilon_F)\gg S(\varepsilon_F)\gg1. \tag{61}$$

The enhancement factor is lower than in the case of one-particle resonances for two reasons: (1) multiparticle edge resonances exist only when impurity levels are sufficinetly deep (at least at $|B|\ll1$), whereas this limitation is not imposed on one-particle resonances, and the peak enhancement factor takes place in the case of a resonance due to relatively shallow levels with $|B|\sim1$; (2) all multiparticle resonances are broader than one-particle [compare Eqs. (49) and (50) with Eq. (44)], and this effect is stronger than that of the larger tunneling width.

Here we must emphasize that the anomalous transparency enhancement with $S(\varepsilon_F)\gg1$ in the quantum structure with a two-dimensional continuum is caused, in addition to the small edge resonance width, by its proximity to the 2D band edge, which yields the additional factor $I^2(z_r)$.

Equations (56), (58)–(60) determine the transparency enhancement due to an elementary (microscopic) tunneling event. The total contribution to the quantum well transparency is known to be $\sigma_{im}=c_{im}\sigma_r$, where c_{im} is the impurity concentration. As follows from our equations, $\sigma_{im}\gg\sigma_0$ holds for reasonable values of impurity concentration.

7. CONCLUDING REMARKS

The results reported in this paper permit a direct experimental observation of a non-Fermi-liquid state using features on current–voltage characteristics (CVC) at zero tempera-

ture. It seems feasible to detect changes in these features as tunneling structure parameters are varied. The most essential property of such structures may be the cross-over between the non-Fermi-liquid and Fermi-liquid tunneling modes as the Fermi level approaches the 2D band edge. The impurity level energy and interaction constant λ_z are determined by Eq. (51). While $|\varepsilon_F-\varepsilon_c|>\Gamma_K$ holds, the tunneling transparency and CVC are controlled by the non-Fermi-liquid resonance at the Fermi level. The transparency in this case is given by Eq. (56) with tunneling widths Γ_c^a from Eq. (53). When the Fermi level and band edge are so close that $|\varepsilon_F-\varepsilon_c|<\Gamma_K$, the transparency and CVC are controlled by Fermi-liquid edge resonances, and the transparency is determined by Eqs. (58) and (59), which contain the characteristic ''Fermi-liquid factor'' $(\gamma_0/\gamma_r)^2$ (compare with the expression for S_0).

But if the Fermi level is fixed near the band edge so that $|\varepsilon_F-\varepsilon_c|<\Gamma_K$, the cross-over between the Fermi-liquid and non-Fermi-liquid tunneling modes takes place when either the impurity level is ''refined'' or the interaction constant λ_z is reduced. In the latter case, changes in the transparency and CVC are caused by changes in the thickness (or height) of barriers around the quantum well. In fact, as follows from the definition of the interaction matrix elements in Eq. (16), they are proportional to γ_0^2, i.e., they are exponential functions of the barrier thickness. Therefore, insignificant changes in the barrier thickness can lead to a cross-over between the non-Fermi-liquid and Fermi-liquid tunneling modes.

Finally, let us make two comments. The two-channel model of GaAlAs/GaAs/GaAlAs heterostructures discussed in this paper applies, most likely, to the case of light impurities, such as V^{2+}, in the GaAs inner layer. As was shown previously,[10] the level closest to the two-dimensional band edge in this system is the e_g doublet. It follows from our results, however, that the equations can be also applied qualitatively to the multichannel tunneling $(n>2)$.

The results concerning the two-channel Kondo model reported above are valid in the case when the magnitude h of Jahn–Teller distortions in an impurity center in Hamiltonian (23) is smaller than all energy parameters of the problem. It is known, however, that the non-Fermi-liquid state is stable against such perturbations, which lift the pseudospin degeneracy of the impurity state.[17,31] It seems that the formation of electron states at the Fermi level and population of orbital states of the impurity center should be treated in a self-consistent manner. This problem deserves a separate consideration.

I am grateful to L. A. Maksimov for reading the manuscript of the paper and critical remarks. The work was supported by the Russian Fund for Fundamental Research.

*)E-mail: manakova@kurm.polyn.kiae.su

[1] L. I. Glazman and M. E. Raĭkh, JETP Lett. **48**, 445 (1988).
[2] T. K. Ng and P. A. Lee, Phys. Rev. Lett. **61**, 1768 (1988); T. K. Ng, Phys. Rev. Lett. **70**, 3635 (1993).
[3] Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **66**, 3048 (1991).
[4] K. A. Matveev and A. I. Larkin, Phys. Rev. B **46**, 15337 (1992).
[5] N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, Phys. Rev. B **40**, 11834 (1989).
[6] L. I. Glazman and R. I. Shekhter, Zh. Éksp. Teor. Fiz. **94**, 292, (1988) [Sov. Phys. JETP **67**, 163 (1988)].
[7] K. A. Kikoin and L. A. Manakova, in *Proceedings of XXIII Int. Conf. Semicond.*, Berlin (1996), p. 894.
[8] K. A. Kikoin and L. A. Manakova, Phys. Rev. B **57**, 4863 (1997).
[9] K. A. Kikoin and L. A. Manakova, JETP Lett. **65**, 484 (1997).
[10] K. A. Kikoin and L. A. Manakova, Fiz. Tekh. Poluprovod. **29**, 291 (1995) [Semiconductors **29**, 145 (1995)].
[11] P. Nozieres and A. Blandin, J. de Phys. **41**, 193 (1980).
[12] N. Andrei and C. Destri, Phys. Rev. Lett. **52**, 364 (1984).
[13] V. J. Emery and S. Kivelson, Phys. Rev. B **46**, 10812 (1992).
[14] M. H. Hettler, J. Kroha, and S. Hershfield, Phys. Rev. Lett. **73**, 1967 (1994); D. C. Ralph *et al.*, Phys. Rev. Lett. **72**, 1064 (1994).
[15] D. C. Ralph and B. A. Buhrman, Phys. Rev. Lett. **72**, 3401 (1994).
[16] D. C. Mattis and F. H. Lieb, J. Math. Phys. **6**, 304 (1965).
[17] D. J. Cox, Phys. Rev. Lett. **59**, 1240 (1987); D. J. Cox *et al.*, Phys. Rev. Lett. **62**, 2188 (1989).
[18] K. Vladar and A. Zawadowski, Phys. Rev. B **28**, 1564 (1983).
[19] A. Murumatsu and F. Guinea, Phys. Rev. Lett. **57**, 2337 (1986).
[20] K. A. Kikoin and V. N. Fleurov, *Transition Metal Impurities in Semiconductors*, World Scientific, Singapore (1994).
[21] K. D. Schotte and U. Schotte, Phys. Rev. **182**, 479 (1969).
[22] P. Nozieres and C. T. de Dominicis, Phys. Rev. **178**, 1097 (1969).
[23] N. Kawakami and A. Okiji, Phys. Rev. B **42**, 2383 (1990).
[24] A. C. Hewson, *The Kondo Problem to Heavy Fermions*, Cambridge University Press, Cambridge (1993).
[25] C. Lacroix, J. Phys. F **11**, 2389 (1981).
[26] A. M. Sengupta and A. Georges, Phys. Rev. B **49**, 1020 (1994).
[27] P. Schlottmann, J. de Phys. **39**, 1486 (1978).
[28] P. B. Vigman and A. M. Finkel'shteĭn, Zh. Éksp. Teor. Fiz. **75**, 204 (1978) [Sov. Phys. JETP **48**, 102 (1978)].
[29] G. D. Mahan, Phys. Rev. **163**, 612 (1967).
[30] J. R. Schrieffer, J. Appl. Phys. **38**, 1143 (1967).
[31] P. D. Sacramento and P. Schlottmann, Phys. Rev. B **43**, 13294 (1991).

Translation provided by the Russian Editorial office.

Stability problem in nonlinear wave propagation

Yu. N. Ovchinnikov

L. D. Landau Institute for Theoretical Physics, 117940 Moscow, Russia
(Submitted 12 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 1487–1499 (October 1998)

An explicit expression for the excitation spectrum of the stationary solutions of a nonlinear wave equation is obtained. It is found that all branches of many-valued solutions of a nonlinear wave equation between the $(2K+1,2K+2)$ turning points (branch points in the complex plane of the nonlinearity parameter) are unstable. Some parts of branches between the $(2K,2K+1)$ turning points are also unstable. The instability of the latter is related to the possibility that pairs of complex conjugate eigenvalues cross the real axis in the κ plane. © *1998 American Institute of Physics.* [S1063-7761(98)02210-0]

1. INTRODUCTION

In the Refs. 1–5 it was found that for a transverse electromagnetic wave (Fig. 1) propagating in a nonlinear medium, many states are possible for a given amplitude of the incident wave. The nonlinear medium was taken in the form of a slab. The reflection and transmission coefficients in this case are functionals of the state. In a linear medium, there exists only one state for a given incident wave, and this state is stable against small perturbations. In a nonlinear medium, some of the states are stable and some are unstable against small perturbations. This property is very important for practical purposes. In this paper we study the problem of stability for all states. The main result is as follows: all solutions of the nonlinear problem can be parametrized by one parameter ρ_1, which is equal to the transparency of the nonlinear medium. This parameter ρ_1 is a multivalued function of the effective nonlinearity μ. The graph of $\rho_1=\rho_1(\mu)$ has turning points (see Fig. 2). All branches between the $2K+1$ and $2K+2$ $(K=0,1,2,...)$ turning points are unstable, and some parts of branches between the $2K$ and $2K+1$ turning points are unstable against small perturbations. In thermodynamics it is also possible to find many solutions for given external conditions. Some of them are stable, some are not. But there always exists a state that yields the absolute minimum of the free energy. All other stable states can be considered metastable. Only quantum or thermal fluctuations can lead to transitions between different metastable states. In a dynamical problem, on the other hand, there is no general principle that distinguishes one solution from all other local stable states. Which state will be realized after a transition from an unstable state is still an unsolved problem.

2. FORMULATION OF STABILITY PROBLEM

We investigate the stability of solutions of the wave equation

$$\frac{\partial^2 E}{\partial x^2}-\frac{1}{2c^2}\frac{\partial^2}{\partial t^2}\frac{\partial}{\partial E}(\epsilon E^2)=0, \quad \epsilon=n^2(1+\mu_0 E^2) \tag{1}$$

in a slab of length d. In Eq. (1), n is the refractive index and c the speed of light. In the general case, μ_0 is a function of position x. In real materials $\mu_0 \sim E_0^{-2}$, where E_0 is the electric field on the atomic scale. Hence, in real materials $\mu_0|E|^2 \ll 1$. Only such a case will be considered below. We investigate the stability of solutions of Eq. (1), that take the form

$$E_0=\mathrm{Re}(e^{-i\omega t}\psi_0(x)) \tag{2}$$

for an incident wave given by

$$A\ \exp\left(\frac{i\omega}{c}\,x\right). \tag{3}$$

It is convenient to use the dimensionless variables

$$y=\frac{\omega n}{c}\,x, \quad \mu=\frac{3}{4}\,\mu_0|A|^2, \quad \psi_0 \to A\psi, \quad b=\frac{\omega n}{c}\,d. \tag{4}$$

Then Eq. (1) takes the form[5] (inside the slab)

$$(1+2\mu|\psi|^2)\psi+\frac{\partial^2\psi}{\partial y^2}=0, \tag{5}$$

with boundary conditions

$$\psi_{(0)}=1+R, \quad \psi'_{(0)}=\frac{i}{n}\,(1-R),$$

$$\psi_{(b)}=Te^{ib/n}, \quad \psi'_{(b)}=\frac{i}{n}\,Te^{ib/n}. \tag{6}$$

The unknown coefficients R and T can be eliminated from Eq. (5), and we obtain

$$\psi_{(0)}-in\psi'_{(0)}=2, \quad \psi_{(b)}+in\psi'_{(b)}=0. \tag{7}$$

We represent ψ in the form[5]

$$\psi=\sqrt{\rho}e^{i\alpha}. \tag{8}$$

The value of the function $\rho_{(y)}$ at point b $(\rho_1=\rho(b))$ completely determines the modulus of the transmission and reflection coefficients:

$$|T|^2=\rho_1, \quad |R|^2=1-\rho_1. \tag{9}$$

Equation (1) for the stationary solutions of type (7) can be reduced to a function ρ only,[5] and its solutions in the

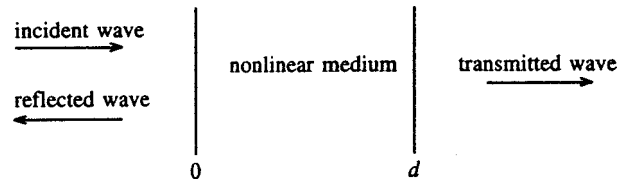

FIG. 1. Geometry of wave propagation.

general case are elliptic functions. The condition $|\mu E^2| \ll 1$ drastically simplifies the investigation of solutions of type (8).[5]

The qualitative dependence of ρ_1 on the effective nonlinearity is shown in Fig. 2 (see also Eq. (48)).

We are able now to formulate the stability problem for solutions of type (8). We seek a solution of Eq. (1) in the form

$$E = E_0 + \tilde{E}, \tag{10}$$

where E_0 is given by Eqs. (2), (8), and

$$E_1 = \mathrm{Re}[e^{-i\omega t}(E_1 e^{\kappa \omega t} + E_2 e^{\kappa^* \omega t})]. \tag{11}$$

The boundary conditions for the function $\tilde{E}$ correspond to the outgoing wave:

$$\left(\frac{E_1'}{E_1}\right)_0 = -\frac{i}{n}(1+i\kappa), \quad \left(\frac{E_1'}{E_1}\right)_b = \frac{i}{n}(1+i\kappa),$$

$$\left(\frac{E_2'}{E_2}\right)_0 = -\frac{i}{n}(1+i\kappa^*), \quad \left(\frac{E_2'}{E_2}\right)_b = \frac{i}{n}(1+i\kappa^*). \tag{12}$$

Inserting Eq. (11) into Eq. (1), we obtain

$$\frac{\partial^2 E_1}{\partial y^2} + (1+i\kappa)^2\{(1+2\mu|\psi|^2)E_1 + 2\mu(|\psi|^2 E_1 + \psi^2 E_2^*)\} = 0,$$

$$\frac{\partial^2 E_2}{\partial y^2} + (1+i\kappa^*)^2\{(1+2\mu|\psi|^2)E_2 + 2\mu(|\psi|^2 E_2 + \psi^2 E_1^*)\} = 0. \tag{13}$$

Note that ψ is the solution of Eq. (5) with boundary conditions given by (6). The system of equations (13) with boundary conditions (12) can be considered an eigenvalue problem for the symmetric operator $\hat{L}$. The explicit form of operator $\hat{L}$ is given by Eq. (13). We easily obtain for the first line

$$\hat{L}_{11} = \frac{\partial^2}{\partial y^2} + (1+4\mu|\psi|^2), \quad \hat{L}_{12} = 0,$$

$$\hat{L}_{13} = 2\mu\cos(2\alpha)|\psi|^2, \quad \hat{L}_{14} = 2\mu\sin(2\alpha)|\psi|^2. \tag{14}$$

As a trivial fact, we note that if κ is an eigenvalue, then so is κ^*.

3. STABILITY IN LINEAR MEDIUM

Consider as a starting point the linear case ($\mu = 0$). In the linear case, the system of equations (13) decouples into two independent subsystems for the quantities E_1 and E_2. Solution of the first of Eqs. (13) yields

$$E_1 = A_1\cos((b-y)(1+i\kappa)) + A_2\sin((b-y)(1+i\kappa)). \tag{15}$$

From the boundary conditions (12), we obtain

$$A_2 = -\frac{i}{n}A_1, \quad \tan(b(1+i\kappa)) = -\frac{2in}{n^2+1}. \tag{16}$$

Solution of the second of Eq. (16) yields

$$\kappa = -\frac{1}{b}\ln\frac{n+1}{n-1} + i\left(1 - \frac{\pi N}{b}\right), \quad N = 0, \pm 1, \pm 2... \tag{17}$$

The second of Eqs. (13) yields

$$\kappa = -\frac{1}{b}\ln\frac{n+1}{n-1} - i\left(1 - \frac{\pi N}{b}\right); \quad N = 0, \pm 1, \pm 2... \tag{18}$$

Hence the full spectrum of the operator $\hat{L}$ in the linear case is given by Eqs. (17) and (18). The real part of all eigenvalues is negative, hence the unique solution of (5) is stable in the linear case.

If the length b of the slab is given by

$$b = \pi N_0 + \varepsilon, \quad |\varepsilon| \ll 1, \tag{19}$$

then for $n \gg 1$, two eigenvalues of the linear problem are closer to zero than all others:

$$\kappa_{1,2} = -\frac{1}{b}\ln\left(\frac{n+1}{n-1}\right) \pm \frac{i\varepsilon}{b}. \tag{20}$$

As we will see below, these two eigenvalues play a special role in the nonlinear problem.

4. STABILITY PROBLEM IN NONLINEAR MEDIUM

Points where $\partial\mu/\partial\rho_1 = 0$ are turning points. It is easy to show from Eqs. (5) and (7) that the function $\partial\psi/\partial\rho_1$ is a solution of the system of equations (13) for $\kappa = 0$. This

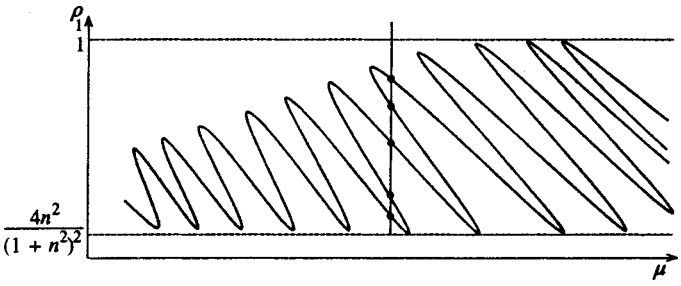

FIG. 2. Dependence of ρ_1 on μ; the number of solutions for given μ.

means that of the two branches entering at the turning point, one is unstable, because in the vicinity of the turning point $\kappa \sim \delta\rho_1$. We prove this statement below.

We use the following simple expression[5] for ρ and α:

$$\rho=\rho_1\left\{\frac{1}{n^2}+\left(1-\frac{1}{n^2}\right)\cos^2\left(\frac{(y-b)}{1-3\mu\rho_1(1+1/n^2)/4}\right)\right\}, \tag{21}$$

$$\alpha=\alpha_{(0)}+\frac{1}{n}\int_0^y \frac{dy_1}{1/n^2+(1-1/n^2)\cos^2\left(\frac{b-y_1}{1-3\mu(1+1/n^2)\rho_1/4}\right)}$$

$$=\alpha_{(0)}+\left(1-\frac{3\mu}{4}\left(1+\frac{1}{n^2}\right)\rho_1\right)\times\left[\arctan\left\{\frac{1}{n}\tan\left(\frac{y-b}{1-3\mu(1+1/n^2)\rho_1/4}\right)\right\}+\arctan\left\{\frac{1}{n}\tan\left(\frac{b}{1-3\mu(1+1/n^2)\rho_1/4}\right)\right\}\right].$$

In Eq. (21), $\alpha_{(0)}$ is the phase of ψ at the point $y=0$:

$$\alpha_{(0)}=\arctan\left[\frac{n\partial\rho/\partial y}{2(\rho_1+\rho)}\right]_{y=0}. \tag{22}$$

Equations (5) and (21) enable us to represent the function ψ in a form that yields an explicit physical picture of nonlinear wave propagation. Inserting the expression (21) into Eq. (5), we find that ψ can be represented in the form

$$\psi=e^{i\lambda_1(b-y)}(Ae^{i\gamma(b-y)}+Be^{-i\gamma(b-y)})+e^{i\lambda_2(b-y)}\times(Ce^{i\gamma(b-y)}+De^{-i(\gamma-\lambda)(b-y)}), \tag{23}$$

where

$$\gamma=\frac{1}{1-3\mu\rho_1(1+1/n^2)/4}.$$

Equation (5) leads to the following expression for the quantities $\lambda_{1,2}$:

$$\lambda_{1,2}=\pm\lambda, \tag{24}$$

where $\lambda=\mu\rho_1/2n$. From the same Eq. (5), we also obtain two equations for the quantities *A,B* and *C,D*. As a result, we have

$$\psi=B\left[\left(1+\frac{1}{n}\right)\exp(-i(\gamma-\lambda)(b-y))+\left(1-\frac{1}{n}\right)\times\exp(i(\gamma+\lambda)(b-y))\right]+C\left[\left(1+\frac{1}{n}\right)\exp(i(\gamma-\lambda)\times(b-y))+\left(1-\frac{1}{n}\right)\exp(-i(\gamma+\lambda)(b-y))\right]. \tag{25}$$

Now from the boundary condition (7) at $y=b$, we obtain one equation for B and C

$$C=B\,\frac{\mu\rho_1}{4}\left(\frac{1}{2}+\frac{3}{2n^2}\right). \tag{26}$$

Finally, recalling that $|\psi_{(b)}|=\sqrt{\rho_1}$, we have

$$\psi=\frac{\sqrt{\rho_1}}{2}\left[\left(1+\frac{1}{n}\right)\exp(-i\gamma(b-y))+\left(1-\frac{1}{n}\right)\times\exp(i\gamma(b-y))\right]\exp(i\tilde{\varphi}_c+i\lambda(b-y)), \tag{27}$$

where $\tilde{\varphi}_c$ is some constant, that is simply related to the phase α_0 given by Eq. (22).

Equation (27) means that the speed of light is slightly different for waves moving in opposite directions; and this is one of the main effects of nonlinearity.

Below, we use expressions (21) and (27) to solve the stability problem.

5. DERIVATION OF THE EQUATION OF THE SPECTRUM κ

The spectral points κ, defined by Eqs. (13) with boundary conditions (12), are the roots of some equation that is an analytic function of κ and the associated parameter ρ_1. Near the turning points, there exists a region where the two eigenvalues are real. This enables us to consider only real-valued κ. The equations for κ obtained under the assumption that κ is real can be analytically continued into the complex κ plane. The roots of this equation also yield the complex values of κ of the initial eigenvalue problem, given by Eqs. (12) and (13).

For real values of κ, we obtain the following eigenvalue problem:

$$\frac{\partial^2 E_1}{\partial y^2}+(1+i\kappa)^2\{(1+2\mu|\psi|^2)E_1+2\mu(|\psi|^2E_1+\psi^2E_1^*)\}=0, \tag{28}$$

$$\frac{E'_{1(0)}}{E_{1(0)}}=-\frac{i}{n}(1+i\kappa),\quad \frac{E'_{1(b)}}{E_{1(b)}}=\frac{i}{n}(1+i\kappa), \tag{29}$$

where $|\psi|$ is given by Eq. (21) and ψ is given by Eq. (27). Equation (28) has four linearly independent solutions. We seek them in the form

$$\tilde{E}_1=Ae^{i(\gamma+\beta)(b-y)}+Be^{i(\gamma-\beta^*+\mu\rho_1/n)(b-y)}+Ce^{-i(\gamma-\beta)(b-y)}+De^{-i(\gamma+\beta^*-\mu\rho_1/n)(b-y)}, \tag{30}$$

where *A*, *B*, *C*, *D* and β are complex numbers. We omitted in expression (30) higher harmonics with small amplitudes of order $O(A\mu)$. We also put

$$E_1=e^{i\tilde{\varphi}_c}\tilde{E}_1. \tag{31}$$

Inserting expression (30) for quantity $\overline{E_1}$ in to Eq. (28), we obtain the following system of equations for the *A,B,C,D*:

$$\hat{A}\begin{pmatrix}A\\B^*\\C\\D^*\end{pmatrix}=0, \tag{32}$$

where the matrix $\hat{A}$ is

$$\hat{A}=\begin{pmatrix} \frac{1}{4}\left(1+\frac{1}{n^2}\right)+\frac{i\kappa-\beta}{\mu\rho_1} & \frac{1}{4}\left(1-\frac{1}{n}\right)^2 & \frac{1}{2}\left(1-\frac{1}{n^2}\right) & \frac{1}{2}\left(1-\frac{1}{n^2}\right) \\ \frac{1}{2}\left(1-\frac{1}{n^2}\right) & \frac{1}{2}\left(1-\frac{1}{n^2}\right) & \frac{1}{4}\left(1+\frac{1}{n^2}\right)+\frac{i\kappa+\beta}{\mu\rho_1} & \frac{1}{4}\left(1+\frac{1}{n}\right)^2 \\ \frac{1}{2n}+\frac{i\kappa-\beta}{\mu\rho_1} & \frac{1}{2n}+\frac{i\kappa-\beta}{\mu\rho_1} & 0 & 0 \\ 0 & 0 & -\frac{1}{2n}+\frac{i\kappa+\beta}{\mu\rho_1} & -\frac{1}{2n}+\frac{i\kappa+\beta}{\mu\rho_1} \end{pmatrix}. \tag{33}$$

The values of β are solutions

$$\det \hat{A}=0. \tag{34}$$

A simple calculation of $\det \hat{A}$ yields

$$\left(\frac{1}{2n}+\frac{i\kappa-\beta}{\mu\rho_1}\right)^2\left(-\frac{1}{2n}+\frac{i\kappa+\beta}{\mu\rho_1}\right)^2=0. \tag{35}$$

We see that each eigenvalue β is doubly degenerate. As a result, we obtain only two linearly independent solutions of Eq. (28) of the form (30):

$$\tilde{E}_1=i[A\ \exp(i(\gamma+\mu\rho_1/2n+i\kappa)(b-y)) + C\ \exp(-i(\gamma-\mu\rho_1/2n-i\kappa)(b-y))], \tag{36}$$

where A and C are real numbers.

Two other solutions of the system of equations (28) can be found in the form

$$\tilde{E}_1=(iy+A_1)\exp(i(\gamma+\mu\rho_1/2n+i\kappa)(b-y)) + C_1\ \exp(-i(\gamma-\mu\rho_1/2n-i\kappa)(b-y)) \tag{37}$$

for $\beta=\mu\rho_1/2n+i\kappa$, and

$$\tilde{E}_1=(iy+C_2)\exp(-i(\gamma-\mu\rho_1/2n+i\kappa)(b-y)) + A_2\ \exp(i(\gamma+\mu\rho_1/2n-i\kappa)(b-y)) \tag{38}$$

for $\beta=\mu\rho_1/2n-i\kappa$.

Inserting expressions (37) and (38) into Eq. (28), we obtain the following system of equations for the coefficients $A_{1,2}$ and $C_{1,2}$:

$$(A_1+A_1^*)\ \frac{\mu\rho_1}{4}\left(1-\frac{1}{n}\right)^2+\frac{\mu\rho_1}{n} \times\left(1-\frac{1}{n^2}\right)(C_1+C_1^*)+1=0,$$

$$(C_1+C_1^*)\ \frac{\mu\rho_1}{4}\left(1+\frac{1}{n^2}\right)+(A_1+A_1^*) \times\frac{\mu\rho_1}{2}\left(1-\frac{1}{n^2}\right)+2i\kappa C_1=0, \tag{39}$$

and

$$(C_2+C_2^*)\ \frac{\mu\rho_1}{4}\left(1+\frac{1}{n^2}\right)+\frac{\mu\rho_1}{2} \times\left(1-\frac{1}{n^2}\right)(A_2+A_2^*)-1=0,$$

$$(A_2+A_2^*)\ \frac{\mu\rho_1}{4}\left(1-\frac{1}{n}\right)^2+\frac{\mu\rho_1}{2} \times\left(1-\frac{1}{n^2}\right)(C_2+C_2^*)+2i\kappa A_2=0. \tag{40}$$

The solutions of the system of equations (39), (40) are

$$A_1=-\frac{2}{\mu\rho_1(1-1/n)^2},\quad C_1=-\frac{i}{\kappa}\frac{1+1/n}{1-1/n}, \tag{41}$$

$$A_2=\frac{i}{\kappa}\frac{1-1/n}{1+1/n},\quad C_2=\frac{2}{\mu\rho_1(1+1/n)^2}. \tag{42}$$

Hence, the general solution of the system of Eqs. (28) is

$$\begin{aligned}\tilde{E}_1=i\Bigg(&A\ \exp\left(i\left(\gamma+\frac{\mu\rho_1}{2n}+i\kappa\right)(b-y)\right)+B \\ &\times\exp\left(-i\left(\gamma-\frac{\mu\rho_1}{2n}+i\kappa\right)(b-y)\right)\Bigg)+Ce^{-\kappa(b-y)} \\ &\times\Bigg[\left(-iy+\frac{2}{\mu\rho_1(1-1/n)^2}\right)\exp\left(i\left(\gamma+\frac{\mu\rho_1}{2n}\right)(b-y)\right) \\ &+\frac{i}{\kappa}\frac{1+1/n}{1-1/n}\exp\left(-i\left(\gamma-\frac{\mu\rho_1}{2n}\right)(b-y)\right)\Bigg]+De^{\kappa(b-y)} \\ &\times\Bigg[\left(iy+\frac{2}{\mu\rho_1(1+1/n)^2}\right)\exp\left(-i\left(\gamma-\frac{\mu\rho_1}{2n}\right)(b-y)\right) \\ &+\frac{i}{\kappa}\frac{1-1/n}{1+1/n}\exp\left(i\left(\gamma+\frac{\mu\rho_1}{2n}\right)(b-y)\right)\Bigg].\end{aligned} \tag{43}$$

In Eq. (41), A, B, C and D are real numbers.

The boundary conditions (29) at $y=b$ yield the first pair of equations for the coefficients A, B, C, D. In the leading approximation for the parameter μ, we have

$$C=D\left(\frac{1-1/n}{1+1/n}\right)^3,$$

$$A\left(\frac{1+1/n}{1-1/n}\right)-B-C\left(\frac{1+1/n}{1-1/n}\right)\left(b+\frac{1}{\kappa}\right)+D\left(-b+\frac{1}{\kappa}\right)=0. \tag{44}$$

The second pair of equations for the coefficients A, B, C, D we obtain from the boundary conditions (29) at $y=0$:

$$\begin{aligned}
&A\left(1-\frac{1}{n^2}\right)e^{-\kappa b}\cos(\gamma_+ b)-B\left(1+\frac{1}{n}\right)^2 e^{\kappa b}\cos(\gamma_- b)\\
&\quad+Ce^{-\kappa b}\left(\frac{1+1/n}{1-1/n}\right)\left(\frac{2\sin(\gamma_+ b)}{\mu\rho_1}-\frac{1}{\kappa}\left(1+\frac{1}{n}\right)^2\right.\\
&\quad\left.\times\cos(\gamma_- b)\right)+De^{\kappa b}\left(\frac{2\sin(\gamma_- b)}{\mu\rho_1}\right.\\
&\quad\left.+\frac{1}{\kappa}\left(1-\frac{1}{n}\right)^2\cos(\gamma_+ b)\right)=0,\\
&A\left(1-\frac{1}{n^2}\right)e^{-\kappa b}\sin(\gamma_+ b)+Be^{\kappa b}\left(1+\frac{1}{n}\right)^2\sin(\gamma_- b)\\
&\quad+Ce^{-\kappa b}\left(\frac{1+1/n}{1-1/n}\right)\left(-\frac{2\cos(\gamma_+ b)}{\mu\rho_1}+\frac{1}{\kappa}\left(1+\frac{1}{n}\right)^2\right.\\
&\quad\left.\times\sin(\gamma_- b)\right)+De^{\kappa b}\left(\frac{2\cos(\gamma_- b)}{\mu\rho_1}\right.\\
&\quad\left.+\frac{1}{\kappa}\left(1-\frac{1}{n}\right)^2\sin(\gamma_+ b)\right)=0,
\end{aligned} \tag{45}$$

where

$$\gamma_\pm=\gamma\pm\frac{\mu\rho_1}{2n}. \tag{46}$$

The condition that the system of Eqs. (44) and (45) has nontrivial solutions leads to an algebraic equation for the spectrum κ of the eigenvalue problem, given by Eq. (28) with boundary conditions (29). To obtain this equation, it is convenient to eliminate the coefficient C from the system of equations (44), (45), and to calculate the determinant of third order. The result of this calculation is

$$\begin{aligned}
&\frac{\sin(2\gamma b)}{\kappa}\left(e^{2\kappa b}\left(1+\frac{1}{n}\right)^2-e^{-2\kappa b}\left(1-\frac{1}{n}\right)^2\right)\\
&\quad-\frac{4}{\mu\rho_1}\cos(2\gamma b)+\frac{2}{\mu\rho_1}\left(e^{2\kappa b}\left(\frac{1+1/n}{1-1/n}\right)^2\right.\\
&\quad\left.+e^{-2\kappa b}\left(\frac{1-1/n}{1+1/n}\right)^2\right)+\left[b\left(\left(1+\frac{1}{n}\right)^2+\left(1-\frac{1}{n}\right)^2\right)\right.\\
&\quad\left.-\frac{1}{\kappa}\left(\left(1+\frac{1}{n}\right)^2-\left(1-\frac{1}{n}\right)^2\right)\right]\sin(2\gamma b)=0.
\end{aligned} \tag{47}$$

To simplify Eq. (47), we use the equation[5] for ρ_1

$$1-\cos(2\gamma b)=\frac{8(1-\rho_1)}{\rho_1}\frac{n^2}{(n^2-1)^2}. \tag{48}$$

Inserting the expression for $\cos(2\gamma b)$ from Eq. (48) into Eq. (47), we obtain the following equation for the spectrum points κ:

$$\begin{aligned}
&\frac{32n^2}{\mu\rho_1^2 b(n^2-1)^2}+\frac{\sin(2\gamma b)}{\kappa b}\left[(e^{2\kappa b}-1)\left(1+\frac{1}{n}\right)^2\right.\\
&\quad\left.-(e^{-2\kappa b}-1)\left(1-\frac{1}{n}\right)^2\right]+\sin(2\gamma b)\left(\left(1+\frac{1}{n}\right)^2\right.\\
&\quad\left.+\left(1-\frac{1}{n}\right)^2\right)+\frac{2}{b\mu\rho_1}\left[(e^{2\kappa b}-1)\left(\frac{1+1/n}{1-1/n}\right)^2\right.\\
&\quad\left.+(e^{-2\kappa b}-1)\left(\frac{1-1/n}{1+1/n}\right)^2\right]=0.
\end{aligned} \tag{49}$$

As noted above, Eq. (49) solves the general eigenvalue problem given by Eqs. (12) and (13).

Near the turning points, Eq. (49) has two real solutions. If $n\gg 1$, then both are in the range $|\kappa b|\ll 1$. In the range $|\kappa b|\ll 1$, we obtain from Eq. (49) the quadratic equation

$$\begin{aligned}
&\left[\frac{16n^2}{\mu\rho_1^2 b(n^2-1)^2}+3\frac{n^2+1}{n^2}\sin(2\gamma b)\right]+\kappa b\left[\frac{16n(n^2+1)}{\mu\rho_1 b(n^2-1)^2}\right.\\
&\quad\left.+\frac{4}{n}\sin(2\gamma b)\right]+4\kappa^2 b^2\left[\frac{1}{\mu\rho_1 b}\frac{n^4+6n^2+1}{(n^2-1)^2}\right.\\
&\quad\left.+\frac{n^2+1}{3n^2}\sin(2\gamma b)\right]=0.
\end{aligned} \tag{50}$$

The first term in Eq. (50) vanishes at the turning points, because the equation for the turning points is precisely the free term in Eq. (50):

$$\frac{16n^2}{\mu\rho_1^2 b(n^2-1)^2}+3\frac{n^2+1}{n^2}\sin(2\gamma b)=0. \tag{51}$$

The last statement immediately follows from Eq. (48). The coefficients of the terms κb and $(\kappa b)^2$ in Eq. (50) are both positive at the turning points, because ρ_1 is bounded from above and below:

$$\frac{4n^2}{(n^2+1)^2}<\rho_1\leqslant 1. \tag{52}$$

The inequality (52) is a consequence of Eq. (48). Hence, near the turning points the two eigenvalues are real. One of them can be found from Eq. (50) for any value of the refractive index n; it changes sign at the turning points. The second eigenvalue is also small ($|\kappa b|\ll 1$) near the turning points only if $n\gg 1$. For all other eigenvalues $|\kappa b|\gtrsim 1$.

We are able now to give a qualitative picture of the movement of these two ''lowest'' eigenvalues. It is presented in Fig. 3. At some values of μ, the two conjugate eigenvalues reach the real axis in the κ plane (point (1,1) in Fig. 2). After collision, they become real. One of them moves along the real axis towards the origin, and reaches it at point 2 (turning point, branch point in the μ-plane). After that, one branch is unstable (see Fig. 1). At point 3, both eigenvalues reach their extreme values and start to decrease in absolute value. At point 4 we go to the next turning point. Note that in the limit $\mu\to 0$, Eq. (49) yields for κ the values of the linear problem (Eqs. (17), (18)).

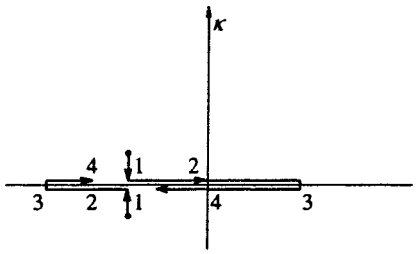

FIG. 3. Trajectory of two «conjugate» eigenvalues, one of them passing through zero.

To complete our investigation, we need to prove the possibility of the complex eigenvalues (with Im $\kappa \neq 0$) crossing the real axis Re $\kappa = 0$ when the nonlinearity parameter μ changes from zero to some finite value.

Suppose that for some value of the parameters $(b\mu\rho_1, \gamma b)$ a purely imaginary solution of Eq. (47) exists with $\kappa = i\tilde{\kappa}$ ($\tilde{\kappa}$ is real). Then from Eq. (47), we obtain two equations:

$$\frac{4}{\gamma\rho_1 b} = -\sin(2\gamma b)\,\frac{\tan(\tilde{\kappa}b)}{\tilde{\kappa}b}\,\frac{(n^2-1)^2}{n^2(n^2+1)}, \tag{53}$$

and

$$\sin(2\gamma b)\,\frac{2(n^2+1)}{n^2}\left(1+\frac{\sin(2\tilde{\kappa}b)}{\tilde{\kappa}b}\right) - \frac{4}{\mu\rho_1 b}\left[\cos(2\gamma b)-\cos(2\tilde{\kappa}b)\,\frac{n^4+6n^2+1}{(n^2-1)^2}\right]=0. \tag{54}$$

Inserting the expression for $\mu\rho_1 b$ from Eq. (53) into Eq. (54), we find one equation for γb and $\tilde{\kappa}b$

$$\frac{2(n^2+1)}{n^2}\left(1+\frac{\sin(2\tilde{\kappa}b)}{\tilde{\kappa}b}\right)+\frac{\tan(\tilde{\kappa}b)}{\tilde{\kappa}bn^2(n^2+1)}\,[(n^2-1)^2 \times\cos(2\gamma b)-\cos(2\tilde{\kappa}b)(n^4+6n^2+1)]=0. \tag{55}$$

The function on the left-hand side of Eq. (55) is an even function of the parameter $\tilde{\kappa}b$. Hence, we can investigate just the region $\tilde{\kappa}b \geqslant 0$. Consider first $\tilde{\kappa}b$ in the range

$$\tilde{\kappa}b \in \frac{\pi}{2}\,[2K_1, 2K+1], \quad K=0,1,... \tag{56}$$

That is,

$$\tilde{\kappa}b = \pi K + X, \quad 0<X<\pi/2. \tag{57}$$

In the range given by Eq. (57), we have

$$f(X)=\frac{2(n^2+1)}{n^2}\left(1+\frac{\sin(2X)}{\tilde{\kappa}b}\right)+\frac{\tan X}{\tilde{\kappa}bn^2(n^2+1)} \times[(n^2-1)^2\cos(2\gamma b)-\cos(2X) \times(n^4+6n^2+1)]=0, \tag{58}$$

$0<X<\pi/2$.

It is easy to prove that the function on the left-hand side of Eq. (58) is positive for $0 \leqslant X < \pi/2$. To check this, we change the $\cos(2\gamma b)$ to -1 and find

$$f(X) \geqslant \frac{2(n^2+1)}{n^2}\left(1+\frac{\sin(2X)}{2\tilde{\kappa}b}\right)+\frac{8\tan X}{\tilde{\kappa}b(n^2+1)}\sin^2 X, \quad 0<X<\pi/2. \tag{59}$$

In the range $0<X<\pi/2$, the expression on the right-hand side of the inequality (59) is positive, and hence in the range

$$\tilde{\kappa}b \in \frac{\pi}{2}\,[2K, 2K+1],$$

Eq. (55) does not have a solution.

Consider now the range

$$\tilde{\kappa}b \in \frac{\pi}{2}\,[2K+1, 2K+2], \tag{60}$$

that is,

$$\tilde{\kappa}b = \pi K + \pi/2 + X; \quad 0<X<\pi/2. \tag{61}$$

In the range (61), we obtain from Eq. (55)

$$\tilde{f}(X)=\frac{2(n^2+1)}{n^2}\left(1-\frac{\sin(2X)}{\tilde{\kappa}b}\right)-\frac{\cot X}{\tilde{\kappa}bn^2(n^2+1)} \times[(n^2-1)^2\cos(2\gamma b)+\cos(2X)(n^4+6n^2+1)]=0. \tag{62}$$

For any value of the parameter γb, the function $\tilde{f}(x)$ given by Eq. (62) varies from $-\infty$ to $2(n^2+1)/n^2$ when x goes from zero to $\pi/2$. Hence, there exists a minimum of one solution of Eq. (55) in the range

$$\tilde{\kappa}b \in \pm\frac{\pi}{2}\,[2K+1, 2K+2], \quad K=0,1,2. \tag{63}$$

Taking Eq. (53) into account, we find, that new unstable modes (with Im $\kappa \neq 0$) always appear outside some neighborhood of the turning points, because the quantity $\mu\sin(2\gamma b)$ is negative at all turning points ($\mu\sin(2\gamma b)<0$). The system of Eqs. (48), (53), (62) can have a solution only if $|\mu b|$ is sufficiently large, so we find that all branches between the turning points ($2K+1, 2K+2$; $K=0,1,2,...$) are unstable. For $|\mu b|$ greater than some critical value $b\mu_{cr}(n)$, some parts of branches between the turning points ($2K, 2K+1$, $K=1,2...$) become unstable. The instability of these branches is related to pairs of conjugate eigenvalues crossing the real axis Re $\kappa = 0$. The crossing points are given by Eqs. (53) and (55).

6. LIMITING CASE OF WEAK NONLINEARITY $|\mu|B \ll 1$

In the case $|\mu|b \ll 1$, the turning point can exist only if the refractive index $n \gg 1$, so $|\mu|bn \gtrsim 1$. In the vicinity of the turning points, γb is close to πN, where N is an integer:

$$\gamma b = \pi N + \epsilon + \frac{3\mu\rho_1}{4}\,b. \tag{64}$$

In the range of parameters considered here, Eq. (48) can be reduced to the cubic equation

$$Y^3 + Y\left(4-\frac{\epsilon^2 n^2}{3}\right)-\left(\frac{2\epsilon^2 n^3}{27}+\frac{8\epsilon n}{3}+3\mu bn\right)=0, \tag{65}$$

where

$$Y=\frac{2\epsilon n}{3}+\frac{3\mu b n\rho_1}{4}. \tag{66}$$

From Eq. (65) we find that at the turning points, Y is given by

$$Y=\pm(\epsilon^2 n^2-12)^{1/2}/3. \tag{67}$$

From Eq. (51), it follows that the turning points exist only if $\mu\epsilon<0$.

Between the turning points, all three solutions of Eq. (65) are given by

$$Y_k=\frac{2}{3}(n^2\epsilon^2-12)^{1/2}\cos\varphi_k, \tag{68}$$

where

$$\varphi_k=\frac{2\pi K}{3}+\frac{1}{3}\cos^{-1}\left(\frac{\epsilon^3 n^3+36\epsilon n+81\mu b n/2}{(n^2\epsilon^2-12)^{1/2}}\right),$$

$$K=0,1,2. \tag{69}$$

Equation (50) for κ can be substantially simplified in this case

$$\kappa^2 b^2+\frac{4\kappa b}{n}+\frac{1}{n^2}\left[4+3Y^2-\frac{n^2\epsilon^2}{3}\right]=0. \tag{70}$$

The solutions of this equation are

$$(\kappa b)_{1,2}=-\frac{2}{n}\pm\frac{1}{n}\sqrt{\frac{n^2\epsilon^2}{3}-3Y^2}. \tag{71}$$

It is easy to see that between the turning points, one mode has positive values of κb, hence the branch between the turning points is unstable.

7. CONCLUSIONS

In this paper we formulate an algebraic equation for the excitation spectrum κ that solves the problem of the stability of the solutions of a nonlinear wave equation in a slab. It is found that all branches between the $(2K+1, 2K+2)$ turning points are always unstable. Some parts of branches between the $(2K,2K+1)$ turning points are also unstable. The instability of the latter is associated with the possibility that pairs of complex conjugate eigenvalues cross the real axis in the κ-plane. Such a phenomenon can take place only if the effective nonlinearity is sufficiently strong ($|\mu|>\mu_{cr}(n)$). In that event, the temporal behavior of transitions between stationary states, when the amplitude of the incident wave varies, can be very complicated.

It was possible to obtain an explicit expression for the excitation spectrum, but only by virtue of the weak nonlinearity of the coefficient in the wave equation. Strong nonlinear effects result from the large length of the nonlinear medium compared to the wavelength scale.

The author thanks Prof. I. M. Sigal for helpful discussions and hospitality during the period in which most of this work was carried out. This research was supported by the CRDF Grant RPI-194, and by the Russian Fund for Fundamental Research.

[1] H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, New York (1985).
[2] Wei Chen and D. L. Mills, Phys. Rev. B **35**, 524 (1987); Phys. Rev. B **36**, 6269 (1987).
[3] H. M. Gibbs, G. Khitrova, and N. Peyghambarian, *Nonlinear Photomics*, Springe-Verlag Berlin-Heidelberg (1990).
[4] A. C. Newell and J. V. Moloney, *Nonlinear Optics*, Addison-Wesley, Reading, MA (1992).
[5] Yu. N. Ovchinnikov and I. M. Sigal, Preprint, University of Toronto (1997).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Two-component model for the growth of porous subsurface layers

A. É. Filippov

Donetsk Institute of Physics and Technology, National Academy of Sciences of Ukraine, 340114 Donetsk, Ukraine

(Submitted 4 July 1997)
Zh. Éksp. Teor. Fiz. **114**, 1500–1515 (October 1998)

This paper studies a kinetic model that describes the interaction of two fluctuating densities. The model makes it possible to stably reproduce the growth of dense, porous, and fractal structures near the surface of solids placed in an active medium. The solutions of local and nonlocal equations of the model are studied, and the results are used to comment on the possible scenarios of the evolution of systems whose behavior can be reduced to such a model. Finally, the exponents of the growth of the front width in a steady-state regime are calculated for various values of the parameters. © © *1998 American Institute of Physics.* [S1063-7761(98)02310-5]

1. INTRODUCTION

In recent years the studies of the growth and morphology of porous layers that form near a variety of flat surfaces have been attracting an ever growing interest from the practical and theoretical standpoints (see, e.g., Refs. 1–5 and the literature cited therein).

In many cases, irrespective of the specific features of the system under investigation, the evolution of the surface layer proceeds in a fairly universal manner. First, in the immediate vicinity of the smooth (flat) interface of the two media in contact there emerges, as a result of a chemical reaction, a dense layer of one or more reaction products. In the process of growing this layer becomes more and more porous. Gradually an essentially inhomogeneous but, as a rule, scale-invariant structure is formed, and the laws governing the growth of this structure are characterized by fractal dimension and growth exponent.

For instance, a smooth to fractal transition of the corrosion front has been directly observed by Balázs[1] in studies of two-dimension corrosion of thin aluminum films sputtered on optically transparent substrates and placed in an electrolyte containing active components: $Fe_2(SO_4)_3$, HCl, Na_2SO_4, and NaCl. A sequence of micrographs shows how an initially perfect round pit, perforated in the aluminum surface, expands, with characteristic dendritic structures forming at the boundary. With the passage of time the growth of the front becomes self-similar with fractal dimension $D_f = 1.33 \pm 0.01$.

The effect of buildup of the front length may manifest itself even stronger. For instance, in the device described in Ref. 2, when a weak current ($J = 0.2$ mA/cm^2) is sent through a copper electrode, ''runaway'' growth of the front develops, i.e., the front moves only along initially small protuberances of the smooth surface of the electrode, while the other sections of the front remain essentially unchanged in the process. As a result there forms a characteristic dendritic structure consisting, in contrast to an ordinary fractal surface,[1] of a collection of essentially isolated ''trees.''

In addition to arousing purely scientific interest, the study of corrosion-front growth attracts attention because of its importance from the standpoint of practice, since in some applications the problem is closely linked to that of raising the efficiency of electric batteries.[3,4] For instance, when a lithium anode is placed into an electrolyte containing $SOCl_2$ as an additive, due to the exceptionally high reactivity of lithium, a porous two-component layer of LiCl and SO_2 is formed at the surface of the anode. The presence of such a layer leads to what is known as the lag effect[4] when the element is stored for a long time. Micrographs of the surface layer show that the layer can be considered a combination of a relatively dense initial layer with a subsequent transition to a fractal structure with an ever increasing porosity.

The highly universal properties manifested by different systems suggest that one can use universal growth models based on a combination of the ideas of continuum field theory and kinetic equations with a random source.[5,6]

Being fairly common in the theory of phase separation and fluctuation phenomena in phase transitions,[7–20] the kinetic equations with a source of noise should be used cautiously in describing front growth, the reason being that, in contrast to phase transitions, where generation of the order parameter occurs in the bulk of the system, a random source cannot be considered additive. The generation of a finite density of the components forming the front occurs only in the immediate vicinity of the already existing boundary. This means that in the case at hand the corresponding source in the equation must be multiplicative (i.e., at least contain the density as a factor). However, in recent publications devoted to theoretical studies of phase diagrams and transitions in systems with multiplicative noise,[21–24] it was noted that the presence of such strong noise can have a dramatic effect on the ordered structure and on the phase diagram, and may lead to the emergence of new nontrivial phases. In our case this means that the model equation should be written in such a way so as to exclude additional difficulties associated with this noise.

From an experimental standpoint, the study of fractal corrosion structures is convenient since the corrosion front is observed directly in micrographs and the corresponding two-dimensional distributions of density can be studied explicitly. At the same time, the processes involved are very complex, and notwithstanding the continuing efforts, the theoretical models still remain extremely simple, although they presuppose a numerical analysis of the kinetic equations. Usually only the density of a single distributed quantity considered the most important in each specific case is involved.[5,10]

This is generally not the case in physicochemical processes, since usually two or more components participate in the reactions. No matter how subtle the description of a system by the single-component approach, it is sure to replace the study of the system by an analysis of purely theoretical models. Given contemporary computer modeling techniques, any attempt to reduce the problem to a single equation is more a tribute to the analytic tradition than a real necessity. The present paper demonstrates the feasibility of moving in this direction by the example of a two-component model formulated for the description of growth and corrosion of a broad class of porous surface layers initiated by chemical reactions.

2. FORMULATION OF THE MODEL

Following the work of Parisi and Zhang Yi-Cheng,[5] we will formulate the model in the form of the field-theoretic continuum variant of a model of growth with self-suppression, which in the case of a single component is the continuum version of the well known Eden model.[6] To put it briefly, Eden's model amounts to the following. Let us suppose that a particle source in d-dimensional space (we will limit ourselves to the case where $d=2$) generates, with a certain probability, new particles at neighboring vacant points of the space, etc.

Following Ref. 5, we define a sequence of order parameters at any stage of generation n by the following quantities:

$$\rho^{(1)}(\mathbf{r};n);\quad \rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;n);\quad \rho^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3;n);\ \ldots;$$
$$\rho^{(j)}(\mathbf{r}_1,\mathbf{r}_2,\ \ldots,\mathbf{r}_j;n), \tag{1}$$

with the first being the probability of finding a particle at point $\mathbf{r}$ in space at time n, the second being the probability of finding two particles, one at point $\mathbf{r}_1$ and the other at point $\mathbf{r}_2$, and so on. Then an Eden process with self-suppression is described by the formula

$$\rho^{(1)}(\mathbf{r};n+1)-\rho^{(1)}(\mathbf{r};n)$$
$$=D\sum_{\mu}\ \rho^{(1)}(\mathbf{r}+\mu;n)+\rho^{(1)}$$
$$\times(\mathbf{r}-\mu;n)/2dn-C\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;n), \tag{2}$$

where D and C are constants, the sum is over the nearest neighbors in the d-dimensional space, and the two-particle probability density in the lowest-order multiplicative approximation becomes a product of two single-particle functions,

$$\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2;n)=\rho^{(1)}(\mathbf{r}_1;n)\rho^{(1)}(\mathbf{r}_2;n), \tag{3}$$

while the corresponding hierarchy of equations for the multiparticle function $\rho^j(\mathbf{r}_1,\mathbf{r}_2,\ \ldots,\mathbf{r}_j;n)$ turns out to be truncated.

Inserting (3) into Eq. (2), passing to the continuum limit for the Laplacian,

$$\sum_{\mu}\ [\rho^{(1)}(\mathbf{r}+\mu;n)+\rho^{(1)}(\mathbf{r}-\mu;n)]/2dn\rightarrow\nabla^2\rho^{(1)}(\mathbf{r};n), \tag{4}$$

and defining the time variable via the substitution $t=\ln n$, we arrive at the simplest continuum version of the equation:

$$\partial\rho(\mathbf{r};t)/\partial t=D\nabla^2\rho(\mathbf{r};t)+C\rho(\mathbf{r};t)[1-\rho(\mathbf{r};t)]. \tag{5}$$

Here and in what follows, by the densities $\rho(\mathbf{r};n)$ we mean single-particle functions $\rho^{(1)}(\mathbf{r};n)$, so that the upper indices are dropped. Clearly, if the bare function $\rho(\mathbf{r};t)$ is equal to zero, $\rho(\mathbf{r};t)$ will remain zero at later moments in time. Generation of a moving front $\rho(\mathbf{r};t)$ emerges at such initial conditions that at $t=0$ the value of $\rho(\mathbf{r};0)$ is finite along one of the system boundaries.

Note that formally the structure of Eq. (5) is such that at negative values of the variable ρ its solutions $\rho(\mathbf{r};t)$ become unstable as $\rho\rightarrow-\infty$. In the theory of phase transitions such instability is removed by higher-order nonlinearities $\sim\rho^3(\mathbf{r};t)$ in the equation and, accordingly, by terms $\sim\rho^4(\mathbf{r};t)$ in the free-energy generating functional. In front growth models, the bare function $\rho(\mathbf{r};0)$ is positive and terms of order ρ^4 are generally not needed.[5] Here, however, one must bear in mind that the initial (and boundary) conditions always belong to the attracting basin of stable bounded solutions $\rho(\mathbf{r})$ as $t\rightarrow\infty$.

The rate of generation of the reaction components fluctuates. In Ref. 21 it was shown that this leads to the emergence of a multiplicative source of noise $\zeta(\mathbf{r},t)$ with a certain intensity D:

$$\langle\zeta(\mathbf{r},t)\rangle=0,\quad \langle\zeta(\mathbf{r},t)\zeta(\mathbf{r}',t)\rangle=D\delta(\mathbf{r}-\mathbf{r}')\delta(t-t'). \tag{6}$$

This source is more likely to model the participation in reactions of those components that have been explicitly allowed for in the equation (or equations) rather than the effect of thermal fluctuations.[21]

The structure of the local term in Eq. (5) shows that it corresponds to the variations of an appropriately chosen effective energy $V(\rho(\mathbf{r};t))$:

$$\delta V(\rho(\mathbf{r};t))/\delta\rho(\mathbf{r};t)=-C\rho(\mathbf{r};t)[1-\rho(\mathbf{r};t)]. \tag{7}$$

We have $V(\rho(\mathbf{r};t))=-C\rho^2(\mathbf{r};t)[1/2-\rho(\mathbf{r};t)/3]$ and hence there is no generation barrier $\rho(\mathbf{r};t)\neq0$. It is important that the noise be multiplicative, since it guarantees that there can be no such process in which a fluctuation $\rho(\mathbf{r};t)\neq0$ at an arbitrary point inside the system becomes a generator of $\rho(\mathbf{r};t)$, which would lead to spontaneous generation of matter far from the contamination front. However, a simple addition to Eq. (5) of a source proportional to density,

$$\partial\rho(\mathbf{r};t)/\partial t=D\nabla^2\rho(\mathbf{r};t)+C\rho(\mathbf{r};t)[1-\rho(\mathbf{r};t)]$$
$$+\rho(\mathbf{r};t)\zeta(\mathbf{r},t), \tag{8}$$

leads to strong side effects (even to a complete transformation of an ordered steady state), which are undesirable for an application model (these effects are described in Refs. 21–24). Qualitatively, the nature of the strong effect of noise proportional to $\rho(\mathbf{r};t)$ on a steady state is clear. As the front passes a given point in space, the density behind the front takes an equilibrium value: $\rho(\mathbf{r};t)=1$. The reaction that leads to front movements dies out. However, the damping process is hindered by noise, whose intensity is at its maximum in this region.

The fluctuations should be the strongest in the region occupied by the front, i.e., where the combination $\rho(\mathbf{r};t)[1-\rho(\mathbf{r};t)]$ is finite, and tends to zero far from this region. In the simplest case this assumption can be used by postulating the form of the equation,

$$\partial\rho(\mathbf{r};t)/\partial t = D\nabla^2\rho(\mathbf{r};t)+\rho(\mathbf{r};t) \times[1-\rho(\mathbf{r};t)][C+\zeta(\mathbf{r},t)]. \tag{9}$$

Despite the presence of self-suppression, $-C\rho^2(\mathbf{r};t)$, Eq. (9) yields a fairly trivial pattern of motion of a dense front with a gradually expanding boundary (due to random walks related to the noise $\zeta(\mathbf{r},t)$). The formation of a realistic porous structure is closely related to the fact that the problem is multicomponent. Below we examine the simplest two-component case, assuming, by way of an example, that we are dealing with chemical reactions that proceed in a system with a contaminated lithium anode (see Refs. 3 and 4 and the literature cited therein).

The complete picture of the reactions in such a system is fairly complicated and can be expressed as follows:

$$\mathrm{Li}\rightarrow\mathrm{Li}^+ + e^-, \tag{10}$$

$$4\mathrm{Li}^+ + 4e^- + 2\mathrm{SOCl}_2 \rightarrow 4\mathrm{LiCl} + \mathrm{SO}_2 + \mathrm{S}.$$

Actually, however, we are interested only in the formation of a front consisting of lithium chloride LiCl contaminated by the reaction product SO_2 that concentrates near the surface. Bearing all this in mind, we can interpret the bare equation (9) as the initial equation for the evolution of the density of LiCl, which we denote by $\rho_1(\mathbf{r};t)$.

The corresponding coefficients and the source of noise will be labeled by the index ''1.'' By $\rho_2(\mathbf{r};t)$ we denote the density of SO_2. We model the local repulsion of the reaction products LiCl and SO_2 by a fixed-sign additional term in the effective energy of the system, $V_{12}(\rho_1,\rho_2)$, which in the lowest order can be written $V_{12}(\rho_1,\rho_2)=B\rho_1^2\rho_2^2/2$. Equation (9) becomes

$$\partial\rho_1/\partial t = D_1\nabla^2\rho_1+\rho_1(1-\rho_1) \times(C_1+\xi_1(\mathbf{r},t))-B\rho_1\rho_2^2. \tag{11}$$

This equation must be augmented with an equation describing the evolution of the second component, $\rho_2(\mathbf{r};t)$. To do this we resort to the line of reasoning that resulted in Eqs. (1)–(5), where we must bear in mind, however, that the second component $\rho_2(\mathbf{r};t)$, just like the first, is generated as a result of the same reactions (10) near the free (i.e., not contaminated by SO_2) LiCl surface. This means that for $\partial\rho_2/\partial t$ we must use the same generating term as for $\partial\rho_1/\partial t$:

$$\partial\rho_2/\partial t = D_2\nabla^2\rho_2+\rho_1(1-\rho_1) \times(C_2+\xi_2(\mathbf{r},t))-\rho_2\rho_1^2-F(\rho_2). \tag{12}$$

Here we have allowed for the fact that although both densities, ρ_2 and ρ_1, emerge as a result of the same reaction, the rate of formation of the dense components in ρ_1 and ρ_2 may differ, so that generally $C_2/C_1\equiv v\neq 1$. Below, in a physically interesting case, we will actually use the quantity $\mu\geqslant 1$ (assuming, however, that $\mu\simeq 1$).

Obviously, the terms linear in ρ_2 cannot ensure that the increase in ρ_2 is stopped and is stabilized ($\rho_2\rightarrow 1$) in the static limit. We must also bear in mind that far from the front there is no spontaneous generation of ρ_2, and hence the effective energy, whose variation yields the function $F(\rho_2)$,

$$\delta V_2(\rho_2(\mathbf{r};t))/\delta\rho_2(\mathbf{r};t) = -F(\rho_2(\mathbf{r};t)), \tag{13}$$

contains a barrier that separates the two similar minima at $\rho_2=0$ and $\rho_2=1$. In the lowest-order nontrivial approximation we have

$$V_2(\rho_2)=B\rho_2^2(1-\rho_2^2)/2, \tag{14}$$

so that the sought equation for $\rho_2(\mathbf{r};t)$ assumes the form

$$\partial\rho_2/\partial t = D_2\nabla^2\rho_2+\rho_1(1-\rho_1)(C_2+\xi_2(\mathbf{r},t)) -\rho_2\rho_1^2-\rho_2(0.5-\rho_2)(1-\rho_2). \tag{15}$$

The transition to the continuum approximation, with only the lowest harmonics in the energy and hence the terms $\nabla^2\rho_{1,2}$ retained in the equations, leads to the well known loss of information about the restriction imposed on the minimum amount of LiCl in a neighborhood $|\mathbf{r}-\mathbf{r}'|\leqslant\sigma$ of a given point $\mathbf{r}$, an amount needed for the reactions (10) to proceed. Formally, such a restriction is equivalent to requiring that the state with $\rho_{1,2}=0$ be stable under small perturbations in the (integrated) density

$$\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma} d\mathbf{r}'\,\rho_1(\mathbf{r}').$$

Within the continuum approach, such stability can be ensured only by introducing essentially nonlocal terms into Eqs. (11) and (15). Note that the transition to the continuum limit does not eliminate the need to allow for higher-order gradients [with respect to $(\nabla\rho(\mathbf{r}))^2$] and nonlocal forms of the type

$$\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma} d\mathbf{r}'\,\rho_1(\mathbf{r}')V(\mathbf{r}-\mathbf{r}')\rho_1(\mathbf{r}')$$

in the generating functional. In the theory of phase transitions, similar nonlocal terms in the free energy are actively used to describe inhomogeneous ordering of the order parameter, critical behavior, blocking of new-phase nuclei, etc. (see, e.g., Refs. 16–20).

In the Monte Carlo method, where the quantity ρ_1 is discrete and equal in each location to 0 or 1, the reaction is ''turned on'' when there is at least one LiCl molecule in at least one location closest to the location considered.[4] In the continuum model, the generation terms in both equations for $\partial\rho_{1,2}/\partial t$ must contain essentially nonlocal factors, in addition to the local factors $\rho_1(1-\rho_1)(C_{1,2}+\xi_{1,2}(\mathbf{r},t))$. The

generation of ρ_1 and ρ_2 is ''turned on'' when the total density in a neighborhood $|\mathbf{r}-\mathbf{r}'|\leqslant\sigma$ of the given point $\mathbf{r}$ exceeds a certain threshold,

$$\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma} d\mathbf{r}'\ \rho_1(\mathbf{r}')>a.$$

The generating term must contain a cutoff factor

$$\Theta(\mathbf{r})=\vartheta\left(\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma} d\mathbf{r}'\ \rho_1(\mathbf{r}')-a\right) \tag{16}$$

such that $\Theta(\mathbf{r})\to 1$ when $\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma}d\mathbf{r}'\ \rho_1(\mathbf{r}')>a$ and $\Theta(\mathbf{r})=0$ when $\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma}d\mathbf{r}'\ \rho_1(\mathbf{r}')<a$. When the threshold is sharp enough, the function $\vartheta(x)$ degenerates into a step function.

If we now use the factor (16), the system of equations becomes essentially nonlocal and assumes the final form

$$\begin{aligned}\partial\rho_1/\partial t &= D_1\nabla^2\rho_1+\rho_1(1-\rho_1)\\ &\quad\times(C_1+\xi_1(\mathbf{r},t))\Theta(\mathbf{r})-\rho_1\rho_2^2,\\ \partial\rho_2/\partial t &= D_2\nabla^2\rho_2+\rho_1(1-\rho_1)(C_2+\xi_2(\mathbf{r},t))\\ &\quad\times\Theta(\mathbf{r})-\rho_2\rho_1^2-\rho_2(0.5-\rho_2)(1-\rho_2).\end{aligned} \tag{17}$$

In the nearest neighbor approximation we have $a=1$, and the integral condition $\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma}d\mathbf{r}'\ \rho_1(\mathbf{r}')>a$ can be reduced to the (discrete) Laplacian, which simply means that higher-order gradients are included in the equations.

3. SOLUTION OF EQUATIONS AND DISCUSSION

The system of equations (17) is complicated and can only be solved numerically. It is a modification of kinetic equations that have been widely used in recent years, in a variety of guises, to model the kinetics of phase transitions (see Refs. 9–20) and critical phenomena[17,18] on the basis of the generalized Landau–Khalatnikov equation[7] or the Cahn equation,[8] and phase separation in a system consisting of interacting subsystems.[14,15,18] Note that (17) contains essentially all ''irritating factors'' that can be encountered in kinetic equations of this type, including nonlocality, multiplicative noise, and interaction of the subsystems.

In addition to all this, in contrast to the theory of phase transitions, the system (17) can be solved under specific initial conditions, when ordering and phase separation occurs not in the form of growth of nuclei over the entire volume[9–17,19,20] but by the motion of the front starting from one of the boundaries of the system and followed by ordering and phase separation of both interacting fields $\rho_{1,2}(\mathbf{r};t)$ behind the front.

Since even a numerical solution of such a system requires at least an idea about the possible scenario of the process so that the parameters can be fixed, it is advisable to start with a simpler local version of the equations. In a certain sense this version must describe the system in the regions where the densities $\rho_{1,2}(\mathbf{r};t)$ are approximately equal and the gradient terms, as well as the cutoff factors $\Theta(\mathbf{r})$, can be dropped.

In this case we simply have a system of two differential equations for the $\mathbf{r}$-independent variables $\rho_{1,2}=\rho_{1,2}(t)$:

$$\begin{aligned}\partial\rho_1/\partial t &= C_1\rho_1(1-\rho_1)-B_1\rho_1\rho_2^2,\\ \partial\rho_2/\partial t &= C_2\rho_1(1-\rho_1)-B_2\rho_2\rho_1^2\\ &\quad-B_2\rho_2(0.5-\rho_2)(1-\rho_2).\end{aligned} \tag{18}$$

The fixed points and the isoclines of the vertical and horizontal lines of this system can be found directly by solving the algebraic equations

$$\begin{aligned}\partial\rho_1^*/\partial t &= C_1\rho_1^*(1-\rho_1^*)-B_1\rho_1^*\rho_2^{*2}=0,\\ \partial\rho_2^*/\partial t &= C_2\rho_1^*(1-\rho_1^*)-B_2\rho_2^*\rho_1^{*2}\\ &\quad-B_2\rho_2^*(0.5-\rho_2^*)(1-\rho_2^*)=0.\end{aligned} \tag{19}$$

In addition to the three obvious fixed points,

$$\begin{aligned}&\mu_0^*:\quad \rho_1^*=0,\quad \rho_2^*=0,\\ &\mu_1^*:\quad \rho_1^*=1,\quad \rho_2^*=0,\\ &\mu_2^*:\quad \rho_1^*=0,\quad \rho_2^*=1,\end{aligned} \tag{20}$$

there can be a pair of nontrivial points, which can be found by solving the system of equations (19) with $\rho_{1,2}\neq 0$ and $\rho_{1,2}\neq 1$.

In particular, in the symmetric case $C_1/B_1=C_2/B_2\equiv C$ we have

$$\begin{aligned}&\rho_1^*=1-\rho_2^{*2}/C,\\ &(1-\rho_2^{*2}/C)[\rho_2^*-(1-\rho_2^{*2}/C)]\\ &\quad-(0.5-\rho_2^*)(1-\rho_2^*)=0,\end{aligned} \tag{21}$$

and for example at $C=5$, this yields

$$\begin{aligned}&\mu_3^*:\quad \rho_1^*=0.53(9),\quad \rho_2^*=1.51(83),\\ &\mu_4^*:\quad \rho_1^*=0.87(2),\qquad \rho_2^*=0.8.\end{aligned} \tag{22}$$

Here it can be verified that μ_4^* is a saddle point, while the stability of the point μ_3^* depends on the position of the point in relation to μ_2^* (specifically, μ_3^* is a stable point if at it $\rho_2^*>1$).

It is expected, however, that the stable fixed points of the local system (18) determine, in the limit $t\to\infty$, the steady-state solution for both densities $\rho_{1,2}$. Bearing in mind the physical meaning of $\rho_{1,2}$, we conclude that one of the following points corresponds to such a solution, μ_1^*: $\rho_1^*=1$, $\rho_2^*=0$, or μ_2^*: $\rho_1^*=0$, $\rho_2^*=1$, or at least the point at which $\rho_1^*+\rho_2^*=1$.

The fixed point μ_3^*, which us stable when $\rho_2^*>1$, does not satisfy this requirement. At this point, in addition to $\rho_1+\rho_2$ being greater than unity, the density ρ_1^* is finite, so that the point is obviously a nonphysical one. The physical point μ_2^*, however, is unstable under the same condition ($\rho_2^*>1$ at point μ_3^*). The phase trajectories leave the neighborhood of this point in the direction of μ_3^*. In the limit $\rho_2^*\to 1$, point μ_3^* approaches μ_2^*, and at $\rho_2^*=1$ the two merge. In the process, point μ_2^* becomes stable.

The requirement that μ_2^* be a stable point imposes restrictions on the physical range of the parameter in numerical modeling of the nonlocal system. In particular, at $C_1/B_1=C_2/B_2\equiv C$ the bifurcation of the points μ_2^* and μ_3^* for the local equations occurs at $C=1$. As a result of diffusion

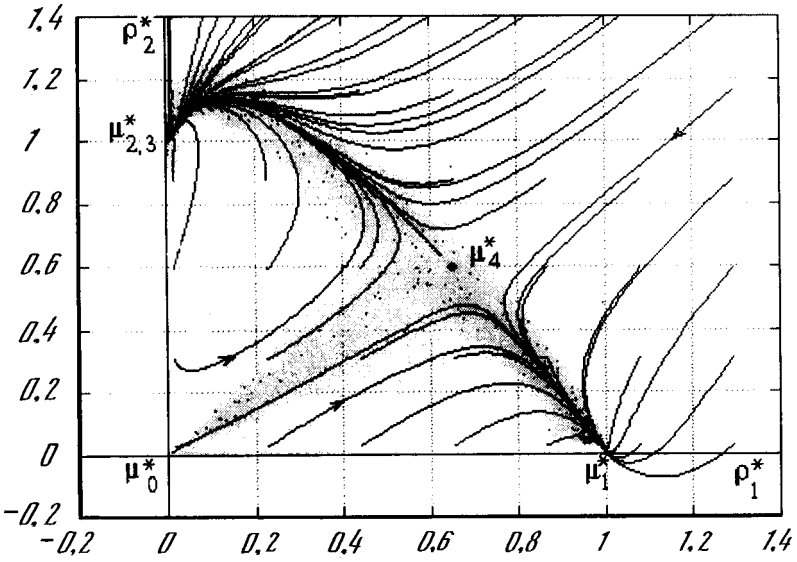

FIG. 1. Structure of the phase portrait for the local system (the case $B/C>(B/C)_{\rm cr}$ is depicted). The phase trajectories of the local system of equations, depicted by solid curves, are associated with the projections of the points of the arrays $\{\rho_1(x,y;t);\rho_2(x,y;t)\}$ onto the (ρ_1^*,ρ_2^*) plane (depicted by a gray area), which constitute the numerical solution of the complete system of nonlocal equations with noise.

($\nabla^2\rho_{1,2}\neq 0$) and ''fading'' of the smaller of the two densities, ρ_1, to zero, this limit for the nonlocal system shifts in the direction in which the corresponding constraint is weakened, so that the (calculated) bifurcation of the points μ_2^* and μ_3^* occurs at $C_{\rm cr}\approx 1.4$.

The global structure of the phase portrait is depicted in Fig. 1, which presents a physically interesting realization of the portrait for $C<C_{\rm cr}$. The fixed points μ_3^* and μ_4^* for the case where $C>C_{\rm cr}$ can be obtained by numerically solving the equation

$$f(\rho_2^*)=(1-\rho_2^{*2}/C)[\rho_2^*-(1-\rho_2^{*2}/C)] -(0.5-\rho_2^*)(1-\rho_2^*)=0.$$

The system of equations (18) actually describes the evolution of the densities $\rho_{1,2}^*$ at each point in space without allowing for interaction between different points. In this approximation the interaction is taken into account only via the initial conditions. Specifically, as the front arrives at a point in space, both densities $\rho_{1,2}^*$ begin to be generated at that point, so that the physical scenario in the phase portrait in Fig. 1 corresponds to the trajectories that emerge in the neighborhood of the trivial point μ_0^*.

The separatrix connecting this point and the saddle point μ_4^* divides the plane (ρ_1^*,ρ_2^*) into attracting basins for the stable fixed points μ_1^* and $\mu_{2,3}^*$. Studying the behavior of the trajectories that start near the separatrix, we can predict several results of numerically modeling the complete equations and, in the final analysis, the properties of real systems. In particular, we can easily predict the role of the source of noise $\xi_{1,2}(\mathbf{r},t)$. If the noise is strong, the phase trajectories can pass both above and below the separatrix, irrespective of the scenario according to which the front arrives at the given point.

Within a certain time interval after the arrival of the front, both densities $\rho_{1,2}^*$ increase essentially simultaneously and very fast, to which the first maximum in the evolution rate

$$W(t)=[(\partial\rho_1^*/\partial t)^2+(\partial\rho_2^*/\partial t)^2]^2 \qquad (23)$$

depicted in Fig. 2 corresponds. Near the saddle point μ_4^* there can be no further increase in $\rho_{1,2}^*$. The rate $W(t)$ rapidly decreases.

At the same time there is phase separation in the system, with one of the densities, ρ_1^* or ρ_2^*, expelled from the given region in space. This is followed by a sharp increase in $W(t)$ accompanied by a rapid buildup of the remaining component, a process that is slowed down near one of the stable fixed states, μ_1^*: $\rho_1^*=1$, $\rho_2^*=0$ or μ_2^*: $\rho_1^*=0$, $\rho_2^*=1$.

The characteristic double-humped curves representing the evolution rate are indeed observed when the complete system of equations is solved numerically. In accordance with the physics of the problem, the initial condition is selected in the form of a narrow strip of density, $\rho_1(x,y;t=0)\neq 0$, near one of the boundaries of the two-dimensional system. Here the process of generation and separation of the densities $\rho_{1,2}(\mathbf{r},t)$ is accompanied by the formation of characteristic dendritic spatial distributions of both densities.

The two densities are generated simultaneously in the vicinity of the front. However, in the absence of noise, $\xi_{1,2}(\mathbf{r},t)=0$, the initial distribution, $\rho_1(x,y;t=0)\neq 0$ and $\rho_2(x,y;t=0)=0$, leads to the formation behind the front of a completely filled region, $\rho_1(x,y;t)\neq 0$, and to $\rho_2(x,y;t)$ remaining equal to zero in the inner part of the system. For patches with $\rho_2(x,y;t)\neq 0$ to appear in the steady-state distribution behind the front, $\rho_2(x,y;t)$ must be larger than $\rho_1(x,y;t)$ at least in some regions of space near the front (where both densities are small). This condition is met if $\xi_{1,2}(\mathbf{r},t)\neq 0$.

At first the distribution $\rho_1(x,y;t=0)\neq 0$ behind the front is essentially homogeneous. The regions with $\rho_2(x,y;t)>\rho_1(x,y;t)$ not only generate patches with $\rho_2(x,y;t)\neq 0$ but, more importantly, stop the generation of both densities at the point of the front. Active growth occurs only within fragments where

$$\Theta(\mathbf{r})=\vartheta\left(\int_{|\mathbf{r}-\mathbf{r}'|\leqslant\sigma} d\mathbf{r}'\ \rho_1(\mathbf{r}')-a\right)>0.$$

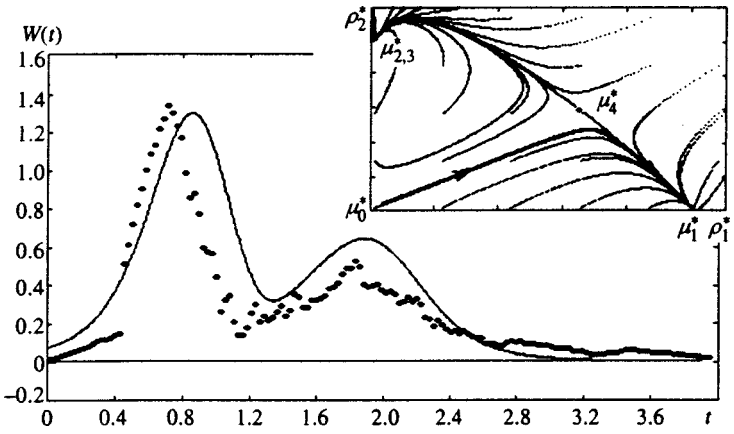

FIG. 2. Two peaks in the rate of variation of both concentrations, $\rho_1^*(x,y;t)$ and $\rho_2^*(x,y;t)$, at a fixed point (x,y) in the evolution in front of and behind the neighborhood of the fixed saddle point μ_4^*. The dots depict the results of a computer experiment. For the sake of comparison, we depict a similar curve (the solid curve) predicted on the basis of an analysis of the local system with a trajectory close to a physical one. The inset depicts the phase portrait of the local system, where such a trajectory is represented by a heavy solid curve.

The front becomes discontinuous, and in time the expanding ordered region is transformed into a fractal.

In Fig. 3a we use a gray scale to show a characteristic fragment of the system (consisting of 512×512 computational points) with a distribution of the total density $\rho(x,y;t)=\rho_1(x,y;t)+\rho_2(x,y;t)$ that emerges at the intermediate stage in the transition from homogeneous growth to fractal growth. Contaminated regions and regions of active growth are clearly visible. These are characterized by intermediate values of the densities, $\rho_1(x,y;t)\cong\rho_2(x,y;t)\leqslant0.8$, which correspond to sections in Fig. 3 with intermediate shades of gray.

The phase trajectories of the local system of equations, depicted by solid lines in Fig. 1, are associated with the projections of the points of arrays $\{\rho_1(x,y;t);\rho_2(x,y;t)\}$ onto the (ρ_1^*,ρ_2^*) plane (depicted by a gray area), which constitute the numerical solution of the complete system of nonlocal equations with noise. Note that such association is necessary in the given case since the effect of multiplicative noise may, in principle, be far from trivial.[21–24]

The good agreement observed in Fig. 1 is achieved because of a special combination of the source of noise $\xi_1(\mathbf{r},t)$ and the density ρ_1 in Eqs. (17), which was written above in the form $\rho_1(1-\rho_1)(C_1+\xi_1)$ derived from physical considerations.

In Fig. 2 there are two peaks in the rate of variation of both concentrations, $\rho_1(x,y;t)$ and $\rho_2(x,y;t)$, with the data obtained in a computer experiment depicted by dots. We see that the data are in good agreement with the results predicted earlier by the analysis of the local system of equations (the solid curve). This is an indirect indication that the interpretation of the model and the results is correct, since such agreement exists only for special trajectories (close to physical trajectories), which emerge in the neighborhood of point μ_0^*. The inset in Fig. 2 depicts the phase portrait of the local system, where such a trajectory is represented by a heavy solid curve.

One novel property of the system (in comparison to studying growth processes with models with only one fluctuating variable[4]) consists in the possibility of the emergence of voids behind the front, i.e., regions filled with neither of the two components. When there is only one field, i.e., $\rho_1(x,y;t)\equiv\rho(x,y;t)$, such lacunae must be thought of as being filled with the other, ''contaminating,'' component of density, $\rho_2(x,y;t)$, which in this case is not explicitly present in the equations. The model (17) contains additional information about the second field $\rho_2(x,y;t)$, which makes it possible to distinguish between the regions occupied by $\rho_2(x,y;t)\neq0$ and the voids.

The mechanism of void formation is clearly seen in Fig. 3b, which for a small fragment of the front depicts a typical growth sequence for $\rho(x,y;t)$. Three characteristic moments in time are singled out: the emergence of a dense initial layer, the emergence of the first dendritic protuberances, and the collapse of the first internal pores in a structure with a total density $\rho(x,y;t)$. Well-formed voids are clearly visible in Fig. 3a. Void formation is closely related to the ability of the ''contaminant'' $\rho_2(x,y;t)$ to block the active sections of the front $\rho_1(x,y;t)$ and, at least in principle, to terminate the growth process.

As the pores collapse, the front usually continues to move in both directions. Naturally, the outer boundary $\rho_1(x,y;t)\neq0$ is almost insensitive to the presence of a pore blocked somewhere inside the system and continues its forward motion. The inner boundary $\rho_1(x,y;t)\neq0$ surrounding a pore is qualitatively similar to the outer boundary and can move ''back,'' up to the point where it is completely blocked by sections with $\rho_2(x,y;t)\neq0$.

The scenario of the evolution of the system turns out to be exceptionally multifaceted and under a slight variation of the coefficients of the model makes possible a reproduction of very realistic configurations of the densities $\rho_1(x,y;t)$ and $\rho_2(x,y;t)$. In this respect the strongest effect is accompanied by a variation of the C_1-to-C_2 ratio, which reflects the possible difference between the local rates of formation of the components $\rho_1(x,y;t)$ and $\rho_2(x,y;t)$ in the reaction.

When $C_2\gg C_1$, the separatrix connecting μ_0^* and μ_4^* lies much lower than most phase trajectories, and in the steady-state limit a large part of the (x,y) plane becomes ordered in such a way that $\rho_1(x,y;t)\to0$ and $\rho_2(x,y;t)\to1$. Here the active sections of the front are rapidly contaminated by a density layer $\rho_2(x,y)=1$, and growth stops. Obviously, in the opposite limit, $C_2\ll C_1$, the front moves off to infinity. There is a critical value of the ratio, $v_{cr}=(C_2/S_1)_{cr}$, which leads to the first blocking of the front. Computer experiments show that this ratio is close to unity and amounts to $v\approx1.03$.

As $v\to v_{cr}$, the front ceases to grow because the fraction

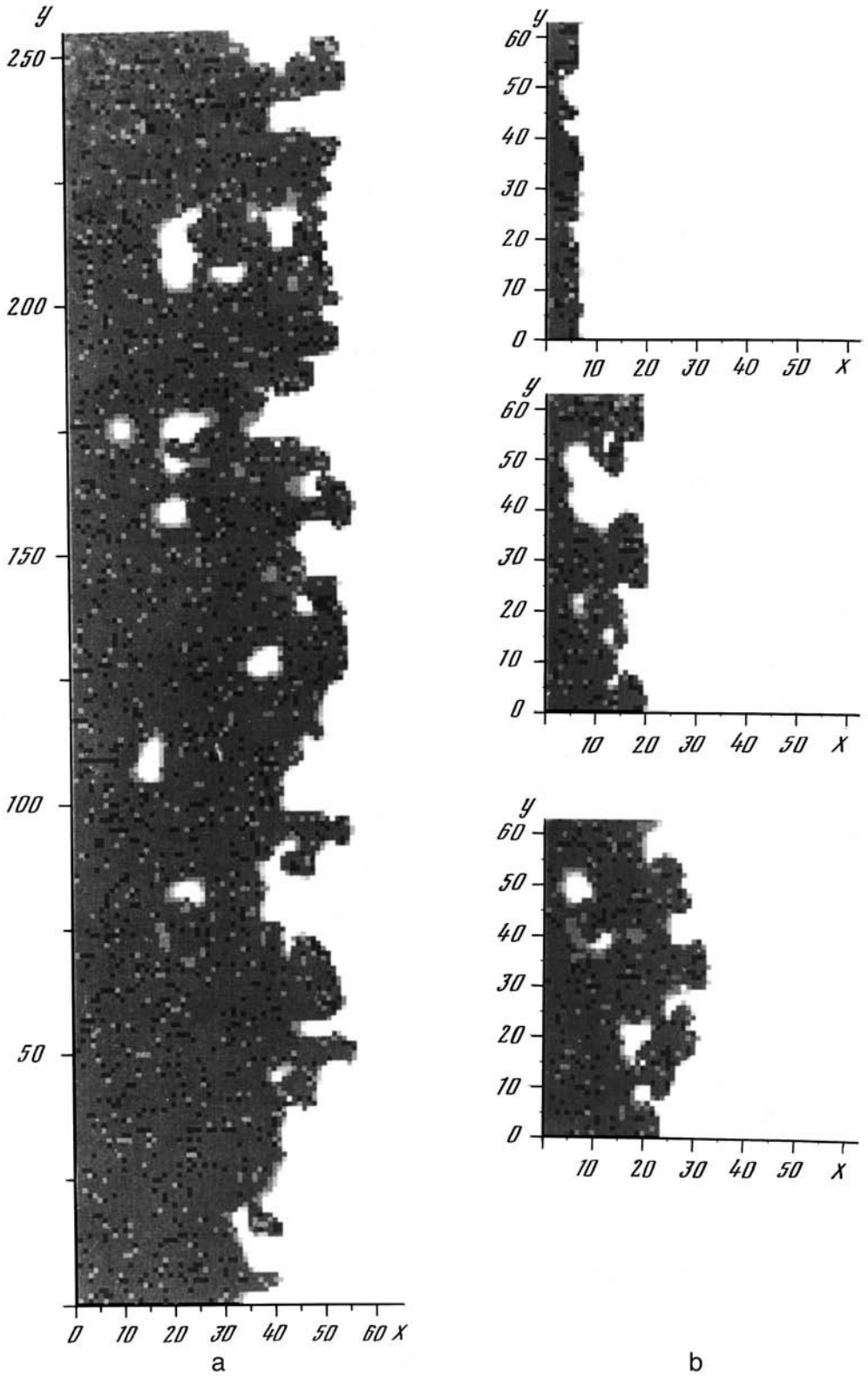

FIG. 3. Typical spatial configurations of the total density $\rho(x,y;t)=\rho_1(x,y;t)+\rho_2(x,y;t)$ with the ratio of parameters C_1 and C_2 close to the critical value. Fragments of the system of 512×512 computational points are depicted. a) The density distribution $\rho(x,y;t)$ that emerges at the intermediate stage in the transition from smooth growth to fractal growth. b) A typical sequence of growth of $\rho(x,y;t)$ for a small segment of the front depicted at three representative moments in time: the emergence of a dense layer, the emergence of the first dendritic protuberances, and the collapse of the first internal pores in a structure with a total density $\rho(x,y;t)$.

of the space (x,y) in which $\rho_1(x,y;t)\neq0$ turns out to be too small. A quantitative characteristic of this fraction can be found by calculating the values of the two densities averaged over the entire sample,

$$h_{1,2}(t)\equiv\langle\langle\langle\rho_{1,2}(x,y;t)\rangle_\Omega\rangle\rangle=\left\langle\left\langle\int_\Omega dx\,dy\,\rho_{1,2}(x,y;t)\right\rangle\right\rangle, \tag{24}$$

the total average density $h=h_1+h_2$, and the ratios $h_{1,2}/h$. The temporal evolution of the averages $h_{1,2}$ and h as $v\to v_{cr}$ is accompanied by changes in the relationship between the h_j as the system goes from the initial regime (in which $h_1>h_2$) to the steady-state regime as $t\to\infty$, the latter characterized by preferential growth of the $\rho_2(x,y;t)$ component.

Note that because the characteristic dimensions of the

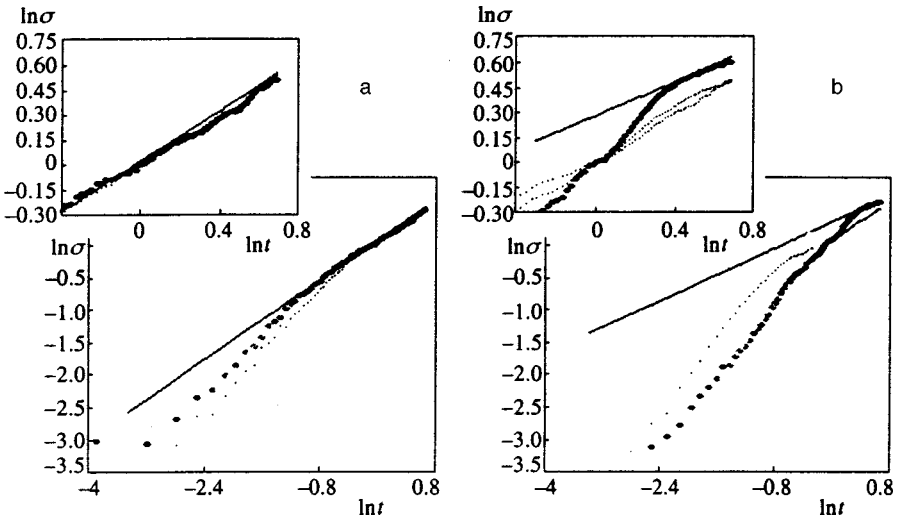

FIG. 4. Construction of the functions $\ln(\sigma \ln t)$ needed to determine the fractal dimensions of the front: a) for a subcritical regime $v \leqslant 1 < v_{cr}$, and b) for a regime close to the critical, $v \simeq v_{cr}$. The big dots correspond to σ and the small dots, to $\sigma_{1,2}$.

structures increase with the passage of time and the arrays are finite, the fluctuations $\delta h_{1,2}(t)$ for a specific realization of the computer experiments become substantial against the background of the averages $h_{1,2}(t)$ for $t \gg 1$. The monotonic asymptotic curves representing $h_{1,2}(t)$ are obtained by averaging the results over several (about ten) realizations. This averaging is denoted in (24) by double angle brackets, $\langle\langle \cdots \rangle\rangle$.

Such calculations show that the front stops growing at $h_1/h \simeq 1/3$ and $h_{1,2}/h \simeq 2/3$. These relationships appear to be fairly ''universal'' and are probably due to the universal laws that govern ''percolation'' along the continuous cluster $\rho_1(x,y;t) \neq 0$. However, establishing the possible relations between the properties of the model and percolation theory is a problem in its own right and lies outside the scope of the present paper.

It is extremely difficult to study the fractal properties of the structures that form in the model. The presence of interacting subsystems and several spatial scales in the initial equations can, at least in principle, lead to a time-varying or multifractal behavior of the solutions.[25] Moreover, for the continuum model, even a static distribution of the densities $\rho_{1,2}(x,y;t \to \infty)$ far behind the front contains values of the densities that vary in space, $0 \leqslant \rho_{1,2}(x,t) \leqslant 1$, which introduces an arbitrariness into the problem of finding the fractal dimensions.

Bearing in mind the above definition of averages,

$$h_{1,2}(t)/\Omega \equiv \langle\langle\langle \rho_{1,2}(x,y;t)\rangle_\Omega\rangle\rangle = \left\langle\left\langle \int_\Omega dx\, dy\ \rho_{1,2}(x,y;t)/\Omega \right\rangle\right\rangle,$$

and allowing for the fact that, except for the vicinity of interphase boundaries, the densities $\rho_{1,2}(x,y;t \to \infty)$ are close to the constant values $\rho_{1,2}(x,y) = 0$ or 1, we find the widths of these boundaries for each component via the relationships

$$\sigma_{1,2}(t) = \left\langle\left\langle\left\{ \int_\Omega \frac{dx\,dy}{\Omega}\left[\rho_{1,2}(x,y;t) - \int_\Omega \frac{dx\,dy}{\Omega}\rho_{1,2}(x,y;t)\right]^2\right\}^{1/2}\right\rangle\right\rangle \equiv \left\langle\left\langle\left\{ \frac{1}{\Omega}\left\langle\left[\rho_{1,2}(x,y;t) - \frac{h_{1,2}(t)}{\Omega}\right]^2\right\rangle\right\}^{1/2}\right\rangle\right\rangle, \quad (25)$$

for the partial densities $\rho_{1,2}$, and

$$\sigma(t) = \left\langle\left\langle\left\{ \int_\Omega \frac{dx\,dy}{\Omega}\left[\rho(x,y;t) - \int_\Omega \frac{dx\,dy}{\Omega}\rho(x,y;t)\right]^2\right\}^{1/2}\right\rangle\right\rangle \equiv \left\langle\left\langle\left\{ \frac{1}{\Omega}\left\langle[\rho(x,y;t) - h(t)/]^2\right\rangle\right\}^{1/2}\right\rangle\right\rangle \quad (26)$$

for the total density ρ. Here averaging over the set of realizations, denoted by double angle brackets $\langle\langle \cdots \rangle\rangle$, is necessary for the same reason as in the calculations of $h_{1,2}(t)$.

Figure 4 depicts the behavior of the functions $\sigma(t)$ on a log–log scale for two typical cases, $v \leqslant 1 < v_{cr}$ and $v \cong v_{cr}$. In both cases, after the initial transient process, as $t \to \infty$ the scaling behavior sets in in $\sigma(t)$, or

$$\sigma(t) \sim t^\alpha, \quad (27)$$

with $\alpha = 0.75$ for $v \leqslant 1 < v_{cr}$ and $\alpha = 0.5$ for $v \simeq v_{cr}$.

The reason the exponent α decreases to $\alpha = 0.5$ as $v \to v_{cr}$ lies in the fact that the active sections of the front rapidly become contaminated from sections with $\rho_2(x,y) = 1$, where the increase of both densities $\rho_{1,2}(x,y)$ gradually stops. As a result we have a more rapid buildup of the width than for $v \leqslant 1 < v_{cr}$:

$$\sigma(t) = \langle\langle\{\langle\langle[\rho(x,y;t) - h(t)/\Omega]^2\rangle/\Omega\}^{1/2}\rangle\rangle$$

in the early stage with a subsequent slowdown in the scaling limit $t \to \infty$.

In the opposite limit, $v \ll v_{cr}$, the front moves off to infinity, while regions with the density $\rho_2(x,y)$ have essentially no effect on the formation of $\sigma(t)$. In the process the model degenerates into the Eden model with $\alpha = 1/3$ (Refs. 5 and 6). Thus, our model leads to a situation in which the front is heavily incised in the intermediate range $v \leqslant 1 < v_{cr}$, where the effects of the interaction of $\rho_1(x,y)$ and $\rho_2(x,y)$ are most clearcut.

The fact that the system of equations (17) degenerates into a one-component model in the limit $v \ll v_{cr}$, where generation of $\rho_2(x,y)$ is negligible, provides an obvious way of verifying the model. At the same time, as is the case with phase transitions in critical-phenomena theory, the lower symmetry of the system with interacting fields makes the model unstable against leaving the (universality) class of a scalar model. There is also a similar instability related to the lowering of the symmetry of the propagator when nonlocalities are taken into account.

All these factors are the formal reason the scaling behavior of σ changes. It is still unclear to what extent the size of α depends on the structure of the terms in Eqs. (17). The study and enumeration of the universality classes of such models constitute an interesting problem, whose solution, however, lies outside the scope of the present paper, which is devoted primarily to the construction of a workable two-component model capable of stably reproducing the growth of porous surface layers with a scale-invariant structure.

I thank J.-P. Badiali and V. Russier for drawing my attention to this problem, for their hospitality at the University of Pierre and Marie Curie (Paris VI), and for many discussions due to which the model acquired its present form. This work was partially supported by the Ukrainian Foundation for Basic Research (F4/72-97, Project No. 2.4/199) and the international program INTAS 96-0410.

[1] L. Balázs, Phys. Rev. E **54**, 1183 (1996).
[2] V. Fleury and D. Barkey, Europhys. Lett. **36**, 253 (1996).
[3] T. I. Evans, T. V. Nguyen, and R. E. White, J. Electrochem. Soc. **136**, 329 (1989).
[4] I. Nainville, L. Lemarchand, and J. P. Badiali, Phys. Rev. E **53**, 2537 (1996).
[5] G. Parisi and Zhang Yi-Cheng, J. Stat. Phys. **41**, 1 (1985).
[6] M. Eden, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. IV, J. Neyman (Ed.), University of California Press, Berkeley (1961), p. 223.
[7] L. D. Landau and I. M. Khalatnikov, in *L. D. Landau: Collected Papers*, Pergamon Press, Oxford (1965), p. 626.
[8] J. W. Cahn, Acta Metall. **8**, 554 (1960).
[9] A. Mazor and A. R. Bishop, Physica D **39**, 22 (1989).
[10] F. Falk, Z. Phys. B **54**, 159 (1984).
[11] F. Falk, J. Phys. C **20**, 2501 (1987).
[12] Ch. Zhulicke, A. S. Mikhailov, and L. Scimansky-Geller, Physica A **163**, 559 (1990).
[13] A. Gordon, Phys. Lett. A **154**, 79 (1991).
[14] T. M. Rogers, K. R. Elder, and R. C. Desai, Phys. Rev. B **37**, 9638 (1988).
[15] K. R. Elder and R. C. Desai, Phys. Rev. B **40**, 243 (1989).
[16] Yu. E. Kuzovlev, T. K. Soboleva, and A. É. Filippov, Zh. Éksp. Teor. Fiz. **103**, 1742 (1993) [JETP **76**, 858 (1993)].
[17] A. É. Filippov, J. Stat. Phys. **75**, 241 (1994).
[18] A. S. Zel'tser and A. É. Filippov, Zh. Éksp. Teor. Fiz. **106**, 1117 (1994) [JETP **79**, 605 (1994)].
[19] A. S. Zel'tser, T. K. Soboleva, and A. É. Filippov, Zh. Éksp. Teor. Fiz. **108**, 356 (1995) [JETP **81**, 193 (1995)].
[20] A. S. Zel'tser and A. É. Filippov, JETP Lett. **62**, 627 (1995).
[21] Yu. Tu, G. Grinstein, and M. A. Munoz, Phys. Rev. Lett. **78**, 274 (1997).
[22] J. Garsia-Ojalvo, A. Hernandez-Machado, and J. M. Sancho, Phys. Rev. Lett. **71**, 1542 (1993).
[23] A. Becker and L. Kramer, Phys. Rev. Lett. **73**, 955 (1994).
[24] G. Grinstein, M. A. Munoz, and Yu. Tu, Phys. Rev. Lett. **76**, 4376 (1996).
[25] F. J. Solis and L. Tao, Phys. Lett. A **228**, 351 (1997).

Translated by Eugene Yankovsky

High-precision measurement of separatrix splitting in a nonlinear resonance

V. V. Vecheslavov*) and B. V. Chirikov†)

G. I. Budker Institute of Nuclear Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 11 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 1516–1531 (October 1998)

We present the results of numerical modeling and a theoretical analysis of the splitting of a nonlinear-resonance separatrix in the intermediate asymptotic region for the standard-map model. Direct measurements of the splitting angle $\alpha(K)$, where K is the small parameter of the system, have been carried out over a huge range, $0.1 \gtrsim \alpha \gtrsim 10^{-208}$ ($1 \geqslant K \geqslant 0.0004$), with a relative accuracy greater than one part in 10^{-25} and an average accuracy of roughly one part in 10^{-30}. This made it possible to compare in detail our results with those of the existing asymptotic theory and to detect a number of new effects. We find a relatively simple empirical expression for the α vs. K dependence in the intermediate asymptotic region, and this region proves to be surprisingly broad: $K \lesssim 10^{-2}$. We also study the effect of noise, in particular, errors in measuring the angle, which proved to be much more significant and complicated than expected. Finally, we point out unresolved questions and possible directions of research involving this problem. © *1998 American Institute of Physics.* [S1063-7761(98)02410-X]

1. INTRODUCTION

The dynamics of Hamiltonian (nondissipative) systems is governed by the interaction of nonlinear resonances, which are the basic structural elements of the modern theory of nonlinear oscillations.[1–4] The Hamiltonian of such a system can be written

$$H(I,\theta,t)=H_0(I)+\varepsilon\sum_{n,m} V_{nm}(I)\exp(in\theta+itm\Omega), \quad (1.1)$$

where ε is the small perturbation parameter, I and θ are action–angle variables, Ω is the vector of the frequencies of the external perturbation with harmonics $m\Omega$, and n labels the harmonics of natural oscillations with the unperturbed frequencies

$$\omega(I)=\frac{\partial H_0(I)}{\partial I}. \quad (1.2)$$

Each perturbation term in (1.1) defines a primary resonance:

$$\omega_{nm}\equiv n\omega(I)+m\Omega\approx 0. \quad (1.3)$$

When the oscillations are linear, the frequencies are parameters of the system, which either does or does not wind up in resonance regardless of the initial conditions of motion. The most important feature of nonlinear oscillations is the fact that the oscillation frequencies change in the process of motion because of their dependence on the action variables. Below we examine the case of strong nonlinearity, i.e., when such dependence is present even for the unperturbed frequencies:

$$\frac{\partial\omega(I)}{\partial I}=\frac{\partial^2 H_0}{\partial I^2}\neq 0. \quad (1.4)$$

In this broad class of dynamical systems, the description of nonlinear resonances and the interaction of such resonances proves to be universal and relatively simple.[2]

In numerical models it is more convenient to replace differential equations in terms of the continuous time variable by a discrete map.[1–4] One model of this type that is simple and yet rich in content is the so-called standard map[2] (for a history of the model and related physical applications see Ref. 5):

$$\overline{p}=p+K\sin x, \quad \overline{x}=x+\overline{p}, \quad (1.5)$$

where p and x are action–angle variables. $K \ll 1$ is the sole parameter in the model, which characterizes the effect of the perturbation over a single map period, which is taken to be unity. In terms of the continuous time variable, this model is described by the Hamiltonian (cf. (1.1))

$$H(p,x,t)=\frac{p^2}{2}+K\sum_{m=-\infty}^{\infty}\cos(x-m\Omega t), \quad (1.6)$$

where $\Omega=2\pi$ is the fundamental frequency of the external perturbation. The model comprises an infinite set of highly nonlinear resonances ($\partial^2 H/\partial p^2=1$), with the motion near each of these resonances being identical to within a shift in the action variable: $p-p_m\to p$, where $p_m=m\Omega$ is the resonant value of the action. The frequency of small natural oscillations in any resonance is $\omega_0=\sqrt{K}\ll 1$. Although all the resonances have the same amplitude (K), all except the fundamental, which is governed by the initial conditions, constitute an extremely weak perturbation. This is explained by the high perturbation frequency compared to the natural oscillation frequency at the main resonance:

$$\lambda=\frac{\Omega}{\omega_0}=\frac{2\pi}{\sqrt{K}}\gg 1. \quad (1.7)$$

Interaction among the these resonances is adiabatic, and the effect of such interaction is exponentially small in the large adiabaticity parameter λ.

Neglecting weak nonadiabaticity, we can use the pendulum Hamiltonian

$$H_1(p,x)=\frac{p^2}{2}+K\cos x \tag{1.8}$$

to describe an isolated resonance, say, the one with $m=0$ (see (1.6)). Such a simple form of the resonance Hamiltonian proves to be universal in the case of strong nonlinearity.[2] The most important feature of this problem, which characterizes the nonlinear resonance (1.8), is the presence of a separatrix,

$$p_s=\pm 2\omega_0 \sin(x/2), \quad H_1^{(s)}=\omega_0^2=K, \tag{1.9}$$

a special trajectory that separates phase oscillations (in resonance) from phase rotation (out of resonance). Clearly, near the separatrix the motion of the system specified by (1.8) is highly unstable, since almost any arbitrarily small perturbation changes the nature of the motion dramatically (from rotation to oscillation and vice versa). It is here that chaos can emerge in nonlinear oscillations. As far as we know, this was first observed via numerical modeling,[6] and was subsequently studied by many researchers (see, e.g., Refs. 1–4 and 7).

A simple example of such weak nonadiabaticity is the interaction of just two nonlinear resonances, say, with $m=0$ and $m=1$, which can be described by the Hamiltonian (see (1.6))

$$H_2(p,x,t)=\frac{p^2}{2}+K\cos x+\epsilon\cos(x-\Omega t), \tag{1.10}$$

where ϵ is the amplitude of the perturbing resonance. Motion near the separatrix of the main resonance ($m=0$) can be approximated by the so-called separatrix map, which was first introduced (implicitly) by Zaslavskiĭ and Filonenko.[7] We write this map in the form used in Ref. 2 (see also Ref. 4):

$$\bar{w}=w+W\sin\phi, \quad \bar{\phi}=\phi+\lambda\ln\frac{32}{|\bar{w}|}. \tag{1.11}$$

Here $w=H_2/K-1$ is the dimensionless (energy) deviation from the unperturbed separatrix, $\phi=\Omega t$ is the perturbation phase at the instant when the ''pendulum'' passes the position of stable equilibrium ($x=\pi$), and W is the nonadiabaticity amplitude. If the perturbing resonance is much weaker than the main resonance ($\epsilon\ll W$), the amplitude W can be calculated relatively simply by the Mel'nikov method[8] (see also Refs. 2–4), and for $\lambda\gg 1$ we can write

$$W\approx 8\pi f\frac{\epsilon}{K}\lambda^2 e^{-\pi\lambda/2}. \tag{1.12}$$

Note that this problem cannot be solved by perturbation techniques, because the dependence of this (or in fact any other) nonadiabatic effect ($W(K)$) on the initial perturbation parameter (K) is not analytic at $K=0$. However, after this effect has been isolated, we can employ perturbation-theory techniques, and this has been successfully done for approximate equations of motion of the separatrix-map type. The first to solve this problem was probably Poincaré.[9] Equation (1.12) also shows that in the given model (1.6) the contribution of other resonances ($|m|>1$) is negligible ($\lambda_m=m\lambda$). In this model, however, all the resonance amplitudes are the same, $\epsilon=K$ (see Eq. (1.6)). This leads to the emergence of an additional factor $f\sim 1$ in (1.12). A qualitative explanation of the emergence of this factor as an effect of higher-order approximations of perturbation theory can be found in Ref. 2, where the value of this factor obtained from numerical models is also given:

$$f=2.15\pm 0.04. \tag{1.13}$$

The low accuracy of the measurements (which in fact turned out to be an overestimate) can be explained by the insufficiently small values of K ($K=0.1-1$) and by the approximate nature of the separatrix map itself.

Considerable progress in solving this problem was made only relatively recently (in 1984) by Lazutkin *et al.*[10,11] The value of the correction factor $f=2.2552\ldots$ was obtained by numerically solving an auxiliary equation from which the exponential factor had been eliminated. In contrast to the general expression (1.12), the factor f is not universal as assumed in Ref. 10 but depends on the specific set of interacting resonances.[12]

In their mathematical work, Mel'nikov,[8] Poincaré,[9] Lazutkin *et al.*,[10] and Gelfreich *et al.*[11,12] and others calculated not the effect of nonadiabaticity in (1.11), which was later studied by physicists,[1–7] but an auxiliary quantity, the separatrix splitting angle α. Although this quantity alone is insufficient for reproducing the detailed dynamics near the separatrix, α is an important characteristic of resonance interaction and, in contrast to W, is rigorously defined and can be calculated to any precision.

Separatrix splitting can be approximately described by the map (1.11) as follows. When there is no perturbation ($W=0$), each of the branches of the separatrix (1.9) is an asymptotic trajectory with an infinitely long period, a trajectory that leaves ($w=0$, $t\to\infty$) the position of unstable equilibrium, $x=0 \bmod 2\pi$, and returns to it ($\bar{w}=0$, $t\to+\infty$). When a perturbation is turned on ($W\neq 0$), two intersecting trajectories emerge: one still originates at $x=p=0$ ($w=0$, $t\to-\infty$) but never returns to that point ($\bar{w}\neq 0$, $t\to\infty$), and the other behaves as the first would under time reversal ($t\to -t$). The free ends of the split separatrix form an infinite number of loops with unboundedly increasing lengths,[8,9] which, however, fill a limited and narrow region along the unperturbed separatrix, forming a so-called chaotic layer.[1–7] One important characteristic of such a layer is the layer half-width $w_s\approx\lambda W\approx 4\alpha/\omega_0$ (see Ref. 2), which is directly related to the separatrix splitting angle (see (1.15)). These layers are the universal, ultimate source of chaos in nonlinear oscillations.

The two branches of the split separatrix intersect, in particular, at $x=\pi$ and a certain $p_s(\pi)\approx p_0=2\omega_0$ (see (1.19)). It is at this intersection that the angle α is usually studied, and this is also true of the present work. The intersection of separatrices corresponds to the value $\phi(\pi)=0$ in (1.11).

Near the intersection point the deviation of the two branches of the separatrix from the unperturbed separatrix is described by a simple approximate formula:

$$q_{\pm}(y)=p_{\pm}(x)-p_s(x)\approx\pm W\frac{\omega_0}{4}\sin\frac{\lambda y}{2}, \tag{1.14}$$

where $y=x-\pi$, and we have employed the fact that $dw/dp=2/\omega_0$ and $\phi=\Omega t=\Omega y/p_0=\Omega y/2\omega_0$. Moreover, the variation of w at the intersection point is half the total variation of w in (1.11), since the latter is symmetric about $x=\pi$ (see Ref. 2). As a result, for the total separatrix splitting angle we have

$$\alpha(\lambda)\approx 2\frac{dq}{dy}\approx\frac{\omega_0\lambda W}{4}=2\pi f\omega_0\lambda^3\frac{\epsilon}{\omega_0^2}e^{-\pi\lambda/2}$$

$$=(2\pi)^4 f\frac{e^{-\pi\lambda/2}}{K}, \tag{1.15}$$

where we have used the standard-map parameter $\epsilon=\omega_0^2=K$ and $\Omega=2\pi$. Note that this simple and important relationship holds only for the separatrix splitting angle at $x=\pi$.

The quantity $\mathscr{S}$ introduced in Refs. 10 and 11, which we call the Lazutkin constant, is related to the correction factor f:

$$\mathscr{S}=16\pi^3 f=1118.82770595\ldots\ . \tag{1.16}$$

According to our data, the most accurate value of $\mathscr{S}$ is given below, in (4.14). Note that the factor $f\ll\mathscr{S}$ more correctly characterizes the order of the effect of the finite amplitude of a perturbing resonance.

The last term in (1.15), with an exact value of $\mathscr{S}$ or f, yields the asymptotic value (as $\lambda\to\infty$) of the separatrix splitting angle: $\alpha_\infty=\alpha(\infty)$. In Ref. 10 there is also an estimate of the correction to α_∞ in the intermediate asymptotic region $0<K\ll 1$:

$$c_\alpha(\lambda)=\frac{\alpha(\lambda)}{\alpha_\infty}-1\gtrsim K^{1/8}. \tag{1.17}$$

Our preliminary numerical models have shown, however, that the correction decreases with K much faster: $c_\alpha\sim K^{1/2}$.

We immediately note that this ''correction'' describes the intermediate asymptotic region of separatrix splitting that is of present interest, and hence the formation of a chaotic layer. Solving this problem is the principal goal of our investigation.

The best-developed theory of the standard map separatrix splitting[11] predicts not only a rapid decrease in this correction, but also makes it possible to calculate the first terms in the power-series expansion by numerically solving auxiliary equations. We believe that to a large extent the success of this theory can be attributed to the felicitous change of variables $(K,\alpha)\to(h,\sigma)$, where

$$h(K)=\ln\left(1+\frac{K}{2}+\sqrt{K+\frac{K^2}{4}}\right)\approx\sqrt{K} \tag{1.18}$$

is the positive characteristic index of the tangential (linearized) map (1.5) at the unstable fixed point $x=p=0$,

$$\sigma(h)=\nu(h)\sin\alpha \tag{1.19}$$

is the simplectic invariant, and $\nu(h)$ is a certain norm of the tangent vectors.

Gelfreich *et al.*[11] found an approximate solution to this problem that can also be written as a correction (similar to (1.17)) to the invariant:

$$c_\sigma(h)=\frac{\sigma(h)}{\sigma_\infty}-1=\sum_{m=1}^{\infty}a_\sigma(m)h^{2m}, \tag{1.20}$$

where $\sigma_\infty=4\alpha_\infty$. Actually, Gelfreich *et al.*[11] calculated only the first four terms in the series (see Table II below). The main limitation of this solution is the implicit function $\sigma(\alpha)$, which can be found only numerically because of the additional unknown function $\nu(h)$.

In the present paper we present the first results of direct measurements of the splitting angle of the separatrix of the standard map (1.5), carried out over a broad range of the parameter K: $1\geqslant K\geqslant 0.0004$ $(1\gtrsim h\gtrsim 0.02)$. Here α is $0.1\gtrsim\alpha\gtrsim 10^{-208}$, with a relative accuracy of better than one part in 10^{-25} and an average accuracy of roughly one part in 10^{-30}. In order to solve this problem we developed a special method for measuring and processing the results that used fully portable software[13] with arbitrary-precision arithmetic. In fact, the accuracy of the numerical model was as high as one part in roughly 10^{-300}.

In addition to measuring the angle α, which is directly related to the nonadiabaticity W, we measured the invariant σ (Eq. (19)) it order to compared to the theoretical value. The function $\nu(h)$, which is needed if we wish to calculate σ but whose analytic expression is not known, was calculated using a special program kindly furnished by V. G. Gelfreich, to whom we are sincerely grateful. All this has made it possible to compare our results with those of the analytic theory of Gelfreich *et al.*[11] in detail, which was fully substantiated, and to detect new effects. We found a relatively simple empirical expression for the K dependence of α in the intermediate asymptotic region, which proved to be surprisingly broad: $K\lesssim 10^{-2}$. We also studied the effect of noise—angle measurement errors, in particular—which turned out to be much more significant and complex than expected.

2. MEASURING THE SEPARATRIX SPLITTING ANGLE

The main difficulty in measuring the separatrix splitting angle α relates to its extremely small magnitude. For instance, when the perturbation parameter K in (1.5) is 0.0004, α amounts to roughly 4.2×10^{-208} rad. As noted in the Introduction, this problem could only be solved thanks to a special-purpose software package[13] that implements all standard FORTRAN capabilities in arbitrary-precision arithmetic, where one can specify the number $\mathscr{N}_c$ of significant figures in the mantissa of the decimal representation of a real number.

We calculated α in the following manner. First and foremost, for any given value of the perturbation parameter K, the ordinate $p_s(\pi)$ of the point of intersection of the separatrix branches was found on the line $x=\pi$. The unperturbed theoretical value $p(\pi)=p_0=2\sqrt{K}$ (Eq. (1.9)) was taken as

the initial ordinate, and the behavior of the phase x (rotation or oscillation) was established for the orbit starting at $(\pi, p(\pi))$. Depending on the result, the value of $p(\pi)$ was successively increased or decreased by a certain step size until the counterpart phase behavior emerged. This made it possible to initially localize the intersection point, with the upper bound $p_r(\pi)$ corresponding to rotation of the phase x, and the lower bound $p_l(\pi)$ to phase oscillations independent of the direction of time. Recall that each split branch of the separatrices remains a boundary between oscillations and rotations of the phase for its own particular direction of time (see the Introduction). The step size in momentum was always chosen larger than the expected width of the chaotic layer ($\approx 4\alpha$), so that the width of this initial interval, $dp(\pi) = p_r(\pi) - p_l(\pi)$, turned out to be much greater than the required value $dp_s(\pi)$, which, in contrast, was always chosen to be much less than the width of the layer (see below). For example, for our smallest value of the perturbation parameter, $K = 0.0004$, the width of the chaotic layer was approximately 2×10^{-207}, while the multiple-precision level $\mathscr{N}_c$ and the precision required to yield $dp_s(\pi) = 10^{-\mathscr{N}_a}$ in locating the ordinate $p_s(\pi)$ were set at $\mathscr{N}_c = 300$ (corresponding to a precision of about one part in roughly 10^{-300}) and $\mathscr{N}_a = 280$, respectively.

After establishing the initial interval $[p_l(\pi), p_r(\pi)]$, we used the bisection method to get to the point $dp(\pi) \leqslant dp_s(\pi)$, and then calculated the desired ordinate of the intersection point, taking $p_s(\pi) = [p_r(\pi) + p_l(\pi)]/2$. We note in passing that this stage of the calculations accounted for most of the computing time (see below): as the interval shrank, the orbit moved closer to the unstable fixed point, and the period of motion (the number of iterations of the map) increased substantially.

After the central point of intersection of the separatrix branches was established, it was adopted as the origin of a new coordinate system ($y = x - \pi$, $q = p - p_s(\pi)$), in which all subsequent steps in calculating the angle α were taken. For two values of the phase, $y = \pm \delta y$, $|\delta y| \ll 1$, to the right and left of the new origin ($y = q = 0$) we found (to the same accuracy $dp_s(\pi) = 10^{-\mathscr{N}_a}$) the boundaries between oscillations and rotations (with time increasing toward the future). This made it possible to approximate a section of the straight branch of the separatrix by a second-degree polynomial $q = Q_2(y)$ and, by determining its coefficients, to find the angle of the slope of the separatrix. Repeating the above procedure with time progressing into the past, we were able to find the slope of the reverse branch of the separatrix and take the difference of the two, yielding the desired intersection angle α_2.

Special measures were taken to ensure the feasibility of this approach.

For the full range of the parameter K we strove to obtain a relative measurement accuracy in α no worse than one part in 10^{-25}. To this end we were forced to choose values for all the multiple-precision parameters mentioned above: the number of significant figures $\mathscr{N}_c$, the roundoff error $dp_s(\pi) = 10^{-\mathscr{N}_a}$ incurred in locating the ordinate $p_s(\pi)$, and the phase offset δy needed to set up the approximating polynomials, with proper selection of the third parameter being the most critical. According to (1.14), the phase separation Δy between the central and neighboring intersection points is $\Delta y \approx 2\pi/\lambda = \sqrt{K}$, and it is convenient to choose as a phase offset some fraction of this distance,

$$\delta y = \eta \sqrt{K}, \tag{2.1}$$

with the value of η for all numerical alternative usually ranging from 10^{-16} to 10^{-12}.

Above all, it was necessary to understand whether the desired relative accuracy of one part in 10^{-25} in the calculated value of α_2 had been achieved (clearly, this accuracy must in no event be identified with the number of significant digits $\mathscr{N}_c$). To this end we repeated the above calculations for one more pair of double phase values, $y = \pm 2\delta y$, for both senses of time. Using the results for the two pairs of points, we were able to construct fourth-degree approximating polynomials $q = Q_4(y)$ and obtain an improved value of α_4.

Generally, some of the $\mathscr{N}_a$ of the leading significant digits (with allowance for round-off) in the expressions for α_2 and α_4 are the same, but then these values diverge. We always achieved $\mathscr{N}_a \geqslant 25$, which yielded a relative angle-measurement accuracy no worse than one part in 10^{-25}. Calculations with higher-degree polynomials and higher precision only supported this criterion. In actual fact, the relative accuracy turned out to be higher, and on the average amounted to roughly one part in 10^{-30} (see Sec. 3).

Note that there is no need to calculate the points of the separatrix branches time proceeding into the past. The standard map (1.5) is symmetric, and this symmetry makes it possible to reconcile any point (x_f, p_f) on the forward branch of the separatrix with the corresponding point (x_b, p_b) on the backward branch via the simple relationship[11]

$$x_b = 2\pi - x_f, \quad p_b = p_f + K \sin x_f. \tag{2.2}$$

Thus, the approximating polynomials for both branches are constructed simultaneously, so that there is no need for additional computer time. However, to verify our results we did check calculations of some of the values of the perturbation parameter with time progressing into the past. These calculations validated the numerical scheme describe above.

The major difficulty is to find the coefficients of the polynomials that approximate the segments of the separatrices that pass through the intersection point $y = q = 0$:

$$q = a_1 y + a_2 y^2 + a_3 y^3 + a_4 y^4. \tag{2.3}$$

Using four reference points, we obtain a set of linear equations represented in matrix form as

$$\begin{pmatrix} \delta y & (\delta y)^2 & (\delta y)^3 & (\delta y)^4 \\ -\delta y & (\delta y)^2 & -(\delta y)^3 & (\delta y)^4 \\ 2\delta y & 4(\delta y)^2 & 8(\delta y)^3 & 16(\delta y)^4 \\ -2\delta y & 4(\delta y)^2 & -8(\delta y)^3 & -16(\delta y)^4 \end{pmatrix}. \tag{2.4}$$

The elements of this matrix differ by many orders of magnitude (in view of the smallness of the phase offset, δy, according to (2.1)), and its determinant is close to computer zero, which makes it impossible to invert the matrix.

The problem was solved through a change of variables of the form $(y,q)\to(Y=Sy,\ Q=Sq)$. In terms of the new variables the polynomial was

$$Q=a_1Y+b_2Y^2+b_3Y^3+b_4Y^4. \tag{2.5}$$

We see that the only coefficient of interest to us is $a_1=\tan\alpha_4$, which remains constant for any value of the scaling factor S. This is obvious, since our change of variables amounts only to magnifying the entire picture S-fold. We usually set $S=1/\delta y$, which balanced the orders of magnitude of the matrix elements and resolved all difficulties with processing such a matrix.

We used the above scheme to find $\alpha(K)$ for sixty values of the perturbation parameter K in the range $1\geqslant K\geqslant 0.0004$. The angle was represented to 35 significant digits, which provided more than the minimum (one part in 10^{-25}; see above) and average (one part in roughly 10^{-30}; see Sec. 3) relative accuracy needed to determine the angle.

The minimum perturbation parameter that we could work with, $K=0.0004$, was limited by the time per run (about 24 hours). This time rapidly increases as K gets smaller primarily because of an increase in the number of necessary iterations of the map (Eqs. 1.5). The latter is the product of the period of motion near the separatrix $(\sim|\ln dp_s(\pi)|/\sqrt{K}\sim\mathscr{N}_a/\sqrt{K}$; see Ref. 2) and the number of successive approximations at the intersection point $(\sim|\ln dp_s(\pi)|\sim\mathscr{N}_a)$. Moreover, the time it takes to calculate (1.5) is proportional to $\mathscr{N}_c$ (in the range of moderate accuracy $\mathscr{N}_c\lesssim 300$ that we are interested in). As a result, we estimate the total processing time T with this approach to be

$$T\approx T_0\frac{\mathscr{N}_c\mathscr{N}_a^2}{\sqrt{K}}, \tag{2.6}$$

where the empirical value of the parameter T_0 is roughly 2×10^{-8} h.

3. PROCESSING THE MEASUREMENTS

The initial empirical data used our analysis of $\alpha(K)$ dependence consisted of 60 measured values of the unknown function (Sec. 2). In the first stage of processing, to calculate the corrections c_α we used the most accurate value of $\mathscr{L}$ (1.16).[10] In order to compare our result with the theoretical result of Ref. 11, we processed in a similar way the empirical dependence of the correction c_σ to the invariant σ calculated by (1.19).

In accordance with the theory developed in Ref. 11, both corrections were sought in the form of a finite series expansion in even powers of h.

$$\tilde{c}(h)=a(0)+\sum_{m=1}^{M}a(m)h^{2m}, \tag{3.1}$$

with least-squares interpolation (see, e.g., Ref. 14). Although formally the coefficient $a(0)$ is zero (see (1.17) and (1.20)), incorporating it into (3.1) makes possible a considerable improvement in the parameter $\mathscr{L}$ in comparison to its known value (1.16).

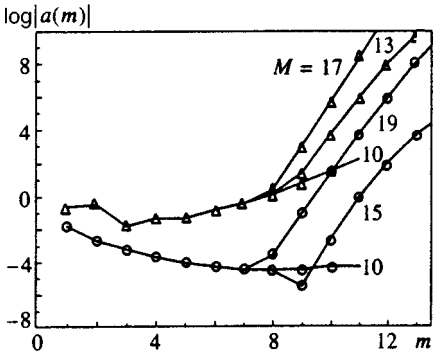

FIG. 1. An example of the dependence of the expansion coefficients in (3.1), found by interpolation of the empirical data, on the total number M of terms in the series for the angle (open triangles) and the invariant (open circles).

The main difficulty in the interpolation of the empirical data via the series (3.1) lies in the fact that the various terms of this series differ by many orders of magnitude. For example, in the typical case where $h\sim0.05$ and $M=10$, the ratio $(a_{10}/a_1)h^{18}$ is of order 10^{-27}. Hence, despite the special methods used in processing the data, even in the initial stage the processing was done with quadruple precision, and in the final stages the required precision amounted to one part in roughly 10^{-100}. But even this was not enough. As in the interpolation of the separatrix branches (Sec. 2), the matrix of the system of linear equations that was so highly inhomogeneous usually became singular during numerical processing. The solution to this problem was found in the same way as in Sec. 2, i.e., by scaling the variables of the problem: $(h,\tilde{c})\to(H=Sh,\ C=S\tilde{c})$, with the scaling factor S much larger than unity. In contrast to the problem of Sec. 2, the coefficients of the polynomial (3.1) are not invariants under such a transformation, and must be restored after interpolation in the new variables: $A(m)\to a(m)=A(m)S^{2m-1}$.

The number of terms in the series (3.1), $M\sim10$, is bounded above by errors in the calculation of the coefficients $a(m)$ primarily because of ''noise'' resulting from the finite precision of the empirical data on $\alpha(h)$. If M is increased beyond 10, we obtain no new coefficients—we even lose some of the old ones. This is especially evident in Fig. 1 from the sharp break in $a(m)$ dependence. We chose the optimum value $M=10$ by trial and error (see also Fig. 3 below).

The accuracy of the empirical formula (3.1) can be characterized by the root-mean-square (rms) error

$$\Delta c=\langle[c(h)-\tilde{c}(h)]^2\rangle^{1/2}, \tag{3.2}$$

where the angle brackets denote averaging over the entire interpolation interval. The latter does not necessarily include all 60 values of $c(h)$. Furthermore, attempts to use the entire empirical interval have revealed the extremely low accuracy of such ''global'' interpolation: $\Delta c(h)\gtrsim10^{-6}$ (cf. Fig. 2). This is quite natural, since the theoretical power-law dependence[11] (3.1) characterizes only the intermediate asymptotic region. For this reason, the deviation

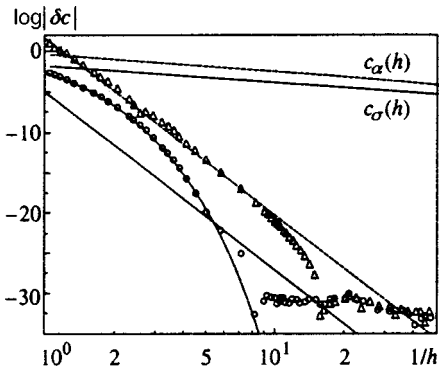

FIG. 2. Interpolation of the empirical data on the separatrix splitting angle (open triangles) and the invariant (open circles); $\delta c(h)$ is the deviation from the intermediate asymptotic behavior specified by (3.1). The oblique straight lines represent the first term of the remainder (4.7), and the curves represent the exponential deviation (4.6). The lines at the top represent the total corrections $c(h)$ for the angle and invariant, respectively.

$$\delta c(h)=c(h)-\tilde{c}(h) \tag{3.3}$$

contains the most valuable and interesting information about additional nonadiabatic effects not in the theory. Thus, in the given set of empirical data, it was also necessary to choose an interpolation range (h_1-h_2), where $h_1=0.02$ is the minimum value in the original data. The basic criteria here were the smallest value of Δc (see Eq. (3.2)) and the precision of the derived coefficients of the series (3.1) (see Tables I and II).

The quality of interpolation decreases not only as h_2 increases (as noted earlier), but also as it decreases, because of the small contribution of high powers of h for small values of h, and also due to the decreasing number of points N_p participating in the interpolation. The interpolation is best when $h_2\approx 0.063$ $(N_p=19)$ and $\Delta c_\alpha\approx 2\times 10^{-31}$, and when $h_2\approx 0.12$ $(N_p=36)$ and $\Delta c_\sigma\approx 2\times 10^{-31}$, but other values of N_p close to those just mentioned were also used (see Sec. 4). Note that the eventual interpolation accuracy yields a rough estimate of the average accuracy of the measured values of $\alpha(K)$.

The basic results were obtained through standard interpolation by minimizing the variance $(\Delta c)^2$ (see (3.2) and (3.1)). To control the procedure, we used forward interpolation, in which for the zeroth-order unknown coefficient $a(0)$ in expressions of type (3.1) we took the total value of the Lazutkin constant. Relative interpolation was also carried out, and with it we minimized the relative variance

$$\frac{(\Delta c)^2}{c^2}=\left\langle\left[1-\frac{\tilde{c}(h)}{c(h)}\right]^2\right\rangle. \tag{3.4}$$

The results of all three interpolation schemes are in good agreement with one another.

The results of processing the data are depicted in Fig. 2 and listed in Tables I and II.

The accuracy of the coefficients was established in two different ways. First, we calculated the standard rms interpolation error[14] (columns headed $\langle\delta\rangle$ in Tables I and II). This quantity characterizes the expected error in calculations of the coefficients for a random rms error in the empirical data. Actually, however, the total error is almost never purely random but contains a systematic error, which leads to a shift (additional variation) in the values of the coefficients. This is clear in Fig. 2, where the accuracy of interpolation is much greater than the remainder in the series (1.20), which is not included in the interpolation of (3.1). This is why another method was used. The values of the coefficients were determined by averaging over several interpolations with various amounts of initial data: $N_p=14-20$ for the angle and $N_p=33-38$ for the invariant. These values are listed in Tables I and II (columns headed $a_\alpha(m)$ and $b_\sigma(m)$, respectively). The relatively weak dependence of the averages on N_p served as the main criterion in selecting these two groups. For the error we took the rms errors of the coefficients in a group (columns headed Δ in the tables). We see that the two estimates are of the same order, although the error in a group is the greatest, and hence dominant, in all cases except the one with $b(10)$. The difference between the two rms errors, which is especially appreciable for the angle, definitely suggests that there is a systematic error. The values of the rms error in a group determine the number of valid significant figures in the coefficients (in our opinion). In the columns headed $a_\alpha(m)$ and $b_\sigma(m)$, we have left two or three ''superfluous'' digits for comparison with future more accurate empirical and/or theoretical values.

TABLE I. Coefficients $a_\alpha(m)$ in the series (3.1) for the angle.

m	$a_\alpha(m)$	Δ	$\langle\delta\rangle$
1	−0.23337 64288 64381 61062 76396 19	0.239×10^{-24}	0.118×10^{-24}
2	−0.29081 81551 24688 86036 776	0.101×10^{-20}	0.453×10^{-21}
3	−0.01482 49555 34894 05088 4	0.240×10^{-17}	0.973×10^{-18}
4	0.04318 21901 48643 630	0.357×10^{-14}	0.130×10^{-14}
5	−0.04151 92394 77208	0.348×10^{-11}	0.115×10^{-11}
5	−0.13137 33101 9	0.227×10^{-08}	0.673×10^{-09}
7	−0.31916 9504	0.983×10^{-06}	0.261×10^{-06}
8	−1.06063 5	0.273×10^{-03}	0.647×10^{-04}
9	−4.3613	0.439×10^{-01}	0.923×10^{-02}
10	−24.02	$0.312\times 10^{+01}$	$0.579\times 10^{+00}$

Note: Here and in Table II, Δ is the rms error in the group, and $\langle\delta\rangle$ is the average rms error in an individual interpolation.

TABLE II. Coefficients $b(m)$ in the series (4.1) for the invariant.

m	$b_\sigma(m)$	Δ	$\langle\delta\rangle$
1	18.59891 19582 09297 35881 71520	0.101×10^{-22}	0.343×10^{-23}
2	−4.34114 12705 68162 53678 60	0.369×10^{-19}	0.125×10^{-19}
3	−4.18326 37590 91894 112	0.971×10^{-16}	0.345×10^{-16}
4	−4.93413 95907 30929	0.186×10^{-12}	0.735×10^{-13}
5	−10.64548 64427 41	0.263×10^{-09}	0.121×10^{-09}
6	−35.86008 1765	0.276×10^{-06}	0.151×10^{-06}
7	−177.60356	0.212×10^{-03}	0.139×10^{-03}
8	−1239.507	$0.114\times10^{+00}$	0.894×10^{-01}
9	−11766.0	$0.386\times10^{+02}$	$0.362\times10^{+02}$
10	−163000	$0.627\times10^{+04}$	$0.693\times10^{+04}$

Note: The underlined figures are the values of the coefficients obtained in Ref. 11.

4. DISCUSSION

We start by examining the behavior of the expansion coefficients in (3.1) on the basis of the data in Tables I and II. We use the representation of the coefficient in the simplest form (3.1) except when we compare our results directly with those of Ref. 11, where the coefficients $b(m)$ are represented by the Taylor series (cf. (1.20))

$$\frac{\sigma}{\sigma_\infty}\mathscr{L}=\mathscr{L}+\sum_{m=1}^{\infty}\frac{b(m)}{m!}h^{2m},\quad b(m)=a(m)m!\,\mathscr{L}. \tag{4.1}$$

First and foremost, we were able to find a relatively simple extrapolation of $a(m)$ outside the range of direct measurements of the separatrix splitting angle. The results of this extrapolation are depicted in Fig. 3 and can be described by the approximate expressions

$$a_\alpha(m)\approx A_\alpha\frac{e^{\gamma m}}{m^p} \tag{4.2a}$$

for the angle, and

$$a_\sigma(m)\approx\frac{A_\sigma}{m^p} \tag{4.2b}$$

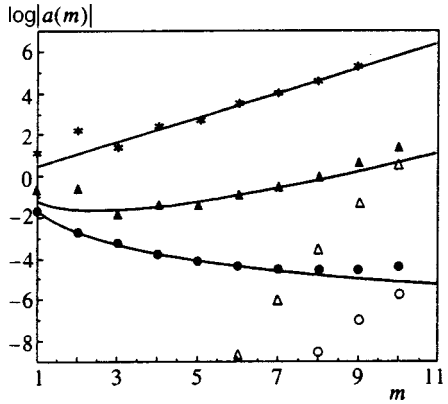

FIG. 3. Variation of the coefficients in the intermediate asymptotic region specified by (3.1) for the angle $a_\alpha(m)$ (filled triangles), the invariant (filled triangles), the invariant $a_\sigma/(m)$ (filled triangles), and their a_α/a_σ ratio (asterisks). The solid curves represent the empirical formulas (4.2a) and (4.2b). The open triangles and circles represent the errors in the coefficients.

for the invariant.

These empirical relationships were built according to the following reasoning. From Fig. 2 and Tables I and II we see that the positive correction $c_\sigma(h)>0$ is much smaller in absolute value than $c_\alpha(h)<0$ over the entire range of h studied here, and the two corrections have opposite signs. This is a consequence of the behavior of the norm $\nu(h)=\nu_\infty(1+c_\nu(h))$ ($\nu_\infty=4$ and $c_\nu(h)>0$) in (1.19). This yields the following relationship between the coefficients:

$$a_\sigma(m)=a_\nu(m)+\sum_{k=1}^{m}a_\alpha(k)a_\nu(m-k). \tag{4.3}$$

All the norm coefficients $a_\nu(m)$ are positive, while almost all $a_\alpha(m)$ and $a_\sigma(m)$ are negative (with the exception of $a_\alpha(4)$ and $a_\sigma(1)$; see Tables I and II). On the other hand, Fig. 3 shows that the coefficients for the angle and invariant behave quite differently. The latter decrease relatively slowly with increasing m, approximately as power-law functions, while the former increase very rapidly, almost exponentially, as the norm coefficients do. This is corroborated by forward interpolation and directly follows from the fact that the norm balances the large correction for the angles almost perfectly: $c_\sigma\ll|c_\alpha|$.

This, then, is the key to building the empirical relations (4.2a) and (4.2b). Indeed, such balance does not occur for a purely exponential dependence of $a_\alpha(m)\approx a_\nu(m)$. One can easily verify, however, that this becomes possible for the combined dependence (4.2a), provided that $p>1$ and $m\gg1$. Figure 3 also depicts the dependence of a_α/a_σ on m, which is perfectly fit by an exponential (see (4.4) below). This is all the more remarkable since the last two or even three coefficients of the invariant clearly deviate from the simple power law (4.2b). The same is true of the combined dependence (4.2a) for the angle.

The nature of these anomalies remains an open question, and requires further study. Note that anomalies appear only in the highest-order coefficients, which cannot be increased in number without a catastrophic increase in the errors (Fig. 1). Furthermore, the ratio of the coefficients of the angle and invariant in the intermediate asymptotic region contains no obvious anomalies. Of course, the exact dependence of the coefficients, provided that it can be expressed explicitly, hardly has the simple form of (4.2a) and (4.2b) even when

$m \gg 1$. This is clear, if only from the fact that the accuracy of the empirical relations can in no way be compared to the precision of the coefficients themselves (see Tables I and II and Eqs. (4.4) below). Nevertheless, even the approximate formulas (4.2a) and (4.2b) help in interpreting the characteristic features of the intermediate asymptotic region and in further studies of this problem. In this paper, however, we have limited ourselves to this approximation.

The parameters of the empirical relations can be obtained by interpolating data on the ratio of the coefficients with $m=7-10$ on a semilogarithmic scale, and data on the invariant with $m=4-6$ on a log–log scale. The results are as follows:

$$\gamma = 1.358 \pm 0.0059, \quad p = 3.51 \pm 0.17,$$
$$A_\alpha = -0.0174(1 \pm 0.28), \quad A_\sigma = -0.0233(1 \pm 0.27), \tag{4.4}$$

where the accuracy is given in terms of absolute and relative rms errors. Note that interpolation with just three coefficients $(m=2,3,4)$, obtained in Ref. 11, yields similar values of the parameters in (4.2b):

$$A_\sigma = -0.0216(1 \pm 0.45), \quad p = 3.38 \pm 0.41. \tag{4.5}$$

The accuracy of the empirical relations (4.2a) and (4.2b) can also be characterized by the relative rms error of the extrapolation itself ($\Delta a/|a|$ for the values of m specified above), which amounts to ± 0.029 for the angle and ± 0.028 for the invariant. Note that the relative smallness of the rms errors themselves, compared to the rms errors in the coefficients (4.4), can be explained in terms of the strong correlation of the latter. In both cases the anomalies of the coefficients are much larger than the reduced errors in the values of (4.2a) and (4.2b) and the two errors of the coefficients themselves in Tables I and II (see Fig. 3).

Using (4.2a) and (4.2b), we can set up an approximate model of the intermediate asymptotic behavior, taking for the initial coefficients, which clearly do not obey (4.2a) and (4.2b), their exact values from Tables I and II. Preliminary experiments along these lines show that the model indeed reproduces the shift of all (and especially the last) coefficients. However, this shift lies within the limits of error (see Tables I and II and Fig. 3) and does not explain the anomalies discussed earlier. Furthermore, even turning on additional noise that is uniform in h and simulates the errors in angle measurements does not help.

The approximate relationships (4.2a) and (4.2b) also make it possible to perceive global behavior in the intermediate asymptotic region. Above all, the series (4.2b) for the invariant is convergent over the entire range $h \leqslant 1$ up to global chaos limit, although it does not describe the actual behavior of the invariant for $h \gtrsim 0.14$. Here the presence of an exponential is clear (see Fig. 2):

$$|\delta c(h)| \approx 63 e^{-\pi^2/h}. \tag{4.6}$$

This function exceeds the remainder (4.7) and describes the perturbation of the separatrix by a distant resonance with a frequency $2\Omega = 4\pi$ (see (1.6)). The simple theory in Ref. 2 predicts a numerical factor of 8, i.e., smaller by a factor of almost 10. Such a discrepancy can easily be explained by another (unknown) value of the factor f in (1.12) and a system of resonances of higher-order approximations, which is much more complicated.[15]

The situation for the separatrix splitting angle is much more interesting, since the series (4.2a) diverges when $h > h_{\rm cr} \approx \exp(-\gamma/2) \approx 0.507$, i.e., within the range under investigation (15 leftmost points in Fig. 2). At the same time, no singularities or anomalies in the behavior of the function $c_\alpha(h)$ or its deviation $\delta c_\alpha(h)$ from interpolation has been observed in this range. Furthermore, the deviation can be described perfectly well by the remainder term in (4.7), which, incidentally, is represented in Fig. 2 only by its first term. This clearly shows that there is a significant change in the behavior of both the angle and the invariant at large $h \sim 1$ as compared with the intermediate asymptotic region.

The approximate relations (4.2a) and (4.2b) make it possible to more accurately (than in Ref. 11) estimate the remainder in the series (1.20), which is not included in the interpolation (3.1):

$$R(h,M) = \sum_{m=M+1}^{\infty} a(m) h^{2m} \approx a(M+1) h^{2M+2}. \tag{4.7}$$

Figure 2 shows that even the first term in R provides a fairly good description of the behavior of the angle deviation over the entire range under investigation. The same can be said of the exponential function (4.6) for the invariant. However, an attempt to add both expressions to the polynomial (3.1) catastrophically reduces the interpolation accuracy: $\Delta c \sim 10^{-6}$. This again demonstrates that the structure of the region where $h \sim 1$ is extremely complicated. In view of the importance of this region in many applications, this problem deserves further study.

Finally, we focus on the most precise and accurate way of finding the Lazutkin constant $\mathscr{L}$. We introduce the correction

$$\delta \mathscr{L}_i = (\mathscr{L}_i - \mathscr{L}_0) \times 10^{23}, \tag{4.8}$$

where i labels the various ways of obtaining $\mathscr{L}$, and

$$\mathscr{L}_0 = 1118.82770\ 59409\ 00778\ 41514\ 639 \tag{4.9}$$

is the zeroth-order value, which we already obtained in preliminary numerical modeling (cf. (1.16)). Formally, the highest accuracy is achieved in individual interpolation with $N_p = 14$ and $M = 10$:

$$\delta \mathscr{L}_\alpha = 0.323560 \pm 0.000017,$$
$$\delta \mathscr{L}_\sigma = 0.323572 \pm 0.000017. \tag{4.10}$$

Since the intermediate asymptotic series (4.2) differ substantially in these two cases, such good agreement is a serious argument favoring the reality of this accuracy.

In view of the importance of this constant, we also used other methods to determine it. First we isolated a group of variants of interpolation schemes with the same value $M = 10$ but different values $N_p = 12-20$ and $N_p = 13-38$ for the angle and invariant, respectively (35 cases in all). The group was chosen to be as broad as possible, the only limitation being that $\delta \mathscr{L}_i$ was supposed to decrease rapidly and

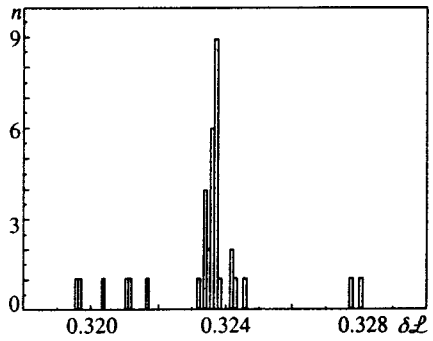

FIG. 4. Bar diagram of the correction (4.8) of the value of the Lazutkin parameter (4.9): n is the number of different values of the correction in a bar with a width of 10^{-4}.

monotonically for reasons discussed earlier. For the average value and the rms error in this group we found

$$\langle \delta\mathscr{L}_i\rangle_{35} = 0.323368 \pm 0.0017. \tag{4.11}$$

We see that the error has increased significantly. However, a histogram of the distribution of $\delta\mathscr{L}_i$ in the group (Fig. 4) shows that the main peak is much narrower. If we discard the wings of the distribution (two cases to the right and six to the left), we obtain

$$\langle \delta\mathscr{L}_i\rangle_{27} = 0.323654 \pm 0.00029. \tag{4.12}$$

Finally, leaving only the nine cases in the rightmost cell of the histogram, we find that

$$\langle \delta\mathscr{L}_i\rangle_{9} = 0.323660 \pm 0.000021. \tag{4.13}$$

Weighing the pros and cons, we conclude that the most accurate value of the Lazutkin constant and the error are

$$\mathscr{L} = \underline{1118.82770\ 59409\ 00778\ 41514\ 63932\ 3566} \pm 3\times 10^{-27}. \tag{4.14}$$

We have underline the value of $\mathscr{L}$ previously obtained with the approximate theory of Lazutkin *et al.*[10]

The results of the present work corroborate the theory of Gelfreich *et al.*[11] both qualitatively (the form of the intermediate asymptotic series (3.1)) and quantitatively (Table II). Furthermore, we have found the intermediate asymptotic region directly for the separatrix splitting angle, which is important in and of itself, and which has made it possible, when combined with the data on the invariant, to obtain approximate empirical relations (4.2) in this region that are not limited by the number of directly derived coefficients.

We are grateful to N. S. Dikanskiĭ for the opportunity to use the ALPHA-4100 computer, to V. G. Gelfreich for useful discussions and for the assistance rendered in the process of comparing our results with the theoretical results, to A. G. Grozin for providing the arbitrary-precision software[13] and helping to master the program, and to L. F. Khaĭlo for considerable assistance with numerical modeling. This work was partially supported by a grant from the Russian Fund for Fundamental Research (Grant No. 97-01-00865).

*)E-mail: vecheslavov@inp.nsk.su
†)E-mail: chirikov@inp.nsk.su

[1] G. M. Zaslavskiĭ and B. V. Chirikov, Usp. Fiz. Nauk **105**, 3 (1971) [Sov. Phys. Usp. **14**, 549 (1972)].
[2] V. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[3] R. Z. Sagdeev, D. A. Usilov, and G. M. Zaslavskiĭ, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos*, Harwood Academic, New York (1988).
[4] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Springer-Verlag, Berlin (1992).
[5] B. V. Chirikov, Zh. Éksp. Teor. Fiz. **110**, 1174 (1996) [JETP **83**, 646 (1996)].
[6] N. N. Filonenko, R. Z. Sagdeev, and G. M. Zaslavsky, Nucl. Fusion **7**, 253 (1967).
[7] G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1968) [Sov. Phys. JETP **27**, 851 (1968)].
[8] V. K. Mel'nikov, Trudy Mosk. Mat. Obsch. **12**, 3 (1963) [Trans. Moscow Math. Soc. **12**, 1 (1965)].
[9] A. Poincaré, *Les méthodes nouvelles de la méchanique céleste*, Gauthier-Villars, Paris (1892), p. 226.
[10] V. F. Lazutkin, I. G. Schachmansky, and M. B. Tabanov, Physica D **40**, 235 (1989).
[11] V. G. Gelfreich, V. F. Lazutkin, and N. V. Svanidze, Physica D **71**, 82 (1994).
[12] V. G. Gelfreich, *Reference System for Splitting of Separatrices*, Universidad de Barcelona, Mathematics Preprint Series No. 163 (1990).
[13] D. H. Bailey, ACM Trans. Math. Softw. **19**, 288 (1993).
[14] B. M. Shchigolev, *Mathematical Analysis of Observations*, Elsevier, New York (1965); D. Hudson, *Statistics*, CERN, Geneva (1964).
[15] V. V. Vecheslavov, Zh. Eksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].

Translated by Eugene Yankovsky