JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Behavior of a bundle of rays forming the image of a source in cosmological models with rotation

V. F. Panov*)

Perm' State University, 614600, Perm', Russia

Yu. G. Sbytov

''Stankin'' Moscow State Technological University, 101472 Moscow, Russia

(Submitted 26 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 769–776 (September 1998)

Our earlier results regarding the absence of Birch anisotropy in cosmological models with rotation were obtained under the assumption that a bundle of rays forming the image of a source is nonrotating. This appears to be at variance with the rotation of space. The detailed analysis in the present paper of the behavior of rays propagating from a source to an observer shows that the congruence of the rays is, in fact, nonrotating. The doubts expressed by some authors in reference to our conclusions that the Birch effect is not attributable to cosmological rotation are thereby put to rest. © *1998 American Institute of Physics.* [S1063-7761(98)00109-7]

1. INTRODUCTION

A debate as to whether it is possible to attribute the Birch effect to cosmological rotation has been continued in the literature for a fairly long time. We recall that Birch[1] claimed on the basis of his own observations of a large set of radio galaxies that there is dipole anisotropy of the position angle Δ (the angle between the dominant polarization of the radiation and the direction of maximum elongation of the radio galaxy) in the form $\Delta \propto \cos\theta$, where θ is the angle between the direction of the ray and the anisotropy axis. He also theorized that this anisotropy could be attributed to cosmological rotation.

It was shown in Ref. 2 that rotation of the polarization vector (relative to a local coordinate basis set) does, in fact, take place in a rotating cosmological model and that the rotation law is similar to the one proposed by Birch. However, no allowance was made for the possible variation of the direction of maximum elongation of the source. Such rotation occurs because of deformation of the bundle of rays forming the image of the source as the rays propagate in curved space. Therefore, the results in Ref. 2 did not answer the question of what happens to the angle Δ in rotating universes. We[3] recently investigated the character of the variation of the position angle Δ in a cosmological model with global Gödel rotation. It was discovered that this rotation cannot induce the Birch type of anisotropy with $\Delta \propto \cos\theta$ and that anisotropy with $\Delta \propto \sin^2\theta$ should be expected. It was concluded in Ref. 3 on this basis that the Birch effect, if it exists at all, cannot be caused by rotation of the universe.

It was shown in Ref. 4 that the anisotropy of space in anisotropic cosmological models leads to the appearance of visible anisotropy in the orientations of distant galaxies. This effect was calculated in first order from the curvature for the most general case of Petrov type-I spaces. At the same time, it was concluded in Ref. 4 that Birch's observations cannot be a consequence of general-relativistic effects accompanying the propagation of radiation.

The conclusions in Ref. 3 were challenged in Ref. 5. The subject of the discursion was the properties of a bundle of rays forming the image of a source. In Ref. 3 it was called a ''representative'' bundle, and it was assumed that it is nonrotating. Objections were raised specifically against this assumption in Ref. 5, where it was stated that a bundle of isotropic geodesics producing an image in a rotating universe must also be rotating. In this paper we examine the properties of a representative bundle of rays in greater detail than in Ref. 3 and show that it has the properties ascribed to it in Ref. 3. The criticism expressed by Korotkiĭ and Obukhov[5] is thereby put to rest.

2. EQUATIONS OF RAYS IN A REPRESENTATIVE BUNDLE

For simplicity, the further calculations are performed for a stationary Gödel-type cosmological model described by the metric

$$ds^2 = dt^2 - dx^2 - ka^2(x)\,dy^2 - dz^2 - 2\sqrt{p}\,a(x)\,dy\,dt. \tag{1}$$

This metric differs from the constant scale factor used in Refs. 3 and 5, which was set equal to unity. In the metric (1) $a(x) = e^{mx}$, and k, p, and m are constants.

Let us consider a system of rays propagating from a source $ABCD$ to an observer O (see Fig. 1). It is assumed that photons of a definite frequency ω reach point O at the same moment in time t_1, being emitted from points $A, B, C, \ldots$ of the source at different moments in time, depending on the coordinates of the points. If the position of a photon on an isotropic geodesic is characterized by the affine parameter s, the value of s for all photons arriving at the time t_1 at point O can be set equal to $s_1 > 0$. The initial values of s for the photons emerging from points A, B, etc.

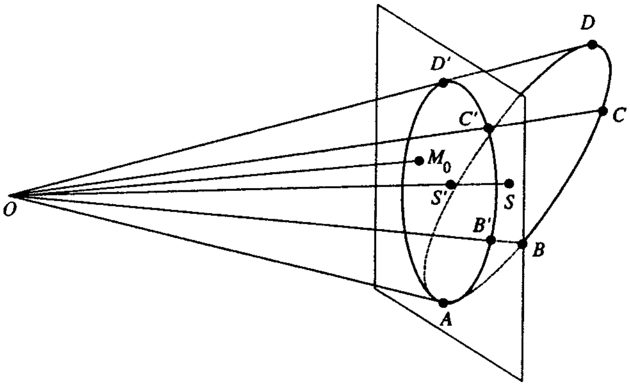


FIG. 1. Behavior of a bundle of rays forming the image of a source in cosmological models with rotation.

at different times (which correspond to their arrival at point O at t_1) will naturally be different and will depend on the coordinates of these points.

We take into account that the observer receives the image of the projection $AB'C'D'$ of the source onto a plane perpendicular to the line of vision, rather than the source itself. In this plane we assign the value $s=s_0$ to the affine parameter s. In flat space a definite value of s_1-s_0 would correspond to a definite distance from the observer. Then the segment of the spherical wave surface which is tangent to the source at point A would correspond to $s=s_0$. Because of the great distance to the source, this segment of the wave surface would be indistinguishable from the flat surface depicted in Fig. 1.

We can do virtually the same thing in curved space: the projection of the source consists of a set of spatial points on isotropic geodesics characterized by the same value of the affine parameter, which is equal, say, to the value of $s_0(A)$ for a photon emerging from point A, to which we assign the value $s_0(A)=0$. We note that in view of the small dimensions of the source in comparison to the cosmological curvature, the relation between the geometry of the source and the image of it derived from the condition $s=s_0(A)=0$ will be practically the same as in flat space (as is shown in Fig. 1). Let $M_0(x_0^\mu)$ be an arbitrary point on the projection of the source. We have

$$x_0^0=t_0,\quad x_0^1=x_0,\quad x_0^2=y_0,\quad x_0^3=z_0.$$

A ray belonging to the representative bundle emerges from point $M_0(t_0,x_0,y_0,z_0)$ and reaches point $O(t_1,x_1=0,$ $y_1=0,z_1=0)$, and s is the affine parameter on the ray. It is assumed for all the rays of the representative bundle that $s=s_0=0$ at points on the projection of the source and $s=s_1$ at point O.

Space (1) has four Killing vectors

$$\zeta_0^\mu=\delta_0^\mu,\quad \zeta_1^\mu=\frac{1}{m}\delta_1^\mu-y\,\delta_2^\mu,\quad \zeta_2^\mu=\delta_2^\mu,\quad \zeta_3^\mu=\delta_3^\mu. \tag{2}$$

Therefore, the equations of the isotropic geodesics have four first integrals, which we denote for convenience as

$$(k_\mu\zeta_0^\mu)=q_0,\quad (k_\mu\zeta_i^\mu)=q_0q_i,\quad i=1,2,3. \tag{3}$$

It is not difficult to obtain the following expressions for the co- and countervariant components of the wave vector from the equations $k^\mu_{\mu;\nu}k^\nu=0$ and the equality $k_\mu k^\mu=0$:

$$k_0=q_0,\quad k_1=-q_0\frac{g(a,q_2,q_3)}{a},$$
$$k_2=-q_0q_2,\quad k_3=-q_0q_3, \tag{4}$$

$$k^0=\frac{q_0k}{k+p}\left(1+\frac{q_2\sqrt{p}}{ak}\right),\quad k^1=\frac{q_0g}{a},$$
$$k^2=\frac{q_0}{a(k+p)}\left(-\sqrt{p}+\frac{q_2}{a}\right),\quad k^3=q_0q_3. \tag{5}$$

Here

$$g(a,q_2,q_3)=\left[a^2\left(\frac{k}{k+p}-q_3^2\right)+\frac{2\sqrt{p}\,q_2a}{k+p}-\frac{q_2^2}{k+p}\right]^{1/2}.$$

Integrating the equations $dx^\mu/ds=k^\mu(x^\alpha,q_a)$ with k^μ from (5) and with the conditions $x^i(s_1)=0$ and $t(s_1)=t_1$, we obtain

$$t=t_1+\frac{kq_0}{k+p}\left(s-s_1+\frac{\sqrt{p}\,q_2}{k}\int_{s_1}^{s}\frac{ds}{a(s,q_2,q_3)}\right),$$
$$x=q_0\int_{s_1}^{s}\frac{g(a,q_2,q_3)}{a}ds, \tag{6}$$
$$y=\frac{q_0}{k+p}\left[-\sqrt{p}\int_{s_1}^{s}\frac{ds}{a}+q_2\int_{s_1}^{s}\frac{ds}{a^2}\right],\quad z=q_0q_3(s-s_1).$$

Motion along an isotropic geodesic is accompanied by the variation of $x(s)$ and, therefore, of $a(x(s))=\exp(mx(s))$. The equation for $a(s)$ has the form

$$mq_0s=\int_{a_0}^{a}\frac{da}{g(a,q_2,q_3)},\quad a_0=\exp(mx_0), \tag{7}$$

and its solution has the form

$$a(s,q_2,q_3)=\frac{q_2\sqrt{1-q_3^2}}{\sqrt{k+p}(k/(k+p)-q_3^2)}\times\cosh\left[\varphi_0+mq_0s\sqrt{\frac{k}{k+p}-q_3^2}\right]-\frac{\sqrt{p}\,q_2}{k-(k+p)q_3^2}. \tag{8}$$

Here φ_0 is related to $a_0=\exp(mx_0)$ by the expression

$$\frac{\sqrt{p+k}}{q_2\sqrt{1-q_3^2}}\left[a_0\left(\frac{k}{k+p}-q_3^2\right)+\frac{\sqrt{p}\,q_2}{k+p}\right]=\cosh\varphi_0. \tag{9}$$

Since $a(s_1)=e^{m0}=1$ for $s=s_1$, we have

$$1=\frac{q_2\sqrt{1-q_3^2}}{\sqrt{k+p}\,(k/(k+p)-q_3^2)}\times\cosh\left[\varphi_0+mq_0s_1\sqrt{\frac{k}{k+p}-q_3^2}\right]-\frac{\sqrt{p}\,q_2}{k-(k+p)q_3^2}. \tag{10}$$

Using (10), we can eliminate φ_0 in (8) and express a in terms of only s, q_2, and q_3:

$$a=(1+B)\cosh\psi-\sqrt{(1+B)^2-A^2}\,\sinh\psi-B, \tag{11}$$

where

$$B=\frac{\sqrt{p}\,q_2}{(k+p)[k/(k+p)-q_3^2]},\qquad A=\sqrt{\frac{k+p}{p}}\sqrt{1-q_3^2}\,B,$$

$$\psi=mq_0\sqrt{\frac{k}{k+p}-q_3^2}\,(s_1-s).$$

3. CONGRUENCE PROPERTIES OF THE RAYS OF A REPRESENTATIVE BUNDLE

In order to obtain the field of isotropic vectors k_μ, for which the rotation tensor $\omega_{\mu\nu}=(k_{\mu,\nu}-k_{\nu,\mu})/2$ characterizing the rotation of a bundle of rays is calculated, s, q_2, and q_3 must be expressed in terms of x, y, and z using Eqs. (6) and then substituted into (4). In light of everything, the lack of this procedure was the reason for the erroneous conclusion drawn in Ref. 5. In any case, the expression given in Ref. 5 for the rotation parameter ω is obtained, if only the explicit dependence on the coordinates is taken into account in formulas (4) and q_2 and q_3 are regarded as constants (in Ref. 5 the constants q_2 and q_3 were replaced by the angles of approach θ and φ of the rays at the point of observation).

The position of a point on an isotropic geodesic is completely defined by assigning the parameters q_2, q_3, s, and t_1. In this case t_1 specifies the temporal sequence of wave fronts reaching point O and is an analog of the phase u. The parameter q_0 remains equal to a constant (at least in the stationary case; incidentally, in the nonstationary case there is no such constant parameter) and does not play any role in the expressions describing the congruence of the wave vectors. If q_0 is assumed to be absolutely constant, it can be set equal to unity, which we shall do in the following. This corresponds to the fact that we are considering a physical situation in which photons of one frequency are emitted from each point on the projection of the source. However, there is a possibility which can lead to rotating congruence. More specifically, when q_0, being an integral of motion, i.e., being independent of s, is a function of two other constants q_2 and q_3: $q_0=q_0(q_2,q_3)$. In this case the frequency of the emitted photons is different for different points of the source. We shall not consider such a physical situation in this paper.

In principle, s, q_2, and q_3 must be expressed in terms of x, y, and z using (6) and then substituted into (4). The congruence of $k_\mu(x^\alpha)$ is thus obtained, and the rotation tensor $\omega_{\mu\nu}$ must be calculated for it. However, we shall proceed differently. Instead of the components k_μ in the (t,x,y,z) system of coordinates, we shall work with the components k'_μ in a new system of the coordinates

$$x'^0=t_1,\quad x'^1=s,\quad x'^2=q_2,\quad x'^3=q_3.$$

We have

$$k'_0=k_\alpha\frac{\partial x^\alpha}{\partial t_1}=k_0\frac{\partial t}{\partial t_1}=k_0=q_0,$$

$$k'_1=k_\alpha\frac{\partial x^\alpha}{\partial s}=k_\alpha k^\alpha=0\left(\text{since}\,\frac{\partial x^\alpha}{\partial s}=k^\alpha\right), \tag{12}$$

$$k'_2=k_\alpha\frac{\partial x^\alpha}{\partial q_2}=k_0\frac{\partial t}{\partial q_2}+k_1\frac{\partial x}{\partial q_2}+k_2\frac{\partial y}{\partial q_2}+k_3\frac{\partial z}{\partial q_2}.$$

To calculate $\partial x^\alpha/\partial x'^\alpha$ we use formulas (6). Then,

$$\begin{aligned}
\frac{\partial t}{\partial q_2}&=\frac{q_2q_0\sqrt{p}}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_2}ds-\frac{q_0\sqrt{p}}{k+p}\int_s^{s_1}\frac{ds}{a(s)},\\
\frac{\partial t}{\partial q_3}&=\frac{\sqrt{p}\,q_0q_2}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_3}ds,\\
\frac{\partial x}{\partial q_2}&=\frac{1}{ma}\frac{\partial a}{\partial q_2},\qquad \frac{\partial x}{\partial q_3}=\frac{1}{ma}\frac{\partial a}{\partial q_3},\\
\frac{\partial y}{\partial q_2}&=-\frac{q_0}{k+p}\int_s^{s_1}\frac{ds}{a^2}-\frac{q_0\sqrt{p}}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_2}ds\\
&\quad+\frac{2q_0q_2}{k+p}\int_s^{s_1}\frac{1}{a^3}\frac{\partial a}{\partial q_2}ds,\\
\frac{\partial y}{\partial q_3}&=-\frac{q_0\sqrt{p}}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_3}ds+\frac{2q_0q_2}{k+p}\int_s^{s_1}\frac{1}{a^3}\frac{\partial a}{\partial q_3}ds,\\
\frac{\partial z}{\partial q_2}&=0,\qquad \frac{\partial z}{\partial q_3}=q_0(s-s_1).
\end{aligned} \tag{13}$$

Using (13) and (4), we obtain

$$\begin{aligned}
k'_2&=\frac{q_0^2}{k+p}\sqrt{p}\left[-\int_s^{s_1}\frac{ds}{a}+2q_2\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_2}ds\right]\\
&\quad+\frac{q_0^2q_2}{k+p}\int_s^{s_1}\frac{ds}{a^2}-\frac{2q_0^2q_2^2}{k+p}\int_s^{s_1}\frac{1}{a^3}\frac{\partial a}{\partial q_2}ds-\frac{g}{ma^2}\frac{\partial a}{\partial q_2},\\
k'_3&=\frac{\sqrt{p}\,q_0^2q_2}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_3}ds-q_0^2q_2\left[-\frac{\sqrt{p}}{k+p}\int_s^{s_1}\frac{1}{a^2}\frac{\partial a}{\partial q_3}ds\right.\\
&\quad\left.+\frac{2q_2}{k+p}\int_s^{s_1}\frac{1}{a^3}\frac{\partial a}{\partial q_3}ds\right]-\frac{g}{ma^2}\frac{\partial a}{\partial q_3}-q_0^2q_3(s-s_1).
\end{aligned} \tag{14}$$

It is not difficult to show that

$$k'_{2,1}=\frac{\partial k'_2}{\partial s}=k'_{3,1}=\frac{\partial k'_3}{\partial s}=0. \tag{15}$$

This can be proved by direct differentiation of expressions (14). However, it can also be demonstrated in a different manner. In fact, $k'_2=k_\mu\partial x^\mu/\partial q_2$ can be regarded as an invariant. Then, using $D/\partial s$ and $D/\partial q_2$ to denote covariant derivatives along s and q_2, we obtain

$$\begin{aligned}
k'_{2,1}&=\frac{\partial k'_2}{\partial s}=\frac{D}{\partial s}\left(k_\mu\frac{\partial x^\mu}{\partial q_2}\right)=k_{\mu;\nu}k^\nu\frac{\partial x^\mu}{\partial q_2}+k_\mu\frac{D}{\partial s}\left(\frac{\partial x^\mu}{\partial q_2}\right)\\
&=k_\mu\frac{Dk^\mu}{\partial q_2}=\frac{1}{2}\frac{D}{\partial q_2}(k_\mu k^\mu)=0.
\end{aligned}$$

Here we have only utilized the fact that k_μ is an isotropic geodesic vector ($k_{\mu;\nu}k^\nu=0$, $k_\mu k^\mu=0$).

Similarly, it can be shown that $k'_{3,1}=\partial k_3/\partial s=0$. Thus, k'_2 and k'_3 are constant along rays and are equal, say, to their values at the point $s=s_1$ (point O). At this point all the integrals in the expressions for k'_2 and k'_3 vanish, and we have

$$k'_2|_{s=s_1}=-\frac{g(1,q_2,q_3)}{m}\left.\frac{\partial a}{\partial q_2}\right|_{s=s_1},$$

$$k'_3|_{s=s_1}=-\frac{g(1,q_2,q_3)}{m}\left.\frac{\partial a}{\partial q_3}\right|_{s=s_1}. \tag{16}$$

On the basis of Eq. (11) we can obtain

$$\begin{aligned} q_2\frac{\partial a}{\partial q_2}&=a-\frac{\sqrt{k+p}\,g(a,q_2,q_3)}{\sqrt{(k+p)(1-q_3^2)-(\sqrt{p}-q_2)^2}}\\ &=a-\frac{g(a,q_2,q_3)}{g(1,q_2,q_3)},\\ \frac{\partial a}{\partial q_3}&=\frac{\partial B}{\partial q_3}(\cosh\psi-1)\\ &-\frac{(1+B)(\partial B/\partial q_3)-A(\partial A/\partial q_3)}{\sqrt{(1+B)^2-A^2}}\\ &\times\sinh\psi+[(1+B)\sinh\psi\\ &-\sqrt{(1+B)^2-A^2}\cosh\psi]\frac{\partial\psi}{\partial q_3}, \end{aligned} \tag{17}$$

where

$$\frac{\partial\psi}{\partial q_3}=-\frac{mq_0q_3(s_1-s)}{\sqrt{k/(k+p)-q_3^2}}.$$

It follows from (17) for $s=s_1$ $(a=1)$ that $\psi=0$ and

$$\frac{\partial a}{\partial q_2}=\frac{\partial a}{\partial q_3}=0.$$

Thus, everywhere

$$k'_2=k'_3=0. \tag{18}$$

Therefore, in the (t_1,s,q_2,q_3) system of coordinates the isotropic congruence of the wave vectors investigated has the form

$$k'_\alpha=(q_0,0,0,0), \tag{19}$$

and the rotation tensor for this congruence equals zero.

It can be concluded from the foregoing that if the frequency q_0 of the light is an absolute integral, which is identical for all rays, the bundle of rays forming the image is nonrotating in the cosmological model with rotation described by the metric (1). A similar result should be expected for all metrics with global rotation.

4. CONCLUSION

We have shown that the ''representative'' bundle of rays forming an image of a source considered in Ref. 3 is, in fact, nonrotating. Thus, the main conclusion in Ref. 3 that the Birch effect cannot be attributed to cosmological rotation within the general theory of relativity holds. The question of the possibility of attributing the observed Birch anisotropy to cosmological rotation in a theory of gravitation with twisting requires a separate treatment. Finally, we note that despite the advances regarding inflationary scenarios of the evolution of the early universe, within which the rotation of the early universes is strongly damped, it would be useful for astrophysicists to continue the efforts to find manifestations of rotation of the universe in the modern epoch and its possible global anisotropy.

*)E-mail: info@mail.psu.ru

---

[1] P. Birch, Nature (London) **298**, 451 (1982).

[2] V. A. Korotkiĭ and Yu. N. Obukhov, Zh. Éksp. Teor. Fiz. **99**, 22 (1991) [Sov. Phys. JETP **72**, 11 (1991)].

[3] V. F. Panov and Yu. G. Sbytov, Zh. Éksp. Teor. Fiz. **101**, 769 (1992) [Sov. Phys. JETP **74**, 411 (1992)].

[4] S. L. Parnovskiĭ, Yu. N. Kudrya, and A. N. Aleksandrov, Zh. Éksp. Teor. Fiz. **106**, 1559 (1994) [JETP **79**, 840 (1994)].

[5] V. A. Korotkiĭ and Yu. N. Obukhov, Zh. Éksp. Teor. Fiz. **108**, 1889 (1995) [JETP **81**, 1031 (1995)].

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Rindler solutions and their physical interpretation

A. I. Nikishov and V. I. Ritus*)

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

(Submitted 9 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 777–785 (September 1998)

We show that the singular behavior of Rindler solutions near horizon testifies to the currents of particles from a region arbitrarily close to the horizon. Besides, the Rindler solutions in right Rindler sector of Minkowski space can be represented as a superposition of only positive- or only negative-frequency plane waves; these states require infinite energy for their creation and possess infinite charge in a finite space interval, containing the horizon. The positive- or negative-frequency representations of Rindler solutions analytically continued to the whole Minkowski space make up a complete set of states in this space, which have, however, the aforementioned singularities. These positive (negative)-frequency states are characterized by positive (negative) total charge, the charge of the same sign in right (left) Rindler sector and by quantum number κ. But in other Lorentz invariant sectors they do not possess positive (negative)-definite charge density and have negative (positive) charge in left (right) Rindler sector. Therefore these states describe both the particle (antiparticle) and pairs, the mean number of which is given by Planck function of κ. These peculiarities make the Rindler set of solutions nonequivalent to the plane wave set and the inference on the existence of thermal currents for a Rindler observer moving in empty Minkowski space is unfounded.
[S1063-7761(98)00209-1]

As is known, it is impossible to create a constant uniform gravitational field with acceleration a in a space length of the order of a^{-1}, or the coordinate system in Minkowski space, imitating such a field.[1] So, in the Rindler coordinate system due to hardness requirement the acceleration is inversely proportional to the space coordinate, $a(z)=z^{-1}$, and becomes infinite at a distance a^{-1} from the plane with acceleration a (event horizon). It is clear that such a system is unrealizable by moving bodies up to horizon. At the same time there are statements in literature that the wave equation solutions with the Rindler system symmetry testify to appearance of currents of particles with thermal spectrum for an observe at rest in the Rindler system, i.e., uniformly accelerated in Lorentz system where no particles are present.[2,3] Moreover, it is stated that this phenomenon imitates black hole evaporation.[2–4] On the other hand, according to the interesting paper by Belinski,[5] the already formed black hole should not create particles. The same is expected if the analogy with QED is warranted:[6] both virtual or real pair components are attracted by black hole and no particles should appear far away.

Rindler solution (the solution of Klein–Gordon equation in Rindler wedge), decreasing for $\zeta\to\infty$, is given by the Bessel function of the second kind (McDonald function),

$$K_{i\kappa}(\zeta)\exp(-i\kappa v+i\mathbf{p}_\perp\mathbf{x}_\perp),$$

$$v=at'=\tanh^{-1}\frac{t}{z},\quad \zeta=m_\perp z'=m_\perp\sqrt{z^2-t^2}, \tag{1}$$

and may be written down[6] as a superposition of positive ($p^0>0$) or negative ($p^0<0$) frequency plane waves with different rapidity θ:

$$K_{i\kappa}(\zeta)\exp(-i\kappa v)=\frac{1}{2}\exp\left(\mp\frac{\pi\kappa}{2}\right)\times\int_{-\infty}^{\infty}d\theta\,\exp[i(p^3z-p^0t)\mp i\kappa\theta],$$

$$p^0=\pm m_\perp\cosh\theta,\quad p^3=m_\perp\sinh\theta. \tag{2}$$

It is an analytic function of variables $x_\pm=t\pm z$ with branch points at $x_\pm=0$. This solution can be analytically continued to the whole Minkowski space from semiaxes $x_-<0$, $x_+>0$ to semiaxes $x_->0$, $x_+<0$; for $p^0>0$ the continuation is performed through the lower and for $p^0<0$ through the upper half-planes of complex $x_\pm$. According to the lower or upper ways of continuation we get the positive- or the negative-frequency solutions $\Phi_\kappa^{(+)}$ or $\Phi_\kappa^{(-)}$ coinciding in Rindler sector R and differing in other sectors:

$$R:\quad \Phi_\kappa^{(\pm)}=K_{i\kappa}(\zeta)\exp(-i\omega t'),\quad \omega=a\kappa,$$

$$F:\quad \Phi_\kappa^{(\pm)}=\exp\left(\mp\frac{\pi\kappa}{2}\right)K_{i\kappa}(\pm i\tau)\exp(-i\omega z'),$$

$$\tau=m_\perp\sqrt{t^2-z^2},$$

$$P:\quad \Phi_\kappa^{(\pm)}=\exp\left(\mp\frac{\pi\kappa}{2}\right)K_{i\kappa}(\mp i\tau)\exp(-i\omega z'),$$

$$az'=\tanh^{-1}\frac{z}{t},$$

$$L:\quad \Phi_\kappa^{(\pm)}=\exp(\mp\pi\kappa)K_{i\kappa}(\zeta)\exp(-i\omega t'). \tag{3}$$

So, the solutions $\Phi_\kappa^{(\pm)}$ are characterized by frequency sign and by real parameter κ (the latter instead of plane wave momentum).

Solutions $\Phi^{(-)}$ may be obtained from $\Phi^{(+)}$ by complex conjugation and changing the sign of κ:

$$\Phi_\kappa^{(-)}=\Phi_{-\kappa}^{(+)*}. \tag{4}$$

If instead of $\Phi^{(\pm)}$ one considers $\phi^{(\pm)}=\exp(\pm\pi\kappa/2)\Phi^{(\pm)}$ then these solutions $\phi^{(\pm)}$ will be ‘‘similarly normalized’’ in whole Minkowski space by values differing only in sign. The solution $\phi^{(+)}$ corresponds to positive charge located mainly in R or L for κ greater or less than 0 and $\phi^{(-)}$ corresponds to negative charge located mainly in L or R for the same κ.

For $\Phi^{(+)}$ we have the following current density components $j_\pm=j^0\pm j^3$:

$$R:\quad j_\pm=\mp\frac{2\kappa}{x_\mp}K_{i\kappa}^2(\zeta),$$

$$F:\quad j_\pm=\frac{e^{-\pi\kappa}}{x_\mp}[\pi\mp2\kappa|K_{i\kappa}(i\tau)|^2],$$

$$P:\quad j_\pm=\frac{e^{-\pi\kappa}}{x_\mp}[-\pi\mp2\kappa|K_{i\kappa}(-i\tau)|^2],$$

$$L:\quad j_\pm=\mp e^{-2\pi\kappa}\frac{2\kappa}{x_\mp}K_{i\kappa}^2(\zeta). \tag{5}$$

In sectors R and L the current flows along hyperbolae with constant ζ and in sectors F and P both along the hyperbolae with constant τ and along the rays outgoing from (in F) and ingoing to (in P) the origin of coordinates:

$$R:\quad j_\alpha=j_\alpha^{\mathrm{hyp}}=-\frac{2\kappa}{z^2-t^2}K_{i\kappa}^2(\zeta)\epsilon_{\alpha\beta}x^\beta, \tag{6}$$

$$F:\quad j_\alpha=j_\alpha^{\mathrm{lin}}+j_\alpha^{\mathrm{hyp}},$$

$$j_\alpha^{\mathrm{lin}}=\frac{\pi e^{-\pi\kappa}}{t^2-z^2}x_\alpha,$$

$$j_\alpha^{\mathrm{hyp}}=\frac{2\kappa e^{-\pi\kappa}}{t^2-z^2}|K_{i\kappa}(i\tau)|^2\epsilon_{\alpha\beta}x^\beta, \tag{7}$$

$\epsilon_{\alpha\beta}$ is the antisymmetric tensor, $\epsilon_{03}=-\epsilon_{30}=1$.

The linear and hyperbolic current densities are orthogonal:

$$j_\alpha^{\mathrm{lin}}j^{\mathrm{hyp}\alpha}=0.$$

Singular behavior of current density at $x_\pm=0$ is evident. Due to current density conservation the current

$$J=\frac{1}{2}\int_C(j_+dx_--j_-dx_+)$$

through a contour C lying in F and having ends on the hyperbola $\tau=\tau_1$ at the points (x_{+1},x_{-2}) and (x_{+2},x_{-1}) is equal to the current through a broken line formed by segments of the straight lines $x_+=x_{+2}$ and $x_-=x_{-2}$ crossing on the hyperbola $\tau=\tau_2\equiv m_\perp\sqrt{x_{-2}x_{+2}}$ (see Fig. 1):

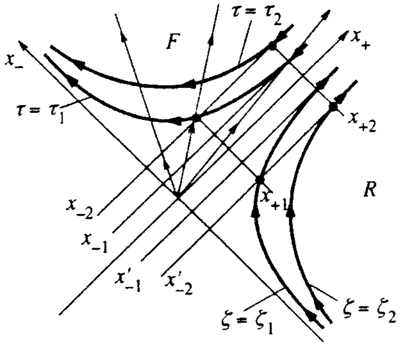


FIG. 1.

$$\frac{1}{2}\int_{x_{-1}}^{x_{-2}}\frac{dx_-}{x_-}e^{-\pi\kappa}[\pi-2\kappa|K_{i\kappa}(i\tau)|^2]$$
$$+\frac{1}{2}\int_{x_{+1}}^{x_{+2}}\frac{dx_+}{x_+}e^{-\pi\kappa}[\pi+2\kappa|K_{i\kappa}(i\tau)|^2]$$
$$=2\pi e^{-\pi\kappa}\ln\frac{\tau_2}{\tau_1}. \tag{8}$$

It is seen that the hyperbolic parts of the currents cancel each other and the linear parts are equal, and yield the result presented in (8). The former is evident beforehand as the contour can be deformed to the arc of hyperbola $\tau=\tau_1$ between the chosen points on it, while the latter follows from (7) and the relation

$$\frac{x_{-2}}{x_{-1}}=\frac{x_{+2}}{x_{+1}}=\left(\frac{\tau_2}{\tau_1}\right)^2. \tag{9}$$

If one fixes the coordinates x_{+1},x_{+2} of the hyperbola arc ends and the parameter τ_1 tends to zero, then the considered arc of hyperbola can be made as close to the segment (x_{+1},x_{+2}) of x_+-axis one wishes. Yet the current through this arc of hyperbola remains constant according to (8), (9); the same is true for its components along x_+- and x_--axes. Similarly the current through the contour lying in sector R with ends on hyperbola $\zeta=\zeta_1$ at points (x_{+1},x'_{-2}) and (x_{+2},x'_{-1}) is equal to the current through the broken line formed by segments of straight lines $x_-=x'_{-2}$ and $x_+=x_{+2}$ crossing on the hyperbola $\zeta=\zeta_2\equiv\sqrt{-x'_{-2}x_{+2}}$:

$$\frac{1}{2}\int_{x_{+1}}^{x_{+2}}\frac{dx_+}{x_+}[-2\kappa K_{i\kappa}^2(\zeta)]+\frac{1}{2}\int_{x'_{-1}}^{x'_{-2}}\frac{dx_-}{x_-}2\kappa K_{i\kappa}^2(\zeta)=0. \tag{10}$$

It is equal to zero as in R the current flows only along hyperbolae and the chosen contour can be deformed to the arc of hyperbola $\zeta=\zeta_1$. So the arcs of hyperbolae $\tau=\tau_1$ and $\zeta=\zeta_1$, lying in F and R and having the projection (x_{+1},x_{+2}) on x_+-axis, for τ_1, $\zeta_1\to\infty$ press themselves to the projection from both sides but the current flowing through the both arcs remains constant and equal to (8). If the ends of contours in F and R are placed not on hyperbolae, but on straight lines $x_-=x_{-2}>0$ and $x_-=x'_{-2}<0$ at the points with coordinates x_{+1},x_{+2}, then the current through these contours will be given by the second and the first integrals in (8) and (10) as these contours can be deformed to the corresponding segments of straight lines. Choosing $x'_{-2}=-x_{-2}$ (what is equivalent to $\zeta_1=\tau_1$) it is possible to show that the total current crossing both segments, when they approach each other (i.e., for $\tau_{1,2}\to 0$, but $\tau_2/\tau_1=\text{const}$), differs from zero and oscillates with increasing frequency:

$$e^{-\pi\kappa}\left\{\pi\ln\frac{\tau_2}{\tau_1}+2\kappa\int_{\tau_1}^{\tau_2}\frac{d\tau}{\tau}\,|K_{i\kappa}(i\tau)|^2\right\} -2\kappa\int_{\tau_1}^{\tau_2}\frac{d\zeta}{\zeta}\,K^2_{i\kappa}(\zeta)\approx\frac{2\pi}{\kappa(e^{\pi\kappa}+1)}\sin\left(\kappa\ln\frac{\tau_2}{\tau_1}\right) \times\cos\left(\beta+\kappa\ln\frac{\tau_1\tau_2}{4}\right), \qquad (11)$$

where $\beta=\arg(\Gamma^2(1-i\kappa))$, $\tau_1\tau_2=m_\perp^2\sqrt{x_{+1}x_{+2}}x_{-1}\to 0$.

So the current through the closed contour, going through the ends of the segment (x_{+1},x_{+2}) on the x_+-axis and lying in F and R, due to singular behavior of current density at the x_+-axis depends on the way of calculating the integrals near the singularity. Calculations of principal values of integrals by means of ''hyperbolic'' or ''linear'' approaches give respectively finite (8) or indefinite (11) values.

Note, that the charge and the energy of the state (2) in any finite volume containing horizon are infinite due to singularities of charge density j^0 and energy density T^{00} on the horizon.

Let us consider the scalar or ''inner'' product of two positive-frequency solutions (3) in Rindler sector R:

$$J^R_{\kappa\kappa'}=\int_{|t|}^{\infty}dzK_{-i\kappa}(\zeta)e^{i\omega t'}i\overleftrightarrow{\partial}_t K_{i\kappa'}(\zeta)e^{-i\omega' t'}. \qquad (12)$$

Replacing the integration variable z by the variable ζ, it is not difficult to see that the integrand is total differential of function

$$-J^R_{\kappa\kappa'}(\zeta)=\frac{\zeta}{\kappa-\kappa'}\exp\left[i(\kappa-\kappa')\sinh^{-1}\frac{m_\perp t}{\zeta}\right] \times[K_{-i\kappa}(\zeta)K'_{i\kappa'}(\zeta)-K'_{-i\kappa}(\zeta)K_{i\kappa'}(\zeta)], \qquad (13)$$

which exponentially decays at the upper limit (for $\zeta\to\infty$) and oscillates with infinitely increasing frequency at the lower limit (for $\zeta\to 0$). So, the integral $J^R_{\kappa\kappa'}$ which is equal to the limit of the function $J^R_{\kappa\kappa'}(\zeta)$ for $\zeta\to 0$ does not exist in ordinary sense. Let us consider the function $J^R_{\kappa\kappa'}(\zeta)$ instead of it, i.e., the integral (12) in which the low limit $z=|t|$ is substituted by $z_1(t)=\sqrt{t^2+(\zeta/m_\perp)^2}$, where ζ is the parameter of hyperbola which intersects the straight line of constant t at the point $z_1(t)$. For this function there is also another representation:

$$J^R_{\kappa\kappa'}(\zeta)=\exp\left[i(\kappa-\kappa')\sinh^{-1}\frac{m_\perp t}{\zeta}\right] \times(\kappa+\kappa')\int_{\zeta}^{\infty}\frac{dx}{x}\,K_{-i\kappa}(x)K_{i\kappa'}(x). \qquad (14)$$

For the special case $\kappa=\kappa'$ the function $J^R_{\kappa\kappa}(\zeta)$ is real, positive and represents the charge of the state (1), pertaining to the region $z_1(t)\leqslant z<\infty$ and related to unit area of plane z_1. For $\zeta\to\infty$ this charge exponentially decreases,

$$J^R_{\kappa\kappa}(\zeta)\approx\frac{\pi\kappa}{2\zeta^2}\,e^{-2\zeta},\quad \zeta\to\infty, \qquad (15)$$

and for $\zeta\to 0$ it increases logarithmically:

$$J^R_{\kappa\kappa}(\zeta)=\frac{\pi}{\sinh\pi\kappa}\left[\ln\frac{2}{\zeta}+\mathrm{Re}(\psi(1-i\kappa))+\frac{1}{2\kappa} \times\sin\left(\beta+2\kappa\ln\frac{\zeta}{2}\right)+\cdots\right],\quad \zeta\to 0. \qquad (16)$$

Here $\psi(x)$ is logarithmic derivative of Γ-function, and $\beta=2\arg(\Gamma(1-i\kappa))$.

So, the charge in any finite volume becomes arbitrarily large when this volume approaches the horizon $z=|t|$.

Using equations (13), (14) and asymptotic expressions for the McDonald functions, it is easy to show that for $\kappa\neq\kappa'$ and $\zeta\to 0$

$$\int_{\zeta}^{\infty}\frac{dx}{x}\,K_{-i\kappa}(x)K_{i\kappa'}(x)\approx\frac{\pi}{2\kappa\kappa'}\sqrt{\frac{\kappa\kappa'}{\sinh(\pi\kappa)\sinh(\pi\kappa')}} \times\left\{\frac{1}{\kappa-\kappa'}\sin\left[(\kappa-\kappa')\times\ln\frac{2}{\zeta}-\frac{\beta-\beta'}{2}\right]-\frac{1}{\kappa+\kappa'}\times\sin\left[(\kappa+\kappa')\ln\frac{2}{\zeta}-\frac{\beta+\beta'}{2}\right]\right\}, \qquad (17)$$

where the omitted terms go to zero for $\zeta\to 0$.

Using the representation

$$\lim_{N\to\infty}\frac{\sin(Nx)}{x}=\pi\delta(x), \qquad (18)$$

let us rewrite the right-hand side of (17) in the form

$$\frac{\pi^2}{2\kappa\kappa'}\sqrt{\frac{\kappa\kappa'}{\sinh(\pi\kappa)\sinh(\pi\kappa')}}\,\{\delta(\kappa-\kappa')-\delta(\kappa+\kappa')\}. \qquad (19)$$

Then

$$\lim_{\zeta\to 0}J^R_{\kappa\kappa'}(\zeta)=\frac{\pi^2}{\sinh(\pi\kappa)}\,\delta(\kappa-\kappa'). \qquad (20)$$

The scalar product of the same two solutions (3) in sector F

$$J^F_{\kappa\kappa'}=e^{-\pi(\kappa+\kappa')/2} \times\int_{-t}^{t}dzK_{-i\kappa}(-i\tau)e^{i\omega z'}i\overleftrightarrow{\partial}_t K_{i\kappa'}(i\tau)e^{-i\omega' z'} \qquad (21)$$

is considered as a limit for $\tau\to 0$ of function $J^F_{\kappa\kappa'}(\tau)$, which differs from the integral (21) in that the limits of integration $z=\mp t$ are substituted by limits $z_{1,2}(t)=\mp\sqrt{t^2-(\tau/m_\perp)^2}$, where τ is the parameter of limiting hyperbola which intersects the straight line of constant t at the points $z_{1,2}$. It can be shown that

$$J^F_{\kappa\kappa'}(\tau)=-\frac{2\exp[-(\pi/2)(\kappa+\kappa')]}{\kappa-\kappa'}\times\sin\left[(\kappa-\kappa')\cosh^{-1}\frac{m_\perp t}{\tau}\right]\tau[K_{-i\kappa}(-i\tau)K'_{i\kappa'}\times(i\tau)+K'_{-i\kappa}(-i\tau)K_{i\kappa'}(i\tau)] \qquad (22)$$

$$=\frac{2\exp[-(\pi/2)(\kappa+\kappa')]}{\kappa-\kappa'}\times\sin\left[(\kappa-\kappa')\cosh^{-1}\frac{m_\perp t}{\tau}\right]\left(\pi-i(\kappa^2-\kappa'^2)\times\int_\tau^\infty\frac{dx}{x}K_{-i\kappa}(-ix)K_{i\kappa'}(ix)\right). \qquad (23)$$

For $\kappa=\kappa'$ the function

$$J^F_{\kappa\kappa}(\tau)=2\pi e^{-\pi\kappa}\cosh^{-1}\frac{m_\perp t}{\tau}. \qquad (24)$$

It gives the charge in region F between planes $z=z_1$ and $z=z_2$ per unit area of one of the planes. As seen from (24) this charge is positive, increases logarithmically in time and also as the limits of integration approach the horizons.

For $\kappa\neq\kappa'$ and $\tau\to 0$ it follows from (22) and (18)

$$\lim_{\tau\to 0}J^F_{\kappa\kappa'}(\tau)=2\pi^2e^{-\pi\kappa}\delta(\kappa-\kappa'). \qquad (25)$$

Equations (22)–(25) also hold for sector P, if one subjects them to complex conjugation and notes that arch x is an even function.[7]

Finally, scalar products of two positive-frequency solutions (3) in sectors L and R are related by the expression

$$J^L_{\kappa\kappa'}(\zeta)=-\exp(-\pi(\kappa+\kappa'))J^{R*}_{\kappa\kappa'}(\zeta). \qquad (26)$$

Hence in the region $-\infty<z\leqslant -z_1(t)$ of sector L the charge is negative, makes up the $\exp(-2\pi\kappa)$ part of the charge in symmetrical region of sector R and increases logarithmically for $\zeta\to 0$, i.e., as the limiting plane approaches the horizon:

$$J^L_{\kappa\kappa}(\zeta)\approx-\frac{\pi\exp(-2\pi\kappa)}{\sinh(\pi\kappa)}\left\{\ln\frac{2}{\zeta}+\mathrm{Re}(\psi(1-i\kappa))+\frac{1}{2\kappa}\sin\left(\beta+2\kappa\ln\frac{\zeta}{2}\right)\right\}. \qquad (27)$$

At the same time for $\kappa\neq\kappa'$ and $\zeta\to 0$

$$\lim_{\zeta\to 0}J^L_{\kappa\kappa'}(\zeta)=-\frac{\pi^2\exp(-2\pi\kappa)}{\sinh\pi\kappa}\delta(\kappa-\kappa'), \qquad (28)$$

Summing (20), (25), and (28) yields

$$\lim_{\zeta,\tau\to 0}[J^R_{\kappa\kappa'}(\zeta)+J^F_{\kappa\kappa'}(\tau)+J^L_{\kappa\kappa'}(\zeta)]=4\pi^2e^{-\pi\kappa}\delta(\kappa-\kappa'), \qquad (29)$$

which may be considered as the orthogonality and normalization condition of two positive-frequency solutions in the whole space.

For negative-frequency solutions $\Phi^{(-)}$ the integrals $J_{\kappa\kappa'}$ may be obtained from (12)–(16), and (20)–(29) by complex conjugation, changing the signs of κ and κ' and changing the overall sign, and for ''similarly normalized'' solutions $\phi^{(\pm)}$ they differ from those obtained for $\Phi^{(\pm)}$ by multipliers $\exp[\pm(\pi/2)(\kappa+\kappa')]$.

All given formulas hold both for positive and negative κ. Changing sign of κ is equivalent to reflection $z\to -z$:

$$\phi^{(\mp)}_{-\kappa}(z)=\phi^{(\pm)}_{\kappa}(-z). \qquad (30)$$

In other words, the sign of κ is not connected with the sign of total charge, but is connected with its spatial distribution.

Note the important relation

$$\phi^{(-)}_{\kappa}=\phi^{(+)*}_{-\kappa}, \qquad (31)$$

which is satisfied also by functions $\Phi^{(\pm)}_{\kappa}$.

Two solutions $\phi^{(+)}_{\kappa}$ and $\phi^{(-)}_{\kappa'}$ with different frequency signs are orthogonal in Minkowski space, moreover, they are orthogonal in $R+L$, F and P separately.

The energy density of the state $\Phi^{(+)}$ in sector R is

$$T_{00}=\frac{m^2_\perp(t^2+z^2)}{z^2-t^2}|K'|^2+\frac{\kappa^2(z^2+t^2)}{(z^2-t^2)^2}|K|^2+m^2_\perp|K|^2, \qquad (32)$$

where $K=K_{i\kappa}(\zeta)$, $\zeta=m_\perp\sqrt{z^2-t^2}$, and in sector F it is

$$T_{00}=e^{-\pi\kappa}\left\{\frac{m^2_\perp(t^2+z^2)}{t^2-z^2}|K'|^2+\frac{\kappa^2(z^2+t^2)}{(t^2-z^2)^2}|K|^2+m^2_\perp|K|^2-\frac{2\pi\kappa zt}{(t^2-z^2)^2}\right\}, \qquad (33)$$

where $K=K_{i\kappa}(i\tau)$, $\tau=m_\perp\sqrt{t^2-z^2}$. Comparison with charge density j^0 shows that T_{00} diverges for $z\to\pm t$ stronger then j^0 both in R and in F.

The infinite charge and infinite energy in finite volume containing the horizon testify to nonequivalence of the set of Rindler solutions to usual complete sets of wave equation solutions.[8] That the horizons act as sources of particle and antiparticle pairs is confirmed also by changing of charge density sign at crossing horizon. Finally, the positive-frequency solution $\phi^{(+)}$ possesses in sectors L and R the charges $Q^{(+)}_L$ and $Q^{(+)}_R$ of opposite signs and the total positive charge $Q^{(+)}$. These charges are connected by the relation

$$Q^{(+)}_L=-e^{-2\pi\kappa}Q^{(+)}_R=-\frac{1}{2}(e^{2\pi\kappa}-1)^{-1}Q^{(+)},\qquad Q^{(+)}>0. \qquad (34)$$

Similar connection between the charges in sectors L, R and total charge for negative-frequency state $\phi^{(-)}$ is

$$Q^{(-)}_R=-e^{-2\pi\kappa}Q^{(-)}_L=-\frac{1}{2}(e^{2\pi\kappa}-1)^{-1}Q^{(-)},\qquad Q^{(-)}<0. \qquad (35)$$

Any ratio of the charges of opposite signs may be taken in Minkowski space as a measure of intensity of pair production in this space.

The general feature of the states $\phi^{(\pm)}$ and their arbitrary superposition (36) is that their total charge is equally divided between the sectors $R+L$ and F (or P).

The fact that the state $\phi_\kappa^{(+)}(\phi_\kappa^{(-)})$, being a superposition of positive (negative)-frequency plane waves, not everywhere possess the positive (negative) charge density, is another manifestation of nonequivalence of $\phi_\kappa^{(\pm)}$-system to plane wave system.

As the functions $\phi^{(\pm)}$ differ in each of the sectors R and L only by constant factor, it is always possible to find such superpositions:

$$\phi_\kappa^R=\alpha\phi_\kappa^{(+)}+\beta\phi_\kappa^{(-)}, \quad \phi_\kappa^L=\beta\phi_\kappa^{(+)}+\alpha\phi_\kappa^{(-)}, \tag{36}$$

which a) are identically zero in L- and R-sectors correspondingly and b) have only positive and only negative charges and their densities in other sectors. For this it is necessary, that

$$\text{a)}\quad \beta=-\alpha e^{-\pi\kappa}, \quad \text{b)}\quad \beta^2=\alpha^2-1=\frac{1}{e^{2\pi\kappa}-1}, \quad \kappa>0. \tag{37}$$

Therefore it is possible to consider the states ϕ_κ^R and ϕ_κ^L, $\kappa>0$, as describing the positively and negatively charged particles, i.e., the particle and antiparticle.

Then in representations

$$\phi_\kappa^{(+)}=\alpha\phi_\kappa^R-\beta\phi_\kappa^L, \quad \phi_\kappa^{(-)}=-\beta\phi_\kappa^R+\alpha\phi_\kappa^L, \quad \kappa>0, \tag{38}$$

inverse to (36), the squares of the coefficients α and β must be interpreted as mean numbers of particles and antiparticles in the state $\phi_\kappa^{(+)}$ with total charge $+1$, and as mean numbers of antiparticles and particles in the state $\phi_\kappa^{(-)}$ with total charge -1. Relation (34) shows that in sector L there is only half (i.e., $\beta^2/2$) of all antiparticles of the state $\phi_\kappa^{(+)}$. The other half is in sector F or P. Similarly, half of all particles of the state $\phi_\kappa^{(-)}$ are in sector R while the other half is in sector F or P.

It is interesting to note that the positivity of charge density j^0 for the state ϕ_κ^R in F and P sectors ($|z|<|t|$),

$$j^0=\frac{\pi}{t^2-z^2}\left[|t|-z\,\frac{\pi\kappa}{\sinh(\pi\kappa)}\,|J_{i\kappa}(\tau)|^2\right], \tag{39}$$

mathematically is the consequence of inequality

$$\frac{\pi\kappa}{\sinh(\pi\kappa)}\,|J_{i\kappa}(\tau)|^2\leqslant 1, \quad \kappa,\tau\geqslant 0, \tag{40}$$

for Bessel function $J_{i\kappa}(\tau)$, which was not found in mathematical literature.

The charge density for the state $\phi_\kappa^{(+)}$ in the same sectors

$$j^0=\frac{\pi}{t^2-z^2}\left[|t|-z\,\frac{2\kappa}{\pi}\,|K_{i\kappa}(i\tau)|^2\right] \tag{41}$$

has both signs.

According to unconvincing arguments in Ref. 3, the vanishing of ϕ^R in L means that for Rindler observer (uniformly accelerated in sector R of Minkowski space, i.e., at rest in Rindler system) the β^2 turns out to be the measure of pair production intensity, and looks like a thermal Bose spectrum with temperature $a/2\pi$ and frequency $\omega=a\kappa$.

In our opinion, if the exotic state $\phi_\kappa^{(+)}$ is created, then two observers, one at rest in Minkowski system and the other in Rindler's one, can measure and receive information about the charges $Q_{\kappa R}^{(+)}$ and $Q_{\kappa P}^{(+)}$ in sectors R and P. The difference of these charges divided by $Q_{\kappa P}^{(+)}$ for each of the observers is given by the expression

$$\frac{Q_{\kappa R}^{(+)}-Q_{\kappa P}^{(+)}}{Q_{\kappa P}^{(+)}}=\frac{1}{e^{2\pi\kappa}-1}. \tag{42}$$

The same formula holds also for the charges in R and P sectors of the complex conjugated state $\phi_\kappa^{(+)*}=\phi_{-\kappa}^{(-)}$, as they differ from considered ones only by the sign. Therefore both observers deal with the same field state in Minkowski space, creation of which needs sources of unlimited intensity.

According to Ref. 9, for the quantization of free field in Rindler space it is necessary to satisfy the boundary condition on the Rindler manifold boundary $z'=0$ with arbitrary t', which corresponds to the point $z=t=0$ in Minkowski space. So, these authors conclude that the quantization of free fields is quite different in Rindler and Minkowski spaces and their analysis can give no ground for any conclusions about the behavior of uniformly accelerated detector.

This work was supported in part by the Russian Fund for Fundamental Research (Grants No. 96-02-17314a and 96-15-96463).

---

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

[1] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation*, Freeman, San Francisco (1973).
[2] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge University Press, Cambridge (1982).
[3] W. Greiner, B. Müller, and J. Rafelski, *Quantum Electrodynamics of Strong Fields*, Springer-Verlag, Berlin (1985).
[4] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
[5] V. A. Belinski, Phys. Lett. A **209**, 13 (1995).
[6] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **94**(7), 31 (1988) [Sov. Phys. JETP **67**, 1313 (1988)].
[7] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York (1964).
[8] A. I. Nikishov and V. I. Ritus, *Proc. of the Second Int. Sakharov Conf. on Physics*, ed. by I. M. Dremin and A. M. Semikhatov, World Scientific, Singapore (1997), p. 522.
[9] V. A. Belinski, B. M. Kamakov, V. D. Mur and N. B. Narozhny, JETP Lett. **65**, 902 (1997).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Heating of ultracold neutrons on beryllium surfaces

V. E. Varlamov,*) V. V. Nesvizhevskiĭ, A. P. Serebrov, R. R. Tal'daev,
and A. G. Kharitonov

B. P. Konstantinov St. Petersburg Institute of Nuclear Physics, Russian Academy of Sciences, 188350 Gatchina, Russia

P. Geltenbort

Institut Max von Laue–Paul Langevin, 38042 Grenoble, France

Ts. Ts. Panteleev, A. V. Strelkov, and V. N. Shvetsov

Joint Institute for Nuclear Research, 141980 Dubna, Russia

M. Pendlebury

University of Sussex, Brighton BNI 9QH, UK

K. Schreckenbach

Technische Universität München, D-85747 Garching, Germany

(Submitted 12 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 786–797 (September 1998)

The temperature dependence of the loss factor for ultracold neutrons owing to heating at thermal energies on the surface of a beryllium sample is studied. The probability of heating ultracold neutrons is anomalously high throughout the entire measured temperature interval, but especially at low temperatures.
[S1063-7761(98)00309-6]

1. INTRODUCTION

The condition of complete reflection of neutrons from matter at arbitrary angles of incidence imposes an upper bound on their energy at a level of 10^{-7} eV determined by the magnitude of the nuclear potential of the material. These neutrons are referred to as ultracold, and are capable of being retained in hermetic vessels for a long time.[1,2]

The theoretical description of the reflection of ultracold neutrons is based on solving the Schrödinger equation for a plane wave interacting with a so-called optical potential. In terms of this theory, the probability of loss of a neutron during reflection is determined by the loss factor $\eta = \mathrm{Im}\, U/\mathrm{Re}\, U$, where U is the optical potential, which should be complex when absorption occurs in the material. The difficulty of confining ultracold neutrons in traps is directly related to the fact that experimental values of the loss factor η are one or two orders of magnitude greater than theoretical estimates obtained assuming that the standard neutron–matter interaction cross sections are valid for ultracold neutrons. For example, the theoretical loss factor owing to inelastic scattering for Be is $\eta = 5\times10^{-6}$ at $T=300$ K, and that owing to radiative capture is 3×10^{-7}, while experiment yields $\eta=(2-4)\times10^{-4}$.

A significant advance in solving this problem occurred[3] when it was shown that about 75% of the neutrons leave a trap as a result of inelastic scattering with heating at thermal energies. It seemed most reasonable to explain this result in terms of the hypothesis that ultracold neutrons are heated on hydrogen contained in the surface layer of the trap material. Hydrogen actually was observed through nuclear reactions.[4] In order to obtain the experimentally observed loss probability, however, the required concentration of hydrogen would have to have been too high. It was shown[5] that high-temperature heating of a Be sample at $T=1000$ K could be used to reduce the loss factor η from $(2-3)\times10^{-3}$ to 2×10^{-4} (η was measured at $T=300$ K), while the gas desorbed from the Be surface was predominantly hydrogen and water.

Deep cooling of a trap with a sputtered Be coating that has initially been outgassed to $\eta=10^{-4}$ ($T=300$ K) leads to a reduction in η to 3×10^{-5}.[6] It should be noted that this value of the loss factor was attained by $T=77$ K and did not vary as the temperature was reduced to 6.5 K.

Interesting results were obtained during some experiments at Gatchina[7] to study the retention of ultracold neutrons in Be traps. The main conclusion of these measurements was that, regardless of the technique for coating the trap (sputtering or entirely of Be) and the extent to which it has been initially outgassed ($\eta=6\times10^{-5}$ to 2×10^{-4} at $T=300$ K), when the trap temperature was reduced to 13 K, the loss factors reached 3×10^{-5}, although as before the experimental result was two orders of magnitude greater than the theoretical estimate.

The possible heating of ultracold neutrons on a beryllium foil surface as the reason for their loss during confinement has been studied before.[8] It was shown that, at room temperature, the heating of ultracold neutrons determines their loss during confinement. At liquid nitrogen temperatures, however, the probability of heating ultracold neutrons on a beryllium foil turned out to be substantially lower than 3×10^{-5}, the probability of loss during confinement.

A detailed analysis of the other possible channels for loss of ultracold neutrons from the trap, such as superweak heating or contamination of the surface by substances with high capture cross sections, did not yield a reason for the limitation of the losses at a level of 3×10^{-5}. These losses have been characterized as anomalous.

2. STATEMENT OF THE PROBLEM

The purpose of this work was to study the possibility of the penetration of ultracold neutrons into a medium. If an ultracold neutron penetrates into beryllium, it will be heated, for the most part, since the heating cross section for beryllium (above liquid nitrogen temperature) exceeds the capture cross section. For stainless steel, on the other hand, capture will predominate. Thus, by studying the heating of ultracold neutrons on an intact beryllium foil and on a stainless steel foil, it is possible to perceive the large difference in the heating probabilities if the anomalous loss process is determined by penetration to a depth exceeding the sputtering depth of $\sim$3000 Å.

One additional purpose of the experiment was to verify the results of Ref. 8, i.e., to measure the heating probability for ultracold neutrons on beryllium surfaces near the temperature of liquid nitrogen.

3. EXPERIMENTAL SETUP

In this paper we study the inelastic scattering (heating) of ultracold neutrons at sample surfaces made of beryllium. We used two types of Be samples. The first was prepared by hot rolling of pressed, powdered, distilled 99.59%-pure beryllium. The major impurities were Fe (0.216%), Cu (0.096%), Cr (0.036%), Ni (0.029%), and Mn (0.025%). After etching in an HNO_3 solution and washing in distilled water, a Be ribbon of thickness 0.1 mm and width 5 cm was crumpled and placed in a special basket made of the same kind of ribbon. The second sample was prepared by magnetron sputtering onto a corrugated substrate of stainless steel 0.1 mm thick and 5 cm wide. The layer of sputtered beryllium was at least 3000 Å thick. The magnetron target was made of Be of the same purity as in the first case. The total area of each sample was 1 m^2.

The measurements were made at the high-flux reactor of the ILL in Grenoble, France (the PF2 instrument); the setup is sketched in Fig. 1. The test samples were positioned inside containment vessel *2*, made of 1.5-mm-thick polished stainless steel and covered with a 3000-Å-thick layer of sputtered Be. Containment vessel *2* was surrounded by ^{3}He counters to detect heated neutrons.

A gas of ultracold neutrons fills containment vessel *2* via valve *5*, which has a 100-μm-thick aluminum separation membrane *4* at its inlet. Containment vessel *2*, in conjunction with valve *5*, forms a so-called clean volume with a separate pumping system *6* that includes a heated vacuum valve and an oil-free (Dry Scroll Pump), which provides a vacuum of 6×10^{-3} Torr. The pure volume was cooled by pumping liquid nitrogen through an outside coil *7* and heated by electrical heaters *8* wrapped around the outside of all its parts.

The samples were outgassed by heating to $T=650$ K with repeated purges of helium gas (helium scrubbing). After completion of outgassing at $T=650$ K the heated vacuum valve on the pure volume was closed and kept closed throughout subsequent measurements in order to ensure vacuum integrity. The temperature was monitored by thermocouples, of which two were located inside the containment vessel near the sample, while the remaining four were located at various points outside the containment vessel.

The thermal insulation *9* on the pure volume was made of asbestos and aluminum foil. Forty-eight vertically mounted cylindrical counters, each 20 mm in diameter and 300 mm long, evenly surrounded the containment vessel and formed the principal element of the device, i.e., the heated neutron counter *11*. Since the characteristics of the individual cylindrical counters differ somewhat, they were selected and grouped in eight sections. The anodes of the counters in a section were joined, and each section of the counter had its own preamplifier.

Filter chamber 10 (a cylindrical cavity 4 cm thick, filled with BF_3 enriched to 10% ^{10}B) was positioned in front of the heated neutron counter *11*. This device measured the average energy of heated neutrons. The velocity of heated neutrons could be derived from the dependence of the filter transmission on BF_3 pressure. A 1.5-mm-thick Cu plate *12* was attached to the cooling/heating system to stabilize the temperature of counter *11* when the containment vessel was heated or cooled.

The heated neutron counter was surrounded by a shield *13* made of borated sheet resin 10–15 mm thick. The lower part of this shield, which was attached firmly to the heated parts of the device, was made of 1-mm sheet cadmium.

4. EXPERIMENTAL TECHNIQUE

The presence of even a small number of neutrons with energies exceeding the cutoff energy of Be in the spectrum of the ultracold neutrons presents a serious problem. Neutrons with energies exceeding the barrier energy will produce a spurious reading when they penetrate the sample.

The experimental setup included a preliminary containment vessel *14* because of the need for reliable suppression of ultracold neutrons with energies exceeding the Be cutoff. This cylinder, 2 m high and 0.6 m in diameter, was made of polished stainless steel. A polyethylene neutron absorber *15* suspended on a shaft makes it possible to remove neutrons from the vessel volume that have enough energy to rise higher than the absorber in the gravitational field. For the chosen absorber height of 1.8 m, a ultracold neutron spectrum with an upper bound of 185 neV is produced.

Neutrons were transported from the ultracold neutron source along a neutron duct made of stainless steel with an aluminum separator foil. The intensity of ultracold neutrons in the preliminary containment vessel *14* was monitored using two ^{3}He monitor counters. Counter *19*, with a 2 cm^2 aluminum input window mounted above the bottom of vessel

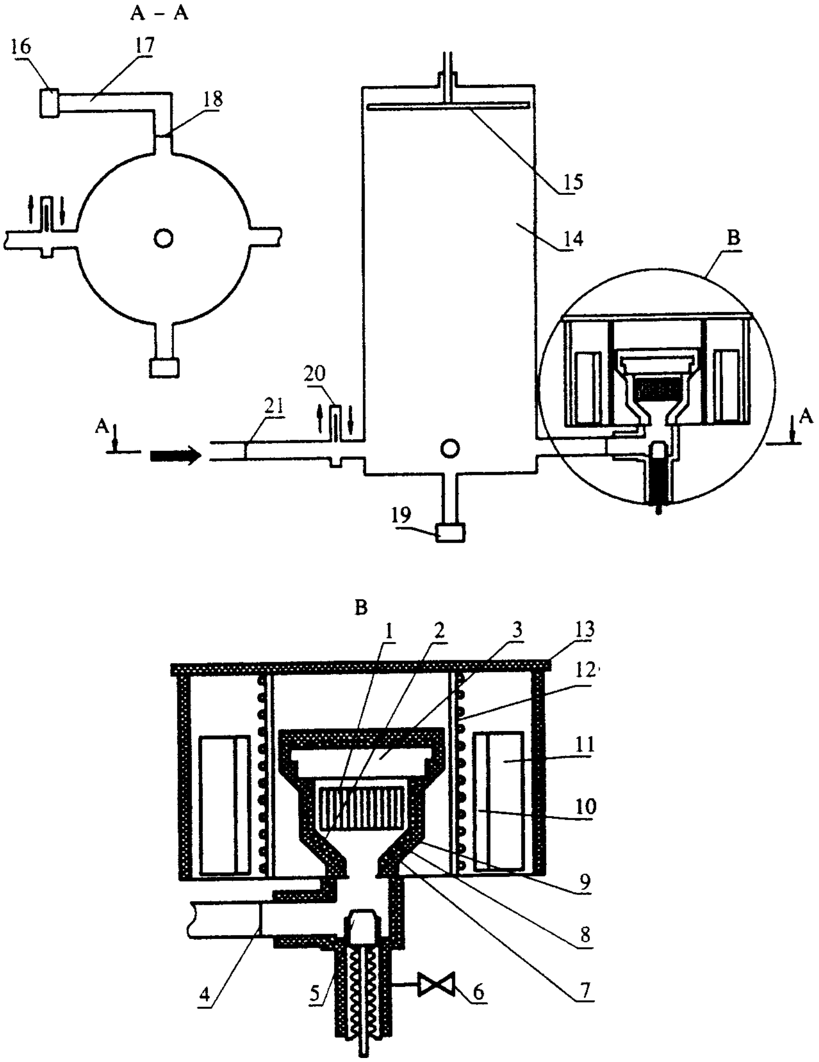


FIG. 1. Setup at the ILL reactor in Grenoble for studying the heating of ultracold neutrons (thick arrow) at beryllium surfaces: (1) Be target foil; (2) containment vessel for ultracold neutrons; (3) neutron counter chamber; (4), (21) Al separator foils; (5), (20) valves; (6) pumpout valve; (7) cooling system; (8) heating system; (9) thermal insulation; (10) BF_3 filter chamber; (11) heated neutron counter; (12) copper thermal plate; (13) detector neutron shield; (14) preliminary containment vessel; (15) polyethylene absorber; (16), (19) ultracold neutron monitor detectors; (17) bent neutron duct; (18) 56-μm-thick beryllium foil.

14, measured the flux of ultracold neutrons with energies exceeding the cutoff energy of aluminum. One of the outlets of vessel *14* was covered with a 56-μm-thick Be foil. Counter *16*, with a 50 cm^2 aluminum inlet window, was located at the end of a neutron duct *17* bent at an angle of 90 °. This configuration made it possible to measure the flux of ultracold neutrons penetrating through the beryllium foil and avoid the inevitable background of residual ultracold neutrons heated at the foil surface. Temporal variations in the detected neutron intensities during a single measurement cycle are shown in Fig. 2. The time trace begins at the time inlet valve *20* is opened. The filling of the containment vessels with neutrons continues for 100 s, whereupon valve 20 is closed. The total cycle time is 310 s.

The ratio of the detection intensities from the beryllium and aluminum monitor counters over the first 100 s of the cycle can be used to estimate the fraction of ultracold neutrons with energies exceeding the Be cutoff. Given the ratio of the areas of the inlet windows of the monitor counters, this fraction is 0.7%. Further analysis shows that the number of neutrons with energies exceeding the barrier energy falls off over a few seconds to negligible levels with a time constant of 0.7 s. Ten seconds after closure of the gate valve (i.e., at the 110 s point in the time evolution of Fig. 2), the fraction of neutrons with energies above the barrier energy is less than 10^{-6}. Thus, working with neutrons detected after the 110th second of the cycle time evolution diagram completely guarantees the lack of any contribution from neutrons with energies exceeding the barrier energy.

As a rule, the total count of heated neutrons between 110 and 210 s was used in the subsequent analysis. The total count of heated neutrons in the interval between 0 and 100 s is roughly five times greater, but about 0.7% of the neutrons with energies exceeding the barrier energy penetrate into the volume from the beryllium foil. An analysis of the data showed that the ratio of the total counts is constant to high accuracy over the full temperature range studied. Thus, the results obtained for 0–100 s are identical to those for 110–210 s, but their statistical accuracy is more than twice that of the latter. In the following analysis we shall, in several cases, use the statistically more accurate data, since the lack of spurious effects in these data was reliably demonstrated experimentally.

The resulting total count curves for the neutrons heated on the sample surface were cross-calibrated in units of the dimensionless parameter η, the loss factor. As a so-called

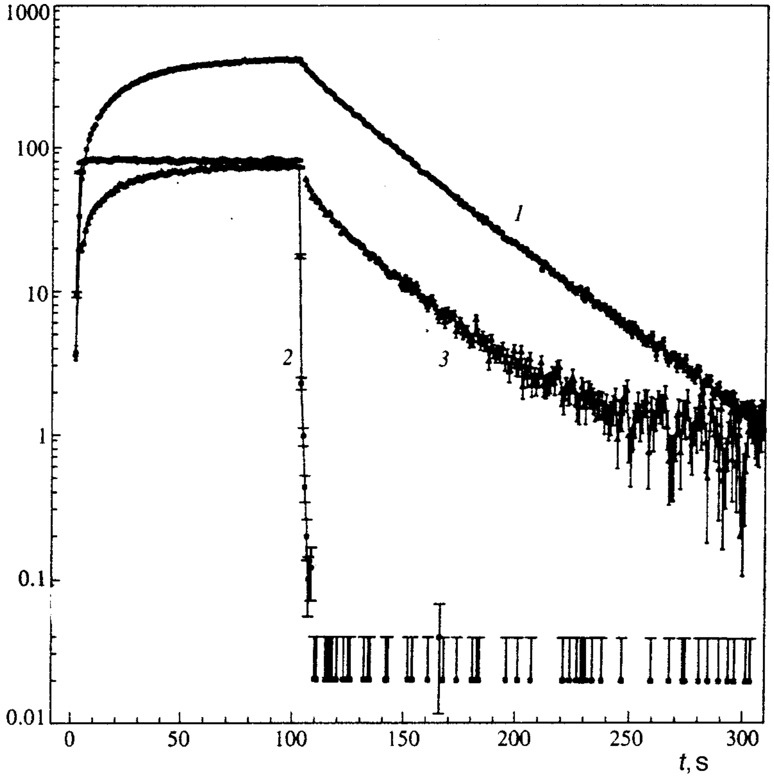


FIG. 2. Detected neutron intensity as a function of time during a single measurement cycle: (*1*) ultracold neutron monitor counter with aluminum foil (19 in Fig. 1); (*2*) ultracold neutron monitor counter (16 in Fig. 1); (*3*) heated neutron counter (11 in Fig. 1).

calibration source of heated neutrons we used high purity gaseous ^{4}He, which is inert and does not capture neutrons. The inelastic interaction of ^{4}He atoms with ultracold neutrons conforms to a simple theory. For practical calculations, we used the parameter $(p\tau)_{\rm He}$, with an experimental value of 467 ± 33 mbar·s.[9]

The technique for calibrating using ^{4}He is as follows: the count rate of heated ultracold neutrons is proportional to the heating probability and the density of neutrons in the containment vessel:

$$N_{\rm upsc}=(\tau_{\rm upsc}^{-1}+\tau_{\rm He}^{-1})\frac{N_0}{1+\tau_f(\tau_{\rm upsc}^{-1}+\tau_{\rm He}^{-1})}\exp\left[-\frac{p_{\rm He}\tau_{\rm eff}}{(p\tau)_{\rm He}}\right], \tag{1}$$

where the first factor is the heating probability and the second describes the dependence of the density of ultracold neutrons in the containment vessel on the total loss factor and the size of the inlet aperture. The third factor accounts for the loss of ultracold neutrons in the section of the neutron duct preceding the containment vessel, where $\tau_{\rm upsc}^{-1}=\eta_{ie}\overline{\gamma}$, $\tau_{\rm He}^{-1}=p_{\rm He}/(p\tau)_{\rm He}$, η_{ie} is the loss factor owing to heating of ultracold neutrons, $\overline{\gamma}$ is the product of the spectral loss function and the collision frequency of ultracold neutrons in the containment vessel averaged over the spectrum, or the so-called effective collision frequency, $p_{\rm He}$ is the pressure of ^{4}He, τ_f is the geometric filling time of the containment vessel, $\tau_{\rm eff}$ is the time to transport ultracold neutrons between the foil *4* and the inlet of the containment vessel *2*, and N_0 is the product of the density of ultracold neutrons in the preliminary containment vessel, the heated neutron detection efficiency, and the effective solid angle. The unknown parameters N_0, τ_f, and $\tau_{\rm eff}$ can be determined by fitting the experimentally measured $N_{\rm upsc}(p_{\rm He})$ curve to the theoretical curve (1).

The heated neutron count rate is plotted as a function of ^{4}He pressure in Fig. 3. The smooth curve corresponds to Eq. (1). The values of the fit parameters are shown in the caption to Fig. 3. Putting $p_{\rm He}=0$ in Eq. (1) and separating out the loss factor η, we obtain an expression for the cross-calibration of $N_{\rm upsc}$ and η:

$$\eta_{ie}=\frac{N_{\rm upsc}}{N_0}\frac{1}{\overline{\gamma}}\frac{1}{1-N_{\rm upsc}\tau_f/N_0}. \tag{2}$$

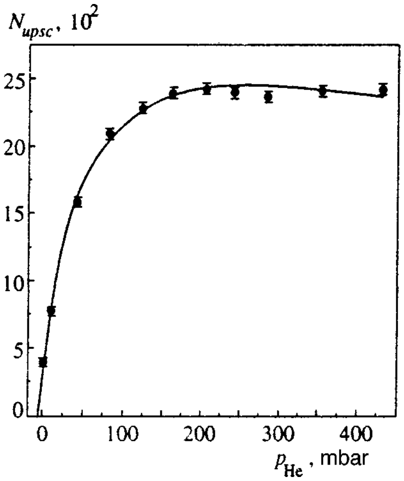


FIG. 3. Count rate of heated neutrons as a function of the ^{4}He pressure in the containment vessel for ultracold neutrons. $N_0=30175(2920)$ neutrons; $\tau_{\rm upsc}^{-1}=0.0145(0.0018)$ s^{-1}; $\tau_f=8.76(1.4)$ s; $\tau_{\rm eff}=0.28(0.07)$ s.

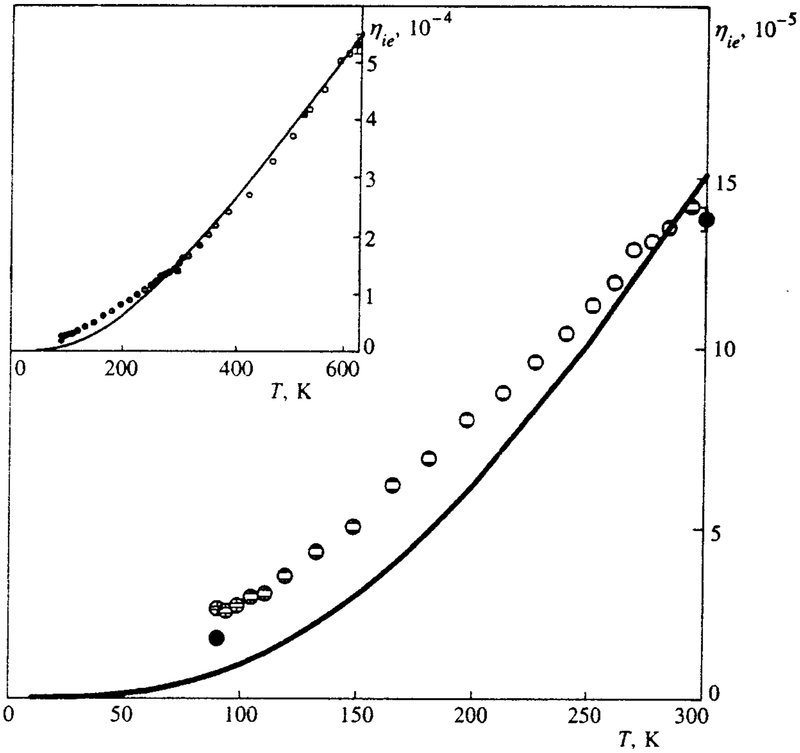


FIG. 4. Temperature dependence of the ultracold neutron loss factor owing to heating at the surface of Be samples. Open circles: rolled beryllium foil; filled circles: beryllium-coated stainless steel foil; smooth curve: theoretical dependence of the ultracold neutron loss factor according to Eq. (4).

The mean effective collision frequency $\overline{\gamma}$ of ultracold neutrons in the containment vessel is a computational parameter which in the present case is $\overline{\gamma}=120\pm 8$. The uncertainty in this calculation is due to imperfect knowledge of the upper bound on the spectrum of ultracold neutrons in the trap.

To properly calibrate the loss factor on a beryllium foil of neutrons heated in helium gas, one must know the final energy of the neutrons heated on the foil and in the helium, inasmuch as the detector efficiency can depend on energy. In the present experiment this circumstance was studied by measuring the transmission of heated neutrons through a filter with varying $^{10}BF_3$ pressure.[10] It was found that the dependence on the $^{10}BF_3$ pressure is essentially identical in the two cases, so there is no need to correct for the energy dependence of the detector when calibrating. Note, however, that no conclusion can be drawn regarding the energy of the neutrons heated on the beryllium, because there is enough material between the containment vessel and the detectors to thermalize the neutrons.

5. RESULTS

The experimentally obtained heated neutron count rates N_{upsc} were cross-calibrated with values of the dimensionless loss factor η in accordance with Eq. (2). Figure 4 shows the temperature variation of the partial ultracold neutron loss factor η_{ie} owing to heating on the surface of rolled beryllium foil (hollow points) and on a sputtered Be surface (solid points). The loss factor owing to heating of ultracold neutrons at room temperature is η_{ie} (300 K)$=(1.39\pm 0.18)\times 10^{-4}$ and $(1.47\pm 0.15)\times 10^{-4}$, respectively, for the sputtered and rolled samples, while at $T=90$ K these quantities are $(1.7\pm 0.2)\times 10^{-5}$ and $(2.6\pm 0.2)\times 10^{-5}$. The uncertainty in these results depends on the accuracy of the calibration technique.

Temperature dependences were measured during cooling and warming of the samples. We show temperature-averaged variations, which make it possible to eliminate so-called temperature hysteresis, a dynamic effect. In Fig. 4 only the statistical measurement errors are shown in the experimental temperature variation of the partial loss factor η_{ie}; the total measurement errors are given in the text.

As will be shown below, the measured heating probability cannot be explained by inelastic scattering on beryllium. When invoking the hypothesis of heating of ultracold neutrons on hydrogen contained in the surface layer, it is important to understand the relationship between the probabilities of capture and heating.

To assess the relative contribution of capture on hydrogen to the total loss factor, we use the data of Bondarenko *et al.*,[11] who determined experimentally the ratio of the inelastic scattering cross section to the cross section for absorption on hydrogen dissolved in the surface layer of a Be sample at $T=300$ K: $\sigma_{ie}^{H}/\sigma_{a}^{H}=20.6\pm 0.5$. We estimate the loss factor owing to capture for our samples to be $\eta_{a}^{H}=1.5\times 10^{-4}/20.6=0.7\times 10^{-5}$. Then the total loss factor at room temperature will be $(1.46\pm 0.18)\times 10^{-4}$ and $(1.54\pm 0.15)\times 10^{-4}$ for the sputtered and rolled samples, respectively, and at $T=90$ K it is $(2.4\pm 0.2)\times 10^{-5}$ and $(3.3\pm 0.2)\times 10^{-5}$. These results are in good agreement with measurements of the total loss in a sputtered beryllium trap of stainless steel[6] and in a trap made entirely of beryllium,[7] where the total loss factors were η (77 K)$=3\times 10^{-5}$ and η (300 K)$=1\times 10^{-4}$. A comparison of the loss factors at $T=90$ K shows that inelastic scattering is the main process limiting the confinement time of ultracold neutrons in traps with a beryllium coating, even at liquid nitrogen temperatures.

6. DEBYE MODEL CALCULATIONS OF INELASTIC SCATTERING

In an experiment on the transmission of very cold (10–15 m/s) neutrons through beryllium samples[12] it was shown that the temperature variation of the inelastic scattering cross section on beryllium is fairly well described by a Debye model in the incoherent approximation. However, a calculation of the loss probability during reflection of ultracold neutrons from beryllium owing to thermal fluctuations of the crystal lattice yields a value much lower than the experimentally observed one, so it becomes necessary to invoke the hypothesis of scattering on hydrogen dissolved in the beryllium, for which the scattering cross section is much larger.

According to Blokhintsev and Plakida[13] the flux of neutrons heated on physically sorbed (i.e., weakly bound to the surface) hydrogen should have a weak fractional-exponent dependence on the temperature. The quite sharp temperature variation in the loss factor suggests that the observed heating of ultracold neutrons is due to hydrogen strongly bound to the surface. An attempt to describe the dependence of inelastic scattering on hydrogen atoms rigidly bound to the crystal lattice of beryllium in terms of a Debye model appears logical; the stiffness of the bond between hydrogen and beryllium suggests that the hydrogen atom will undergo oscillations at frequencies characteristic of beryllium. Then, for the Debye approximation of the single-phonon inelastic scattering cross section we can obtain[14]

$$\sigma_{ie}^{\mathrm{H}}(E_0)=\sigma_0^{\mathrm{H}}\frac{1}{M}\int_0^{\theta}\sqrt{\frac{E}{E_0}}\exp\left(-\frac{E}{\epsilon M}\right)\frac{3E^2dE}{\theta^3(e^{E/T}-1)}, \quad (3)$$

where $\sigma_0^{\mathrm{H}}=80$ b is the cross section for incoherent scattering on a bound hydrogen atom, E_0 and E are the respective energies (in Kelvins) of the incident and inelastically scattered neutrons, $M=1$ is the relative mass of the hydrogen atom, and θ and ϵ are the parameters of the Debye model. The smooth curve in Fig. 4 is the theoretical dependence of the loss factor given by Eq. (3). The theoretical cross section for inelastic scattering on hydrogen calculated using Eq. (3) and reduced to the thermal velocity of a neutron is 4.4 b at $T=300$ K. For comparison, the cross section calculated from the experimental data of Ref. 11 is $\sigma_{ie}(300\text{ K})=6.8\pm0.2$ b.

Since the contribution of the cross section for inelastic scattering on hydrogen to the loss factor η_{ie} is determined by the concentration of hydrogen in the surface layer,

$$\eta_{ie}(T)=\frac{c_{\mathrm{H}}\sigma_{ie}^{\mathrm{H}}(T)+\sigma_{ie}^{\mathrm{Be}}(T)}{2\lambda b_{\mathrm{coh}}^{\mathrm{Be}}}\approx\frac{c_{\mathrm{H}}\sigma_{ie}^{\mathrm{H}}(T)}{2\lambda b_{\mathrm{coh}}^{\mathrm{Be}}}, \quad (4)$$

where c_{H} is the concentration of hydrogen atoms normalized to the concentration of Be atoms, λ is the wavelength of the ultracold neutrons, and $b_{\mathrm{coh}}^{\mathrm{Be}}$ is the coherent scattering length on Be. Fitting the experimental values of $\eta_{ie}(T)$ to a theoretical dependence of the form (4) enables us to determine the hydrogen concentration c_{H}. In the present case, the required hydrogen concentration is 96%, or $\sim10^{17}$ hydrogen atoms per cm^2 in a 100-Å-thick surface layer.

The existence of so much hydrogen in the surface layer seems unlikely. No direct measurements of the amount of surface hydrogen were made. However, data obtained by a nuclear reaction technique[15] on a number of other materials (Cu, Ni, Pb, Al) suggest that the hydrogen concentration in the surface layer is $(2-4)\times10^{15}$ cm^{-2}. This is much lower than the concentration required to explain our measured loss factors for the Be samples.

Since the loss factor η_{ie} is proportional to the product of the hydrogen concentration and the inelastic scattering cross section, there can in principle be two possible explanations for such a large loss factor. Indeed, the first is that the concentration of hydrogen in the surface layer is close to 100%. This must be tested by direct measurements, which will be conducted in the near future. The second is that the hydrogen concentration is low, but the cross section for inelastic scattering on hydrogen is enhanced. A possible mechanism for such an enhancement has been discussed by Serebrov and Romanenko,[16–18] who have shown that the total cross section for ultracold neutrons interacting with impurities in a medium can be greatly enhanced when the neutrons have subbarrier energies relative to the optical potential of the medium.

Note that the temperature dependence of the heating of ultracold neutrons given by the Debye model is satisfactory only at high temperatures. Near 100 K the calculated and experimental cross sections differ by a factor of two or more (see Fig. 4).

The difference in heating probability between the all-beryllium and sputtered foils is not significant, and is at most 30–40%. Thus, if ultracold neutrons do penetrate a material, the penetration falls off quite rapidly with depth.

A comparative experiment to study the heating of ultracold neutrons on all-beryllium foils and on beryllium-coated stainless steel foils would make it possible to estimate the probability of minor heating of ultracold neutrons, if such a process exists. For minor heating of ultracold neutrons, for example up to 10 m/s, a neutron will penetrate the foil material. It will be heated to thermal energies in an all-beryllium foil, but captured in a beryllium-coated stainless steel foil. The lack of any difference in the count rate of heated neutrons (to within 10%) in experiments with the different foils makes it possible to place an upper bound of 4×10^{-5} per collision on the probability of minor heating.

7. CONCLUSIONS

We have experimentally studied the behavior of the partial loss factor for ultracold neutrons heated by interaction with Be sample surfaces.

1. Our results suggest that inelastic scattering on hydrogen is the dominant factor in the loss of ultracold neutrons when they are confined in beryllium traps (or in traps with a beryllium coating), even at liquid nitrogen temperatures. Thus, the process responsible for anomalous losses at tem-

peratures around 80 K has been found experimentally. This result failed to confirm the results of Ref. 8.

2. The probability of heating of ultracold neutrons is anomalously high throughout the entire temperature range that was measured, but especially at low temperatures.

3. We propose that the observed loss factor at Be surfaces can be accounted for by an enhancement in the inelastic scattering and absorption cross sections on hydrogen, which is a small impurity relative to the beryllium,[16–18] rather than by an anomalously high concentration of hydrogen in the surface layer of the Be sample.

This experiment was supported by the Russian Fund for Fundamental Research (Grant No. 92-02-18663) and INTAS (Grant No. 93-0298). The authors thank S. Neumaier for assistance during the experiment at Grenoble, and all our colleagues for useful discussions and their interest in this work.

*)E-mail: varlamov@hep486.pnpi.spb.ru

---

[1] V. I. Lushchikov, Yu. N. Pokotilovskiĭ, A. V. Strelkov *et al.*, JETP Lett. **9**, 23 (1969).

[2] A. Steyerl, Phys. Lett. B **29**, 33 (1969).

[3] A. V. Strelkov and M. Khettsel't, Zh. Éksp. Teor. Fiz. **74**, 23 (1978) [Sov. Phys. JETP **47**, 11 (1978)].

[4] W. A. Lanford and R. Golub, Phys. Rev. Lett. **39**, 1509 (1977).

[5] Yu. Yu. Kosvintsev, V. I. Morozov, and G. I. Terekhov, Atomnaya Énergiya **55**, 288 (1983).

[6] P. Ageron, W. Mampe, and A. I. Kilvington, Z. Phys. B **59**, 261 (1985).

[7] V. P. Alfimenkov, V. V. Nesvizhevskiĭ, A. P. Serebrov*et al.*, JETP Lett. **55**, 84 (1992).

[8] V. K. Ignatovich, Kim Zun Bok, V. I. Lushchikov *et al.*, Report No. R3-82-811 [in Russian], OIYaI (Joint Institute for Nuclear Research) (1982).

[9] Yu. Yu. Kosvintsev, Yu. A. Kushnir, V. I. Morozov *et al.*, in *Neutron Physics* [in Russian], Vol. 1, TsNII Atominform, Moscow (1980), p. 130.

[10] S. D. Kalchev, A. V. Strelkov, and G. I. Terekhov, Report No. R2-90-120 [in Russian], OIYaI (Joint Institute for Nuclear Research) (1990).

[11] L. Bondarenko, S. Chemyavsky, A. Fomin *et al.*, Physica B **234–236**, 1189 (1997).

[12] I. S. Altarev, A. I. Egorov, V. F. Ezhov *et al.*, Preprint No. 246, LIYaF AN SSSR (Leningrad Inst. of Nuclear Physics, USSR Academy of Sciences) (1976).

[13] D. I. Bolkhintsev and N. M. Plakida, Phys. Status Solidi B **82**, 627 (1977).

[14] V. F. Turchin, in *Slow Neutrons*, Israel Program for Scientific Translations, Jerusalem (1965).

[15] P. H. La Marche, W. A. Lanford, and R. Golub, Nucl. Instrum. Methods **189**, 533 (1981).

[16] A. P. Serebrov, in *Proceedings of International Seminar on Interactions of Neutrons with Nuclei*, Dubna (1997) p. 67.

[17] A. P. Serebrov, Preprint No. 2193, PIYaF RAN (St. Petersburg Inst. of Nuclear Physics, Russian Academy of Sciences) (1997).

[18] A. P. Serebrov and N. V. Romanenko, Preprint No. 2194, PIYaF RAN (St. Petersburg Inst. of Nuclear Physics, Russian Academy of Sciences) (1997).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Relativistic version of the imaginary-time formalism

V. D. Mur and B. M. Karnakov*)

Moscow Engineering-Physics Institute, 115409 Moscow, Russia

V. S. Popov

Institute of Theoretical and Experimental Physics, 117218 Moscow, Russia

(Submitted 23 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 798–820 (September 1998)

A relativistic version of the quasiclassical imaginary-time formalism is developed. It permits calculation of the tunneling probability of relativistic particles through potential barriers, including barriers lacking spherical symmetry. Application of the imaginary-time formalism to concrete problems calls for finding subbarrier trajectories which are solutions of the classical equations of motion, but with an imaginary time (and thus cannot be realized in classical mechanics). The ionization probability of an *s* level, whose binding energy can be of the order of the rest energy, under the action of electric and magnetic fields of different configuration is calculated using the imaginary-time formalism. Besides the exponential factor, the Coulomb and pre-exponential factors in the ionization probability are calculated. The Hamiltonian approach to the tunneling of relativistic particles is described briefly. Scrutiny of the ionization of heavy atoms by an electric field provides an additional argument against the existence of the ''Unruh effect.'' [S1063-7761(98)00409-0]

1. The imaginary-time formalism was proposed[1,2] in connection with the calculation of the probability of the multiphoton ionization of atoms by the field of a strong light wave. Subbarrier trajectories which satisfy the classical equations of motion, but have an imaginary ''time'' t, are introduced to describe the tunneling process. The imaginary part of the action function calculated along such a ''classical'' trajectory determines the tunneling probability of a particle in quantum mechanics.[1–3]

The imaginary-time formalism has recently been used to investigate the influence of a magnetic field on the ionization of atoms and ions,[4–6] as well as on Lorentzian ionization,[7] which occurs when atoms move in a constant magnetic field. In these studies the subbarrier motion of the electron was considered to be nonrelativistic, as is the case for valence electrons in all atoms from hydrogen to uranium. However, in the case of ionization of the K shell in heavy atoms, the relativistic effects become significant, and systematic allowance for them (in the quasiclassical approximation) requires generalization of the imaginary-time formalism to the relativistic case, which can also be useful in a number of questions in relativistic nuclear physics and quantum chromodynamics.

We shall demonstrate the possibility of such a generalization in the specific problem of the ionization of a bound state, whose binding energy $E_b = mc^2 - E_0$ is comparable to the rest energy mc^2, and find the leading (exponential) factor in the ionization probability. We shall also consider the allowance for the Coulomb interaction in the tunneling process and the calculation of the pre-exponential factor.

This paper is organized in the following manner. Sections 2–5 consider the cases of pure electric fields and parallel and mutually perpendicular fields $\mathscr{E}$ and $\mathscr{H}$ (the potential binding an electron to an atomic core is assumed to be short-range). Section 4 is devoted to the special case of crossed fields, i.e., $\mathscr{E} \perp \mathscr{H}$ and $\mathscr{E} = \mathscr{H}$. In these examples we describe the procedure for determining the extremum subbarrier trajectory, which specifies the most probable tunneling path of a particle and thus the exponential factor in the ionization probability. Consideration of a bundle of subbarrier trajectories close to the extremum one also permits finding the pre-exponential factor. Section 6 describes the method for taking into account (within the imaginary-time formalism) the Coulomb interaction between an escaping electron and the atomic core. The introduction of a Coulomb correction enables us to consider the case of the ionization of neutral atoms and positive ions, which is of practical importance. The barrier width and the conditions for applicability of the imaginary-time formalism are discussed in Sec. 7. The Hamiltonian approach to the tunneling of relativistic particles is described briefly in Sec. 8. In Sec. 9 we use the solution of the ionization problem to discuss the Unruh effect,[8] and we offer some remarks to supplement the arguments previously advanced in Ref. 9 that the response of a uniformly accelerated detector is not universal, but depends on its structure. The concluding section, Sec. 10, enumerates the main conclusions of the work, and details of the calculations and some cumbersome formulas are presented in Appendices A–C.

In the following $\hbar = c = 1$, but in the final formulas we restore the dimensions of the quantities appearing in them. Some of the results of this work were announced in Ref. 10.

2. We start out from the ionization of an s level bound by short-range forces under the action of an electric field $\mathscr{E}$. In this case the subbarrier trajectories have the form

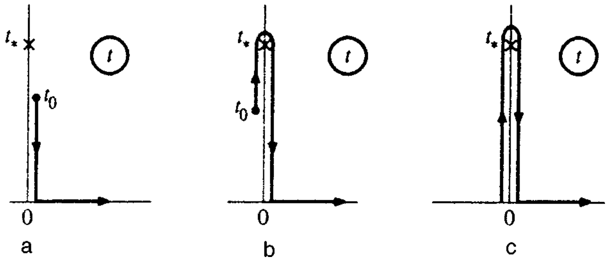


FIG. 1. Variation of the imaginary time t during subbarrier motion: a) energy of the level $E_0>0$; b) $0>E_0>-m$; c) $E_0=-m$ (level on the boundary of the lower continuum). The asterisk marks the branching point of the function $\sqrt{p(t)^2+m^2}$.

$$x=\frac{ip_\perp}{e\mathcal{E}}(\arcsin\tau_0-\arcsin\tau),\qquad y=0,$$

$$z=\frac{M}{e\mathcal{E}}\left(\sqrt{1-\tau^2}-\sqrt{1-\tau_0^2}\right). \tag{1}$$

In addition, $p_x(t)=p_\perp=\text{const}$, and $p_z(t)=e\mathcal{E}t=-iM\tau$. Here $\tau=ie\mathcal{E}t/M$ [τ is a real quantity, which is related to the intrinsic time s by the expression $s=-i(m/e\mathcal{E})\arcsin\tau$], $M=\sqrt{m^2+p_\perp^2}$, the z axis is parallel to $\mathcal{E}$, and $p_\perp$ is the transverse momentum of the particle. The initial time t_0 of subbarrier motion is determined from the boundary conditions[2]

$$\mathbf{r}(t_0)=0,\qquad \frac{m}{\sqrt{1-\dot{\mathbf{r}}^2(t_0)}}=E_0,\qquad \operatorname{Im}\mathbf{r}(0)=\operatorname{Im}\dot{\mathbf{r}}(0)=0 \tag{2}$$

[in the gauge where $\varphi(0,t)=0$ and $\mathbf{A}(0,t)=0$ and in the approximation of a zero radius for the forces binding the level], whence

$$t_0=\frac{im}{e\mathcal{E}}\sqrt{1-\epsilon_0^2+q^2},\qquad q=p_\perp/m. \tag{3}$$

Here $E_0=m\epsilon_0$ is the energy of the bound state ($-1\leqslant\epsilon_0<1$, and the values $\epsilon_0=\pm1$ correspond to the boundaries of the upper and lower continua). The probability of the tunneling of an electron along trajectory (1) equals[2]

$$dw(\mathbf{p}_\perp)=\frac{\text{const}}{\hbar m}\exp\left\{-\frac{2}{\hbar}\operatorname{Im}W(p_\perp)\right\}d^2p_\perp, \tag{4}$$

where W is the reduced action:

$$W=\int_{t_0}^{0}(L+E_0)\,dt-(\mathbf{p}\cdot\mathbf{r})_{t=0},$$

$$L=-m\sqrt{1-v^2}+e(\mathbf{A}\cdot\mathbf{v})-e\varphi \tag{4$'$}$$

($t=0$ is the time when the particle emerges from under the barrier). When the level deepens, the point t_0 moves in a complex plane, traveling around the branch point t_*, as is shown in Fig. 1. Taking this into account, we can write

$$\tau_0=\sqrt{1-\epsilon_0^2}+\frac{\epsilon_0^2q^2}{2\sqrt{1-\epsilon_0^2}}+\dots,$$

$$W=\frac{im^2}{2e\mathcal{E}}\left[(1+q^2)\arccos\frac{\epsilon_0}{\sqrt{1+q^2}}-\epsilon_0\sqrt{1-\epsilon_0^2+q^2}\right]$$

$$=\frac{im^2}{2e\mathcal{E}}[\Phi(\epsilon_0)+q^2\arccos\epsilon_0+O(q^4)], \tag{5}$$

where $\Phi(\epsilon)=\arccos\epsilon-\epsilon\sqrt{1-\epsilon^2}$. We note that $\Phi(-\epsilon)=\pi-\Phi(\epsilon)$ and

$$\Phi(\epsilon)=\left\{\begin{array}{ll}\frac{2^{5/2}}{3}(1-\epsilon)^{3/2}\left[1-\frac{3}{20}(1-\epsilon)+\dots\right], & \epsilon\to1,\\ \frac{\pi}{2}-2\epsilon+\frac{1}{3}\epsilon^3+\dots, & \epsilon\to0,\\ \pi-\frac{2^{5/2}}{3}(1+\epsilon)^{3/2}+\dots, & \epsilon\to-1.\end{array}\right\} \tag{5$'$}$$

Integrating (4) and (5) over the transverse momentum, we find the probability (per unit time) of the ionization of the s level in an electric field:

$$w(\mathcal{E},\epsilon_0)=\frac{mc^2}{2\hbar}|A_\kappa|^2\frac{\mathcal{E}/F_{\text{cr}}}{\arccos\epsilon_0}\exp\left\{-\frac{F_{\text{cr}}}{\mathcal{E}}\Phi(\epsilon_0)\right\}, \tag{6}$$

where A_κ is the asymptotic (as $r\to\infty$) coefficient of the wave function of the bound state in the absence of the external field $\mathcal{E}$ [compare with Eq. (9) in Ref. 4), and $F_{\text{cr}}=m^2c^3/e\hbar$ is the critical, or Schwinger, field, which is characteristic of quantum electrodynamics.[11,12]

In the nonrelativistic limit ($\epsilon_0\to1$) this formula transforms into the known expression[13,14] for the ionization probability of negative ions (H^-, Na^-, etc.). When $\epsilon_0=-1$, i.e., for a level which has sunken to the boundary of the lower continuum [the critical charge of the nucleus $Z_{\text{cr}}(1s_{1/2})=173$ (Refs. 15–18)], the exponential factor in (6) becomes equal to $\exp(-\pi F_{\text{cr}}/\mathcal{E})$ and coincides with the corresponding factor in Schwinger's formula[11] for the probability of the production of electron-positron pairs from a vacuum in a constant electric field.

3. If the fields $\mathcal{E}$ and $\mathcal{H}$ are parallel, the trajectory of a relativistic particle has the form of a spiral of variable pitch. The subbarrier trajectory is obtained from known formulas[19] using the analytical continuation with respect to t:

$$z=\frac{M}{e\mathcal{E}}\left(\sqrt{1-\tau^2}-\sqrt{1-\tau_0^2}\right),$$

$$\boldsymbol{\rho}=x+iy=\frac{i\mathbf{p}_\perp}{e\mathcal{H}}\left(e^{-i\vartheta}-e^{-i\vartheta_0}\right), \tag{7}$$

$$\tau=i\frac{e\mathcal{E}t}{M}=i\sinh\left(\frac{\mathcal{E}}{\mathcal{H}}\vartheta\right),\qquad M=\sqrt{m^2+p_\perp^2}$$

TABLE I. Accuracy of the approximation (29).

ϵ_b	$\rho=0.5$	0.75	1.0	∞
0.1	0.116	0.021	1.2(−3)	−0.014
0.2	0.467	0.087	5.3(−3)	−0.058
0.25	0.729	0.137	8.5(−3)	−0.092
0.50	2.80	0.580	0.040	−0.404
0.75	5.69	1.36	0.102	−1.01
1.0	8.76	2.49	0.206	−2.04
1.2	11.1	3.63	0.326	−3.27

Note. The values of the error δ (in percent) are given [see Eq. (B4)]; $\epsilon_b = E_b/mc^2$, $\rho=\mathcal{E}/\mathcal{H}$, and $a(b)\equiv a\cdot 10^b$.

(the fields $\mathcal{E}$ and $\mathcal{H}$ are parallel to the z axis). The quantities t and ϑ in the subbarrier motion are imaginary, and τ is real. For a scalar (spin-free) particle the action S equals

$$\begin{aligned} S(t) &= \int^t \left(-m\sqrt{1-v^2} + \frac{1}{2} e\mathcal{H}(x\dot{y} - y\dot{x}) + e\mathcal{E}z \right) dt \\ &= \frac{M^2}{4e\mathcal{E}} \sinh 2\psi - \frac{m^2}{2e\mathcal{E}}\psi + \frac{p_\perp^2}{2e\mathcal{H}} \sin\vartheta + \text{const}, \\ \vartheta &= \frac{e\mathcal{H}}{m}s, \quad \psi = \frac{e\mathcal{E}}{m}s, \end{aligned} \tag{8}$$

where s is the intrinsic time of the particle (which is purely imaginary). We ultimately obtain1)

$$\operatorname{Im} W(p_\perp) = \frac{m^2}{2e\mathcal{E}}\Phi(\epsilon_0) + \sinh\left(\frac{\mathcal{H}}{\mathcal{E}} \arccos\epsilon_0\right) \frac{p_\perp^2}{2e\mathcal{H}} + \dots \tag{9}$$

[when $\mathcal{H}=0$, this formula transforms into (5)]. Integrating (4) and (9) over $p_\perp$, we obtain

$$\frac{w(\mathcal{E},\mathcal{H})}{w(\mathcal{E},0)} = \frac{\sigma}{\sinh\sigma}, \quad \sigma = \frac{\mathcal{H}}{\mathcal{E}} \arccos\epsilon_0 . \tag{10}$$

For nonrelativistic bound states $\epsilon_0 = 1 - \alpha^2\kappa^2/2 \to 1$ ($\alpha = e^2/\hbar c = 1/137$, and $\kappa \sim 1$; see Table I in Ref. 6), and $\arccos\epsilon_0 = \alpha\kappa + (\alpha\kappa)^3/24 + \dots$. Therefore, $\sigma = \alpha\kappa\mathcal{H}/\mathcal{E}$ coincides with the parameter γ introduced in Ref. 4, and Eq. (10) yields the correct expression for the pre-exponential factor $P_0(\gamma) = \gamma/\sinh\gamma$ in the case of the ionization of a negative ion.[5,20] In the other limit, $\epsilon_0 = -1$, we have $\sigma = \pi\mathcal{H}/\mathcal{E}$, and Eq. (10) is consistent with the first term of the Schwinger expansion[11] for the imaginary part of the effective Lagrange function in scalar electrodynamics:

$$w_0(\mathcal{E},\mathcal{H}) = \frac{\alpha}{2\pi} \frac{\mathcal{E}\mathcal{H}}{\sinh(\pi\mathcal{H}/\mathcal{E})} \exp\left(-\frac{\pi F_{\text{cr}}}{\mathcal{E}}\right) \tag{11}$$

[under the condition2) $\mathcal{E}, \mathcal{H} \ll F_{\text{cr}}$ the ensuing terms of this expansion are exponentially small compared with (11)].

The imaginary-time formalism enables us to obtain a formula like (9) for fermions as well: we need only add the spin expression

$$\begin{aligned} &\frac{ie}{2m}\epsilon_{\alpha\beta\mu\nu} \int F^{\alpha\beta} u^\mu s^\nu \, ds \\ &= \frac{e}{m} \int \{(\mathbf{s}\cdot\mathcal{H}) - (\mathbf{v}\cdot\mathbf{s})(\mathbf{v}\cdot\mathcal{H}) + (\mathbf{v}\times\mathbf{s})\cdot\mathcal{E}\} \, dt \end{aligned} \tag{12}$$

to the action function, whose contribution (as t varies along the loop in Fig. 1c) is calculated using the Bargmann–Michel–Telegdi equations[21] for the four-spin s^ν in an external field. We can ultimately obtain[22] a formula like (11) for $w_s(\mathcal{E},\mathcal{H})$, in which the pre-exponen tial factor $(\sinh\sigma)^{-1}$, where $\sigma = \pi\mathcal{H}/\mathcal{E}$, should be replaced by $2\coth\sigma$ in the case of electrons ($s=1/2$ $g=2$) and by $(\sinh\sigma)^{-1} + 2\coth\sigma$ in the case of vector bosons ($s=1$) with a gyromagnetic ratio $g=2$, i.e., in the case where the theory is renormalized.[23]

In the special case of constant uniform fields $\mathcal{E}$ and $\mathcal{H}$, we can take into account all (exponentially small) corrections to (11) by introducing subbarrier trajectories that correspond to the n-fold wandering of the particle between the lower and upper continua.[24] Thus, the quasiclassical imaginary-time formalism enables us to obtain not only the leading term (11) in the probability of pair production, but also to exactly restore the entire series for $w = 2\operatorname{Im}\mathcal{L}$ previously calculated (in a more complicated way) by Schwinger[11] for scalar and spinor particles and by Vanyashin and Terent'ev[23] for vector bosons. This coincidence between the results, however, is accidental and is similar to the coincidence between the exact and quasiclassical spectra of the Schrödinger equation for several simple potentials: a harmonic oscillator, a Coulomb potential, the Morse potential $U(x) = U_0(e^{-2x} - 2e^{-x})$, etc.

To complete this section, we note the difference between (6) and (11). In the case of an electric field and $\epsilon_0 = -1$ ($\mathcal{H}=0$, the level on the boundary of the lower continuum) these formulas, while completely coinciding with respect to the exponential factor specified by the value of $\operatorname{Im} S$ along the extremum ($p_\perp = 0$) trajectory, differ with respect to the dependence of the pre-exponential factor on electric field: according (6), $P(\mathcal{E}) \propto \mathcal{E}$, while for (11) $P(\mathcal{E}) \propto \mathcal{E}^2$. This is not surprising, since (6) and (11) refer to different physical processes and have different dimensions: the probability (11) refers to the invariant four-volume of the vacuum $VT=1$ and has the dimensions m^4 (or $\text{cm}^{-3}\cdot\text{s}^{-1}$), and (6) refers to an individual atom and has the dimensions s^{-1}.

4. Crossed fields. We go over to more complicated cases, in which the extremum trajectory is not just one-dimensional. Let $\mathcal{E}\perp\mathcal{H}$; the gauge

$$\mathbf{A} = (-\mathcal{H}y, 0, 0), \quad \varphi = -\mathcal{E}y \tag{13}$$

corresponds to the fact that the y axis is directed along the field $\mathcal{E}$ and the z axis is directed along $\mathcal{H}$. For crossed fields ($\mathcal{E}=\mathcal{H}$) the classical trajectories are assigned in parametric form:[19]

$$x = \frac{m}{2e\mathcal{E}}\left[(\lambda^2 - 1)q + \frac{1}{3}\lambda^2 q^3\right] + C_1,$$

$$y = \frac{m}{2e\mathcal{E}}\lambda q^2 + C_2, \quad z = \frac{1}{e\mathcal{E}}\lambda q p_z + C_3,$$

$$t=\frac{m}{2e\mathscr{E}}\left[(\lambda^2+1)q+\frac{1}{3}\lambda^2q^3\right],$$

where $q=p_y/m$ is a parameter, and $\lambda=m/(\sqrt{p^2+m^2}-p_x)$ is the (dimensionless) integral of motion, which is related to the integral indicated in Ref. 19 $\alpha=E_{\rm kin}-p_x$ by the expression $\lambda=m\alpha^{-1}$. In subbarrier motion the "time" t and the momentum component p_y are purely imaginary:

$$t=-\frac{im}{e\mathscr{E}}\tau,\quad q=i\lambda^{-1}u \tag{14}$$

$(-\tau_0<\tau<0)$. Using the boundary conditions (2), we find the integration constants C_i and determine the extremum subbarrier trajectory:

$$x=\frac{im}{2e\mathscr{E}\lambda}\left[(\lambda^2-1)u-\frac{1}{3}u^3\right]=\frac{im}{6e\mathscr{E}\lambda}(u_0^2-u^2)u, \tag{15}$$

$$y=\frac{m}{2e\mathscr{E}\lambda}(u_0^2-u^2),\quad z=0,\quad \tau=\frac{1}{2\lambda}\left[\frac{1}{3}u^3-(\lambda^2+1)u\right],$$

whence

$$r_0(t)\equiv\sqrt{x^2+y^2+z^2}=\frac{m}{2e\mathscr{E}\lambda}(u_0^2-u^2)\sqrt{1-\frac{u^2}{9}},$$

$$\sqrt{1-v^2}=\frac{2\lambda}{\lambda^2+1-u^2}, \tag{15$'$}$$

where $u_0\equiv u(-\tau_0)$, the value $\tau=-\tau_0$ corresponds to the beginning of subbarrier motion, and $\tau=u=0$ corresponds to the time when the particle emerges from under the barrier. In subbarrier motion the velocity component v_x is real, and v_y (along the electric field) is purely imaginary. The boundary conditions (2) are satisfied, if

$$(\lambda^2-1)u_0-\frac{1}{3}u_0^3=0,\quad \lambda+(1-u_0^2)\lambda^{-1}=2\epsilon_0.$$

Hence it follows that u_0, τ_0, and λ are uniquely determined by the energy of the bound state: $u_0=\sqrt{1+\xi^2}\tau_0=\sqrt{3}\xi$, and

$$\xi=\sqrt{\lambda^2-1}=\sqrt{1-\frac{1}{2}\epsilon_0(\sqrt{\epsilon_0^2+8}-\epsilon_0)}. \tag{16}$$

Here we have introduced the parameter ξ $0<\xi\leqslant\sqrt{3}$, which is convenient for the further treatment.

Equations (14)–(16) completely determine the extremum trajectory. We note that u is proportional to the intrinsic time s of the particle:

$$s=\int^t\sqrt{1-v^2}\,dt=\frac{im}{e\mathscr{E}}u,$$

which, like t, is imaginary on the subbarrier portion of the trajectory. In particular, at the initial time[3)] we have

$$t_0=\frac{s_0}{\sqrt{1+\xi^2}}=\frac{im}{e\mathscr{E}}\sqrt{\frac{3\xi^2}{1+\xi^2}}.$$

Going over to integration over u in (4$'$) and using the gauge (13), we have

$$L=m(\epsilon_0-\sqrt{1-v^2})+e\mathscr{E}(1-\dot{x})y$$
$$=-\frac{m}{\sqrt{1+\xi^2}}\,\frac{\xi^4-(\xi^2-2)u^2}{\xi^2+2-u^2},$$

and we find (to within exponential accuracy) the ionization probability in crossed fields:

$$w(\mathscr{E}=\mathscr{H},\epsilon_0)\propto\exp(-2\,\mathrm{Im}\,W)$$
$$=\exp\left(-2\sqrt{3}\frac{\xi^3}{1+\xi^2}\,\frac{F_{\rm cr}}{\mathscr{E}}\right). \tag{17}$$

In the nonrelativistic limit

$$\xi=\frac{1}{\sqrt{3}}\alpha\kappa\left(1+\frac{7}{72}\alpha^2\kappa^2+\ \dots\right)$$

and

$$w\propto\exp\left\{-\frac{2}{3\varepsilon}\left(1-\frac{1}{24}\alpha^2\kappa^2\right)\right\},\quad \varepsilon\ll1, \tag{18}$$

where $\varepsilon=\mathscr{E}/\kappa^3\mathscr{E}_a$ and $\mathscr{E}_a=\alpha^3F_{\rm cr}$. The correction of order α^2 slightly increases the ionization probability compared with the corresponding nonrelativistic formula.[13] The factor $2\sqrt{3}\xi^3/(1+\xi^2)$ in the exponential function in (17) increases monotonically with increasing depth of the level (for example, it equals $\sqrt{3}$ and 9/2 when $\epsilon_0=0$ and -1), causing a sharp drop in the probability w.

To conclude this section, we offer several comments.

1) As we know, the probability of the production of pairs from a vacuum vanishes in crossed fields. This follows both from the exact expression for $\mathrm{Im}\,\mathscr{L}$ (Refs. 11 and 12) and from the following simple argument. When we go over to reference frame K, which moves with the velocity V in the direction perpendicular to $\mathscr{E}$ and $\mathscr{H}$, the intensities of the crossed fields decrease by a factor of $\sqrt{(c+V)/(c-V)}$ and can be made extremely small as $V\to c$. Pairs, of course, are not generated in an extremely weak electric field.

In our case the probability w is nonzero: according to (17) $w\propto\exp(-9F_{\rm cr}/2\mathscr{E})$ when $\epsilon_0=-1$. This difference is attributed to the fact that there is a preferential reference frame K_0 (in which the atom is at rest), and the transition from K_0 to the Lorentz frame K qualitatively alters the formulation of the problem (unlike the vacuum, which is Lorentz-invariant).

2) As is seen from (18), the relativistic subbarrier trajectory led to a result which differs only slightly from the result of the nonrelativistic theory, if $E_0\approx m$ (or $\kappa\sim1$). Nevertheless, the motion of charged particles in crossed fields is always relativistic,[19] since the drift velocity $v_d=c\mathscr{E}/\mathscr{H}\to c$. This apparent paradox is explained by recalling that the electron accelerates to a velocity of the order of the velocity of light soon after it emerges from under the barrier:

$$\frac{p_x}{mc}\approx k_1\left(\frac{t}{T_0}\right)^{2/3},\quad \frac{p_y}{mc}\approx k_2\left(\frac{t}{T_0}\right)^{1/3},$$

$$\frac{v}{c}=1-k_3\left(\frac{t}{T_0}\right)^{-4/3},\quad t\to\infty,$$

where $T_0 = mc/e\mathscr{E}$ is the time during which the electron achieves a velocity $v \sim c$, and the k_i are coefficients of order unity. On the other hand, the characteristic tunneling time $T_t = me\kappa/\hbar\mathscr{E}$; therefore, $T_0/T_t = (\alpha\kappa)^{-1} \gg 1$.

3) Unlike the one-dimensional quasiclassical approximation, here the point of emergence from under the barrier is not a stopping point of the particle (even for the extremum trajectory). For example, at $t=0$ from (15) we obtain

$$v_x(0) = c\,\frac{\xi^2}{\xi^2+2}, \qquad p_x(0) = mc\,\frac{\xi^2}{2\sqrt{\xi^2+1}}, \tag{19}$$

the escape velocity being directed perpendicularly to $\mathscr{E}$ and $\mathscr{H}$.

5. Let $\mathscr{E} \perp \mathscr{H}$, but let the ratio $\rho = \mathscr{E}/\mathscr{H}$ deviate from unity. Such a field configuration appears, in particular, in the rest frame of or an atom or ion moving in a constant magnetic field [so-called Lorentzian ionization (see, for example Refs. 7 and 25); in this case $\rho < 1$].

The classical trajectories can be obtained using the Lorentz transformation from a reference frame in which only one of the fields, $\mathscr{H}$ or $\mathscr{E}$, is present.[26,27] For example, when $\mathscr{E} < \mathscr{H}$, the extremum subbarrier trajectory has the form

$$x = i\,\frac{m}{e\mathscr{H}}\,\frac{a\rho}{(1-\rho^2)^{3/2}}\left(\tau - \frac{\tau_0}{\sinh\tau_0}\sinh\tau\right),$$

$$y = \frac{m}{e\mathscr{H}}\,\frac{a\rho}{1-\rho^2}(\cosh\tau_0 - \cosh\tau)\,\frac{\tau_0}{\sinh\tau_0}, \qquad z=0, \tag{20}$$

$$\omega_c t = i\,\frac{a}{(1-\rho^2)^{3/2}}\left(\tau - \rho^2\,\frac{\tau_0}{\sinh\tau_0}\sinh\tau\right), \qquad -\tau_0 < \tau < 0,$$

where $\omega_c = e\mathscr{H}/m$ is the cyclotron or Larmor frequency, and $a = (\sqrt{p^2+m^2} - \rho p_x)/m$ is the integral of motion.[4)] The equations for determining the constants a and τ_0 follow from (2):

$$\frac{\tanh\tau_0}{\tau_0} = \frac{\rho^2 a}{a-(1-\rho^2)\epsilon_0}, \qquad \cosh\tau_0 = \frac{a-(1-\rho^2)\epsilon_0}{\rho\sqrt{a^2+\rho^2-1}} \tag{21}$$

(here $0 < \rho < 1$, and $a > \sqrt{1-\rho^2}$). In this case the intrinsic time of the particle equals

$$s = \int_0^t \sqrt{1-v^2}\,dt = \frac{i}{\omega_c}\,\frac{a}{1-\rho^2}\sqrt{1-\rho^2\left(\frac{\tau_0}{\sinh\tau_0}\right)^2}\;\tau. \tag{22}$$

In the initial moment of subbarrier motion

$$s_0 = t_0\sqrt{\frac{1-\rho^2(\tau_0/\sinh\tau_0)^2}{1-\rho^2}},$$

so that $|s_0| > |t_0|$, as in the preceding case. The system (21) can be reduced to a single equation:

$$\frac{1-\rho^2\tau_0\coth\tau_0}{\sqrt{1-\rho^2(\tau_0/\sinh\tau_0)^2}} = \sqrt{1-\rho^2}\,\epsilon_0, \tag{23}$$

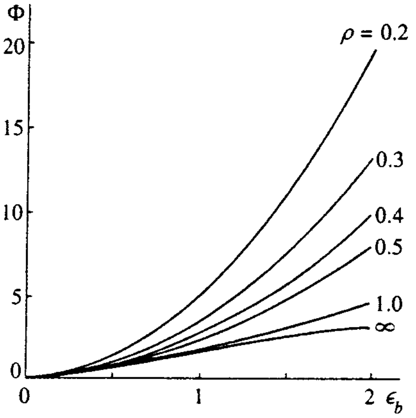


FIG. 2. Plots of $\Phi(\epsilon_0,\rho)$, which determines the exponential factor in the ionization probability (24), as a function of the binding energy of the level $\epsilon_b = (m-E_0)/m$, $0 < \epsilon_b < 2$. The values of $\rho = \mathscr{E}/\mathscr{H}$ are indicated near the curves.

which specifies the parameter $\tau_0 = \tau_0(\epsilon_0,\rho)$. In the gauge (13) we have the Lagrangian

$$L = -m\sqrt{1-v^2} + e\mathscr{E}(1-\rho^{-1}\dot{x})y.$$

From (4′) and (20) we find (to within exponential accuracy) the ionization probability:

$$w \propto \exp\{-F_{\mathrm{cr}}\mathscr{E}^{-1}\Phi(\epsilon_0,\rho)\}, \tag{24}$$

where

$$\Phi = \frac{\rho\tau_0(1-a\epsilon_0)}{\sqrt{1-\rho^2}} = \frac{\rho\tau_0}{\sqrt{1-\rho^2}}\left[1 - \frac{(1-\rho^2)\epsilon_0^2}{1-\rho^2\tau_0\coth\tau_0}\right] \tag{25}$$

[when $\rho = 1$, i.e., in the case of crossed fields, it is more convenient to use Eqs. (17) and (16) from the preceding section].

Equations (23)–(25) solve the problem posed. Let us discuss some limiting cases and results of numerical calculations.

a) As can be seen from Fig. 2, the values of $\Phi(\epsilon_0,\rho)$ increase with both increasing depth of the level and increasing strength of the magnetic field (at a fixed value of $\mathscr{E}$). This fact is easily explained within the imaginary-time formalism: when $\mathscr{H}=0$, the extremum trajectory is one-dimensional (and is directed along $\mathscr{E}$), and as $\mathscr{H}$ increases, it is ‘‘twisted,’’ and the barrier width increases (see Sec. 7 below).

b) In the nonrelativistic limit $\epsilon_0 \to 1$ it is convenient to go over to atomic units:

$$\Gamma = \hbar w = \frac{me^4\kappa^2}{2\hbar^2}|A_\kappa|^2\frac{\mathscr{E}}{\mathscr{E}_a} \times\exp\left\{-\frac{2\kappa^3\mathscr{E}_a}{3\mathscr{E}}[g(\gamma) - \alpha^2\kappa^2 g_1(\gamma) + O(\alpha^4)]\right\}, \tag{26}$$

where $\gamma = \omega_c/\omega_t = \alpha\kappa\mathscr{H}/\mathscr{E}$, Γ is the width of the level, w is the ionization probability, $\mathscr{E}_a = \alpha^3 F_{\mathrm{cr}} = 5.14\times10^9$ V/cm is the electric field intensity, $\kappa = \sqrt{E_b/I_{\mathrm{H}}}$ [see also Eq. (B1)], E_b is the binding energy of the level, I_{H} is the ionization potential of the hydrogen atom,

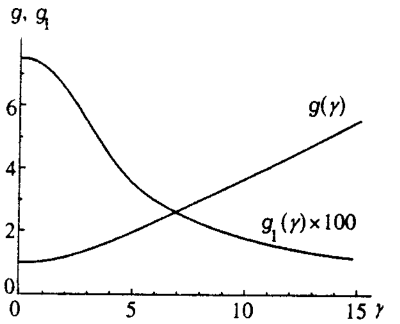


FIG. 3. Plots of $g(\gamma)$ and $g_1(\gamma)$ from (26). The scale along the vertical axis has been magnified 100 times for g_1.

$$g(\gamma)=\frac{3\tau_0}{2\gamma}\left[1-\frac{\sqrt{\tau_0^2-\gamma^2}}{\gamma^2}\right]=\frac{1}{\gamma^3}\sum_{k=1}^{\infty}c_k\tau_0^{2k+1},$$

$$c_k=3\cdot 2^{2k-1}\frac{B_{2k}}{(2k-1)!} \tag{27}$$

($c_1=1$, $c_2=-2/15$, $c_3=2/105$, etc., the B_{2k} are Bernoulli numbers), $\tau_0=\tau_0(\gamma)$ is determined from the equation

$$\tanh\tau_0=\frac{\tau_0}{1+\sqrt{\tau_0^2-\gamma^2}}, \tag{28}$$

and the expression for $g_1(\gamma)$ is fairly cumbersome and has been transferred to Appendix A. The inequality $g_1(\gamma)\ll g(\gamma)$ always holds. For example, $g(0)=1$, and $g_1(0)=3/40$. As γ increases, i.e., as the magnetic field increases, the relative value of the relativistic correction only decreases (Fig. 3).

c) If $\rho\gg\alpha=1/137$, i.e., if the electric field is not very small, then $\gamma\ll 1$, and the expression (26) can be simplified:

$$w\propto\frac{\mathscr{E}}{\mathscr{E}_a}\exp\left\{-\frac{2\kappa^3\mathscr{E}_a}{3\mathscr{E}}\left[1+\frac{\alpha^2\kappa^2}{30\mathscr{E}^2}\left(\mathscr{H}^2-\frac{9}{4}\mathscr{E}^2\right)\right]\right\}. \tag{29}$$

This simple approximation has surprisingly good accuracy even for deep levels (but, of course, not when $\epsilon_0\approx-1$). For example, when $\epsilon_0=0$ (i.e., for a level whose binding energy equals mc^2 and $\epsilon_b=1$) the difference between the exponent in (29) and the exact function $\Phi(0,\rho)$ from (24) amounts to 2% in the case of a pure electric field and only 0.2% for crossed fields (for further details, see Table I, as well as Appendix B).

d) The limit $\rho\to 0$ corresponds to removal of the electric field:

$$w\propto\exp\left\{-\left[(1-\epsilon_0)^2\frac{F_{\mathrm{cr}}\mathscr{H}}{\mathscr{E}^2}+\frac{1}{2}(1-\epsilon_0^2)\frac{F_{\mathrm{cr}}}{\mathscr{H}}+O\left(\rho^2\frac{F_{\mathrm{cr}}}{\mathscr{H}}\right)\right]\right\} \tag{30}$$

($\mathscr{E}\ll\mathscr{H}\ll F_{\mathrm{cr}}$). The proportionality $w\propto\exp(-\mathrm{const}/\mathscr{E}^2)$ shows that the ionization probability is extremely small in this case.

e) When $\rho\to 1$, an uncertainty appears in (20)–(25). Removing it, we arrive at Eq. (17) for crossed fields.

f) We assumed above that $\rho<1$. The equations for the case of $\mathscr{E}>\mathscr{H}$ can easily be obtained from the preceding relations using the analytical continuation: $\sqrt{1-\rho^2}\to i\sqrt{\rho^2-1}$, $\tau\to i\tau$. In this case the ‘‘time’’ t and the intrinsic time s in subbarrier motion remain purely imaginary.

6. The Coulomb correction. We have hitherto neglected the Coulomb interaction between the escaping electron and the atomic core; therefore, the equations obtained above refer to the case of the ionization of negative ions (like H^-, Na^-, etc.). To take into account the Coulomb interaction, we use perturbation theory within the imaginary-time formalism[4,28] and a procedure for matching the asymptotic expansions, introducing the matching point r_1 such that $\langle r\rangle\ll r_1\ll b$, where $\langle r\rangle$ is the mean radius of the bound state and b is the barrier width (under the conditions for applicability of the quasiclassical approximation, the choice of such a point r_1 is always possible; see the next section). Proceeding in analogy to Refs. 4 and 5, we obtain the Coulomb factor Q in the ionization probability:

$$Q=\exp\left\{2\left[\eta\ln(\mu r_1)+iZ\alpha\int_{t_1}^{0}[\mathbf{r}_0^2(t)]^{-1/2}dt\right]\right\}, \tag{31}$$

where $\mu=m\sqrt{1-\epsilon_0^2}$, $\eta=Z\alpha\cdot\epsilon_0(1-\epsilon_0^2)^{-1/2}$ is the relativistic analog of the Sommerfeld parameter, $r_1=[\mathbf{r}_0^2(t_1)]^{1/2}$ is the matching point, $\mathbf{r}_0(t)$ is the extremum subbarrier trajectory, t is the imaginary time, and Z is the charge of the atomic core, so that at large distances from the atom, $\langle r\rangle\ll r\lesssim b$, the electron moves in the potential $V(r)=-Z\alpha/r+o(r^{-2})$ (in a free atom, i.e., where $\mathscr{E}=\mathscr{H}=0$). We note that $Z=1$, 2, and 0 in the cases of the ionization of a neutral atom, a singly charged positive ion, and a singly charged negative ion.

Let us examine some special cases. The extremum trajectory for an electric field $\mathscr{E}$ can be obtained from (1) by setting $p_\perp=0$ and is one-dimensional. An analytical calculation of the integral in (31) gives

$$Q=[2(1-\epsilon_0^2)^{3/2}F_{\mathrm{cr}}/\mathscr{E}]^{2\eta}\exp(2Z\alpha\arccos\epsilon_0) \tag{32}$$

(the details of the calculations are discussed in Appendix C). The Coulomb correction increases the ionization probability significantly. For example, in the nonrelativistic case

$$Q=(2\kappa^3\mathscr{E}_a/\mathscr{E})^{2Z/\kappa}\gg 1. \tag{33}$$

Although the magnitude of this correction decreases as the binding energy of the level $E_b=m(1-\epsilon_0)$ increases, it remains significant. For example, $Q=\exp(\pi Z\alpha)\sim 25$ when $E_b=mc^2$.

Multiplying the expressions (6) and (32), we find the ionization probability w (we stress that in the present case, i.e., when only an electric field is present, both the exponential expression and the Coulomb and pre-exponential factors are calculated; therefore, the formula for w is asymptotically exact in the weak-field limit). In particular, if $\epsilon_0\approx 1$, using the expansions

$$\epsilon_0=1-\frac{1}{2}\alpha^2\kappa^2,\quad \mu=m\alpha\kappa\left(1-\frac{3}{8}\alpha^2\kappa^2+\ \dots\right),$$

$$\eta=\eta_0\left(1-\frac{3}{8}\alpha^2\kappa^2+\ \dots\right)$$

$(\eta_0 = Z/\kappa)$, with allowance for corrections of order α^2 we have

$$w(\mathscr{E},\epsilon_0)=\frac{me^4\kappa^2}{2\hbar^2}|A_\kappa|^2\left(\frac{\varepsilon}{2}\right)^{1-2\eta_0}\exp\left\{-\frac{2}{3\epsilon}\times[1-\alpha^2\kappa^2(c_0+c_1\varepsilon\ln\varepsilon+c_2\varepsilon)+\dots]\right\}, \quad (34)$$

where

$$c_0=\frac{3}{40},\quad c_1=\frac{9}{8}\eta_0,\quad c_2=-\left(3-\frac{9}{8}\ln 2\right)\eta_0+\frac{1}{16},$$

$$\varepsilon=\mathscr{E}/\kappa^3\mathscr{E}_a \ll 1.$$

In analogy to (29) it can be expected that the applicability region of this ''semirelativistic'' formula extends up to $E_b \sim mc^2$.

In the case of crossed fields the integral in (31) can be expressed in quadratures [see (15′) and Appendix C], enabling us to obtain the Coulomb correction in closed form:

$$Q=\left[\frac{2\xi^3(3-\xi^2)^2}{\sqrt{3}(1+\xi^2)}\frac{F_{\text{cr}}}{\mathscr{E}}\right]^{2\eta}\exp\left(6Z\alpha\arcsin\frac{\xi}{\sqrt{3}}\right), \quad (35)$$

where $\xi=\xi(\epsilon_0)$ was defined in (16). When $\epsilon_0 \to 1$, this expression transforms into (33), and when $\epsilon_0=0$, we obtain $Q=\exp(3.8Z\alpha)\sim 45$. In both cases (32) and (35), the Coulomb factor $Q \gg 1$ when $\epsilon_0=0$.

Finally, for the subbarrier trajectory (21) we obtain

$$r_0(t)=\frac{m}{e\mathscr{E}}\frac{a\rho^2\tau_0}{(1-\rho^2)^{3/2}}\left[(1-\rho^2)\left(\frac{\cosh\tau_0-\cosh\tau}{\sinh\tau_0}\right)^2-\left(\frac{\sinh\tau}{\sinh\tau_0}-\frac{\tau}{\tau_0}\right)^2\right]^{1/2}. \quad (36)$$

In this case the integral in (31) is no longer taken analytically. As is shown in Appendix C, it can be brought into the regularized form (C9), which does not contain an arbitrary matching point, and thereafter it is not difficult to find the Coulomb correction Q numerically.

The examples considered show that formula (31) is fully effective for calculations. We note that it is similar to the corresponding formula of the nonrelativistic theory [see Eqs. (6)–(8) in Ref. 4]. This is because the Coulomb interaction $\delta V(r)=-Z\alpha/r$ is the temporal component of the four-potential A_μ and appears in the Lagrangian (4′) precisely as in the nonrelativistic case.

7. Barrier width and condition for applicability of the imaginary-time formalism. The tunneling probability is relative to the barrier width b. Setting $t=0$ in Eqs. (1), (15), and (20), we find

$$b=\frac{m}{e\mathscr{E}}d(\epsilon_0,\rho)=\frac{F_{\text{cr}}}{\mathscr{E}}d(\epsilon_0,\rho)\lambdabar_c, \quad (37)$$

where $\lambdabar_c=\hbar/mc$. For a pure electric field $d=1-\epsilon_0$, and for crossed fields

$$d=\frac{3\xi^2}{2\sqrt{1+\xi^2}}=\frac{3}{8}\left(\sqrt{\epsilon_0^2+8}-3\epsilon_0\right),\qquad \rho=1. \quad (38)$$

In the more general case (20), we have

$$d(\epsilon_0,\rho)=\frac{\rho^2\epsilon_0\tau_0\tanh(\tau_0/2)}{1-\rho^2\tau_0\coth\tau_0}, \quad (39)$$

where τ_0 is determined from Eqs. (21) or (23). The barrier width increases with the magnetic field; therefore, in this problem the perturbative formula (31) is applicable for all values of γ (in contrast to multiphoton ionization,[1–3] where b decreases proportionally to γ^{-1} for $\gamma \gg 1$). In addition, in the range $\rho \gg 1$, where the electric field dominates, the dependence on $\rho=\mathscr{E}/\mathscr{H}$ is insignificant:

$$\frac{d(\epsilon_0,\rho=1)}{d(\epsilon_0,\rho=\infty)}=\begin{cases}1+\dfrac{1}{18}(1-\epsilon_0)+\dots, & \epsilon_0\to 1,\\ 3\cdot 2^{-3/2}=1.061, & \epsilon_0=0,\\ 1.125, & \epsilon_0=-1.\end{cases} \quad (40)$$

If $E_b \sim m$, the mean radius of the bound state $\langle r\rangle \sim \lambdabar_c\sqrt{1-\epsilon_0^2}$, and $b/\langle r\rangle \sim F_{\text{cr}}/\mathscr{E} \gg 1$, ensuring applicability of the imaginary-time formalism.

If a Coulomb interaction takes place, the estimate given above for $\langle r\rangle$ ceases to be valid when $\epsilon_0 \to -1$. However, in this case, too, the bound state on the boundary of the lower continuum remains localized, and $\langle r\rangle \sim \lambdabar_c$. Thus, for the ground $1s_{1/2}$ level in the Coulomb field $V(r)=-Z\alpha/r$ we have[16–18]

$$\langle r\rangle=\frac{(1+0.3\zeta^2)(\zeta^2-3/4)}{\zeta^2(\zeta^2-3/4)}\lambdabar_c\approx\frac{1}{3}\lambdabar_c, \quad (41)$$

where $\zeta=Z_{\text{cr}}\alpha$, and Z_{cr} is the critical nuclear charge, at which the ground-state level of the electron spectrum sinks to the boundary of the lower continuum ($Z_{\text{cr}}=169-173$ and $\zeta^2=1.52-1.59$, depending on whether the nucleus is bare or the outer electronic shells are filled).

For nonrelativistic bound states $\epsilon_0=1-\alpha^2\kappa^2/2\to 1$, $d=(\alpha\kappa)^2/2 \ll 1$, and the barrier width equals

$$b(\mathscr{E})=\frac{1}{2}\kappa^2\frac{\mathscr{E}_a}{\mathscr{E}}a_B,\qquad a_B=(m\alpha)^{-1}, \quad (42)$$

where a_B is the Bohr radius. For neutral atoms $\langle r\rangle \sim \kappa^{-2}$ ($\kappa \sim 1/n$, where n is the principal quantum number), and for negative ions $\langle r\rangle \sim \kappa^{-1}$. Ultimately, $b/\langle r\rangle \sim \epsilon^{-1} \gg 1$.

8. Hamiltonian approach. There is a possibility for a somewhat different approach to the tunneling of relativistic particles, which we shall illustrate in the example of crossed fields with $\mathscr{E}=\mathscr{H}$. The integrals of motion [in the gauge (13)] are

$$P_x,\ P_z,\quad \text{and}\quad H=\sqrt{m^2+(P_x+e\mathscr{E}y)^2+P_y^2+P_z^2}-e\mathscr{E}y=E_0, \quad (43)$$

where H is the Hamiltonian, E_0 is the initial energy of the level, $\mathbf{P}$ is the generalized momentum (for the extremum trajectory $P_z=0$), and $P_y^2=E_0^2-m^2-P_x^2-2e\mathscr{E}(E_0-P_x)y$. Thus, the tunneling problem has been reduced to a one-dimensional problem; therefore, to within exponential accuracy

$$w\propto\exp\left\{-2\int_0^{y_0}\sqrt{-P_y^2}\,dy\right\}=\exp\left(-\frac{2m^2}{e\mathscr{E}}J\right), \tag{44}$$

where y_0 is the turning point and

$$J=\frac{(q^2+1-\epsilon_0^2)^{3/2}}{3|q-\epsilon_0|}. \tag{45}$$

The minimization of $J(q,\epsilon_0)$ with respect to q is equivalent to isolating the extremum trajectory from the entire bundle of subbarrier trajectories. There are two minimum points

$$q_\pm=\frac{1}{4}(3\epsilon_0\mp\sqrt{\epsilon_0^2+8}\,),$$

of which q_+ corresponds to ionization of the electronic level:

$$J(q_+,\epsilon_0)=\frac{\sqrt{3}\,\xi^3}{1+\xi^2},\qquad \xi=\sqrt{(q_+-\epsilon_0)^{-2}-1}, \tag{46}$$

in complete agreement with (16) and (17).

To calculate the pre-exponential factor P we write the asymptote of the unperturbed wave function of the s level in the form

$$\psi_0(r)\approx A\sqrt{\frac{\mu}{2\pi}}\,\frac{\exp(-\mu r)}{r}=\frac{A_\kappa\sqrt{\mu}}{2^{3/2}\pi^{5/2}}\int\frac{\exp(i\mathbf{p}\cdot\mathbf{r})}{p^2+\mu^2}\,d^3p,\qquad r\gg R, \tag{47}$$

where $\mu=m\sqrt{1-\epsilon_0^2}$, R is the action radius of the forces (the potential binding the s level is assumed to be short-range), and we ignore the spin of the particle. Near the y axis (the direction of the electric field) we represent ψ_0 in the form

$$\psi_0\approx\frac{A_\kappa}{(2\pi)^{3/2}}\int\frac{d^2p_\perp}{\sqrt{|p_y|}}\exp(-|p_y|y+i\mathbf{p}_\perp\cdot\boldsymbol{\rho}), \tag{48}$$

which is convenient for matching with the quasiclassical solution of the Klein–Gordon equation in an external field, if $\mathbf{p}_\perp$ is the integral of motion [here $y\approx r\gg\rho$, $|p_y|\approx\mu(1+p_\perp^2/2\mu^2$, and $\boldsymbol{\rho}$ and $\mathbf{p}_\perp$ are two-dimensional vectors in the xz plane]. Continuing (48) through the turning point y_0 and calculating the particle fluxes at $y\to\infty$, we obtain

$$w=\frac{A_\kappa^2}{\sqrt{4\pi}}\int\frac{dP_x}{\sqrt{J_1(P_x)}}\exp\left\{-\frac{2(\mu^2+P_\perp^2)^{3/2}}{3e\mathscr{E}(E_0-P_x)}\right\}, \tag{49}$$

where $P_\perp=(P_x,P_z)$,

$$J_1(P_x)=\int_0^{y_0}\frac{dy}{\sqrt{\mu^2+P_x^2-2e\mathscr{E}(E_0-P_x)y}}=\frac{\sqrt{\mu^2+P_x^2}}{e\mathscr{E}(E_0-P_x)} \tag{48$'$}$$

and $E_0-P_x>0$. Expanding here with respect to $P_\perp^2$ and applying the saddle-point method, we arrive at a formula like (6) for w, in which the exponential function should be replaced by (17), and $1/\arccos\epsilon_0$ in the pre-exponential factor should be replaced by

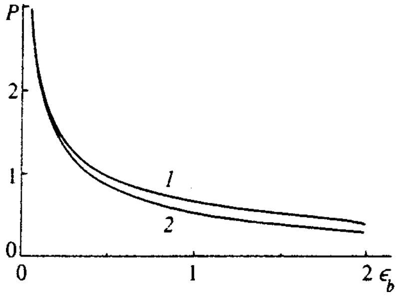


FIG. 4. Dependence of the pre-exponential factor P on the binding energy of the level: *1* — in the case of only an electric field; *2* — for crossed fields.

$$P(\epsilon_0,\rho=1)=(\xi\sqrt{\xi^2+3})^{-1}. \tag{50}$$

The dependences of the pre-exponential factor P on the energy ϵ_0 in these two cases have similar forms (see Fig. 4). In the nonrelativistic limit $P(\epsilon_0,\rho)=1/\alpha\kappa+\ldots$ for any $\rho>0$.

The Hamiltonian approach can be applied to fields of more complex configuration, but this question is beyond the scope of the present paper.

9. Comments on the Unruh effect. It was claimed more than 20 years ago that from the standpoint of an observer moving along a straight line with a constant intrinsic acceleration g (as a consequence of a force of nongravitational origin on him), an ordinary vacuum state in Minkowski space is a mixed state and can be described by a thermal density matrix with the effective Fulling–Unruh temperature[8,29,30]

$$T=\hbar g/2\pi c. \tag{51}$$

This claim has been termed the Unruh effect in the literature (see Refs. 31–33 and the references cited therein). It has become folklore that this effect is due to the fact that in the rest frame of a Rindler (i.e., uniformly accelerating) observer the metric5)

$$ds^2=\rho^2d\sigma^2-d\rho^2-dy^2-dz^2,\qquad -\infty<\sigma<\infty,\qquad 0\leqslant\rho<\infty \tag{52}$$

has a horizon. Therefore, some of the information available to an inertial observer (relative to which the Minkowski vacuum is defined) is not accessible to a Rindler observer. This results in the appearance of a mixed state.

Recently, however, arguments[9] were advanced against the existence of the Unruh effect. Their essential point is as follows. A free quantum (scalar) field φ in Rindler space should vanish not only when $\rho\to\infty$, but also when $\rho\to0$, i.e., it should satisfy the boundary condition $\varphi(\rho,\sigma)|_{\rho=0}=0$ (which corresponds to an impermeable wall at $\rho=0$, i.e., on the boundary of the Rindler manifold). This means that the problems of quantizing φ in Rindler and Minkowski spaces are totally different [see Eqs. (18)–(20) in Ref. 9].

A Boltzmann distribution $p_n\propto\exp(-E_n/T)$ among the energy levels of a uniformly accelerated detector would be an observable manifestation of the Unruh effect. Let us examine the ionization of a heavy atom in a constant uniform electric field $\mathscr{E}$ from this standpoint. In this case the intrinsic acceleration g of the detector (i.e., an atom or ion in the present case) is constant, since the longitudinal component of the electric field remains unchanged after Lorentz transformations:

$$g=\frac{Z-1}{A}\frac{\mathcal{E}}{\mathcal{E}_a}g_0, \quad g_0=\frac{m^2e^6}{\hbar^4 m_p}=4.93\times 10^{21}\,\mathrm{cm\cdot s^{-2}}. \tag{53}$$

Here, as above (Sec. 6), Z denotes the charge of the ''atomic core'' obtained from the atom (ion) when one electron is removed, $A=M/m_p$, and $m\equiv m_e$, m_p, and M are the masses of the electron, the proton, and the atom. For the temperature (51) we obtain

$$T=\frac{|Z-1|}{A}\frac{\mathcal{E}}{\mathcal{E}_a}T_0, \tag{54}$$

where $\mathcal{E}_a=m^2e^5/\hbar^4=5.14\times 10^9$ V/cm and $T_0=m^2e^4/\hbar^2 m_p=1.72\times 10^{-5}$ eV (for example, when $\mathcal{E}=\mathcal{E}_a$ and $A\approx 200$, the temperature $T\approx 10^{-7}\,\mathrm{eV}\approx 10^{-3}$ K).

By virtue of the principle of detailed equilibrium the ionization probability of an atom in a thermal bath at the temperature (54) equals

$$w^{(T)}\propto \exp\left(-\frac{\kappa^2}{2T}\right). \tag{55}$$

On the other hand, according to quantum mechanics the ionization probability of an atomic level is6)

$$w^{(i)}(\mathcal{E},\kappa)\propto \exp\left(-\frac{2\kappa^3}{3\mathcal{E}}\right)=\exp\left(-k\frac{\kappa^3}{T}\right), \tag{56}$$

where $E_b=\kappa^2/2$ is the binding energy of the level and k is an extremely small coefficient

$$k=\frac{\alpha}{3\pi}\frac{|Z-1|m}{Am_p}\simeq 2\times 10^{-9}. \tag{57}$$

A comparison of (55) and (56) reveals that (56) is not a universal Boltzmann distribution, since even the dependence on κ (i.e., on the energy of the level) in these two formulas is functionally different.

TABLE II. Successive stages in the ionization process (for $\mathcal{E}=0.02\mathcal{E}_a\approx 10^8$ V/cm).

	Atom or ion					
	U	U^+	U^{2+}	Fe^-	Fe	Fe^+
I_k, eV	6.194	11.9	20	0.398	7.900	16.19
κ_k	0.674	0.935	1.2	0.171	0.762	1.091
Z_k	1	2	3	0	1	2
$\tau_k^{(i)}$, s	7.4(−16)	5.5(−12)	1.3	2.4(−16)	6.5(−14)	1.8(−4)
θ_k	0	7.6(−5)	3.6(7)	−1.4(−8)	−1.4(−8)	1.1(4)
Δl_k, cm	0	6.2(−6)	∞	5.1(−14)	4.2(2)	∞

Note. The lifetimes $\tau_k\equiv\tau_k^{(i)}$ were calculated using formula (13) from Ref. 6, and the ionization potentials I_k were taken from Ref. 35.

Let us consider a mental experiment. As we know,[19] under the effect of a constant acceleration g a classical particle moves along the trajectory

$$x=x_0+\frac{c^2}{g}\left(\sqrt{1+u^2}-\sqrt{1+u_0^2}\right), \quad t=t_0+\frac{c}{g}(u-u_0),$$

$$v=\frac{cu}{\sqrt{1+u^2}}, \quad s=\frac{c}{g}(\operatorname{arcsinh} u-\operatorname{arcsinh} u_0)$$

(t is the laboratory time, and s is the intrinsic time). As a consequence of ionization the atom (ion) alters its charge ($Z\rightarrow Z+1$) at the times $s=s_0, s_1, s_2, \ldots$, where $s_0=\tau_0$, $s_1-s_0=\tau_1$, $s_2-s_1=\tau_2$, etc., $\tau_k=1/w_k^{(i)}(\mathcal{E},\kappa_k)$, and its intrinsic acceleration g_k varies accordingly. The motion of an atom following N-fold ionization is assigned by the equations

$$x=l_N+\frac{c^2}{g_N}(\cosh\theta-\cosh\theta_N),$$

$$t=t_N+\frac{c}{g_N}(\sinh\theta-\sinh\theta_N), \quad \theta>\theta_N,$$

where $\theta=\ln(u+\sqrt{1+u^2})$ is the velocity, and l_N and t_N are, respectively, the total path and time (for a static observer) until the time of N-fold ionization:7)

$$\begin{gathered}
l_N=\sum_{k=1}^{N}\Delta l_k=\sum_{k=0}^{N-1}\frac{c^2}{g_k}(\cosh\theta_{k+1}-\cosh\theta_k),\\
t_N=\sum_{k=0}^{N-1}\frac{c}{g_k}(\sinh\theta_{k+1}-\sinh\theta_k),\\
v_N=c\tanh\theta_N, \quad s_N=\sum_{k=0}^{N}\tau_k,\\
\theta_k=\theta_0+\frac{1}{c}\sum_{j=0}^{k-1}g_j\tau_j,
\end{gathered} \tag{58}$$

Δl_k is the path traversed by the kth ion and θ_0 corresponds to the initial velocity of the atom. Using the quasiclassical formula for the ionization probability in an electric field $\mathcal{E}$, we obtain (to within accuracy to constants of order unity)

$$\frac{g_k\tau_k}{c}\approx 4\times 10^{-6}\frac{|Z-1|}{A}\left(\frac{\mathcal{E}}{\mathcal{E}_a}\right)^{2\eta_k}\exp\left(\frac{2\kappa_k^3\mathcal{E}_a}{3\mathcal{E}}\right), \tag{59}$$

where $\eta_k=Z_k/\kappa_k$ is the Sommerfeld parameter for the sub-barrier motion of an electron.

Table II lists estimates of the lifetime $\tau_k=1/w_k^{(i)}$ in the accompanying reference frame and the paths Δl_k for two characteristic cases, in which there is a neutral uranium atom or a negative iron ion at $t=0$ (we assume that $\theta_0=0$, i.e., that the atom is initially at rest). We note that Fe^- and Fe move against the field $\mathcal{E}$ (therefore, the first two θ_k have negative signs), while the positive Fe^+ ion is at first slowed by the electric field and is only subsequently accelerated along $\mathcal{E}$. It is seen from Table II that the values of the κ_k increase as the extent of ionization increases, leading [see the exponential function in (59)] to a sharp increase in the lifetime τ_k. The paths Δl_k increase even more rapidly, since, apart from the increase in lifetime, the relativistic slowing of time also affects them:

$$s\sim\frac{c}{g}\ln\frac{gt}{c},\quad \Delta l_k\sim\exp\theta_k \tag{60}$$

for $\theta_k\gg 1$. Therefore, only a small extent of ionization is achievable in an assigned field. For example, for the doubly charged uranium ion U^{2+} and $\mathscr{E}=0.02\mathscr{E}_a$ the path $\Delta l_2\sim\exp(1.8\times10^7)$ cm (!), which, of course, means only that this ion always remains stable (in the field $\mathscr{E}$ cited).

On the other hand, the Unruh effect gives $\tau^{(T)}=1/w^{(T)}$ for the lifetime, which amounts to $\sim\exp(3\times10^9)$ s in $\mathscr{E}=0.02\mathscr{E}_a$ for the neutral uranium atom and equally large values in other cases.[8)] Thus, the ''thermal'' ionization time of the atom (detector) is immeasurably greater than the time for its destruction by the electric field. The Unruh effect[8] refers to detectors of any nature; therefore, the mental experiment is at variance with its counter experiment.

We have hitherto considered the nonrelativistic case, $E_b\ll m$. Without going into the fairly fine question of the quantum-field description of relativistic bound states, we note that if formula (6) is used for $E_b=m(1-\epsilon_0)\sim m$, instead of (56) we obtain

$$w^{(i)}\propto\exp\left\{-\frac{E_b}{T}f(E_b)\right\}, \tag{61}$$

where $f(E_b)=(m/2\pi M)\Phi(\epsilon_0)/(1-\epsilon_0)$. Unruh's claim would correspond to $f(E_b)\equiv1$, which clearly does not hold.

It is noteworthy that Nikishov and Ritus[36] showed that if elemental particles are considered as detectors, the energy spectrum of their radiation, generally speaking, does not correspond to a universal Unruh law. A heavy atom, to which the quasiclassical treatment is applicable, satisfies the physical requirements imposed on a detector to a considerably greater extent. As is seen from the foregoing, the electric field accelerating the atom destroys the detector itself, which was intended to detect the thermal radiation in the accompanying reference frame.

10. Conclusion. A generalization of the imaginary-time formalism to the relativistic case has been developed. A calculation of subbarrier trajectories satisfying the classical equations of motion (which, however, have an imaginary time *t* and are thus impossible in classical physics) enables us to use the well developed machinery of analytical mechanics and to find both the exponential function and the Coulomb and pre-exponential factors in the ionization probability of a level with an arbitrary binding energy ($0<E_b<2mc^2$) under the action of electric and magnetic fields. The equations obtained cover, as limiting cases, both the theory of the ionization of nonrelativistic bound systems (atoms and ions) and the case of $E_b=2mc^2$ (a level on the boundary of the lower continuum, $Z=Z_{\text{cr}}$), where this probability is comparable in value to the probability of the production of electron-positron pairs from a vacuum in an external field. We note that the imaginary-time formalism was previously employed in the problem of the instability of a vacuum and the production of pairs in a strong field in quantum electrodynamics,[24] as well as in the case of non-Abelian gauge theory.[37]

A system of two particles with strongly differing masses (an electron in the field of a heavy nucleus can serve as an example) was considered above. The possibility of applying the imaginary-time formalism to relativistic systems consisting of particles with commensurate masses ($q\bar{q}$, qqq, etc.) remains open.

We thank the participants in the theoretical seminars of the Institute of Theoretical and Experimental Physics and the Moscow Engineering-Physics Institute for some interesting discussions, S. G. Pozdnakov for his assistance in the numerical calculations, and M. N. Markina for her assistance in composing this article The work was performed with partial support from the Russian Fund for Fundamental Research (Grants Nos. 95-02-05417 and No. 98-02-17 007).

APPENDIX A:

The function g_1 appearing in (26) equals

$$g_1(\gamma)=\frac{3\tau_0}{8\gamma^5}\,\frac{s^4-3s^3-(s^3-6s^2+6s)\gamma^2-(s-2)\gamma^4}{s^2-(s-2)\gamma^2}, \tag{A1}$$

where $s=1-(\tau_0/\sinh\tau_0)^2$, and $\tau_0=\tau_0(\gamma)$ is determined from Eq. (28). Taking into account the expansions

$$\begin{aligned}s&=\frac{1}{2}\tau_0^2-\frac{1}{15}\tau_0^4+\dots\\&=\frac{1}{3}\gamma^2\left(1-\frac{4}{45}\gamma^2-\frac{8}{2835}\gamma^4+\dots\right),\quad\gamma\to0,\end{aligned} \tag{A2}$$

$$\begin{aligned}s&=1-4\tau_0^2e^{-2\tau_0}+\dots=1-\exp[-(\gamma^2+1)]\gamma^4\\&\quad\times(1+2\gamma^{-2}+\gamma^{-4}+\dots),\quad\gamma\to\infty,\end{aligned} \tag{A3}$$

we arrive at the following asymptotes:

$$\begin{aligned}g(\gamma)&=1+\frac{1}{30}\gamma^2+\frac{11}{7560}\gamma^4+\dots,\\g_1(\gamma)&=\frac{3}{40}\left(1-\frac{11}{378}\gamma^2+\dots\right),\quad\gamma\to0,\end{aligned} \tag{A4}$$

$$\begin{aligned}g(\gamma)&=\frac{3}{8}\gamma(1+2\gamma^{-2}+\gamma^{-4}+\dots),\\g_1(\gamma)&=\frac{3}{16\gamma}(1-\gamma^{-2}-2\gamma^{-4}+\dots).\end{aligned} \tag{A5}$$

In both cases $g_1(\gamma)\ll g(\gamma)$. This inequality is also confirmed by a numerical calculation (see Fig. 3) and holds for all γ.

APPENDIX B:

Let us discuss the question of the applicability region of the approximation (29). Defining the parameter κ on the basis of the relation[9)] $\epsilon_0=E_0/m=1-\alpha^2\kappa^2/2$, we have

$$\kappa=\alpha^{-1}\sqrt{2(1-\epsilon_0)},\quad\alpha=e^2/\hbar c=1/137. \tag{B1}$$

The exponential factor in (29) can be rewritten in a form similar to (24):

$$w(\mathscr{E},\mathscr{H})\propto\exp\left\{-\frac{F_{\text{cr}}}{\mathscr{E}}\tilde{\Phi}(\epsilon_0,\rho)\right\}, \tag{B2}$$

where

$$\tilde{\Phi}=\frac{\sqrt{32}}{3}\epsilon_b^{3/2}\left[1-\frac{3}{20}\left(1-\frac{4}{9\rho^2}\right)\epsilon_b\right], \qquad \epsilon_b=1-\epsilon_0. \tag{B3}$$

This simple approximation has a high accuracy even in the case of fairly deep levels; see Table I, in which δ is the relative error of (B3):

$$\delta\equiv\delta(\epsilon_0,\rho)=(\Phi-\tilde{\Phi})/\Phi. \tag{B4}$$

A comparison of formulas (29) and (30) shows that the power of $\mathcal{E}$ varies as $\rho\to 0$; therefore, the approximation (29) is inapplicable when $\mathcal{H}\gg\mathcal{E}$. If $\rho=\mathcal{E}/\mathcal{H}\gtrsim 0.5$, the applicability region of formula (29), which was obtained in the nonrelativistic ($\kappa\lesssim 1$) limit, ''extends'' up to energies $E_0\approx -0.5mc^2$, i.e., up to $\kappa\sim 200$ (see Table I). Of course, formula (29) ceases to be applicable when the level approaches the boundary of the lower continuum; in this case the error already amounts to tens of percent. For example, $\delta=0.123$ and -0.188 for $\rho=1$ and ∞, respectively (when $\epsilon_0=-1$).

APPENDIX C:

Here we consider the calculation of the Coulomb correction (31).

a) In writing the extremum ($p_\perp=0$) trajectory in the case of an electric field in the form

$$z=\frac{m}{e\mathcal{E}}(\cos\varphi-\cos\varphi_0), \qquad 0<\varphi<\varphi_0=\arccos\epsilon_0, \tag{C1}$$

we use the value of the integral

$$J(\varphi,\varphi_0)=\int_0^{\varphi}\frac{\cos\varphi\,d\varphi}{\cos\varphi-\cos\varphi_0}=\varphi+\cot\varphi_0\ln\left[\frac{\sin[(\varphi_0+\varphi)/2]}{\sin[(\varphi_0-\varphi)/2]}\right]. \tag{C2}$$

When $\varphi\to\varphi_0$,

$$J(\varphi,\varphi_0)=-a\ln(\varphi_0-\varphi)+a_0+a_1(\varphi_0-\varphi)+\ \ldots, \tag{C3}$$

where $a=\cot\varphi_0$, $a_0=\varphi_0+\cot\varphi_0\ln(2\sin\varphi_0)$, ... Also taking into account that $\eta=Z\alpha\cot\varphi_0$,

$$z(t)=i\sqrt{1-\epsilon_0^2}\,\epsilon_0^{-1}(t-t_0)+\ \ldots$$

$$=\frac{m}{e\mathcal{E}}(\varphi_0-\varphi)\sin\varphi_0+\ \ldots, \qquad t\to t_0,$$

$$\mu z_1=(F_{\rm cr}/\mathcal{E})(\varphi_0-\varphi_1)\sin^2\varphi_0+\ \ldots,$$

we find

$$\eta\ln(\mu z_1)+iZ\alpha\int_{t_1}^{0}\frac{dt}{z(t)}=\eta\left\{\varphi_0\tan\varphi_0+\ln\left[\frac{F_{\rm cr}}{\mathcal{E}}\,\frac{(\varphi_0-\varphi_1)\sin^3\varphi_0}{\sin[(\varphi_0-\varphi_1)/2]}\right\}\right.. \tag{C4}$$

Assuming here that $\varphi_1\to\varphi_0$ [in this case the matching point $z_1=z(t_1)$ drops out of the result], we arrive at formula (32).

In the relativistic limit (34) follows from (6) and (32). If $\epsilon_0=0$, then $\eta=0$ and $Q=\exp(\pi Z\alpha)=21.1$ for $Z=137$ [the values of Z and Q for $E(1s_{1/2})=0$ increase somewhat further when the finite dimensions of the nucleus are taken into account[17,18]].

b) For crossed fields the Coulomb integral (31) can also be expressed in elementary functions:

$$\int_0^{\varphi}\frac{d\varphi}{a^2-\sin^2\varphi}=\frac{1}{\sin 2\varphi_0}\ln\frac{\sin(\varphi_0+\varphi)}{\sin(\varphi_0-\varphi)}, \tag{C5}$$

where $u=3\sin\varphi$, $a=u_0/3=\xi/\sqrt{3}$, and $\varphi_0=\arcsin a$.

Taking into account that at the beginning of subbarrier motion

$$r_1=\frac{\sqrt{3}m}{e\mathcal{E}}\frac{\xi(3-\xi^2)}{\sqrt{1+\xi^2}}(\varphi_0-\varphi_1)+\ \ldots, \qquad \varphi_1\to\varphi_0, \tag{C6}$$

and the value of the integral (C5), we arrive at (35).

c) In the case of mutually perpendicular fields, it follows from (36) that

$$r_0=(\tau_0-\tau)\frac{ma\rho\tau_0}{e\mathcal{H}(1-\rho^2)^{3/2}}\sqrt{1-\rho^2\left(\frac{1}{\tau_0}-\coth\tau_0\right)^2}+\ \ldots, \qquad \tau\to\tau_0$$

(we assume that $\rho<1$). Using J_s to denote the singular part of the integral in (31), we have

$$J_s=Z\alpha\int_{\tau_1}\frac{A}{\tau-\tau_0}d\tau,$$

$$A=\frac{1-\rho^2\tau_0\coth\tau_0}{\rho\tau_0\sqrt{1-\rho^2-(\tau_0^{-1}-\coth\tau_0)^2}}. \tag{C7}$$

Here τ_0 is defined by Eq. (25), from which we find

$$\frac{1}{\sqrt{\epsilon_0^{-2}-1}}=\frac{1-\rho^2\tau_0\coth\tau_0}{\rho[2\tau_0\coth\tau_0-1-(\tau_0/\sinh\tau_0)^2-\rho^2\tau_0^2]^{1/2}}\equiv A. \tag{C8}$$

Hence $\eta=Z\alpha A$ and $J_s=-\eta\ln(\tau_1-\tau_0)+O(1)$. The term $\ln[\mu r_1/(\tau_1-\tau_0)]$, which has a finite limit for $\tau_1\to\tau_0$, ultimately appears in (31). After some manipulations we obtain the regularized expression for the Coulomb correction:

$$Q=\exp\left\{2\eta\left[\ln\left(\frac{m^2}{e\mathcal{E}}\frac{1-\epsilon_0^2}{\epsilon_0}\right)+\int_0^{\tau_0}\left(\varphi(\tau)-\frac{1}{\tau_0-\tau}\right)\right]d\tau\right\}, \tag{C9}$$

$$\varphi(\tau)=\frac{1}{\rho\tau_0}\left(1-\rho^2\tau_0\frac{\cosh\tau}{\sinh\tau_0}\right)\left[(1-\rho^2)\left(\frac{\cosh\tau_0-\cosh\tau}{\sinh\tau_0}\right)^2-\left(\frac{\sinh\tau}{\sinh\tau_0}-\frac{\tau}{\tau_0}\right)^2\right]^{-1/2}, \tag{C10}$$

in which the pole singularities are mutually canceled, and it is not difficult to find the integral numerically. This example shows that exclusion of the matching point r_1 in (31) is sometimes associated with fairly cumbersome transformations, but it can always be accomplished, if the condition $\langle r\rangle \ll b$ holds.

*) E-mail: karnak@theor.mephi.msk.su

1) In performing the integration in (4′) it is convenient to go over to the variables ϑ and ψ with allowance for the relation $dt/m\sqrt{1-v^2} = d\vartheta/e\mathcal{H} = d\psi/e\mathcal{E}$.

2) This condition is necessary for applicability of the quasiclassical approximation and is known to hold in experiment (for electrons $F_{\rm cr}=1.32\times 10^{16}$ V/cm, or 4.41×10^{13} G).

3) We note that, unlike the classical trajectories, here $|t_0|<|s_0|$; this corresponds to imaginary velocities in the subbarrier motion.

4) The presence of this integral is characteristic of the case of mutually perpendicular fields. When $\rho=1$, it coincides with the integral indicated in Ref. 19 $\alpha=E_{\rm kin}-cp_x$.

5) This is the so-called Rindler metric,[34] for which $x=\rho\cosh\sigma$ and $t=\rho\sinh\sigma$. We note that the transition from the global coordinates x and t to the Rindler coordinates ρ and σ is a transformation that is singular for $\rho=0$ [which corresponds to the apex of the light cone, $x=t=0$, in a flat $(1+1)$-dimensional space].

6) Here we have gone over to atomic units, and we have ignored the preexponential factors in (55) and (56), since these formulas differ from one another only in the exponential function.

7) For a neutral atom in the kth segment of a trajectory the corresponding terms in (58) must be replaced by $\Delta l_k = c\tau_k\sinh\theta_k$ and $\Delta t_k=\tau_k\cosh\theta_k$.

8) The enormous difference between $\tau^{(i)}$ and $\tau^{(T)}$ is attributed to the small value of the coefficient k appearing in the exponential function (56).

9) For any energy E_0 of the level, in the nonrelativistic case $E_0=mc^2-\kappa^2me^4/2\hbar^2$ and $\kappa\sim 1$ (see Table I in Ref. 6).

---

[1] A. M. Perelomov, V. S. Popov, and M. V. Teren t'ev, Zh. Éksp. Teor. Fiz. **51**, 309 (1966) [Sov. Phys. JETP **24**, 207 (1967)].

[2] V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1968)].

[3] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay in Nonrelativistic Quantum Mechanics*, transl. of 1st Russ. ed., Israel Program for Scientific Translations, Jerusalem (1969) [Russ. original, 2nd ed., Nauka, Moscow (1971)].

[4] V. S. Popov and A. V. Sergeev, JETP Lett. **63**, 417 (1996).

[5] V. S. Popov, B. M. Karnakov, and V. D. Mur, Phys. Lett. A **229**, 306 (1997).

[6] V. S. Popov, B. M. Karnakov, and V. D. Mur, Zh. Éksp. Teor. Fiz. **113**, 1579 (1998) [JETP **86**, 860 (1998)].

[7] B. M. Karnakov, V. D. Mur, and V. S. Popov, JETP Lett. **65**, 405 (1997).

[8] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).

[9] V. A. Belinskiĭ, B. M. Karnakov, V. D. Mur, N. B. Narozhnyĭ, JETP Lett. **65**, 902 (1997); JETP Lett. **67**, 96 (1998).

[10] V. S. Popov, V. D. Mur, and B. M. Karnakov, JETP Lett. **66**, 229 (1997); Preprint No. 44, Institute of Theoretical and Experimental Physics, Moscow (1997).

[11] J. Schwinger, Phys. Rev. **82**, 664 (1951).

[12] C. Itzykson and J. B. Zuber, *Quantum Field Theory*, Vol. 1, McGraw-Hill, New York (1980).

[13] Yu. N. Demkov and G. F. Drukarev, Zh. Éksp. Teor. Fiz. **47**, 918 (1964) [Sov. Phys. JETP **20**, 614 (1965)].

[14] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977) [Russ. original, Nauka, Moscow (1974)].

[15] W. Pieper and W. Greiner, Z. Phys. **218**, 327 (1969).

[16] V. S. Popov, JETP Lett. **11**, 162 (1970); Yad. Fiz. **12**, 429 (1970) [Sov. J. Nucl. Phys. **12**, 235 (1971)].

[17] Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys. Usp. **14**, 673 (1972)].

[18] W. Greiner, B. Müller, and J. Rafelski, *Quantum Electrodynamics of Strong Fields*, Springer, Berlin (1985).

[19] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon, New York (1975) [Russ. original, Nauka, Moscow (1973)].

[20] S. P. Andreev and V. A. Polunin, JETP Lett. **42**, 190 (1985).

[21] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).

[22] M. S. Marinov and V. S. Popov, Yad. Fiz. **15**, 1271 (1972) [Sov. J. Nucl. Phys. **15**, 702 (1972)]; M. S. Marinov and V. S. Popov, Fortschr. Phys. **25**, 373 (1977).

[23] V. S. Vanyashin and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **48**, 565 (1965) [Sov. Phys. JETP **21**, 375 (1965)].

[24] V. S. Popov, JETP Lett. **13**, 185 (1971); Zh. Éksp. Teor. Fiz. **61**, 1334 (1971) [Sov. Phys. JETP **34**, 709 (1972)].

[25] G. F. Drukarev and B. S. Monozon, Zh. Éksp. Teor. Fiz. **61**, 956 (1971) [Sov. Phys. JETP **34**, 509 (1972)].

[26] W. R. Smythe, *Static and Dynamic Electricity*, McGraw-Hill, New York (1950).

[27] V. V. Batygin and I. N. Toptygin, *Problems in Electrodynamics*, Academic Press, London (1964) [Russ. original, Fizmatgiz, Moscow (1962)].

[28] A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **52**, 514 (1967) [Sov. Phys. JETP **25**, 336 (1967)].

[29] S. A. Fulling, Phys. Rev. D **7**, 2850 (1973).

[30] P. C. W. Davies, J. Phys. A **8**, 609 (1975).

[31] N. D. Birrell and P. C. Davies, *Quantum Fields in Curved Space*, Cambridge University Press (1982).

[32] V. L. Ginzburg and V. P. Frolov, Usp. Fiz. Nauk **153**, 633 (1987) [Sov. Phys. Usp. **30**, 1073 (1987)].

[33] Ya. B. Zel'dovich, L. V. Rozhanskiĭ, and A. A. Starobinskiĭ, JETP Lett. **43**, 523 (1986).

[34] W. Rindler, Am. J. Phys. **34**, 1174 (1966).

[35] A. A. Radtsig and B. M. Smirnov, *Parameters of Atoms and Atomic Ions* [in Russian], Énergoatomizdat, Moscow (1986).

[36] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **94**(7), 31 (1988) [Sov. Phys. JETP **67**, 1313 (1988)].

[37] Sh. S. Agaev, A. S. Vshivtsev, and V. Ch. Zhukovskiĭ, Yad. Fiz. **36**, 1023 (1982) [Sov. J. Nucl. Phys. **36**, 599 (1982)].

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Stabilization of a Rydberg atom and competition between the Λ and *V* transition channels

N. P. Poluéktov and M. V. Fedorov*)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

(Submitted 10 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 821–836 (September 1998)

The photoionization and stabilization of a Rydberg atom in a strong laser field are investigated theoretically. The role of Raman-type transitions between neighboring Rydberg levels via the continuum (Λ transitions) and via lower-energy resonant Rydberg levels (*V* transitions) is analyzed. The conditions under which this phenomenon can be observed experimentally are determined. The characteristics of stabilization due to *V*-type transitions are described. © *1998 American Institute of Physics.* [S1063-7761(98)00509-5]

1. INTRODUCTION

One of the most interesting and widely discussed phenomena involving the interaction of atoms with a strong laser field is stabilization of atoms, i.e., an increase in their stability against photoionization as the laser field strength increases. In Refs. 1 and 2 (and in a large number of subsequent works) two considerably different stabilization mechanisms were proposed and described: adiabatic or high-frequency stabilization[1] and interference stabilization of Rydberg atoms.[2] According to the theory of interference stabilization, when a Rydberg atom interacts with the field of a light wave, efficient coherent redistribution of the populations of the atomic levels E_n close to the initially populated level E_{n_0} occurs as a result of Λ-type Raman transitions (via the continuum). The resulting coherent superposition of Rydberg states is found to be stable against photoionization, as a result of which transitions from various Rydberg states into the continuum interfere and quench one another, impeding ionization of the atom. The experimental observation of interference stabilization of Rydberg atoms and coherent redistribution of the populations of Rydberg levels is described in Refs. 3 and 4, respectively.

In principle, besides Λ-type transitions, coherent repopulation of Rydberg levels can also result from *V*-type Raman transitions through lower-lying resonant atomic levels, if the latter exist (see Fig. 1). In Ref. 4 it was concluded that *V* transitions do not play a role in the redistribution of the populations of Rydberg levels, which is unlikely to be true in general.

It should be noted that in Refs. 5–7, redistribution of the populations of the Rydberg levels was investigated theoretically, taking account of both Λ- and *V*-type transition channels. However, the importance of the *V* channel and the conditions under which this channel is dominant were not found and were not fully investigated, specifically because of the lack of a clearly understood relationship between the Λ and *V* transition matrix elements.

In the present paper, interference stabilization of Rydberg atoms is investigated theoretically, taking account of both Λ and *V* transition channels, using the well-known expressions for matrix elements calculated in the semiclassical (WKB) approximation.[8–10] Section 2 is devoted to the mathematical formulation of the problem and the approximations used in the solution process. Section 3 presents a very simple three-level model that can be solved analytically, and under certain conditions makes it possible to investigate qualitatively the relation between and the role of both Λ- and *V*-type transitions in interference stabilization. The three-level model becomes incorrect for large detunings. The simplest generalization of the three-level model to the nonresonant case is the four-level model studied in Sec. 4. Section 5 presents the results of a numerical analysis of the problem that incorporates up to 22 levels, which makes it possible to validate conclusions based on the analytic solutions, and to make quantitative estimates of the experimentally measurable parameters. The concluding section briefly summarizes the status of the problem and reviews the conditions under which stabilization of a Rydberg atom in a laser field by *V*-type Raman transitions might be observed experimentally.

2. STATEMENT OF THE PROBLEM AND GENERAL EQUATIONS

We consider the interaction of an atom with the classical field of a light wave whose electric field strength in the dipole approximation is

$$\varepsilon(t)=\varepsilon_0(t)\cos(\omega t), \tag{1}$$

where ω is the frequency and $\varepsilon_0(t)$ is the time-dependent field strength (envelope) of a pulse; $\varepsilon_0(t)\to 0$ as $t\to\pm\infty$. Prior to the arrival of the laser pulse ($t\to-\infty$), let the atom be in an excited (Rydberg) *s* state with energy $E_{n_0}=-1/2n_0^2$, where $n_0\gg 1$ (we employ units $\hbar=c=1$ throughout). Let ω be greater than the electron binding energy in the state φ_{n_0}, $\omega>|E_{n_0}|$, i.e., a one-photon transition is possible from this state to the continuum. Raman transitions of the Λ type are transitions via the continuum (for example, $\varphi_{n_0}\to\text{continuum}\to\varphi_n$), which are accompanied by virtual absorption and emission of a photon ω and excitation

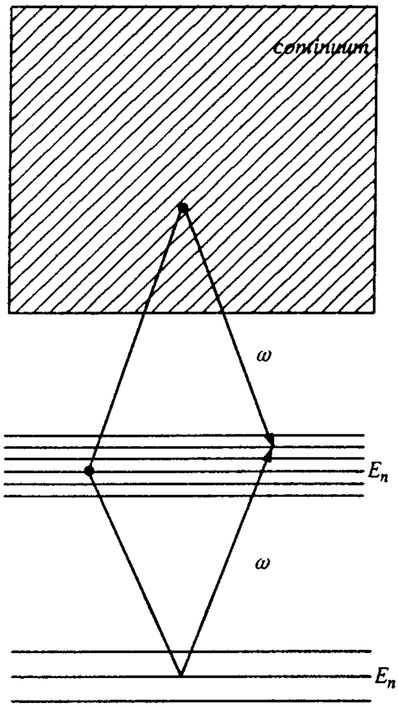


FIG. 1. Overall transition scheme taking account of both Λ and V population–redistribution channels.

of Rydberg levels E_n close to the initially populated level E_{n_0}. Since in general $E_n \neq E_{n_0}$, such transitions can be effective only in a sufficiently strong field.

Let the structure of the atomic spectrum be such that together with Λ-type transitions (via the continuum), efficient resonant (or almost resonant) transitions between states φ_n and Rydberg p states $\varphi_{n'}$ with smaller values of the principal quantum number ($n' < n_0$) and energy ($E_{n'} < E_{n_0}$) are also possible. We call Raman transitions between these groups of states (for example, $\varphi_{n_0} \to \varphi_{n'} \to \varphi_n$) V-type transitions (see Fig. 1).

The wave function $\Psi(t)$ of an atom in the field $\varepsilon(t)$ (1) can be expanded in a basis of free-atom wave functions. The projection of $\Psi(t)$ on the bound states of the atom, $\Psi_{\text{bound}}(t)$, can be represented as a superposition of the functions φ_n and $\varphi_{n'}$,

$$\Psi_{\text{bound}}(t) = \sum_n a_n(t)\varphi_n + \sum_{n'} a_{n'}(t)\varphi_{n'}, \tag{2}$$

where $a_n(t)$ and $a_{n'}(t)$ are the probability amplitudes for finding the atom in the levels E_n and $E_{n'}$, respectively. Since the atom undergoes ionization, the norm of $\Psi_{\text{bound}}(t)$ is not conserved, and determines the probability of ionization by the pulse:

$$w_i = 1 - \langle \Psi_{\text{bound}}(t) | \Psi_{\text{bound}}(t) \rangle |_{t\to\infty}. \tag{3}$$

By stabilization of the atom we mean a situation such that when the field exceeds a certain threshold, the ionization probability w_i becomes a decreasing function of the peak field strength of the pulse, $\varepsilon_{0\,\max}$, or it becomes equal to a constant less than 1.

The Schrödinger equation for the complete function $\Psi(t)$ of the atom in a field can be reduced to an equation for $\Psi_{\text{bound}}(t)$ or to an equivalent system of equations for the probability amplitudes $a_n(t)$ and $a_{n'}(t)$ by a procedure known as adiabatic elimination of the continuum.[11,12] In this approach the Λ transitions between nearby Rydberg levels (for example, E_n and E_m) are described by the tensor $\Gamma_{n,m}$ of ionization widths

$$\Gamma_{n,m} = 2\pi V_{n,E} V_{E,m} |_{E=E_m+\omega}, \tag{4}$$

where E is the energy of an electron in an intermediate p state of the continuum, $V_{a,b} = \langle \varphi_a | -\mathbf{d}\cdot\boldsymbol{\varepsilon}_0/2 | \varphi_b \rangle$ are the transition matrix elements, and $\mathbf{d}$ is the dipole moment of the atom.

On this basis, and taking account of both Λ and V transitions, we write the equations for the probability amplitudes $a_n(t)$ and $a_{n'}(t)$ in the rotating-wave approximation (resonance approximation[11,12]) in the form

$$i\dot{a}_{n'}(t) = (E_{n'} + \omega) a_{n'}(t) + \sum_n V_{n',n} a_n(t), \tag{5}$$

$$i\dot{a}_n(t) = E_n a_n(t) + \sum_{n'} V_{n,n'} a_{n'}(t) - i \sum_m \frac{\Gamma_{n,m}}{2} a_m(t).$$

As noted above, the transition matrix elements $V_{a,b}$ have very simple and convenient analytic expressions obtained in the semiclassical (WKB) approximation[8–10]:

$$V_{n,n'} \sim \frac{\varepsilon_0}{(nn')^{3/2}\omega^{5/3}}, \quad V_{n,E} \sim \frac{\varepsilon_0}{n^{3/2}\omega^{5/3}}. \tag{6}$$

For large n and n', the dependence of the matrix elements $V_{n,n'}$ and $V_{n,E}$ on n and n' becomes quite slow and can be approximately neglected, setting $n \approx n_0$ and $n' \approx n_0'$, where n_0' is the principal quantum number of the level $E_{n_0'}$ (from the series $E_{n'}$) closest to resonance with the level E_{n_0}, i.e., the level corresponding to minimum detuning of the resonance

$$\delta = E_{n_0'} + \omega - E_{n_0}. \tag{7}$$

In the approximation $n \approx n_0$ and $n' \approx n_0'$ we find from Eqs. (4) and (6)

$$\Gamma_{n,m} \approx \Gamma = \text{const} \sim \frac{\varepsilon_0^2}{n_0^3 \omega^{10/3}},$$

$$V_{n,n'} \approx \Omega_R = \text{const} \sim \frac{\varepsilon_0}{(n_0 n_0')^{3/2}\omega^{5/3}}, \tag{8}$$

where Ω_R is the analog of the Rabi frequency in a two-level system.[13]

The ionization width Γ and the Rabi frequency Ω_R (8) are the basic parameters that characterize the system under study. Additional but also important parameters of the system are the detuning δ (7) of resonance between the levels E_{n_0} and $E_{n_0'}$, and the spacings Δ and Δ' between neighboring Rydberg levels near the energies $\sim E_n$ and $E_{n'}$, respectively,

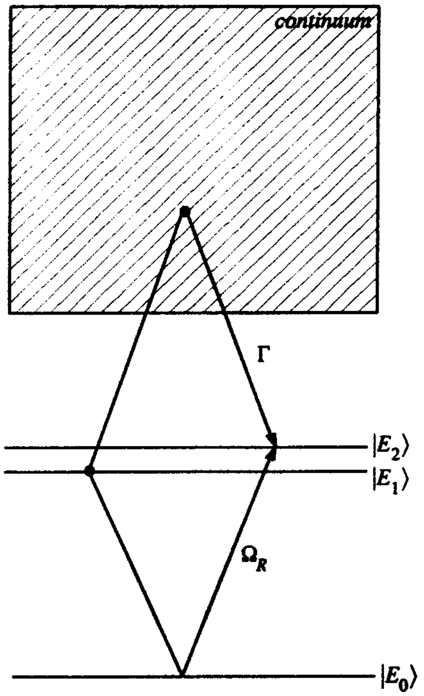


FIG. 2. Transition scheme for a three-level model.

$$\Delta = E_{n_0+1} - E_{n_0} \approx \frac{1}{n_0^3}, \quad \Delta' = E_{n'_0+1} - E_{n'_0} \approx \frac{1}{n'^3_0}. \tag{9}$$

3. THREE-LEVEL MODEL AND THE ROLE OF THE MAIN PARAMETERS CHARACTERIZING THE INTERACTION OF AN ATOM WITH A FIELD

The simplest physical model for the system under study is a model in which only two Rydberg levels E_1 and E_2 in the series $\{E_n\}$ (for example, E_{n_0} and E_{n_0+1}) and one level $E_0(E_{n'_0})$ from a lower-lying series $\{E_{n'}\}$ are taken into account (Fig. 2). The main advantage of this model is its simplicity, which makes it possible to obtain analytic solutions that most clearly show which of the parameters introduced at the end of the preceding section govern the behavior of the system in various field ranges.

In the three-level model the system (5) consists of three equations:

$$\begin{aligned} i\dot{a}_0(t) &= (E_0+\omega)a_0(t) + \Omega_R[a_1(t)+a_2(t)], \\ i\dot{a}_1(t) &= \Omega_R a_0(t) + E_1 a_1(t) - i\frac{\Gamma}{2}[a_1(t)+a_2(t)], \\ i\dot{a}_2(t) &= \Omega_R a_0(t) + E_2 a_2(t) - i\frac{\Gamma}{2}[a_1(t)+a_2(t)]. \end{aligned} \tag{10}$$

In a model where the interaction is turned on and off instantaneously (i.e., a model with rectangular pulses), the pulse envelope is constant while the pulse is on, just like Ω_R and Γ. In this case the system (10) is a set of linear differential equations with constant coefficients, which has solutions of the form

$$a_k(t) = b_k \exp(-i\gamma t), \tag{11}$$

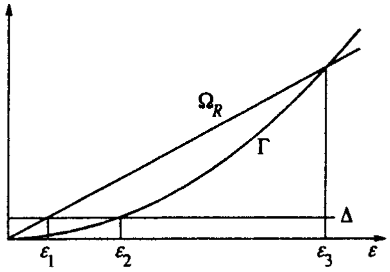


FIG. 3. Relations between the parameters of the problem in various ranges of the field strength. The characteristic fields are shown.

where b_k are constants satisfying the algebraic equations

$$\begin{aligned} &(E_0+\omega)b_0 + \Omega_R(b_1+b_2) = \gamma b_0, \\ &\Omega_R b_0 + \left(E_1 - i\frac{\Gamma}{2}\right)b_1 - i\frac{\Gamma}{2}b_2 = \gamma b_1, \\ &\Omega_R b_0 - i\frac{\Gamma}{2}b_1 + \left(E_2 - i\frac{\Gamma}{2}\right)b_2 = \gamma b_2. \end{aligned} \tag{12}$$

The constant γ is the complex quasienergy of the system,[14,15] and its values are given by the eigenvalues of the matrix of coefficients on the left-hand side of Eqs. (12). Since Eqs. (12) are a set of homogeneous equations, the existence criterion for a solution is that the determinant of the system vanish. This is the characteristic equation of the system, which in the present case can be written in the form

$$\Omega_R^2(2x-\Delta) = (x-\delta)\left[x(x-\Delta) + i\frac{\Gamma}{2}(2x-\Delta)\right], \tag{13}$$

where x is the quasienergy relative to E_1 $x=\gamma-E_1$, $\Delta = E_2 - E_1$, and $\delta = E_0 + \omega - E_1$.

Figure 3 shows Ω_R and Γ as functions of the amplitude ε_0. The range of fields in which a deviation from perturbation theory occurs is the region where Ω_R or Γ exceeds the splitting Δ between the levels E_1 and E_2 (horizontal line in Fig. 3). Let ε_1 and ε_2 be solutions of the equations $\Omega_R(\varepsilon) = \Delta$ and $\Gamma(\varepsilon) = \Delta$, respectively. The third characteristic value of the field strength, ε_3, noted in Fig. 3, is given by the solution of the equation $\Gamma(\varepsilon) = \Omega_R(\varepsilon)$, i.e., this is the field at which the Rabi frequency and the ionization width become equal. Using Eqs. (6)–(9), we find the explicit form of the characteristic fields ε_1, ε_2, and ε_3 in the semiclassical approximation to be

$$\varepsilon_1 = \left(\frac{n'_0}{n_0}\right)^{3/2} \omega^{5/3}, \quad \varepsilon_2 = \omega^{5/3}, \quad \varepsilon_3 = \left(\frac{n_0}{n'_0}\right)^{3/2} \omega^{5/3}. \tag{14}$$

These values correspond to the following values of $V = \varepsilon/\omega^{5/3}$:

$$V_1 = \left(\frac{n'_0}{n_0}\right)^{3/2}, \quad V_2 = 1, \quad V_3 = \left(\frac{n_0}{n'_0}\right)^{3/2}. \tag{15}$$

These values are arranged in increasing order, and each differs from the preceding one by a factor $(n_0/n'_0)^{3/2}$. This factor can become fairly large, and therefore the points ε_1, ε_2, and ε_3 can correspond to considerably different field ranges. For example, under the experimental conditions of

Ref. 4, the quantum numbers n_0 and n_0' were 26 and 5, which yields $(n_0/n_0')^{3/2} \sim 12$. The corresponding characteristic fields are

$$\varepsilon_1 = 7\times10^5, \quad \varepsilon_2 = 8\times10^6, \quad \varepsilon_3 = 9\times10^7 \ \text{V/cm}. \tag{16}$$

These fields clearly differ from one another each by an order of magnitude. Here we focus on the situation corresponding to a large value of $(n_0/n_0')^{3/2}$. In so doing, we also analyze the development of the ionization dynamics when the ratio n_0/n_0' becomes of order 1.

As follows from the foregoing estimates and from Fig. 3, if $(n_0/n_0')^{3/2}$ is large, then there exists a very wide range of fields ε_0, $\varepsilon_1 < \varepsilon_0 < \varepsilon_3$, to which perturbation theory is inapplicable $(\Omega_R > \Delta)$, and the Rabi frequency Ω_R is then much greater than the ionization width Γ. Under these conditions, V-type resonant transitions can play a decisive role in stabilizing an atom. Taking account of resonant coupling with the lower-lying level E_0 has the effect that the ''threshold of nonlinearity,'' i.e., the field at which deviations from perturbation theory (ε_1) first appear, is much lower than when there is no such coupling (in which case the field ε_2, i.e., the root of the equation $\Gamma(\varepsilon) = \Delta$, is the threshold field for stabilization).

We now analyze Eq. (13) to determine the complex quasienergies of the system. We consider first the case in which the detuning δ is half the spacing Δ between the Rydberg levels, $\delta = \Delta/2$. It is easy to see that Eq. (13) can then be solved analytically, and the roots of Eq. (13) take the form

$$x_1 = \frac{\Delta}{2}, \quad x_{3,4} = \frac{\Delta}{2} - i\frac{\Gamma}{2} \pm \beta,$$

$$\beta = \sqrt{2\Omega_R^2 + \left(\frac{\Delta}{2}\right)^2 - \left(\frac{\Gamma}{2}\right)^2}. \tag{17}$$

The first of these equations shows that the level x_1 is stable $(\mathrm{Im}(x_1(\varepsilon_0)) \equiv 0)$, and its position does not depend on the field $(\mathrm{Re}(x_1(\varepsilon_0)) = \Delta/2 = \text{const})$. As for the roots $x_{2,3}(\varepsilon_0)$, the functions $|\mathrm{Im}(x_{2,3}(\varepsilon_0))|$ increase monotonically right up to field strength $\varepsilon_0 \approx 2\varepsilon_3$, for which $\Gamma = 2\Omega_R$ and $\beta = 0$. The monotonic growth of the functions $|\mathrm{Im}(x_{2,3}(\varepsilon_0))|$ means that the quasienergy levels $x_{2,3}$ do not narrow anywhere in the range of fields $0 < \varepsilon_0 < 2\varepsilon_3$. This behavior of the quasienergy levels is in striking contrast to the behavior of the quasienergy levels in the absence of V transitions, i.e., for $\Omega_R = 0$: when only Λ transitions are taken into account, the quasienergy levels start to narrow at field strength $\varepsilon_0 \sim \varepsilon_2$, which is the reason for Λ-type interference stabilization. Therefore, the scheme considered here (Fig. 2) differs qualitatively from the one that takes account of only Λ transitions: in such a scheme, for $\delta = \Delta/2$, the level widths do not decrease (right up to $\varepsilon_0 \approx 2\varepsilon_3$) and there is no field-induced stabilization of the atom, but nevertheless one quasienergy level (x_1) arises that is stable at any field strength between 0 and ε_3.

The ''absolute'' stability of the level x_1 is a result of the special choice of the detuning δ. For $\delta \sim \Delta$ but $\delta \neq \Delta/2$, the level x_1 acquires a width that is small compared with the ionization width Γ, and is proportional to $(\delta - \Delta/2)^2$. Depending on the field strength ε_0, the width $|\mathrm{Im}(x_1(\varepsilon_0))|$ of the level has a maximum at $\varepsilon_0 \sim \varepsilon_1$, and for $\varepsilon_0 > \varepsilon_1$ it decreases, suggesting narrowing of the level and stabilization of the atom.[16] Thus, in the general case $\delta \neq \Delta/2$ (but $\delta \sim \Delta$), in the three-level system under study (Fig. 2), as the field increases, one of the three quasienergy levels (x_1) first undergoes the usual ionization broadening, and then it undergoes narrowing, which starts at anomalously low fields (as compared to a scheme without V transitions).

For large detunings the solutions of Eq. (13) can be found, in principle, by representing them as expansions in powers of $1/\delta$. However, we do not present the solutions here, since the applicability of such an expansion in itself seems formally justified only for very large detunings,

$$|\delta| > \frac{\Omega_R^2}{\Gamma} = \frac{1}{n_0'^3} = \Delta', \tag{18}$$

i.e., detunings exceeding the spacing between the neighboring Rydberg levels $E_{n'}$. Here the four-level model is better suited.

The solution of the photoionization problem is not complete when the quasienergies have been found. In principle, it is necessary to solve the initial value problem, i.e., to find the total electron wave function $\Psi(t)$ in a field or, at least, the projection $\Psi_{\text{bound}}(t)$ of this wave function on the bound states. When the interaction is turned on and off instantaneously, this problem can be solved by the method of quasienergies and quasienergy states. In this method, the function $\Psi_{\text{bound}}(t)$ is represented as a superposition of quasienergy functions

$$\Psi_{\text{bound}}(t) = \sum_k C_k \exp(-i\gamma_k t)\psi_k, \tag{19}$$

where the ψ_k are quasienergy functions corresponding to quasienergies γ_k and are obtained by solving equations like (12), while the expansion coefficients C_k are determined from the initial conditions.[17]

For a three-level system (Fig. 2), an exact analytic solution of the initial value problem can be found when the detuning δ is half the spacing between levels E_1 and E_2, $\delta = \Delta/2$. Omitting the cumbersome calculations, we present the result:

$$\begin{aligned} w_i(t) = 1 - \langle\Psi_{\text{bound}}(t)|\Psi_{\text{bound}}(t)\rangle &= \frac{\Omega_R^2 + (\Delta/2)^2}{2\Omega_R^2 + (\Delta/2)^2} \\ &\times[1 - \exp(-\Gamma t)] + \frac{\Gamma}{2\beta}\frac{\Omega_R^2}{2\Omega_R^2 + (\Delta/2)^2} \\ &\times\exp(-\Gamma t)\sin(2\beta t) - \frac{\Gamma^2}{2\beta^2}\frac{\Omega_R^2 + (\Delta/2)^2}{2\Omega_R^2 + (\Delta/2)^2} \\ &\times\exp(-\Gamma t)\sin^2(\beta t), \end{aligned} \tag{20}$$

where β is defined by Eq. (17).

According to Eq. (20), the atom is ionized in a time $t_1 \sim 1/\Gamma$. However, even in the asymptotic limit $t \gg t_i$, there exists a finite, nonzero, residual probability of finding the atom in the discrete levels w_{res}:

$$w_{\mathrm{res}}=\frac{\Omega_R^2}{2\Omega_R^2+(\Delta/2)^2}. \tag{21}$$

Since $\Omega_R \propto \varepsilon_0$, in weak fields ($\Omega_R < \Delta/2$) the residual probability $w_{\mathrm{res}}(\varepsilon_0)$ of (21) is an increasing function of the field strength ε_0. For $\Omega_R > \Delta/2$ the ionization probability $w_{\mathrm{res}}(\varepsilon_0)$ saturates at 1/2.

We emphasize that this limit is typical only of a three-level model. In a model with a large number of levels, the limiting residual probability $w_{\mathrm{res}}(\varepsilon_0)$ for large ε_0 will be different, but will nevertheless be finite, so that the ionization probability will never reach 1. In Sec. 4, we examine in more detail the effect that the number of levels included in the analysis has on the results by solving the problem numerically.

Note that the finiteness of the residual probability (the probability that the atom is not ionized in the limit of arbitrarily long laser pulse duration) is closely related to the inclusion of V transitions: when there is no V channel, the asymptotic residual probability vanishes. This result follows, specifically, from Eq. (21), where the lack of a V channel corresponds to $\Omega_R \equiv 0$, and therefore $w_{\mathrm{res}} \equiv 0$.

It follows from Eq. (20) that V stabilization requires a laser pulse of substantial duration. Indeed, if the pulse duration t is so short that $\Gamma t \ll 1$ and $\beta t \ll 1$, then the right-hand side of Eq. (20) can be expanded in a Taylor series in the time t, acquiring in the linear approximation a form identical to the result of perturbation theory (Fermi's golden rule)[18]:

$$w_i(t)=\Gamma t. \tag{22}$$

Obviously Eq. (22) does not describe any sort of stabilization, since according to Eq. (22) $w_i \propto \varepsilon_0^2$, i.e., the ionization probability increases monotonically as the field strength increases.

In the intermediate range of pulse durations, $\beta^{-1} \ll t \ll \Gamma^{-1}$, the first term on the right-hand side of Eq. (20) can easily be shown to make the main contribution to the ionization probability $w_i(t)$. Assuming that $\Gamma t \ll 1$, the exponential $\exp(-\Gamma t)$ can be expanded in series, which yields

$$w_i(t) \approx \frac{\Omega_R^2+(\Delta/2)^2}{2\Omega_R^2+(\Delta/2)^2}\,\Gamma t. \tag{23}$$

This expression likewise in no way describes stabilization, as the corresponding ionization probability $w_i(\varepsilon_0)$ is again a monotonically increasing function of the field strength ε_0.

Finally, only when the pulse duration t is greater than the ionization time of the atom $1/\Gamma$, $\Gamma t > 1$, does the ionization probability $w_i(\varepsilon_0)$ saturate and reach the level

$$w_i=1-w_{\mathrm{res}}, \tag{24}$$

where the residual probability of finding the atom in the discrete levels w_{res} is given by Eq. (21). According to our definition, saturation of the function $w_i(\varepsilon_0)$ at a level less than 1 can be interpreted as stabilization of the atom. For pulse durations of the order of the classical Kepler period, $t \sim T_K$, the resulting stabilization criterion $\Gamma t \sim 1$ holds for field strengths $\varepsilon_0 \sim \varepsilon_2$, i.e., under the same conditions as for interference stabilization due to Λ-type Raman transitions. However, if much weaker fields are of interest, $\varepsilon_0 \sim \varepsilon_1$, the stabilization criterion $\Gamma t > 1$ will impose a very stringent lower bound on the pulse duration t:

$$t > (n_0/n_0')^3 T_K. \tag{25}$$

Just how stringent (25) actually is can be assessed, for example, with parameter values corresponding to the experimental conditions of Ref. 4: $T_K \approx 3$ ps, $n_0 = 26$, $n_0' = 5$, and $(n_0/n_0')^3 \approx 10^2$. Then (25) holds if $t > 50$ ps. It should be noted that according to the numerical calculations performed below (Sec. 4), the inequality (25) softens somewhat when the number of levels in the model system is large. Nevertheless, stabilization due to V-type transitions can again be observed in comparatively weak fields, $\varepsilon_0 \sim \varepsilon_1$, only if the pulse duration t is sufficiently long and exceeds at least several Kepler periods. In this respect, stabilization due to V-type transitions differs radically from stabilization due to Λ-type transitions, which is most pronounced at $t < T_K$.

It was assumed everywhere in this section that $n_0 \gg n_0'$. We note here that such a strong inequality may not be satisfied. For example, at a frequency ω close to the electron binding energy in the level E_{n_0}, $\omega \approx 1/2n_0^2$, and as follows from energy conservation in Λ and V transitions, $n_0' \approx n_0/\sqrt{2}$. Hence, generally speaking, it also makes sense to study the case $n_0 \sim n_0'$. It is easy to see that in such a situation all three characteristic fields ε_1, ε_2, and ε_3 given by Eq. (14) and by the diagram in Fig. 3 are similar to one another, $\varepsilon_1 \sim \varepsilon_2 \sim \varepsilon_3$. This means that for $n_0 \sim n_0'$, the existence region for V stabilization $\varepsilon_1 < \varepsilon_2 < \varepsilon_3$ degenerates essentially to a point, and the only feasible stabilization mechanism is that due to Λ-type transitions.

Thus, the existence of a nontrivial range of parameters in which stabilization of the atom by V-type transitions can occur requires that the frequency of the field exceed the threshold value by a sufficient margin: $\omega \gg |E_{n_0}|$. On the other hand, in order for resonances with low-lying Rydberg levels to come into play, ω must not be too high, $\omega \ll |E_g|$, where E_g is the ground-state energy of the atom. On the whole, the constraints on the frequency of the laser field that ensure the existence of stabilization of an atom by V-type transitions can be written

$$|E_{n_0}| \ll \omega \ll |E_g|. \tag{26}$$

4. FOUR-LEVEL MODEL OF AN ATOM

As noted in the preceding section, the three-level model considered above becomes inapplicable at large detunings, $\delta \gg \Delta'$ (18). A natural and very simple generalization of the three-level model is a model with four levels, $E_{n_0} \equiv E_1$, $E_{n_0+1} \equiv E_2$ and, for example, $E_{n_0'-1} \equiv E_{1'}$ and $E_{n_0'} \equiv E_{2'}$ (see Fig. 4).

For a four-level model, the set of equations (5) assumes the form

$$i\dot{a}_{1'}(t)=(E_{1'}+\omega)a_{1'}(t)+\Omega_R[a_1(t)+a_2(t)],$$

$$i\dot{a}_{2'}(t)=(E_{2'}+\omega)a_{2'}(t)+\Omega_R[a_1(t)+a_2(t)],$$

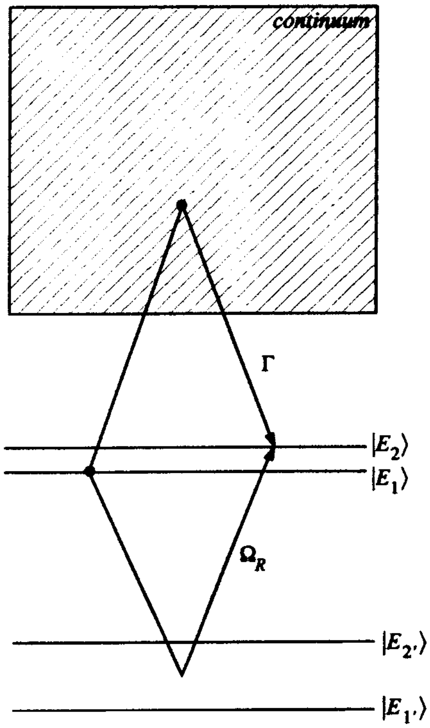


FIG. 4. Transition scheme for a four-level model.

$$\begin{aligned} i\dot{a}_1(t) &= \Omega_R[a_{1'}(t)+a_{2'}(t)] \\ &\quad + E_1a_1(t) - i\frac{\Gamma}{2}[a_1(t)+a_2(t)], \\ i\dot{a}_2(t) &= \Omega_R[a_{1'}(t)+a_{2'}(t)] \\ &\quad + E_2a_2(t) - i\frac{\Gamma}{2}[a_1(t)+a_2(t)], \end{aligned} \tag{27}$$

where Γ and Ω_R are given, as before, by Eqs. (8).

The characteristic equation of the system (27) is a fourth-order equation that generalizes Eq. (13):

$$\begin{aligned} (x-\delta)(x-\delta+\Delta')\left[x(x-\Delta)+i\frac{\Gamma}{2}(2x-\Delta)\right] \\ = \Omega_R^2(2x-\Delta)[2(x-\delta)+\Delta'], \end{aligned} \tag{28}$$

where x is, as before, the quasienergy measured from E_1, $x=\gamma-E_1$, and δ, Δ, and Δ' are given by Eq. (7) and (9).

To analyze large detunings $\delta \geqslant \Delta'$ analytically on the basis of simple equations, we examine the simplest case, in which $\delta=\Delta'/2$. Assuming Δ' to be a large parameter of the problem ($\Delta' \gg \Delta$), we note that for $\delta=\Delta'/2$ the right-hand side of Eq. (28) does not contain Δ'. Therefore, in the zeroth approximation in $1/\Delta'$, the solutions of Eq. (28) are identical to the solutions of the same equation with a vanishing right-hand side:

$$x_1^{(0)}=\delta, \quad x_2^{(0)}=\delta-\Delta, \quad x_{3,4}^{(0)}=\frac{\Delta}{2}-i\frac{\Gamma}{2} \pm\sqrt{\left(\frac{\Delta}{2}\right)^2-\left(\frac{\Gamma}{2}\right)^2}. \tag{29}$$

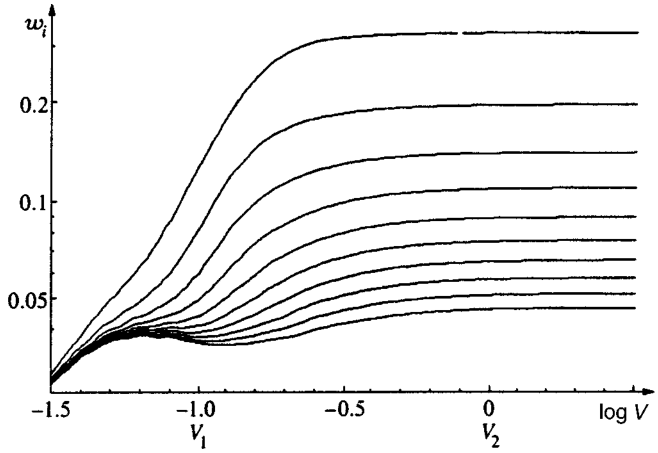


FIG. 5. Ionization probability as a function of the field for various numbers of levels in the series $\{E_n\}$: 3,5,...,21. The laser pulse is rectangular. The calculations were carried out with $t=8T_K$, $\delta=0$, and $(n_0/n_0')^{3/2}=10$.

The solutions (29) are completely independent of the Rabi frequency Ω_R. This means that in zeroth order in $1/\Delta'$ with $\delta=\Delta'/2$, interactions of the levels $E_{1,2}$ with the levels $E_{1',2'}$ are suppressed by virtue of interference. Therefore, the solutions (29) are the quasienergies that would be obtained for a system of two levels, E_1 and E_2, coupled to the continuum ($x_{3,4}$), and two lower-lying levels $E_{1'}$ and $E_{2'}$ that are neither coupled to one another, nor to the upper levels, nor to the continuum ($x_{1,2}$).

In this regard, it is understandable that the solutions $x_{3,4}$ of (29) describe narrowing of one of the levels for $\Gamma>\Delta$ ($\varepsilon_0>\varepsilon_2$), corresponding to interference stabilization of the atom by Λ-type transitions.[2] The requirement that the corrections to the solution (29) introduced by a nonvanishing right-hand side of (28) be small leads to the condition

$$|\delta|<\Omega_R, \tag{30}$$

which holds up to fields $\varepsilon_0\sim\varepsilon_3$. Therefore Λ-type stabilization occurs when the detuning δ is of the order of half the spacing Δ' between the lower-lying levels.

This result cannot be obtained in models containing only one level in the series of levels $\{E_{n'}\}$ (see Fig. 1), since it is a consequence of the partial mutual quenching of the contributions from various levels of the series (in the case at hand, the levels $E_{1'}$ and $E_{2'}$). The range of detunings over which V transitions cancel one another to a substantial degree is fairly wide, and is comparable to Δ'. Thus, it is possible to formulate constraints on the magnitude of the detuning under which stabilization of one or another type arises: near resonance, when

$$|\delta| \ll \Delta', \tag{31}$$

V stabilization occurs in fields $\varepsilon_0>\varepsilon_1$ that are weak compared with the fields $\varepsilon_0\sim\varepsilon_2$ typical of Λ stabilization in the absence of a V-transition channel. In turn, Λ stabilization arises far from resonance, with $|\delta|\sim\Delta'/2$, in fields $\varepsilon_0>\varepsilon_2$. These two types of stabilization transform into one another as the detuning δ varies over a scale $\sim\Delta'$.

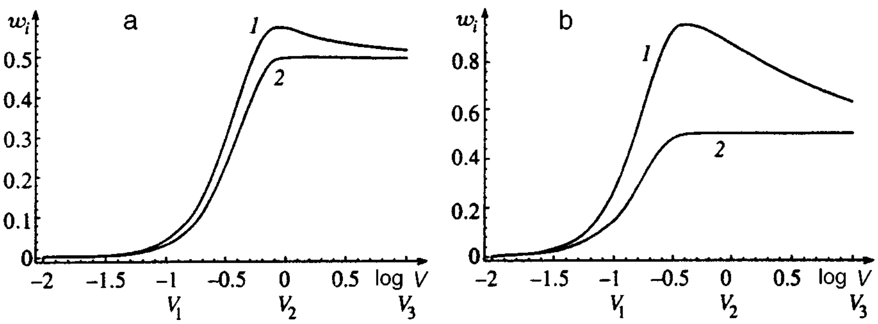


FIG. 6. Ionization probability calculated on the basis of a three-level model as a function of the field for a laser pulse turned on in two different ways—gradually (*1*) and instantaneously (*2*). a) narrow pulses ($t = T_K$); b) significantly longer pulses ($t = 5T_K$). Other parameters: $\delta = \Delta/2$, $(n_0/n_0')^{3/2} = 10$.

5. NUMERICAL RESULTS

For all their attractiveness and simplicity, the models studied above cannot claim to yield a satisfactory quantitative description of the spectra and ionization dynamics of a Rydberg atom. To better approximate reality, we examined more complicated multilevel models, solving the photoionization problem numerically. The numerical solution also made it possible to analyze the dependence of the results on other previously employed approximations, such as, for example, an approximation in which the interaction is turned on instantaneously, and others.

Above all, the existence of stabilization of an atom due to V-type transitions was checked against complicated multilevel models, which must be used to produce a quantitative description of the photoionization of a real Rydberg atom. Figure 5 displays the ionization probability curves corresponding to a systematically and symmetrically increasing number of levels near the initially populated level E_{n_0} (we study exact resonance, $\delta = 0$). It is assumed that these levels are equally spaced and possess equal transition matrix elements (see Eq. (8)). As one can easily see, adding new levels only strengthens the stabilization effect.

We also compared results obtained with two laser fields, one turned on gradually and the other instantaneously (Fig. 6). The results for pulses with identical peak field strength and energy per pulse were compared for the two turn-on modes. Gradual turn-on was modeled by an envelope of the form $f(t) = \sin^2(t/\tau)$. It was found that for narrow pulses, the numerical results for instantaneous and gradual turn-on are similar (Fig. 6a). For longer pulses it is clear (Fig. 6b) that the two ionization probability curves, for both gradual and instantaneous switching, are identical over the range of applicability of perturbation theory, but beyond this range they diverge. Significantly, gradual switching does not introduce any qualitative changes and does not destroy the stabilization effect, though it is less pronounced than when the interaction of the atom with the field is turned on instantaneously.

The dependence of the solutions on laser pulse duration is illustrated by numerical results plotted in Figs. 7a and 7b. These figures differ from one another in the number of atomic levels taken into account: three close levels $(E_{n_0}, E_{n_0+1}, E_{n_0-1})$ + one level $E_{n_0'}$ (Fig. 7a) as opposed nine levels E_n + one level $E_{n_0'}$ (Fig. 7b). One can see from the figure that the rise in ionization probability with increasing field strength, i.e., stabilization of the atom, ceases in fields that weaken as the pulse duration increases. However, the more levels the model contains, the less sensitive the results are to pulse duration. Thus, the condition (25) obtained in a three-level model relaxes somewhat in models comprising more levels. Nevertheless, even in such multilevel models, V-type transitions can only stabilize an atom if the pulse is at least several Kepler periods long.

Note that the numerical results obtained using models with small and large numbers of levels differ appreciably from one another. Nevertheless, all qualitative assertions made on the basis of the simplest models with a small number of levels still apply to more complicated models. We shall therefore not take up the question here of how many levels should be taken into account in types of models studied in order to obtain a correct quantitative description of the ionization of a real Rydberg atom. It is possible that such a problem will actually require not only quantitative but also qualitative modification of the computational models. This question requires further analysis, and despite its indisputable relevance, lies outside the scope of the present paper.

Figure 8 shows the ionization probability as a function of detuning δ for various $V = \varepsilon/\omega^{5/3}$. The numerical scheme comprises three levels $E_{n_0'}$ and fifteen levels E_n. The resulting curves confirm completely the assertion of the preceding

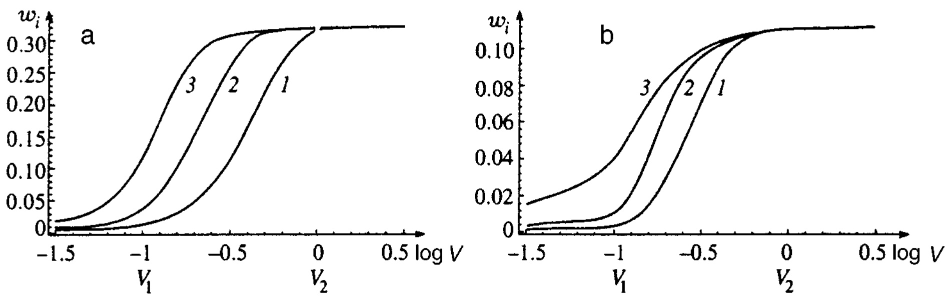


FIG. 7. Ionization probability as a function of field strength for a laser pulse turned on instantaneously. Pulse duration $t = T_K/2$ (*1*), $2T_K$ (*2*), $8T_K$ (*3*). Other parameters: $\delta = 0$, $(n_0/n_0')^{3/2} = 10$. The computational scheme contains $3+1$ (a) and $9+1$ (b) levels.

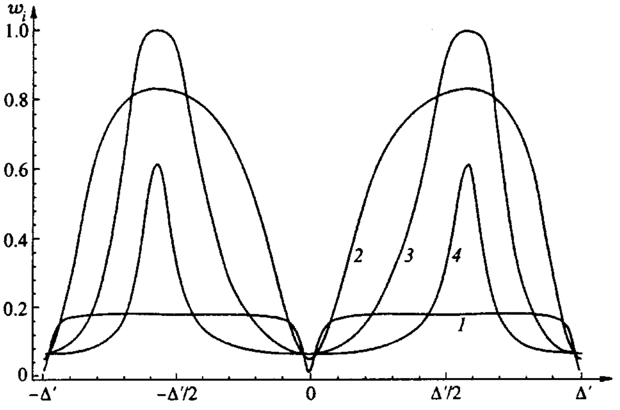


FIG. 8. Ionization probability as a function of detuning for various laser field strengths ($V=0.1$ (*1*), 0.3 (*2*), 1 (*3*), 3 (*4*)). The number of levels is 15 in the series $\{E_n\}$ and 3 in the series $\{E_{n'}\}$. The laser pulse is rectangular. Pulse parameters: $t=3T_K$, $(n_0/n_0')^{3/2}=10$.

section that the dependence of the solutions on the detuning appears on a scale $\sim\Delta'$. V-type stabilization, characterized by strong suppression of ionization, starting with relatively weak fields $V\sim V_1$, occurs when there is strong resonant coupling ($\delta=0,\pm\Delta'$). In contrast, Λ-type stabilization occurs in the interresonance region[2] ($\delta\sim\pm\Delta'/2$), showing up only in much stronger fields $V\geqslant V_2\sim 1$. Indeed, in the region $\delta\sim\pm\Delta'/2$, the curve $w_i(\delta)$ in Fig. 8 for $V=3$ lies below the curve $w_i(\delta)$ for $V=1$, indicating stabilization of the atom in fields $V>1$.

6. CONCLUSIONS

We have undertaken a systematic theoretical description of the photoionization of a Rydberg atom in a strong laser field, taking account of both Λ- and V-type Raman transitions, and investigated the conditions and mechanisms leading to stabilization of the atom under such conditions. We have shown that V-type resonant transitions determine a new type of interference stabilization of Rydberg atoms that differs from stabilization due to Λ-type transitions. The principal feature of V stabilization is that it can set in at very moderate field strength ($\varepsilon_0\sim\varepsilon_1$), much weaker than for Λ stabilization. For example, for $n_0/n_0'\sim 5$, the laser field strength at which deviations from perturbation theory first appear is two orders of magnitude lower than the field strength at which nonlinear effects would be observed in the absence of V transitions.

As shown above, the prerequisites for V-type stabilization are given by Eqs. (25), (26), and (31): the pulse duration must be long enough (25); the field frequency ω must be much greater than the binding energy $|E_{n_0}|$ of an electron in the Rydberg level, but much less than the binding energy $|E_g|$ in the ground state (26); and the detuning δ of the resonance must be much less than the spacing between neighboring Rydberg levels in the series $E_{n'}$ (with energy $E_{n'}\sim E_{n_0}-\omega$) (31). As shown above, the constraint on the pulse duration (25) is most stringent in the very simple three-level model, and softens somewhat in more realistic multilevel models. However, the requirement that the pulse duration be much greater than the classical Kepler period T_K holds as before.

Note also that none of the previous experiments on stabilization and population redistribution[3,4,19] satisfied all of the pertinent constraints for stabilization of a Rydberg atom due to V transitions. For example, in Ref. 4, no changes were observed in the redistribution of the populations of the Rydberg levels as the field frequency ω was varied. The detuning δ varied over a range of the order of 10Δ, where Δ is the spacing between neighboring Rydberg levels of the group $\{E_n\}$ (see Fig. 1). However, on the scale of the spacing Δ' between neighboring levels of the group $\{E_{n'}\}$ this range of variation of δ is very small—less than $0.1\Delta'$. On the other hand, in accordance with the results obtained above, large changes in the mechanism of ionization and stabilization of an atom and in the pattern of the distribution of the population of the atomic levels can be observed for detuning δ varying over a range of the order of Δ'. Moreover, in the experiment of Ref. 4, the laser pulse duration was less than one Kepler period, which, according to Eq. (25), is insufficient for stabilization of an atom by V-type Raman transitions.

Thus, on the whole, it must be acknowledged that to date stabilization of an atom by V-type Raman transitions has not been observed experimentally. At the same time, the conditions necessary for such an experiment are easily realizable. In our opinion, it would definitely be of interest to perform such an experiment. V-type stabilization can be observed, for example, under the following conditions: $n_0=25$, $n_0'=5$, $\omega\approx 8\times 10^{14}\ \mathrm{s}^{-1}$, $t>15$ ps, and $\varepsilon\geqslant 10^6$ V/cm, which corresponds to intensities $I\geqslant 10^9\ \mathrm{W/cm}^2$.

We thank the participants of the scientific seminar led by N. B. Delone (IOFAN) for valuable remarks during a discussion of this problem.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 9602-17649) and the U.S. Civilian Research and Development Foundation (CRDF) (Grant No. RP1-244).

*)E-mail: fedorov@theor.msk.ru

---

[1] M. Gavrila and J. Z. Kaminski, Phys. Rev. Lett. **52**, 613 (1984).
[2] M. V. Fedorov and A. M. Movsesian, J. Phys. B **21**, L155 (1988).
[3] J. H. Hoogenraad, R. B. Vrijen, and L. D. Noordam, Phys. Rev. A **50**, 4133 (1994).
[4] L. D. Noordam, H. Stapelfeldt, and D. I. Duncan, Phys. Rev. Lett. **68**, 1496 (1992).
[5] M. Yu. Ivanov, Phys. Rev. A **49**, 1165 (1994).
[6] A. Wojcik and R. Parzinski, Phys. Rev. A **50**, 2475 (1994).
[7] A. Wojcik and R. Parzinski, J. Opt. Soc. Am. B **12**, 369 (1995).

[8] I. Ya. Bersons, Zh. Éksp. Teor. Fiz. **80**, 1727 (1981) [Sov. Phys. JETP **53**, 891 (1981)].
[9] N. B. Delone, S. P. Goreslavsky, and V. P. Krainov, J. Phys. B **22**, 2941 (1989).
[10] M, S. Adams, M. V. Fedorov, V. P. Krainov, and D. D. Meyerhofer, Phys. Rev. A **52**, 125 (1995).
[11] M. V. Fedorov, *Electrons in a Strong Optical Field* [in Russian], Nauka, Moscow (1991).
[12] M. V. Fedorov, Laser Phys. **3**, 219 (1993).
[13] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, 3rd, ed., Pergamon Press, New York (1977) [current Russian edition, Nauka, Moscow (1989)].
[14] Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **51**, 1492 (1966) [Sov. Phys. JETP **24**, 1006 (1967)].
[15] V. I. Ritus, Zh. Éksp. Teor. Fiz. **51**, 1544 (1966) [Sov. Phys. JETP **24**, 1041 (1967)].
[16] M. V. Fedorov and N. P. Poluektov, Laser Phys. **7**, 299 (1997).
[17] M. V. Fedorov, M.-M. Tegranchi, and S. M. Fedorov, J. Phys. B **29**, 2907 (1996).
[18] M. V. Fedorov and A. E. Kazakov, Prog. Quantum Electron. **13**, 1 (1989).
[19] R. B. Vrijen, J. H. Hoogenraad, and L. D. Noordam, Phys. Rev. A **52**, 2279 (1995).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Generation of the line radiation of argon added to DT gas in Iskra-5 experiments

S. A. Bel'kov,*) A. V. Bessarab, A. V. Veselov, V. A. Gaĭdash, G. V. Dolgoleva,
N. V. Zhidkov, V. M. Izgorodin, G. A. Kirillov, G. G. Kochemasov,†) D. N. Litvin,
S. P. Martynenko, E. I. Mitrofanov, V. M. Murugov, L. S. Mkhitar'yan, S. I. Petrov,
A. V. Pinegin, V. T. Punin, and N. A. Suslov

Russian Federal Nuclear Center, All-Russia Research Institute for Experimental Physics, 607190 Sarov, Nizhegorod Region, Russia

(Submitted 25 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 837–848 (September 1998)

The first experiments measuring the density of a compressed deuterium and tritium mixture in microtargets of indirect irradiation (x-ray targets) were performed at the Iskra-5 facility. The density was determined according to the broadening of the lines of hydrogen- and helium-like argon added to the DT gas as a diagnostics material. A series of three experiments was performed with x-ray targets in which the central capsule filled with a DT+Ar mixture over a range of shell thicknesses. In two of the experiments, argon emission spectra were recorded and the density of the compressed gas was determined. For a microtarget approximately 280 μm in diameter with a wall approximately 7 μm thick, an analysis of the experimental results yielded an estimated density in the compressed gas of ~ 1 g/cm^3. Gas-dynamic calculations using the SNDA (spectral nonequilibrium diffusion with absorption) program show that argon emission takes place just after reaching maximum temperature, but much sooner than maximum compression. The results of a calculation for an experiment with low relative Ar concentration are in overall agreement with the experimental data. Additional investigations are needed to interpret experiments at a relatively high concentration. © *1998 American Institute of Physics.* [S1063-7761(98)00609-X]

1. INTRODUCTION

Determining the density of compressed deuterium–tritium (DT) fuel is one of the most difficult problems of laser-induced thermonuclear fusion. Information about DT-fuel parameters can be obtained by analyzing the characteristics of the x-rays and thermonuclear particles emerging from the volume occupied by the compressed and heated fuel (see, for example, Ref. 1). An effective method of determining the density is based on introducing medium- and high-*Z* additives into the DT fuel, with spectroscopy of the resulting ionic lines of the diagnostics material. In practice, preference is given to inert gases (Ne, Ar, Xe).[2–5] The density of the DT mixture is monitored on the basis of measurements of the width of the emission lines of H- and He-like ions.

Calculations show[6] that temperatures of the ionic component of DT $T_i \simeq 2-3$ keV (Ref. 7) and electron temperatures $T_e \simeq 1-1.5$ keV have been obtained in experiments performed at the Iskra-5 facility using $\simeq 280$ μm diameter glass targets with walls 3–7 μm thick positioned inside a spherical gold box 2 mm in diameter. At such temperatures, a substantial number of Ar ions are in hydrogen- and helium-like states. Thus, we are able, using the Iskra-5 facility, to determine the density of compressed DT gas on the basis of Ar ion spectroscopy.

For the measurements, a spectrometer that detects line radiation in the range $\simeq 3-4$ Å with resolution $\lambda/\Delta\lambda = 10^3$ and a technology for filling shells with a mixture of DT and Ar were developed. The results of the first successful Iskra-5 experiments in which the line spectra of He- and H-like Ar-ions were recorded are reported below.

2. EXPERIMENTAL LAYOUT AND PRINCIPAL RESULTS

The target consists of a thin-wall spherical copper shell whose inner surface is coated with a gold layer ~ 1 μm thick. Shells with a diameter of 2 mm and with apertures 0.6 mm in diameter for introducing laser radiation were used in the experiments. A glass capsule 280–300 μm in diameter was positioned at the center.

The glass capsule was filled with a mixture of Ar (~ 1 atm) and DT (10–20 atm). The capsule was filled with argon through an opening made in the capsule using a periodic-pulse laser. After the opening was sealed, the capsule was filled with the DT mixture by diffusion.[8]

The amount of argon in the capsule was checked either via mass spectrometry or x-ray fluorescence, based on excitation of the x-ray lines of Ar by tritium β-radiation.

After the capsules were filled and checked for the presence of argon and DT gas, they were coated with an additional layer of SiO_2 to make up the required wall thickness.[9,10]

The Iskra-5 diagnostics system is described in detail in Refs. 7 and 11.

The principal experimental results are presented in Table I. In all experiments, the radiation-to-background contrast in both energy and power was at least 10^8. Schlieren images showed that the target sustained no damage before the arrival of a monopulse.

TABLE I.

Experiment No.	ϕ_{sh}, μm	ΔR_{sh}, μm	P_{DT}, atm	E_L, J	$\tau_{0.5}$, ns	$\tau_{\gamma\gamma}$, ns	$\tau_{\gamma n}$, ns	N, 10^9	T_{DT}, keV
1	296	4	11	6700	0.41	0.45 ± 0.15	0.66 ± 0.2	0.8	-
2	300	6.7	20	7400	0.34	0.58 ± 0.15	0.52 ± 0.12	2.5	2.3
3	272	10	9	7300	0.36	-	-	0.07	-

Notation: $\varnothing_{sh}$, ΔR_{sh}—diameter and thickness, respectively, of a glass microsphere; P_{DT}—pressure of the DT gas; E_L, $\tau_{0.5}$—energy and duration of the laser pulse introduced into the interaction chamber; $\tau_{\gamma\gamma}$—delay of the onset of x-ray generation from the compressed core relative to the onset of x-ray generation at the converter wall; $\tau_{\gamma n}$—delay of neutron generation onset relative to x-ray generation onset at the converter wall; N—total neutron yield; T_{DT}—DT fuel temperature determined by the time-of-flight method.

Images of the target obtained with a camera obscura with $\approx 25\ \mu$m spatial resolution behind various filters are presented in Fig. 1. The line of sight of the camera obscura is off-axis relative to the aperture for the laser radiation, so that the image of the microtarget lies near the edge of that aperture. It is evident from the figure that in experiment No. 1 the compressed region is circular, with a bell-shaped brightness distribution (in experiments without the Ar additive the compressed region is annular). The size of the compressed region is $\leqslant 55-60\ \mu$m. In experiment No. 3, the obscurogram consists of a brightly luminescing region surrounded by a darker aureole; the size of the region is at most $\approx 30\ \mu$m. The diameter of the aureole is 50–60 μm, just as in experiments Nos. 1 and 2.

We used a spectrograph based on a flat Si crystal for spatially-resolved detection of x-ray line radiation to measure the density of the compressed core according to the Stark broadening of the hydrogen- and helium-like Ar lines. The measurement scheme is presented in Fig. 2.

The spectrograph parameters are as follows:

range 1.1 Å (from 2.6–2.9 Å to 3.7–4.0 Å, depending on the placement of the spectrograph in the chamber);

spectral resolution (with a 50 μm core) ≈ 0.003 Å (≈ 3 eV), determined entirely by the size of the compressed core;

spatial resolution $\approx 70\ \mu$m;

magnification 4;

target–photographic film distance 10 cm.

Figure 3 displays spectrophotograms and reconstructed x-ray intensities for experiment No. 1. The spectrum appears approximately the same in experiment No. 2. The number of the array element obtained by digitizing the image is plotted on the x axis in the spectrogram. One element in this array

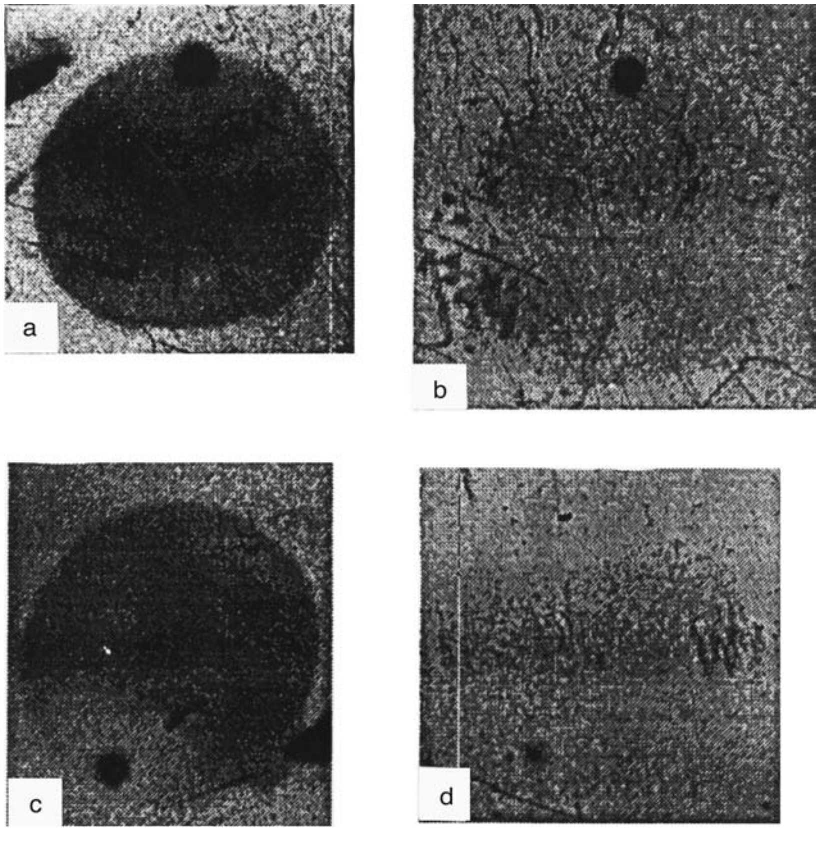


FIG. 1. Obscurograms of the central capsule in experiments No. 1 (a, b) and No. 3 (c, d). The images were obtained with the following filters: a and c—20 μm lavsan, b and g—10 μm polyparaxymelene+2.1 μm Cd.

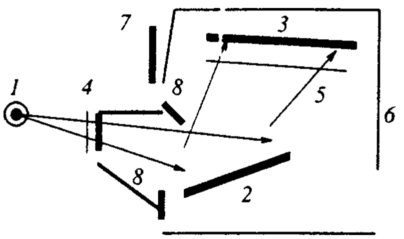


FIG. 2. Diagram of Ar x-ray line radiation measurements: *1*) Target; *2*) Si crystal; *3*) UFSH-S photographic film; *4*) entrance filter (Be 20–40 μm); *5*) filter for shielding from scattered and fluorescence x rays and light (Al 2 μm); *6*) spectrograph case; *7*) filter for shielding the film from hard radiation (Pd 3 mm); *8*) diaphragms for shielding from scattered radiation.

corresponds to 30 μm on the photographic film, i.e., ≈0.00158 Å.

The lines detected are marked in the spectrograms. Their wavelengths [in Å] are: He_α [3.949], L_α [3.722], He_β [3.366], He_γ [≈3.20], L_β [3.155], He_δ [≈3.13] L_γ [2.987]. Satellites are observed about the He_α and He_β lines.

A preliminary analysis of the spectrograms showed that 1) the ratio of the L_β to He_β line intensity is approximately 0.8 in both experiments; 2) the width of the He_β line at half intensity in both experiments is ≈20 mÅ (≈22 eV); 3) the width of the L_β line is ≈13.5 mÅ (≈17 eV) and ≈9 mÅ (≈11 eV) for the first and second experiments, respectively.

3. DISCUSSION

We now summarize the principal experimental results.

1. In two$^{1)}$ of the three experiments, lines of the 2–1, 3–1, and in part the 4–1 transitions in hydrogen- and helium-like Ar ions with linewidth up to ≈20 eV, which is much greater than the resolution of the spectrograph (≈3 eV), were detected at a high confidence level.

2. The structures of the line spectrum in experiments No. 1 and 2 are similar to one another. The linewidths of the 3–1 transitions, which were of principal interest for density diagnostics, are also essentially the same.

3. Ar line radiation was not detected in experiment No. 3, where the glass shell was 10 μm thick. The image of the luminescing compressed ''core'' in the obscurogram in this experiment is much smaller than in the first two experiments.

We first analyze experiment No. 2 in some detail. The He_β, He_γ, and L_β lines are clearly seen in the spectrum of this experiment. The two-peak structure at the position of the L_β line can be interpreted in two ways. First, it can be interpreted as a single L_β line with a central intensity dip typical of 3–1 transitions, and width ≈30 eV. The second and more likely interpretation is that the structure consists of a L_β line with a width ≈11 eV and a He_β line. The ratio of the L_β to He_β line intensity is approximately 0.8. A preliminary estimate of the density can be made using the Inglis–Teller formula $\log n_e = 23.26 - 7.5 \log n + 4.5 \log Z$, where n is the number of the last line that can be resolved in the series and Z is the charge of the nucleus (in this case $Z=17$ for H-like and $Z=16$ for He-like ions). Setting $n=4$ and $Z=16$, we obtain $n_e = 2.75\times10^{23}$ cm^{-3} ($\rho\approx1$ g/cm^3).

A more accurate estimate of the electron density can be obtained by analyzing the lineshapes—specifically, the line-

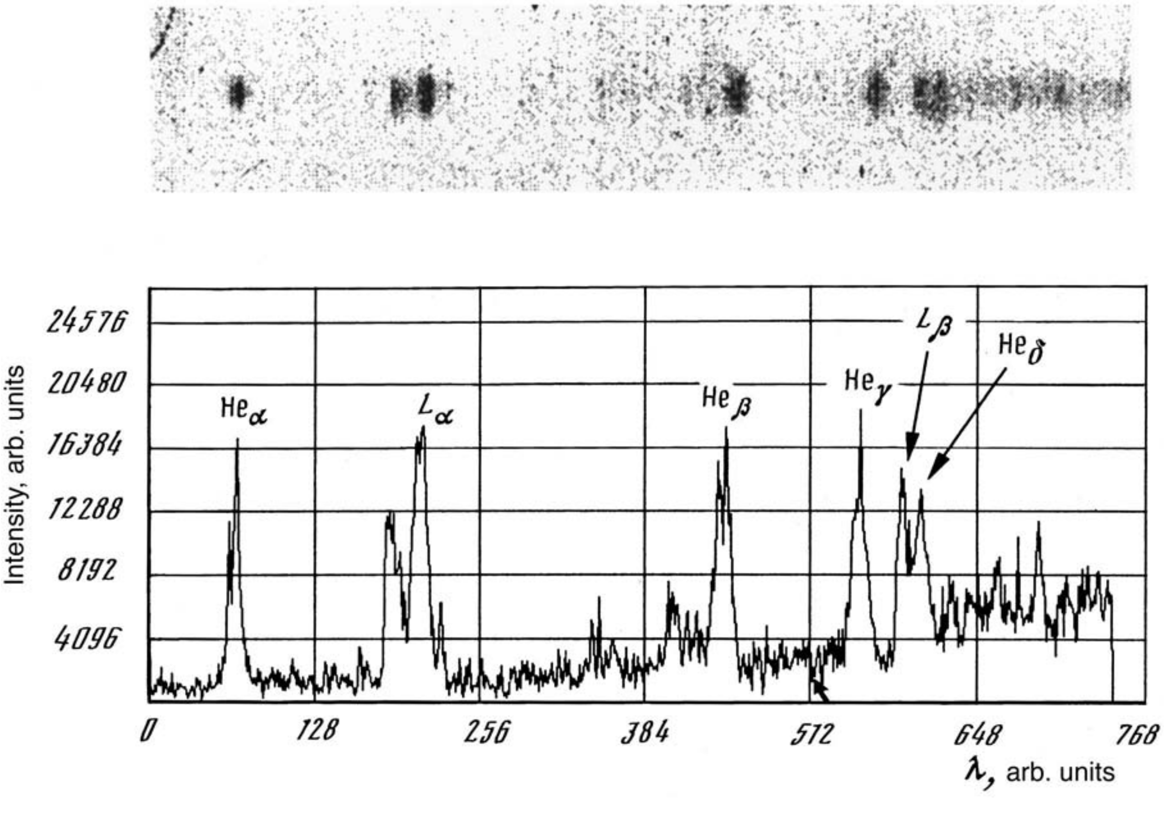


FIG. 3. Spectrophotogram and reconstructed intensity of x rays from the target in experiment 1.

widths. Both ions and electrons contribute to broadening of the Ar lines. As is often the case, under the conditions that we analyzed, it is the ions, whose effect is quasistatic, that make the main contribution.[12]

In the quasistatic regime, the ions are assumed to be immobile. Line broadening is due to the Stark shift of the levels induced by the electric field of the ions surrounding the emitting ion. Linewidths in an ideal plasma can be estimated using the relation[12]

$$\Delta E=\frac{5.55(n^2-1)ea_0}{Z_{\mathrm{Ar}}}F_{0(\Sigma)}, \tag{1}$$

where n is the principal quantum number of the level, a_0 is the radius of the Bohr orbit, and $Z_{\mathrm{Ar}}=18$ is the charge of the argon nucleus. In the present case, both ions of hydrogen isotopes ($Z=1$) and argon ions contribute to the field $F_{0(\Sigma)}$.

In the presence of two types of perturbing particles, the characteristic Holtsmark field $F_{0(\Sigma)}$ is

$$F_{0(\Sigma)}=(F^{3/2}_{0(\mathrm{DT})}+F^{3/2}_{0(\mathrm{Ar})})^{2/3}=2.6e(n_{\mathrm{H}}+Z^{3/2}_{\mathrm{Ar}}n_{\mathrm{Ar}})^{2/3}$$
$$=2.6en_e^{2/3}(\alpha+\beta Z^{3/2}_{\mathrm{Ar}})^{2/3}. \tag{2}$$

Here $\alpha=n_{\mathrm{H}}/n_e$, $\beta=n_{\mathrm{Ar}}/n_e$, and n_e, n_{H}, and n_{Ar} are the electron, hydrogen ion, and argon densities. The quantities α and β can be expressed in terms of n_{H} and n_{Ar}. Introducing

$$p=\frac{n_{\mathrm{H}}}{n_{\mathrm{H}}+n_{\mathrm{Ar}}},\quad q=1-p=\frac{n_{\mathrm{Ar}}}{n_{\mathrm{H}}+n_{\mathrm{Ar}}},$$

we obtain

$$\alpha=\frac{p}{p+Z_{\mathrm{Ar}}q},\quad \beta=\frac{q}{p+Z_{\mathrm{Ar}}q}.$$

The relation (1) is valid if the number of particles in the Debye sphere is large, $N_D\gg 10^3$. In a nonideal plasma, the presence of screening and correlation of charged particles effectively reduces the mean active field. Extending the result of Ref. 13 to a multicomponent plasma yields a correction associated with the departure from ideal behavior:

$$F_{\mathrm{nonid}}=F_{0(\Sigma)}(1-0.7N_D^{-1/3})$$
$$\simeq\exp\left(-\frac{2}{3}\frac{r_1}{r_D}\right)\cdot 2.6en_e^{2/3}(\alpha+\beta Z^{3/2}_{\mathrm{Ar}})^{2/3}, \tag{3}$$

where r_1 is the mean separation between particles, N_D is the total number of particles in the Debye sphere, and r_D is the Debye radius, given by

$$\frac{1}{r_D^2}=4\pi r_0\frac{m_ec^2}{T}n_e(1+\alpha+\beta Z^2_{\mathrm{Ar}}).$$

Here $r_0=2.8\times 10^{-13}$ cm is the classical radius of the electron.

To check the relation (1) with F from Eq. (3), we compared it with calculations from Ref. 14, where the shape of the Ar L_β line was calculated for $n_e=2\times 10^{23}$ cm^{-3}, $T_e=1$ keV, and various values of p by the Baranger–Mozer method. Figure 4 compares the calculations[14] of the linewidth ΔE and the widths estimated using Eq. (1), with $F_{0(\Sigma)}$ replaced by F_{nonid} from Eq. (3). The calculations agree well with the simple estimate.

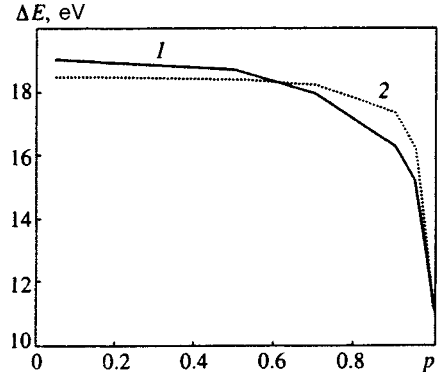


FIG. 4. Dependence of the ArXVIII L_β linewidth on the percentage content of DT in the mixture Ar+DT ($n_e=2\times 10^{23}$ cm^{-3}, $T_e=T_i=1$ keV): *1*) Data from Ref. 12; *2*) estimate from Eq. (1).

The estimates reproduce two important circumstances also found in Ref. 15. In the first, a 10% addition of Ar to DT gas produces almost the same broadening as pure Ar. In the second, in a nonideal plasma the sign of the inequality between the Ar linewidths due to the action of the argon ions themselves and the deuterium and tritium ions with $n_e=\mathrm{const}$ remains unchanged: $\Delta E_{\mathrm{Ar}}>\Delta E_{\mathrm{DT}}$.

We also checked the approach based on Eqs. (1) and (3) by calculating the width of the L_β line for extremely low Ar concentrations in DT at various values of n_e, and compared the results with data from Ref. 4. The results of the detailed calculations agree with the simple estimate, for n_e ranging from 10^{23} cm^{-3} to 10^{24} cm^{-3}.

Using Eqs. (1) and (3), the behavior of the L_β linewidth (and also the He$_\beta$ linewidth[4]) can be calculated with a high degree of confidence at an Ar concentration $q=0.025$. Thus, $n_e=2.5\times 10^{23}$ cm^{-3} corresponds to a linewidth $\Delta E\simeq 17$ eV.

Estimates of the contribution of other broadening mechanisms, such as Doppler, impact, and so on, show that they play a minor role compared to quasistatic broadening by ions; they contribute 2–3 eV at most.

Estimates of the optical depth of the lines of $3\rightarrow 1$ transitions show that for our conditions these lines can be treated as optically thin to a first approximation.

We now proceed to an analysis of experiment No. 1. Comparing the spectra in experiments No. 1 and 2 shows that they are very similar. This seems strange, as in all likelihood, the density in the second experiment should be higher than in the first. Indeed, the initial densities of the DT+Ar mixture in these experiments are $\rho^{(0)}\approx 4.2$ mg/cm^3 and $\rho^{(0)}\approx 6.2$ mg/cm^3. The degree of compression (assuming that compression is governed primarily by gas dynamics, while radiation and electronic heat conduction play a small role) is governed by the ratio M/m, where M and m are the mass of the shell and gas. The ratio of these quantities in the two experiments was ≈ 0.88, while the ratio of the densities of the compressed gas should be ≈ 0.6 (when heat conduction and the transition to the ablation regime are taken into account, this number should become even smaller). The L_β linewidths should therefore have differed by at least a factor of 1.37.

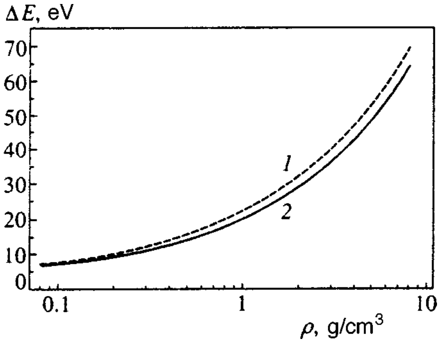


FIG. 5. Density dependence of the He_β and L_β linewidths with various initial DT pressures ($p_{\mathrm{Ar}}=1$ atm): *1*—$p_{\mathrm{DT}}=10$ atm, *2*—$P_{\mathrm{DT}}=20$ atm (instrumental line broadening ≈ 3 eV is taken into account).

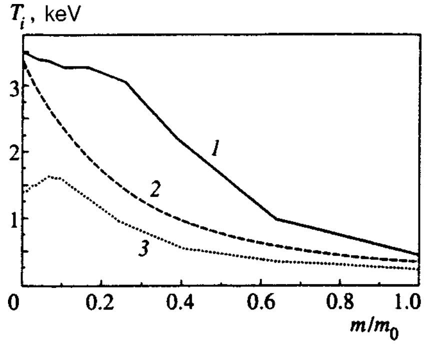


FIG. 6. Computed ion temperature distribution as a function of the mass coordinate in gas at the moment of generation of half of the neutron yield: *1*) Experiment 1; *2*) experiment 2; *3*) experiment 3; m_0 is the total gas mass in the corresponding calculation.

It was shown above (see Fig. 4) that the linewidth at fixed electron density n_e increases with the relative Ar concentration in the mixture. As q increases from 0.025 (experiment No. 2) to $q=0.044$ (experiment No. 1), the linewidth increases by approximately 30%. If the fact that the density of the DT+Ar mixture at fixed n_e varies as a function of p according to

$$\rho=\frac{n_e}{N_A}(A_{\mathrm{DT}}\alpha+A_{\mathrm{Ar}}\beta)$$

is also taken into account (here N_A is Avogadro's number, and A_{DT} and A_{Ar} are the atomic weights of the gas components), then we find that for the second experiment the density should be ≈ 0.8 g/cm^3 (see Fig. 5).

4. COMPARISON OF EXPERIMENTAL DATA WITH GAS-DYNAMIC CALCULATIONS

For a more complete interpretation of the experimental results, calculations were performed using the one-dimensional gas-dynamic program SNDA (spectral nonequilibrium diffusion with absorption).[6] The calculations were performed for parameters of the target and the laser pulse introduced into the converter box that correspond to experiment. The equation of state and transport coefficients of DT+Ar were calculated using the average-ion model. The numerical results are presented in Table II.

We see from Table II that the computed neutron yield for experiments 1 and 2 is two to four times the yield measured in the corresponding experiment. For experiment 3, the computed neutron yield is essentially identical to the experimental value. Comparing the energy balance for the DT gas at the moment of maximum compression shows that in the calculations the principal energy losses from the compressed, hot gas are mediated by argon x-ray emission.

TABLE II.

Experiment No.	R_0, μm	E_{DT}, J	E_γ, J	E_e, J	$R_{\min}$, μm	$\delta_{N/2}$	δ_{Ar}	δ	N, 10^9
1	147	10	5.5	3.1	31	47.9	67	107	3.8
2	146.9	13	8.6	0.71	19	173	175	463	5.6
3	133.6	6.6	4.4	0.58	9.36	153	269	2904	0.062

Notation: E_{DT}—energy introduced into the DT gas during the compression process up to the moment of maximum compression of the central capsule; E_γ, E_e—energy losses of the DT gas due to radiation and electronic heat conduction up to the same moment in time; R_0, $R_{\min}$—initial and minimum interior radii of the glass shell.

Figure 6 shows the computed distribution of ion temperature as a function of normalized mass coordinate of the gas at the moment of maximum neutron production rate. We see from the figure that the temperature distribution in the gas is highly nonuniform.

An important characteristic for determining gas density is the instant at which argon line emission begins. Since the kinetics of gas ionization is calculated in the SNDA program in the average-ion approximation, it is impossible to make a direct, accurate calculation of the temporal form of the emission of the corresponding lines, since we know only the average populations of the ionic levels in the various charge states of argon. However, if it is assumed that the density distributions in the excited states of hydrogen- and helium-like argon are close to equilibrium,[2)] we can reconstruct the concentration of excited states in terms of the average populations, using the binomial distribution

$$C_k=\prod_n C_{g_n}^{N_n^k} f_n^{N_n^k}(1-f_n)^{g_n-N_n^k},$$

where n is the principal quantum number of the level, the N_n^k are the occupation numbers of level n for a prescribed state of the ion k, g_n is the statistical weight of a level ($g_n=2n^2$), $f_n=P_n/g_n$, P_n is the average population of level n, the and

$$C_{g_n}^{N_n}=\frac{g_n!}{N_n!(g_n-N_n)!}$$

are binomial coefficients.

For a calculation of the emission of the hydrogen- and helium-like argon ions of interest—L_α, L_β, L_γ, He_α, He_β, and He_γ—the occupation numbers of the corresponding states are

$$N_n^{L\alpha}=\begin{cases}1, & n=2,\\ 0, & n\neq 2,\end{cases}\qquad N_n^{L\beta}=\begin{cases}1, & n=3,\\ 0, & n\neq 3,\end{cases}$$

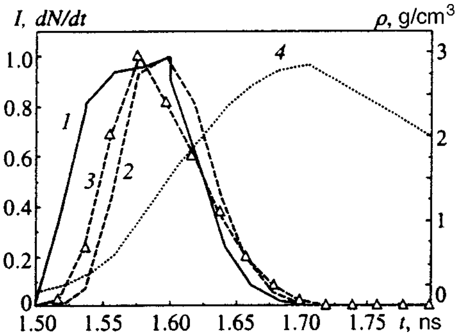


FIG. 7. Computed time dependence for the conditions of experiment 2: *1*) Neutron generation rate dN/dt; *2, 3*) emission intensities of the L_β, He_β argon lines; *4*) average gas density. The maximum L_α, L_β, L_γ, He_α, He_β, and He_γ line intensities are in the ratio 8.2:2.3:1:15:4.4:1.9.

$$N_n^{L\gamma}=\begin{cases}1, & n=4,\\ 0, & n\neq 4,\end{cases}\qquad N_n^{\mathrm{He}\alpha}=\begin{cases}1, & n=1,2,\\ 0, & n\neq 1,2,\end{cases}$$

$$N_n^{\mathrm{He}\beta}=\begin{cases}1, & n=1,3,\\ 0, & n\neq 1,3,\end{cases}\qquad N_n^{\mathrm{He}\gamma}=\begin{cases}1, & n=1,4,\\ 0, & n\neq 1,4.\end{cases}$$

The concentrations of the excited states of the argon atoms calculated in this manner were used to calculate the temporal shape of the argon lines. In so doing, the absorption of the lines as they pass both through the compressed gas and through the glass shell was taken into account. Figure 7 shows the computed results for the temporal characteristics of the radiating mixture of DT and Ar in experiment 2. For experiments 1 and 3, the behavior looks approximately the same. Moreover, as one can see from the figures, in all calculations the various lines are emitted at virtually the same instant in time (to within ±25 ps). The duration of emission is also essentially the same. The lines are emitted essentially immediately after the peak of the neutron production rate (approximately 25–50 ps later, i.e., somewhat later than the time at which the gas temperature reaches its maximum value), and much earlier than the time of maximum compression.

Figure 8 shows the spatial profiles of the density, electron temperature, and luminosity of the L_β and He_β lines for the conditions of experiment 2.

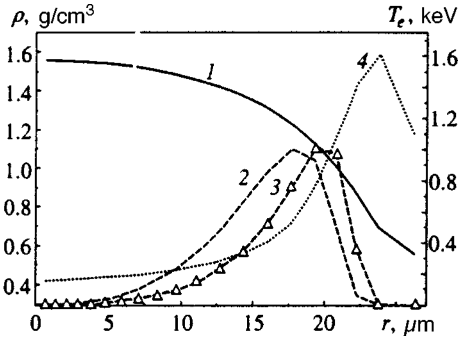


FIG. 8. Computed spatial profiles of the electron temperature (curve *1*), luminosities of the lines L_β (curve *2*) and He_β (curve *3*), and density of the DT+Ar mixture (curve *4*) at the moment of maximum argon line emission for the conditions of experiment 2.

TABLE III.

Experiment No.	N, 10^9	T_i, keV	$\Delta E(\mathrm{He}_\beta)$, eV	$\Delta E(L_\beta)$, eV	$I_{\mathrm{H}}/I_{\mathrm{He}}$	$\varnothing_{\min}$, μm	$\tau_{\gamma n}$, ns	$\tau_{\gamma\gamma}$, ns
1	0.8	-	22	17	0.8	60	0.66	0.45
1	3.8	-	≈10	7	0.2	70	0.35	0.4
2	2.5	2.3	22	11	0.8	55	0.52	0.58
2	5.6	2.6	≈20	13	0.52	50	0.57	0.6

For experiment 2, the maximum luminosity of the He_β line corresponds to a density ≈1.1 g/cm^3, which agrees with the estimate based on the linewidth, presented earlier in this section. The maximum luminosity of the L_β line corresponds to a density ≈0.7 g/cm^3, which is also close to the estimate ≈0.6 g/cm^3. The indicated ranges correspond to electron temperatures $T_e\approx 1$ keV. The computed ratio of the L_β and He_β line intensities is 0.52, and the experimental value is ≈0.8. The diameter of the line emission region in the calculation is approximately 50 μm, and the experimental estimate obtained with the camera obscura is ≈55 μm. The computed times $\tau_{\gamma n}$ and $\tau_{\gamma\gamma}$ in experiment 2 are 0.57 ns and 0.6 ns, and the experimental values are 0.52 ns and 0.58 ns.

The computed and experimental values of the various parameters for experiments 1 and 2 are compared in Table III (top row—experiment, bottom row—calculation).

The agreement between calculation and experiment for experiment 2 is very good. For experiment 1, the discrepancy is greater.

The main difference between the experiments is that the relative Ar concentration is higher in the first experiment. This suggests that the model of thermodynamic and radiation properties of DT+Ar plasma with high relative Ar concentrations must be improved.

This is also suggested by experiment 3, in which the relative Ar concentration is approximately the same as in the first experiment. Here it is observed that the computed and experimental neutron yields are in agreement, and the size of the emitting region (≈30 μm) obtained from the obscurogram is correlated with the computed size. The luminosity of the lines obtained in the optical transparency approximation is comparable to the luminosities in experiments 1 and 2, but no line radiation is observed in the experiment. Additional investigations are required to understand the reason for the discrepancy.

5. CONCLUSIONS

1. The first successful series of experiments with indirectly illuminated targets, in which Ar was added to DT gas for diagnostic purposes, has been conducted at the Iskra-5 experimental station. We detected the emission-line spectrum of hydrogen- and helium-like Ar ions—specifically, lines corresponding to the transitions 2→1, 3→1, and possibly 4→1. The density of the compressed gas at the moment of maximum line emission was estimated from the

width of the He$_\beta$ $(3 \rightarrow 1)$ lines: $\rho_{\mathrm{DT+Ar}} \approx 0.8$ g/cm^3 for experiment 1 and $\rho_{\mathrm{DT+Ar}} \approx 1.1$ g/cm^3 for experiment 2.

2. The experiments were analyzed with the aid of spherically symmetric calculations using the radiative gas dynamics program SNDA. It was shown that for experiment 2, in which the partial Ar concentration was $q = 2.44\%$, highly satisfactory agreement is observed with all measured characteristics: neutron yield, ion temperature, time intervals between the laser pulse and the pulsed neutron and x-ray (generated by the capsule with DT+Ar) emission, the sizes of the compressed and heated regions, and densities of the compressed and heated DT+Ar gas. In experiments 1 and 3, where the Ar concentration is $q \approx 5\%$, behavior appreciably different from the calculations is observed. To determine the reasons for such behavior and to formulate a more detailed interpretation of the resulting spectral structure, additional (numerical-) theoretical and experimental investigations are needed. Time-resolved measurements of the argon line spectrum would be especially important here.

This work was carried out at the Iskra-5 laser thermonuclear facility (registration number 01-50) with financial support from the State Committee of the Russian Federation on Science and Technology, and from the Russian Fund for Fundamental Research (Grant No. 96-01-00046).

---

[*] E-mail: belkov@otd13.vniief.ru

[†] E-mail: kochemasov@otd13.vniief.ru

[1] With shells ≈ 4 μm and 7 μm thick.

[2] This is not that bad an assumption at high gas densities, when local thermodynamic equilibrium is established as a result of collision processes.

[1] N. G. Basov, Yu. A. Zakharov, A. A. Rupasov, G. V. Sklizkov, and A. S. Shikanov, *Diagnostics of Dense Plasma* [in Russian], Nauka, Moscow (1989).

[2] S. A. Bel'kov, A. V. Bessarab, G. G. Kochemasov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **51**, 1263 (1987).

[3] B. Yakobi, R. Epstein, F. J. Marshall *et al.*, Rev. Sci. Instrum. **66**, 728 (1995).

[4] C. J. Keane, B. A. Hammel, D. R. Kania *et al.*, Phys. Fluids B **5**, 3328 (1993).

[5] H. Nishimura, T. Kiso, H. Shirada *et al.*, Phys. Plasmas **2**, 1 (1995).

[6] S. A. Bel'kov and G. V. Dolgoleva, VANT, Ser. Matematicheskoe Modelirovanie Fizicheskikh Protsessov, No. 1, 59 (1992).

[7] F. M. Abzaev, V. I. Annenkov, V. G. Besuglov *et al.*, JETP Lett. **58**, 28 (1993).

[8] A. V. Veselov, V. S. Drozhin, V. M. Izgorodin *et al.*, Fusion Technol. **28**, 1838 (1995).

[9] Yu. V. Andramanova, A. V. Veselov, V. S. Drozhin *et al.* in *Report to Microspheres–Microcapsules and ICE Targets Technology Specialists Workshop*, Moscow, Russia (1997).

[10] Yu. V. Andramanova, A. A. Aushev, N. L. Zolotukhina *et al.* in *Report to Microspheres–Microcapsules and ICE Targets Technology Specialists Workshop*, Moscow, Russia, 1997.

[11] F. M. Abzaev, A. V. Bessarab, G. A. Kirillov *et al.*, VANT, Ser. Matematicheskoe Modelirovanie Fizicheskikh Protsessov, No. 4, 68 (1992).

[12] I. I. Sobelman, L. A. Vainshtein, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines*, Springer-Verlag, Berlin (1981) [Russ. original, Nauka, Moscow (1979)].

[13] G. Bekefi, C. Deutsche, and B. Yakobi in *Plasma in Lasers* [in Russian], Énergoatomizdat, Moscow (1982).

[14] B. Held, C. Deutsch, and M. M. Gombert, Phys. Rev. A **28**, 3134 (1983).

[15] R. J. Tighe and C. F. Hooper, Phys. Rev. A **15**, 1173 (1977).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Magnetic field generated in a plasma by a short, circularly polarized laser pulse

L. M. Gorbunov*) and R. R. Ramazashvili

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

(Submitted 27 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 849–863 (September 1998)

We study the generation of a quasistatic magnetic field by a short, circularly polarized laser pulse in a tenuous cold uniform plasma. It is shown that two physical mechanisms are responsible for the generation of the various components of the magnetic field. One mechanism is due to the ponderomotive forces and governs the generation of the azimuthal component of the magnetic field. The other is similar to the inverse Faraday effect (IFE) in a nonuniform plasma and gives rise to axial and radial components of the magnetic field. At moderate radiative intensities, all magnetic field components are proportional to the squared intensity. The spatial structure of the magnetic field depends strongly on the pulse shape and the plasma density. © *1998 American Institute of Physics.* [S1063-7761(98)00709-4]

1. INTRODUCTION

In the last few years, substantial progress has been made in the generation of subpicosecond laser pulses with energies of tens of Joules and radiative intensity $10^{18}-10^{20}$ W/cm^2. Such pulses propagating in plasma excite plasma oscillations, in whose electric field electrons can be accelerated to high energies (see review in Ref. 1). Together with a quasistatic electric field that results from charge separation in the plasma, a laser pulse generates quasistatic transverse currents and magnetic fields. This question has begun to attract attention recently, especially for circularly polarized laser pulses, where an inverse Faraday effect (IFE) is possible. In Ref. 2 it was found on the basis of perturbation theory with respect to the parameter $a=v_E/c<1$ (v_E and c are the average velocity of an electron in the laser radiation field and the speed of light) that in a cold uniform nondissipative plasma the IFE arises in the quadratic approximation. However, in Ref. 3, it was shown under the same restrictions as in Ref. 2 that irrespective of its polarization, a laser pulse generates a quasistatic magnetic field only in fourth order in the parameter a. In Refs. 4 and 5 the axial magnetic field generated in plasma by a laser pulse on account of the IFE was studied without using perturbation theory in the parameter a.

However, these results are mutually inconsistent. According to Ref. 4, in the limit $a<1$ the magnetic field is proportional to a^2, which corresponds to the result of Ref. 2. According to Ref. 5, for $a<1$ the magnetic field is proportional to a^4, which agrees with Ref. 3. The error that led the authors of Refs. 2 and 4 to overestimation of the IFE considerably is pointed out in Ref. 5. However, both in Refs. 2 and 4 and in Ref. 5 only the magnetic field associated with the IFE is discussed, while other physical mechanisms which under certain conditions lead to generation of a stronger magnetic field also exist.

In the present paper, using perturbation theory in the parameter a, we systematically study all components of the magnetic field that a short, circularly polarized laser pulse generates in a cold uniform nondissipative plasma. We show that quasistatic transverse currents and magnetic fields arise only in the fourth-order approximation in a. There are two types of currents and magnetic fields. One (azimuthal magnetic field, axial and radial current components) is generated by a laser pulse irrespective of the polarization of the pulse and exists not only inside the pulse but also in the wake field region behind it.[3] The other (azimuthal current, axial and radial magnetic field components) is generated only by a circularly polarized laser pulse (IFE). Such a magnetic field is localized in a longitudinal direction inside the pulse. The absolute values of all three components of the magnetic field depend on both the pulse size and the plasma density. The IFE can appear only for sufficiently long or short pulses propagating in a relatively dense plasma. We discuss these questions here for a pulse with a concrete shape.

2. BASIC RELATIONS

To study the propagation of a short laser pulse in plasma we employ Maxwell's equations for the electric **E** and magnetic **B** fields, together with the system of relativistic hydrodynamic equations for a cold electron fluid (see, for example, Ref. 2):

$$\operatorname{curl}\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \qquad (2.1)$$

$$\operatorname{curl}\mathbf{B}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}en\mathbf{v}, \qquad (2.2)$$

$$\operatorname{div}\mathbf{E}=4\pi e(n-n_0), \qquad (2.3)$$

$$\operatorname{div}\mathbf{B}=0, \qquad (2.4)$$

$$\frac{\partial\mathbf{p}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{p}=e\mathbf{E}+\frac{e}{c}\mathbf{v}\times\mathbf{B}, \qquad (2.5)$$

$$\frac{\partial n}{\partial t}+\operatorname{div}(n\mathbf{v})=0, \qquad (2.6)$$

where n, $\mathbf{v}$, and $\mathbf{p}$ are the density, velocity, and momentum of the electrons, and $\mathbf{p}=m\mathbf{v}\gamma$, where $\gamma=\sqrt{1+(\mathbf{p}/mc)^2}$ is the relativistic factor. The plasma ions are assumed to be immobile and singly charged, and their density is assumed to be uniform and equal to n_0.

From Eqs. (2.5) and (2.1) there follows an equation for $\mathbf{\Omega}=\text{curl }\mathbf{p}+(e/c)\mathbf{B}$, called the generalized vorticity,

$$\frac{\partial\mathbf{\Omega}}{\partial t}=\text{curl }\mathbf{v}\times\mathbf{\Omega}. \tag{2.7}$$

According to Eq. (2.7) the flux of generalized vorticity through an arbitrary surface bounded by a contour moving together with the fluid is constant (see, for example, Ref. 6). Therefore the quantity $\mathbf{\Omega}$ at a given point vanishes provided that it was vanished at the same point before the laser pulse arrived there, and therefore

$$\mathbf{B}=-\frac{c}{e}\text{ curl }\mathbf{p}. \tag{2.8}$$

Substituting (2.8) into Eq. (2.5) we find for the electric field intensity

$$\frac{e\mathbf{E}}{mc}=e\nabla\gamma+\frac{1}{mc}\frac{\partial\mathbf{p}}{\partial t}. \tag{2.9}$$

According to Eq. (2.3), the electron density is

$$\frac{n-n_0}{n_0}=\frac{c}{\omega_p^2}\nabla\left(\frac{\partial}{\partial t}\frac{\mathbf{p}}{mc}+c\nabla\sqrt{1+\left(\frac{p}{mc}\right)^2}\right), \tag{2.10}$$

where $\omega_p=\sqrt{4\pi n_0e^2/m}$ is the electron plasma frequency expressed in terms of the ion density.

Taking the time derivative of Eq. (2.9) and making use of (2.2), (2.8), and (2.10), we find an equation containing only the dimensionless electron momentum $\mathbf{q}=\mathbf{p}/mc$:[7]

$$\frac{\partial^2\mathbf{q}}{\partial t^2}+c^2\text{ curl curl }\mathbf{q}+\frac{\omega_p^2\mathbf{q}}{\sqrt{1+q^2}}=-c\frac{\partial}{\partial t}\nabla\sqrt{1+q^2}-\frac{c\mathbf{q}}{\sqrt{1+q^2}}\left(\text{div }\frac{\partial\mathbf{q}}{\partial t}+c\nabla^2\sqrt{1+q^2}\right). \tag{2.11}$$

In what follows, we are interested in the weak-nonlinearity approximation, where the velocity of the electrons is small compared to the velocity of light. This restriction is equivalent to the condition $|\mathbf{q}|<1$ and enables us to seek a power-series solution of Eq. (2.11):

$$\mathbf{q}=\sum_{n=1}^{\infty}\varepsilon^n\mathbf{q}_n,$$

where ε is a small parameter. Expanding the expressions in Eq. (2.11) in powers of ε and collecting terms with like powers of the small parameter, we obtain a coupled set of equations for $\mathbf{q}_n$.

In the first approximation $(n=1)$ we have, according to Eq. (2.11),

$$\frac{\partial^2\mathbf{q}_1}{\partial t^2}+c^2\text{ curl curl }\mathbf{q}_1+\omega_p^2\mathbf{q}_1=0. \tag{2.12}$$

According to the Helmholtz theorem,[8] the vector $\mathbf{q}_1$ can be split into two parts $(\mathbf{q}_1=\mathbf{q}_1^l+\mathbf{q}_1^{\text{tr}})$, such that $\text{div }\mathbf{q}^{\text{tr}}=0$ and $\text{curl }\mathbf{q}_1^l=0$. For the transverse part $\mathbf{q}_1^{\text{tr}}$, Eq. (2.12) has the form

$$\frac{\partial^2\mathbf{q}_1^{\text{tr}}}{\partial t^2}-c^2\nabla^2\mathbf{q}_1^{\text{tr}}+\omega_p^2\mathbf{q}_1^{\text{tr}}=0. \tag{2.13}$$

For the irrotational (potential) part $\mathbf{q}_1^l$, Eq. (2.12) assumes the form

$$\frac{\partial^2\mathbf{q}_1^l}{\partial t^2}+\omega_p^2\mathbf{q}_1^l=0. \tag{2.14}$$

Specifically, Eqs. (2.13) and (2.14) describe transverse and longitudinal waves, which in a linear approximation propagate in a uniform medium independently of one another.

A laser pulse propagating in a nonmagnetized plasma must be treated in the linear approximation as a transverse electromagnetic wave with a slowly varying amplitude. Therefore $\mathbf{q}_1$ is a solenoidal vector, so that $\text{div }\mathbf{q}_1=0$. To simplify the notation, in what follows we drop the symbol tr.

Since $\text{div }\mathbf{q}_1=0$, in the second approximation $(n=2)$ we find from Eq. (2.11)

$$\frac{\partial^2\mathbf{q}_2}{\partial t^2}+c^2\text{ curl curl }\mathbf{q}_2+\omega_p^2\mathbf{q}_2=-\frac{c}{2}\nabla\frac{\partial q_1^2}{\partial t}. \tag{2.15}$$

Since an irrotational vector stands on the right-hand side of Eq. (2.15), the solution of this equation satisfies the condition $\text{curl }\mathbf{q}_2=0$. According to Eq. (2.8), this means that the magnetic field $\mathbf{B}_2$ vanishes in the second approximation $(\mathbf{B}_2=0)$.

Thus, in a uniform, cold nondissipative plasma, a solenoidal electromagnetic field does not generate any magnetic field in the second approximation.[3] This result is at variance with the results of Refs. 2 and 4, where a quadratic inverse Faraday effect was obtained for circularly polarized laser pulses. The reason that the authors of Refs. 2 and 4 were led to incorrect results is indicated in Ref. 5 and consists in the fact that the condition $\text{div }\mathbf{q}_1=0$ was ignored.

A second-order quasistatic magnetic field is generated when factors that are of no consequence for short laser pulses propagating in a tenuous plasma are taken into account (collisions and thermal motion of electrons, plasma nonuniformity, irrotational high frequency field). Concerning this point, see, for example, the review articles Refs. 9 and 10.

When the condition $\text{curl }\mathbf{q}_2=0$ is taken into account, Eq. (2.5) becomes

$$\left(\frac{\partial^2}{\partial t^2}+\omega_p^2\right)\mathbf{q}_2=-\frac{c}{2}\frac{\partial}{\partial t}\nabla q_1^2. \tag{2.16}$$

In the third approximation $(n=3)$ we find from Eq. (2.11)

$$\frac{\partial^2\mathbf{q}_3}{\partial t^2}+c^2\text{ curl curl }\mathbf{q}_3+\omega_p^2\mathbf{q}_3=-c\nabla\frac{\partial}{\partial t}(\mathbf{q}_1\mathbf{q}_2)+\mathbf{q}_1\left(\frac{1}{2}(\omega_p^2-c^2\nabla^2)q_1^2-c\nabla\frac{\partial\mathbf{q}_2}{\partial t}\right), \tag{2.17}$$

where ∇^2 is the Laplacian operator. Equation (2.17) makes it possible to investigate the generation of the third harmonic (for short laser pulses this question has been discussed, for example, in Refs.[11]), as well as quadratic corrections in the dispersion relation for laser radiation.

Returning to the issue of generating a magnetic field, it is necessary to find the quantity div $\mathbf{q}_3$, the equation for which, according to Eq. (2.17), has the form

$$\left(\frac{\partial^2}{\partial t^2}+\omega_p^2\right)\mathrm{div}\,\mathbf{q}_3=\mathbf{q}_1\nabla\left(\frac{1}{2}(\omega_p^2-c^2\nabla^2)q_1^2-c\nabla\,\frac{\partial\mathbf{q}_2}{\partial t}\right)-c\,\frac{\partial}{\partial t}\nabla^2(\mathbf{q}_1\cdot\mathbf{q}_2). \tag{2.18}$$

There is no great difficulty in writing an equation for $\mathbf{q}_4$. However, we confine ourselves here only to an equation for curl $\mathbf{q}_4$, since this quantity (see Eq. (2.8)) is proportional to the fourth-order magnetic field ($\mathbf{B}_4=-(mc^2/e)\mathrm{curl}\,\mathbf{q}_4$):

$$\frac{\partial^2\mathbf{B}}{\partial t^2}-c^2\nabla^2\mathbf{B}+\omega_p^2\mathbf{B}=4\pi c\ \mathrm{curl}(\mathbf{j}_4^{(1)}+\mathbf{j}_4^{(2)}), \tag{2.19}$$

where we have dropped the subscript 4 in the magnetic field ($\mathbf{B}_4\equiv\mathbf{B}$) and we have written the fourth-order current as a sum of two terms. The first current term specifies the current in the direction of $\mathbf{q}_2$:

$$\mathbf{j}_4^{(1)}=-\frac{mc}{4\pi e}\,\mathbf{q}_2\left(\frac{1}{2}(\omega_p^2-c^2\nabla^2)q_1^2-c\,\frac{\partial}{\partial t}\nabla\cdot\mathbf{q}_2\right). \tag{2.20}$$

The direction of the current $\mathbf{j}_4^{(2)}$ is specified by $\mathbf{q}_1$:

$$\mathbf{j}_4^{(2)}=-\frac{mc}{4\pi e}\,\mathbf{q}_1\left((\omega_p^2-c^2\nabla^2)(\mathbf{q}_1\cdot\mathbf{q}_2)-c\,\frac{\partial}{\partial t}\nabla\cdot\mathbf{q}_3\right). \tag{2.21}$$

Bearing in mind the general expression for the current density

$$\mathbf{j}=en\mathbf{v}=ecn_0\mathbf{q}\,\frac{n/n_0}{\gamma}, \tag{2.22}$$

and introducing the notation $g=n/n_0\gamma$, we rewrite the fourth-order current one more time:

$$\mathbf{j}_4=ecn_0(\mathbf{q}_2g_2+\mathbf{q}_1g_3), \tag{2.23}$$

where g_2 and g_3 specify the perturbations of the ratio of the electron density to the relativistic factor γ in the second and third approximations,

$$g_2=\frac{1}{\omega_p^2}\left[c\,\frac{\partial}{\partial t}\,\mathrm{div}\,\mathbf{q}_2-\frac{1}{2}(\omega_p^2-c^2\nabla^2)q_1^2\right], \tag{2.24}$$

$$g_3=\frac{1}{\omega_p^2}\left[c\,\frac{\partial}{\partial t}\,\mathrm{div}\,\mathbf{q}_3-(\omega_p^2-c^2\nabla^2)(\mathbf{q}_1\cdot\mathbf{q}_2)\right]. \tag{2.25}$$

The expressions (2.20), (2.24), and (2.25) can in general form be put into a form that is more convenient for analysis. Since there are no perturbations at a given point in the plasma prior to the arrival of a laser pulse there ($t\rightarrow-\infty$), we write the solution of Eq. (2.16) in the form

$$\mathbf{q}_2(\mathbf{r},t)=-\frac{c}{2\omega_p^2}\nabla\,\frac{\partial\phi(\mathbf{r},t)}{\partial t}, \tag{2.26}$$

where

$$\phi=\omega_p\int_{-\infty}^{t}dt'\ \sin[\omega_p(t-t')]q_1^2(t',\mathbf{r}). \tag{2.27}$$

Substituting (2.26) into Eqs. (2.24) and (2.25), we find

$$g_2=-\frac{1}{2\omega_p^2}\left(\frac{\partial^2}{\partial t^2}+\omega_p^2-c^2\nabla^2\right)\phi, \tag{2.28}$$

$$\left(\frac{\partial^2}{\partial t^2}+\omega_p^2\right)g_3=-c\,\frac{\partial}{\partial t}(\mathbf{q}_1\cdot\nabla g_2)+\left(c^2\nabla^2-\frac{\partial^2}{\partial t^2}-\omega_p^2\right)(\mathbf{q}_1\cdot\mathbf{q}_2). \tag{2.29}$$

The current (2.20) can be written, using Eqs. (2.26) and (2.28), in the form[3]

$$\mathbf{j}_4^{(1)}=ecn_0\left(\frac{c}{2\omega_p^2}\nabla\,\frac{\partial\phi}{\partial t}\right)\frac{1}{2\omega_p^2}\left(\frac{\partial^2}{\partial t^2}+\omega_p^2-c^2\nabla^2\right)\phi. \tag{2.30}$$

3. QUASISTATIONARY NONLINEAR SOLENOIDAL CURRENT

For the quantity $\mathbf{q}_1(\mathbf{r},t)$, which characterizes a laser pulse in the linear approximation, we use the expression

$$\mathbf{q}_1(\mathbf{r},t)=\frac{1}{2}[\mathbf{a}(\mathbf{r},t)\exp(-i\omega t+ikz)+\mathrm{c.c.}], \tag{3.1}$$

where it is assumed that the pulse propagates along the z axis, ω and k are the frequency and wave number, which according to Eq. (2.13) satisfy the dispersion relation $k^2c^2=\omega^2-\omega_p^2$, and $\mathbf{a}$ is the complex amplitude (envelope), which is assumed to be slowly varying in both time (on a scale $2\pi/\omega$) and space (on a scale $2\pi/k$).

To a first approximation in the spatial and time derivatives, an equation for the envelope $\mathbf{a}$ follows from Eq. (2.13):

$$\frac{\partial\mathbf{a}}{\partial t}+v_g\frac{\partial\mathbf{a}}{\partial z}=0, \tag{3.2}$$

where $v_g=c^2k/\omega$ is the group velocity of the pulse. The solution of Eq. (3.2) is an arbitrary function of the variable $\xi=-z+v_gt$, representing an axial variable in a coordinate frame comoving with the pulse, and the variable $r_\perp$, which specifies the position in a plane perpendicular to the z axis. In other words, Eq. (3.2) describes the propagation of a pulse of constant shape.

In reality, both the shape and energy of the pulse vary, both on account of linear effects such as diffraction and dispersion, for which the second spatial and temporal derivatives of the slowly varying amplitude $\mathbf{a}$ are responsible, and nonlinear effects. If the laser pulse is short compared with the distance over which the shape and energy of the pulse vary (the so-called quasistatic approximation[1]), then the variation can be neglected when studying the nonlinear effect of the pulse on the plasma. For this reason, to calculate the magnetic field we assume that the characteristics of a pulse, such as its width, duration, and energy, are given (constant-pulse approximation).

Using Eq. (3.1), we find

$$q_1^2=\frac{1}{2}|\mathbf{a}|^2+\frac{1}{4}[\mathbf{a}^2\exp(-2i\omega t+2ikz)+\text{c.c.}], \tag{3.3}$$

where the first term varies slowly in both space and time, while the second term characterizes the second harmonic of the laser radiation.

Since we have in mind laser radiation that is both circularly and linearly polarized, we express $\mathbf{a}$ in the form

$$\mathbf{a}(\mathbf{r},t)=a_0(\mathbf{r},t)(\mathbf{e}_x+i\lambda\mathbf{e}_y)+\mathbf{e}_z a_z, \tag{3.4}$$

where $\mathbf{e}_{x,y,z}$ are unit vectors in a plane perpendicular to the direction of propagation of the pulse (xy) and in the direction of the propagation axis (z), a_0 is the scalar complex amplitude in the (x,y) plane, and a_z is the axial component of the envelope, which by virtue of $\operatorname{div}\mathbf{q}_1=0$ can be expressed in terms of a_0:

$$a_z\simeq\frac{1}{ik}\left(\frac{\partial a_0}{\partial x}+i\lambda\frac{\partial a_0}{\partial y}\right).$$

The quantity λ equals 0 for linearly polarized radiation and ± 1 for circularly polarized radiation (the sign specifies the rotation direction of the polarization vector).

Substituting (3.4) into Eq. (3.3), we take account of the fact that $|a_z|\ll|a_0|$. As a result, neglecting small terms proportional to the squared ratio of the wavelength λ_0 of the laser radiation to the transverse pulse size r_L, we obtain

$$q_1^2=\frac{1}{2}(1+\lambda^2)|a_0|^2+\frac{1}{4}(1-\lambda^2)\times[a_0^2\exp(-2i\omega t+2ikz)+\text{c.c.}]. \tag{3.5}$$

It follows from (3.5) that there is no second harmonic for circularly polarized laser radiation (in reality, it is small, of the order of $(\lambda_0/r_L)^2$).

We note that the radiative intensity is proportional to $\langle\mathbf{q}_1^2\rangle$, where the brackets denote averaging over time. We introduce the quantity

$$I=\frac{1}{2}(1+\lambda^2)|a_0|^2,$$

to characterize the dimensionless intensity. Obviously, for fixed amplitude $|a_0|$ the intensity for linearly polarized radiation is half that for circularly polarized radiation.

According to the definition (2.27) and Eq. (3.5), ϕ can be written in the form

$$\phi=\phi_0+\phi_2,$$

where

$$\phi_0=\omega_p\int_{-\infty}^{t}dt'\,\sin[\omega_p(t-t')]\,I(\mathbf{r},t'), \tag{3.6}$$

$$\phi_2=\frac{1-\lambda^2}{4}\,\omega_p\int_{-\infty}^{t}dt'\,\sin[\omega_p(t-t')][a_0^2(\mathbf{r},t')\times\exp(-2i\omega t'+2ikz)+\text{c.c.}]. \tag{3.7}$$

The integrand in Eq. (3.7) contains the product of a rapidly varying function $\exp(\pm 2i\omega t')$ and slowly varying functions. Integrating by parts, the function ϕ_2 can be represented as a series in powers of ω^{-1}. The first two terms of the series have the form

$$\phi_2=\frac{1-\lambda^2}{16}\,\frac{\omega_p^2}{\omega^2}\left\{-\left(a_0^2-\frac{i}{\omega}\frac{\partial a_0^2}{\partial t}\right)\times\exp(-2i\omega t+2ikz)+\text{c.c.}\right\}. \tag{3.8}$$

It is simplest to find the quasistationary part of the current (2.30), which can arise from terms quadratic in ϕ_0, and in linearly polarized radiation from terms quadratic in ϕ_2. In Ref. 3 it is shown that the contribution of the second harmonics to the quasistationary current is $(\omega_p/\omega)^2$ times smaller than that of the zeroth harmonics. Neglecting such small terms, the quasistationary part of the current (2.30) is[3]

$$\langle\mathbf{j}_4^{(1)}\rangle=\frac{en_0c^2}{4\omega_p^2}\left(\nabla\frac{\partial\phi_0}{\partial t}\right)\left(I-\frac{c^2}{\omega_p^2}\nabla^2\phi_0\right), \tag{3.9}$$

where the brackets indicate that the current varies slowly with time over the interval ω^{-1}.

For a linearly polarized laser pulse, the quasistationary part of the current (2.21) was studied in Ref. 3, where it is shown to be $(\omega_p/\omega)^2$ times smaller than the current (3.9). In the case of circular polarization, as will be shown below, the slowly varying part of the current (2.21) is much larger.

As follows from Eqs. (3.5) and (3.7), in the case of circular polarization the second harmonics of the quantities $\mathbf{q}_2$ and g_2 vanish, and Eq. (2.29) assumes the form

$$\left(\frac{\partial^2}{\partial t^2}+\omega_p^2\right)g_3=-c\frac{\partial}{\partial t}(\mathbf{q}_1\cdot\nabla g_{20})+\left(c^2\nabla^2-\frac{\partial^2}{\partial t^2}-\omega_p^2\right)(\mathbf{q}_1\cdot\mathbf{q}_{20}), \tag{3.10}$$

where g_{20} and $\mathbf{q}_{20}$ are the zeroth harmonics of the corresponding quantities.

Neglecting the derivatives of these quantities and taking account of (2.13), Eq. (3.10) assumes the form

$$\left(\frac{\partial^2}{\partial t^2}+\omega_p^2\right)g_3=-c(\nabla g_{20})\frac{\partial\mathbf{q}_1}{\partial t}. \tag{3.11}$$

Bearing in mind that $\omega\gg\omega_p$, we neglect the second term on the left-hand side of Eq. (3.11) and write it in the form

$$\frac{\partial g_3}{\partial t}=-c\mathbf{q}_1\cdot\nabla g_{20}. \tag{3.12}$$

Actually, Eq. (3.12) reflects the fact that the high-frequency third-order perturbations of the electron density (g_3) result from the high-frequency motion of electrons with velocity $c\mathbf{q}_1$ in a nonuniform plasma. In addition, the nonuniformity of the plasma is produced by the laser pulse and depends quadratically on the radiation amplitude (g_{20}).

Solving Eq. (3.12), we find for the quasistatic part of the second term in Eq. (2.23), corresponding to the current (2.21),

$$\langle \mathbf{j}_4^{(2)}\rangle = \frac{ien_0c^2}{8\omega}\left[\nabla\left(\frac{|a_0|^2}{2}-\frac{c^2}{\omega_p^2}\nabla^2\phi_0\right)\mathbf{a}^*\times\mathbf{a}\right]. \tag{3.13}$$

Obviously, the current (3.13) is nonzero only for a circularly polarized wave, where $\mathbf{a}^*\times\mathbf{a}\neq 0$.

4. QUASISTATIC MAGNETIC FIELD OF A CIRCULARLY POLARIZED LASER PULSE

Consider an axisymmetric laser pulse, for which the quantity I depends only on the variables $\xi=ct-z$ and r. For definiteness, we assume that

$$I=I_0f_1(\xi)f_2(r), \tag{4.1}$$

where I_0 characterizes the maximum dimensionless radiative intensity, while the functions f_1 and f_2 specify the variation in the intensity in the longitudinal and transverse directions, respectively. In the accordance with the definition (4.1), (3.6) then takes the form

$$\phi_0=2f_2(r)I_0\varphi(\xi), \tag{4.2}$$

where

$$\varphi(\xi)=\frac{k_p}{2}\int_{-\infty}^{\xi}d\xi'\,\sin[k_p(\xi-\xi')]\,f_1(\xi'), \tag{4.3}$$

in which $k_p=\omega_p/c$ is the characteristic wave number. We note that (4.3) is nonvanishing not only in the vicinity of the laser pulse, but also behind it, where it describes the wake field.[12]

For circularly polarized radiation, all three components of the solenoidal quasistationary current are present. Two of them, specified by Eq. (3.9), are unrelated to the polarization of the radiation, and using (4.2) they can be represented in the form

$$\langle j_z\rangle=-ecn_0\frac{\partial^2\varphi(\eta)}{\partial\eta^2}f_2(\rho)\varphi(\eta)[(1-\nabla_\rho^2)f_2(\rho)]I_0^2, \tag{4.4}$$

$$\langle j_r\rangle=ecn_0\frac{\partial^2\varphi(\eta)}{\partial\eta}\frac{\partial f_2}{\partial\rho}\varphi(\eta)[(1-\nabla_\rho^2)f_2]I_0^2, \tag{4.5}$$

where the indices 4 and 1 in the current have been dropped and the dimensionless longitudinal and transverse coordinates $\eta=k_p\xi$ and $\rho=k_pr$ have been introduced; ∇_ρ^2 is the transverse part of the Laplacian.

Making the same changes in notation and coordinates, we transform the azimuthal angle (3.13), using Eqs. (4.1) and (4.2), to the form

$$\langle j_\varphi\rangle=ecn_0\frac{\omega_p}{2\omega}I_0^2\lambda f_1(\eta)\varphi(\eta)f_2(\rho)\frac{\partial}{\partial\rho}[(1-\nabla_\rho^2)f_2]. \tag{4.6}$$

By comparison to the other components, the current (4.6) contains an extra small factor $\omega_p/2\omega$, and fails to vanish only in the vicinity of the pulse, where the functions $f_1(\eta)$ and $f_2(\rho)$ are also nonvanishing.

The axial and radial current components (4.4) and (4.5) determine the azimuthal magnetic field of the laser pulse, the equation for which, according to Eq. (2.19), can be reduced to the form

$$\left(\nabla_\rho^2-\frac{1}{\rho^2}-1\right)\mathscr{B}_\varphi=I_0^2\mathscr{F}_\varphi(\rho,\eta), \tag{4.7}$$

where

$$\mathscr{B}=e\mathbf{B}/mc\omega_p,$$

$$\mathscr{F}_\varphi=\left(\frac{d\varphi}{d\eta}\right)^2\frac{df_2}{d\rho}[(1-\nabla_\rho^2)f_2]-\varphi\frac{d^2\varphi}{d\eta^2}f_2\frac{d}{d\rho}[(1-\nabla_\rho^2)f_2]. \tag{4.8}$$

The azimuthal current (4.6) is responsible for the generation of the axial and radial magnetic field components, which are governed by equations similar to (4.7). We write the solutions of the equations for all three magnetic field components, satisfying the boundary conditions

$$\mathscr{B}(\rho\to\infty)=0,\quad \mathscr{B}_\varphi(0)=\mathscr{B}_\rho(0)=\frac{\partial\mathscr{B}_z(0)}{\partial\rho}=0,$$

in the form

$$\mathscr{B}_\varphi=I_0^2\left\{\left(\frac{d\varphi}{d\eta}\right)^2G_1(\rho)-\left(\varphi\frac{d^2\varphi}{d\eta^2}\right)G_2(\rho)\right\}, \tag{4.9}$$

$$\mathscr{B}_\rho=I_0^2\frac{d}{d\eta}(f_1\varphi)G_3(\rho), \tag{4.10}$$

$$\mathscr{B}_z=I_0^2f_1\varphi G_4(\rho), \tag{4.11}$$

where

$$G_3(\rho)=-\lambda(\omega_p/2\omega)G_2(\rho),$$

$$G_1(\rho)=I_1(\rho)\int_\rho^\infty dx\,xK_1(x)\frac{df_2}{dx}[(1-\nabla_x^2)f_2]+K_1(\rho)\int_0^\rho dx\,xI_1(x)\frac{df_2}{dx}[(1-\nabla_x^2)f_2], \tag{4.12}$$

$$G_2(\rho)=I_1(\rho)\int_\rho^\infty dx\,xK_1(x)f_2(x)\frac{d}{dx}[(1-\nabla_x^2)f_2]+K_1(\rho)\int_0^\rho dx\,xI_1(x)f_2(x)\frac{d}{dx}[(1-\nabla_x^2)f_2], \tag{4.13}$$

$$G_4(\rho)=-\lambda\frac{\omega_p}{2\omega}\left\{I_0(\rho)\int_\rho^\infty dx\,K_0(x)\frac{d}{dx}\left\{xf_2\frac{d}{dx}\times[(1-\nabla_x^2)f_2]\right\}+K_0(\rho)\int_0^\rho dx\,I_0(x)\frac{d}{dx}\times\left\{xf_2\frac{d}{dx}[(1-\nabla_x^2)f_2]\right\}\right\}, \tag{4.14}$$

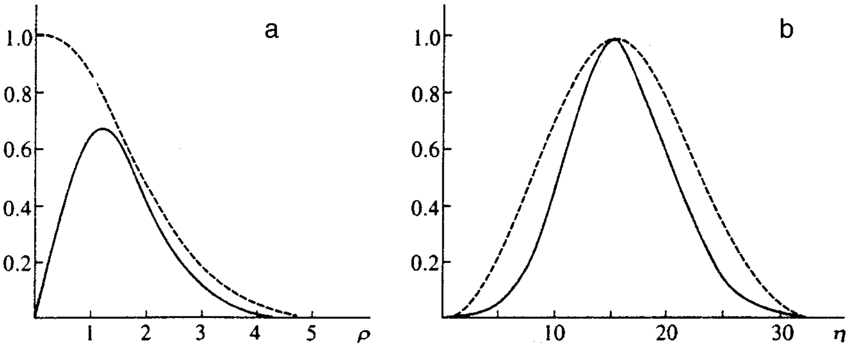


FIG. 1. Dimensionless azimuthal current density versus the transverse coordinate ρ (a) and longitudinal coordinate η (b) for a wide ($\alpha^2=0.2$) and long ($\kappa_L=0.2$) pulse. The dashed line shows the shape of the laser pulse in the corresponding direction.

and $I_{0,1}$ and $K_{0,1}$ are modified Bessel functions of the first and second kind, respectively.

To obtain actual expressions for the components of the magnetic field, it is necessary to know the explicit form of the functions $f_1(\eta)$ and $f_2(\rho)$. Just as in Ref. 3, we take them in the form

$$f_1(\eta)=\sin^2\left(\frac{\pi\eta}{L}\right),\quad f_2(\rho)=\exp(-\alpha^2\rho^2),\tag{4.15}$$

where $0<\eta<L$, L is the pulse length in units of k_p^{-1}, $\alpha^2=2(k_p r_L)^{-2}$, and r_L is the effective pulse width.

The expressions for the functions $G(\rho)$ can be represented in a relatively simple form in two limiting cases: $\alpha\gg1$ (short pulse) and $\alpha\ll1$ (wide pulse). We consider below a wide pulse, for which

$$\begin{gathered}G_1\simeq G_2\simeq-2\alpha^2\rho\exp(-2\alpha^2\rho^2),\\ G_3\simeq\lambda\frac{\omega_p}{2\omega}2\alpha^2\rho\exp(-2\alpha^2\rho^2),\\ G_4\simeq\lambda\frac{\omega_p}{2\omega}4\alpha^2(1-2\alpha^2\rho^2)\exp(-2\alpha^2\rho^2).\end{gathered}\tag{4.16}$$

The dependence of the magnetic field component on the longitudinal coordinate η is given by (4.3), which according to Eq. (4.15) assumes the form

$$\varphi(\eta)=\frac{1}{2(\kappa_L^2-1)}\left[\kappa_L^2\sin^2(\eta/2)-\sin^2\left(\frac{\kappa_L\eta}{2}\right)\right],\tag{4.17}$$

where $\kappa_L=2\pi/L$. We confine our attention to long pulses ($\kappa_L^2<1$), for which $\varphi(\eta)\simeq(1/2)f_1(\eta)$.

Thus, in the vicinity of a circularly polarized laser pulse whose width and length are greater than c/ω_p, the components of the quasistatic magnetic field are

$$\mathscr{B}_\varphi=-2I_0^2\alpha^2\kappa_L^2\rho\exp(-2\alpha^2\rho^2)\sin^2(\kappa_L\eta),\tag{4.18}$$

$$\mathscr{B}_\rho=I_0^2\lambda\frac{\omega_p}{2\omega}\alpha^2\kappa_L\rho\exp(-2\alpha^2\rho^2)\sin(\kappa_L\eta)\sin^2\left(\frac{\kappa_L\eta}{2}\right),\tag{4.19}$$

$$\mathscr{B}_z=I_0^2\lambda\frac{\omega_p}{\omega}\alpha^2(1-2\alpha^2\rho^2)\exp(-2\alpha^2\rho^2)\sin^4\left(\frac{\kappa_L\eta}{2}\right).\tag{4.20}$$

As follows from Eqs. (4.18)–(4.20), the azimuthal and radial components of the magnetic field vanish on the pulse axis $(\rho=0)$, while the axial component does not. Comparing the maxima of the various components of the magnetic field, it follows from Eqs. (4.18)–(4.20) that that the effects due to the circular polarization of the radiation become dominant only for sufficiently long pulses, such that $\kappa_L^2<(\omega_p/\omega)\alpha$.

We note that the azimuthal current (4.6) is responsible for the generation of the radial and axial components of the magnetic field. For a sufficiently wide and long laser pulse, such as the one considered here, whose shape is given by (4.15), this current has the form

$$-\langle j_\varphi\rangle\left[\lambda ecn_0I_0^2\alpha^2\frac{\omega_p}{2\omega}\right]^{-1}\simeq\sin^4\left(\frac{\kappa_L\eta}{2}\right)\rho\exp(-2\alpha^2\rho^2).\tag{4.21}$$

Figure 1 shows the dimensionless azimuthal current (4.21) as a function of the longitudinal and transverse coordinates η and ρ for $\alpha^2=0.2$ and $\kappa_L=0.2$. The current reaches its maximum value at $\eta_0=\pi/\kappa_L$ and $\rho_0=1/2\alpha$.

Figure 2 shows, for the same parameters α^2 and κ_L, the flux lines of the magnetic field generated by the current (4.21), which according to Eqs. (4.18) and (4.19) are given by

$$\rho^2\exp(-2\alpha^2\rho^2)\sin^4\left(\frac{\eta\kappa_L}{2}\right)=C,\tag{4.22}$$

where C is an arbitrary constant that ranges from 0 to $(2\alpha^2e)^{-1}$. The direction of the flux lines is determined by the rotation sense of the polarization vector.

We note that for $\rho=0$, (4.20) is identical to the expression obtained in Ref. 5, if in the latter the limit $|a_0|<1$ is taken.

5. CONCLUSIONS

In summary, when a short, circularly polarized laser pulse propagates in a cold, uniform, nondissipative plasma,

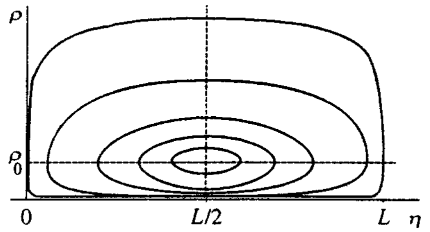


FIG. 2. Flux lines of the magnetic field resulting from the circular polarization of the radiation for $\alpha^2=0.2$ and $\kappa_L=0.2$.

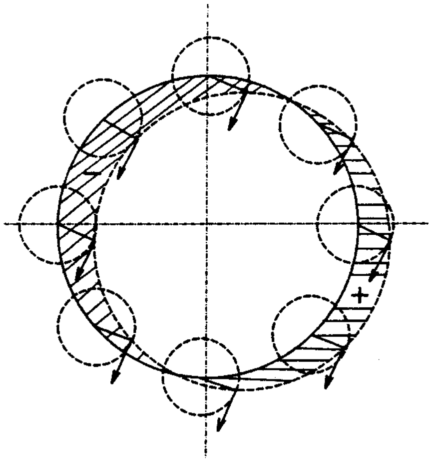


FIG. 3. Generation of a quasistationary azimuthal current by circularly polarized radiation in a radially nonuniform plasma.

two mechanisms give rise to a quasistatic magnetic field (as noted in Ref. 13). One is not associated with the polarization of the radiation, and is governed by ponderomotive forces. This mechanism generates an azimuthal magnetic field, which emerges not only within the pulse but also in the wake field region behind it.[3] The second mechanism is associated with the circular polarization of the radiation. It is responsible for the generation of an azimuthal quasistationary current, and both axial and radial magnetic field components. In this case, it can be interpreted as the inverse Faraday effect.

Figure 3 illustrates the physical mechanism that gives rise to an azimuthal quasistatic current in a radially nonuniform plasma. The solid curve is a curve of constant electron density in the absence of the pulse. The small dashed circles represent the trajectories of electrons in the field of a circularly polarized wave propagating transverse to the gradient of the nonuniformity (perpendicular to the plane of the figure). The dots show the instantaneous positions and the arrows the velocities of the electrons. The dashed curve connecting the electrons is a curve of constant electron density at a certain moment in time. If the plasma density increases with increasing radius, the electron density in the hatched regions with a plus or minus sign is either less than or greater than the density in the absence of the radiation. From Fig. 3, it is then clear that a clockwise component of the instantaneous electron velocity is present in the low-density region, and a counterclockwise component in the high-density region. The instantaneous azimuthal electron current, which is given by the product of the rapidly varying velocity and the rapidly varying perturbation of the electron density, has the same sign everywhere along the circle. This means that the current has a constant, time-independent azimuthal component.

In the example considered, the initial plasma electron density unperturbed by the pulse was assumed to be nonuniform. For this reason, the quasistationary current given by the product of two rapidly varying quantities is proportional to the squared amplitude of the high-frequency radiation. However, in the problem considered above, the initial plasma density was uniform. The slowly time-varying redistribution of the electrons and therefore the nonuniformity of the plasma were produced by the ponderomotive forces. The radial density gradient itself was therefore proportional to the squared amplitude of the laser radiation, and the slow azimuthal current was proportional to the fourth power of the amplitude.

As noted in Ref. 3, the azimuthal magnetic field that emerges not only inside the pulse but also in the wake field region can influence the focusing of electrons accelerated in the wake wave. However, the IFE can appear only inside the laser pulse, where excitation of the wake wave occurs. It stands to reason that for $|a_0|\sim 1$ the electron cyclotron frequency will become comparable to the plasma frequency, and this will be reflected in the excitation of a wake wave. However, this question lies beyond the scope of the approximation $|a_0|<1$ considered here.

For $|a_0|>1$, numerical results can be used as a guide. Such results are presented in Ref. 3 for the axial magnetic field component. It is shown there that for $|a_0|=2$, $\alpha^2=5\times 10^{-2}$, and $L=10$ the maximum dimensionless magnetic field $eB_\varphi/mc\omega_p$ is 0.1. For plasma with electron density 10^{18} cm^{-3} ($\omega_p=5.64\times 10^{13}$ s^{-1}), these parameters correspond to a focal spot radius 30 μm, pulse duration 175 fs, and pulse energy about 5 J. The maximum magnetic field B_φ reaches 350 kG. In order for the other magnetic field components B_r and B_z induced by the inverse Faraday effect to become comparable to B_φ, the duration of the laser pulse must be increased to 1.4 ps, and accordingly the pulse energy must be increased to 42 J. These parameters are all attainable with existing technology.

This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 97-02-16537).

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

---

[1] E. Esarey, P. Sprangle, J. Krall, and A. Ting, IEEE Trans. Plasma Sci. **24**, 252 (1996).
[2] V. Yu. Bychenkov, V. I. Demin, and V. T. Tikhonchuk, Zh. Éksp. Teor. Fiz. **105**, 118 (1994) [JETP **78**, 62 (1994)].
[3] L. M. Gorbunov, P. Mora, and T. M. Antonsen, Jr., Phys. Rev. Lett. **76**, 2495 (1996); Phys. Plasmas **4**(12), 4358 (1997).
[4] V. Ya. Bychenkov and V. T. Tikhonchuk, Laser Part. Beams **14**, 52 (1996).
[5] V. I. Berezhiani, S. M. Mahajan, and N. L. Shatashvili, Phys. Rev. E **55**, 995 (1997).
[6] N. E. Kochin, I. A. Kibel', and N. V. Roze, *Theoretical Hydromechanics* [in Russian], Fizmatgiz, Moscow (1963).
[7] V. I. Berezhiani and I. G. Murusidze, Phys. Lett. **148**, 338 (1990).
[8] N. E. Kochin, *Vector Calculus and Elements of Tensor Calculus* [in Russian], ONTI GTTN, Moscow (1934).
[9] I. V. Sokolov, Usp. Fiz. Nauk **161**, 175 (1991) [Sov. Phys. Usp. **34**, 925 (1991)].
[10] Yu. M. Aliev, V. Yu. Bychenkov, M. S. Javonovic, and A. A. Frolov, J. Plasma Phys. **48**, 167 (1992).
[11] P. Sprangle, E. Esarey, and A. Ting, Phys. Rev. Lett. **64**, 2011 (1990); J. M. Rax and N. J. Fisch, Phys. Rev. Lett. **69**, 772 (1992); W. B. Mori, C. D. Decker, and W. P. Leemans, IEEE Trans. Plasma Sci. **21**, 1 (1993); E. Esarey *et al.*, IEEE Trans. Plasma Sci. **21**, 95 (1993); B. Sken, W. Yu, P. Sprangle, G. Zeng, and Z. Xu, Phys. Plasmas **2**, 4631 (1995).
[12] L. M. Gorbunov and V. I. Kirsanov, Zh. Éksp. Teor. Fiz. **93**, 509 (1987) [Sov. Phys. JETP **66**, 290 (1987)].
[13] T. Lehner, Phys. Scr. **49**, 704 (1994).

Translated by M. E. Alferieff

On coherent harmonic bremsstrahlung in laser plasma

V. P. Silin*)

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

(Submitted 27 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 864–880 (September 1998)

We give a theoretical description of the mechanisms underlying the generation of harmonics of plasma-warming radiation, which depend on the polarization of the latter. We show that there is a striking anomaly in the polarization of the harmonics in the presence of a low degree of circular polarization of the pump, and that harmonic generation efficiency increases. Variations in the polarization of the pump accompanying inverse bremsstrahlung absorption are discussed. © *1998 American Institute of Physics.* [S1063-7761(98)00809-9]

1. Interest in the generation of higher-order harmonics of laser radiation has increased appreciably in connection with the transition to femtosecond pulses, which has made it easier to obtain strong laser fields.[1–3] Such generation opens up the possibility of producing compact sources of hard coherent ultraviolet and x-ray emission.[4,5] In this regard, interest in the generation of higher-order harmonics in plasma, predicted in 1964 in Ref. 6 (see also Refs. 7–9), is natural. This phenomenon is due to the bremsstrahlung of electrons, coherently oscillating in a coherent plasma-warming electromagnetic field, in the process of scattering of electrons by the Coulomb field of the ions.

In the present paper, we discuss conditions produced by the polarization of the warming radiation, under which, first, harmonic generation is more efficient than discussed heretofore, and second, new and unique phenomena arise. The essence of these phenomena lies in the fact that the coherent current generating the harmonics in accordance with the standard bremsstrahlung law[6,8] is proportional to the velocity vector and inversely proportional to the cubed modulus of the velocity. For circular polarization, the modulus of the velocity of the electron oscillations does not depend on time. Therefore, harmonic generation is suppressed. For elliptically polarized warming pump radiation, unique mechanisms arise. If the degree of circular polarization is low, then at a certain moment in time the modulus of the velocity of the electron oscillations will be very small. At approximately the same moment the projection of the electron velocity on the direction of approximately linear polarization will be small. Conversely, the perpendicular projection of the velocity is not so small. The phenomenon, discussed below, of anomalous polarization of harmonics in a plane almost perpendicular to the polarization plane of the pump is connected with this. The anomalous increase, discussed below, of the harmonic generation efficiency is also connected with this. The mechanisms found are similar to the polarization dependence of the static conductivity of plasma in a powerful radiation field.[10] We note that for exact plane polarization of the pump the modulus of the velocity, just like the velocity vector of the oscillations itself, vanishes at the same moment in time. A consistent description could previously only be obtained by taking account of the thermal motion.[6] This produced an effect of the order of the logarithm of the ratio of the large amplitude of the velocity of the oscillations of an electron to its thermal velocity. This way of taking account of the thermal motion and analysis of its competition with the polarization of the radiation made it possible to construct in the present paper a systematic theory of the increase in the harmonic generation efficiency and to describe the polarization properties of the harmonics. In connection with the importance of the influence of the polarization of the warming radiation, a theory of the nonlinear variation of its polarization as a result of inverse bremsstrahlung absorption is given.

2. We study plasma in the field of elliptically polarized radiation $\mathbf{E}=(E_x,E_y,0)$, where

$$E_x=e_xE\cos(\omega t-\varphi_x),\quad E_y=-e_yE\sin(\omega t-\varphi_x). \tag{2.1}$$

Here E is the real electric field amplitude, the e_α $(\alpha=x,y)$ characterize the polarization of the plasma-warming radiation (2.1), $e_x^2+e_y^2=1$, and for simplicity it is assumed that $e_x\geqslant e_y\geqslant 0$. The polarization tensor[11] of such radiation has the form

$$R_{\alpha\beta}=\begin{vmatrix} e_x^2 & ie_xe_y \\ -ie_xe_y & e_y^2 \end{vmatrix}\equiv\frac{1}{2}\begin{pmatrix} 1+\xi_3 & \xi_1-i\xi_2 \\ \xi_1+i\xi_2 & 1-\xi_3 \end{pmatrix}. \tag{2.2}$$

This obviously means that the corresponding Stokes parameters are $\xi_1=0$, $\xi_2=-2e_xe_y$, and $\xi_3=e_z^2-e_y^2$. For what follows, it is helpful to use the degree of circular polarization $A\equiv\xi_2$ as well as the degree of maximum linear polarization $\rho^2\equiv L=\sqrt{\xi_3^2}=\sqrt{1-A^2}$.[11]

In the electric field (2.1) an electron oscillates with velocity $\mathbf{u}_E=(u_{Ex},u_{Ey},0)$, where

$$u_{Ex}=-v_Ee_x\sin(\omega t-\varphi_x),\quad u_{Ey}=-v_Ee_y\cos(\omega t-\varphi_x). \tag{2.3}$$

Here

$$v_E=(|e|E/m\omega) \tag{2.4}$$

characterizes the amplitude of the velocity oscillations, and e and m are the electron charge and mass. We say that the field is weak when the velocity (2.3) is small compared with the thermal velocity $v_T=\sqrt{\kappa_B T/m}$ of the electrons. Conversely, we shall say that the radiation field is strong if

$$v_E \gg v_T. \tag{2.5}$$

We assume that the frequency of the plasma-warming radiation is much higher than the effective electron–ion collision frequency. For a weak field the latter frequency can be expressed in the conventional manner (compare, for example, Ref. 12), which yields

$$\nu_{ei}=\frac{4\sqrt{2\pi}\,Ze^4 n_e\Lambda}{3m^2 v_T^3}. \tag{2.6}$$

Here

$$Z=\sum_i (e_i^2 n_i/e^2 n_e),$$

where the summation extends over all types of ions, e_i is the ion charge, n_e and n_i are the electron and ion number densities, and Λ is the Coulomb logarithm. In the opposite, strong-field limit we have for the effective electron–ion collision frequency[6–8,10]

$$\nu(E)=\frac{8\pi\sqrt{2}\,Ze^4 n_e\Lambda}{m^2 v_E^3}. \tag{2.7}$$

This collision frequency is much lower than the frequency (2.6) on account of the inequality (2.5). We emphasize that in the case (2.7) the Coulomb logarithm by no means coincides with the corresponding expression in the weak-field limit (see below).

In contrast to Ref. 6, we employ the simplified approach of Ref. 8. Following the latter, we use the Fokker–Planck–Landau form of the collision integral to describe electron–ion collisions:

$$J_{ei}[f]=\frac{2\pi Ze^4 n_e\Lambda}{m^2}\frac{\partial}{\partial v_r}\left(\frac{v^2\delta_{rs}-v_r v_s}{v^3}\frac{\partial f}{\partial v_s}\right), \tag{2.8}$$

where $f(\mathbf{v})$ is the electron distribution function. In Eq. (2.8), small terms of the order of the ratio of the electron and ion masses are neglected. In using this approach below, we must introduce the dependence of the Coulomb logarithm on the electric field heuristically. At the same time, an indisputable positive aspect of this approach is the comparative simplicity and obviousness of the results obtained.

The interaction of the radiation with plasma in the dipole approximation corresponds to neglect of the spatial dependence of the field and the distribution function. Accordingly, the Boltzmann equation can be written in the form

$$\frac{\partial f}{\partial t}+\frac{e}{m}\mathbf{E}(t)\cdot\frac{\partial f}{\partial \mathbf{v}}=J_{ee}+J_{ei}[f]. \tag{2.9}$$

Here J_{ee} is the electron–electron collision integral, the actual form of which will not be required.

The assumption that the frequency of the plasma-warming radiation is high makes it possible to neglect in the zeroth approximation the right-hand side of Eq. (2.9). The zeroth-approximation equation

$$\frac{\partial f_0}{\partial t}+\frac{e}{m}\mathbf{E}\cdot\frac{\partial f_0}{\partial \mathbf{v}}=0 \tag{2.10}$$

has the solution

$$f_0(\mathbf{v},t)=F(\mathbf{v}-\mathbf{u}_E(t)). \tag{2.11}$$

We assume below that $F(\mathbf{u})=f_M(u)$, where $f_M(u)$ is a Maxwellian distribution with temperature T. The following expression for the electric current density in the zeroth-approximation follows from Eq. (2.11):

$$\mathbf{j}_0=en_e\mathbf{u}_E(t). \tag{2.12}$$

We multiply the kinetic equation

$$\frac{\partial \delta f}{\partial t}+\frac{e}{m}\mathbf{E}\cdot\frac{\partial \delta f}{\partial \mathbf{v}}=J_{ee}+J_{ei}[f_0(\mathbf{v},t)] \tag{2.13}$$

for the correction to the electron distribution function by the electron charge and velocity vector $\mathbf{v}$ and integrate over velocity space. Then, since $\int d\mathbf{v}\,\delta f=0$, we obtain for the perturbation $\delta\mathbf{j}$ of the electric-current density

$$\frac{\partial \delta\mathbf{j}}{\partial t}=-\frac{4\pi Ze^4 n_e\Lambda}{m^2}\int d\mathbf{v}\,\frac{e\mathbf{v}}{v^3}F(\mathbf{v}-\mathbf{u}_E(t)). \tag{2.14}$$

The right-hand side of Eq. (2.14) is due to electron–ion collisions. The electron–electron collision integral does not contribute because the collisions conserve momentum. The remainder of the present paper is given over to analysis of the consequences of Eq. (2.14).

3. To describe the generation of the harmonics of the plasma-warming radiation, (2.14) must be represented as a Fourier series expansion. To do so, we apply first the relation

$$\frac{\mathbf{v}}{v^3}=-\int\frac{d\mathbf{q}}{(2\pi)^3}\frac{4\pi i\mathbf{q}}{q^2}\exp(i\mathbf{q}\cdot\mathbf{v}). \tag{3.1}$$

Using a Maxwellian distribution for $F(\mathbf{u})$ this formula makes it possible to represent Eq. (2.14) in the form

$$\frac{\partial \delta\mathbf{j}}{\partial t}=\frac{Ze^4 n_e\Lambda}{2\pi^2 m^2}en_e\int d\mathbf{q}\,\frac{i\mathbf{q}}{q^2}\exp\left[-\frac{1}{2}v_T^2 q^2+i\mathbf{q}\cdot\mathbf{u}_E(t)\right]. \tag{3.2}$$

Using spherical coordinates $\mathbf{q}=(q,\theta,\varphi)$ in $\mathbf{q}$-space we have

$$\begin{aligned}\mathbf{q}\cdot\mathbf{u}_E=-qv_E\sin\theta[&\delta_+\sin(\omega t-\varphi_x+\varphi)\\&+\delta_+\sin(\omega t-\varphi_x-\varphi)],\end{aligned} \tag{3.3}$$

where $\delta_\pm=(1/2)(e_x\pm e_y)$. In accordance with Ref. 13 (p. 987, Eq. 8.511.3), this form makes it possible to write

$$\begin{aligned}\exp(i\mathbf{q}\cdot\mathbf{u}_E)=\sum_{k,l=-\infty}^{+\infty}&J_l(qv_E\delta_+\sin\theta)J_k(qv_E\delta_-\sin\theta)\\&\times\exp\{i(k+l)(\omega t-\varphi_x)+i(l-k)\varphi\},\end{aligned} \tag{3.4}$$

where the $J_l(z)$ are Bessel functions. Hence it is obvious that Eq. (3.2) can be expressed in the form of a Fourier series. The required transformations can be found in Appendix 1. The result is

$$\frac{\partial \delta j_x}{\partial t}=\sum_{l=0}^{\infty}\sigma_{xx}^{(2l+1)}\frac{\partial}{\partial t}\,e_xE\,\cos[(2l+1)(\omega t-\varphi_x)], \tag{3.5}$$

$$\frac{\partial \delta j_y}{\partial t}=\sum_{l=0}^{\infty}\sigma_{yy}^{(2l+1)}\frac{\partial}{\partial t}\,(-e_yE)\sin[(2l+1)(\omega t-\varphi_x)], \tag{3.6}$$

$$\sigma_{xx}^{(2l+1)}=\frac{\omega_{Le}^2\nu(E,l)}{4\pi(2l+1)\omega^2}\left[A_l\left(\rho^2,\,\frac{v_E}{2v_T}\right)-A_{l+1}\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right], \tag{3.7}$$

$$\sigma_{yy}^{(2l+1)}=\frac{\omega_{Le}^2\nu(E,l)}{4\pi(2l+1)\omega^2}\left[A_l\left(\rho^2,\,\frac{v_E}{2v_T}\right)+A_{l+1}\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right]. \tag{3.8}$$

Here $\omega_{Le}=\sqrt{4\pi e^2 n_e/m}$ is the electron Langmuir frequency. According to Eq. (A1.8), the functions $A_l(\rho^2,N)$ are given by

$$A_l(\rho^2,N)=\frac{2}{\sqrt{\pi}}\int_0^{N^2}dz\sqrt{z}\;e^{-z}I_l(\rho^2 z). \tag{3.9}$$

The latter formula, as we shall see below, is very convenient for the required analysis.

We must now discuss the collision frequency $\nu(E,l)$, which for

$$(2l+1)\omega>\omega_{Le} \tag{3.10}$$

can depend on the number of the harmonic on account of the Coulomb logarithm. The Coulomb logarithm in the Landau logarithmic approximation is determined by the ratio of the maximum and minimum impact parameters $\Lambda=\ln(r_{\max}/r_{\min})$, which bound the region where the contribution of collisions is considerable (see Ref. 12). The maximum impact parameter is determined by the ratio of the electron velocity to the characteristic frequency. For a weak low-frequency field this is the ratio of the thermal velocity to the Langmuir frequency, equal to the Debye screening radius of the Coulomb field. For a weak field, but such high frequencies that $\omega>\omega_{Le}$, as was established in Refs. 14–16, the ratio of the thermal velocity to the frequency arises for the maximum impact parameter. In our strong-field limit it is natural to assume that $r_{\max}\simeq[v_E/(2l+1)\omega]$. The minimum impact parameter is determined by the larger of the two values $r_{\min,cl}=(Ze^2/mv_E^2)$—the classical limiting impact parameter, bounding the region of small momentum transfer, and $r_{\min,qu}=\hbar/mv_E$—the quantum lower limit on the impact parameters of Landau's logarithmic approximation. Here $\hbar$ is Planck's constant. Thus, in light of (3.10), we have

$$\Lambda=\ln\frac{mv_E^2}{\hbar\omega(2l+1)},$$

$$v_E>\frac{Ze^2}{\hbar}\quad\text{or}\quad\Lambda=\ln\frac{mv_E^3}{Ze^2\omega(2l+1)},\quad v_E<\frac{Ze^2}{\hbar}. \tag{3.11}$$

The decrease in the Coulomb logarithm at large harmonic numbers can be one of the reasons for the cutoff of the series (3.5) and (3.6).

We now discuss certain general consequences of Eqs. (3.5) and (3.6). We note that for $l=0$ we can write the dissipative conductivity tensor determined by the relation

$$\delta j_\alpha^{(1)}=\sigma_{\alpha\beta}^{(1)}E_\beta. \tag{3.12}$$

We have $\sigma_{xy}^{(1)}=\sigma_{yx}^{(2)}=0$, and

$$\sigma_{xx}^{(1)}=\frac{e^2n_e}{m\omega^2}\nu(E,0)\left[A_0\left(\rho^2,\,\frac{v_E}{2v_T}\right)-A_1\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right], \tag{3.13}$$

$$\sigma_{yy}^{(1)}=\frac{e^2n_e}{m\omega^2}\nu(E,0)\left[A_0\left(\rho^2,\,\frac{v_E}{2v_T}\right)+A_1\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right]. \tag{3.14}$$

Aside from the nonlinear dependence of such a conductivity tensor on the intensity of the plasma-warming radiation, the nonlinear dependence on the polarization is also important. This latter dependence is manifested in the anisotropy of the dissipative conductivity tensor.

We now dwell on the intensity of the generated harmonics ($l>0$). We assume that

$$\varphi_x=kz, \tag{3.15}$$

and accordingly,

$$k^2=\omega^2-\omega_{Le}^2. \tag{3.16}$$

Neglecting dissipation, from Maxwell's equations we have for the field of the harmonics (compare Ref. 8)

$$E_x^{(2l+1)}=-e_xE\,\frac{\nu(E,l)}{4l(l+1)\omega}\left[A_l\left(\rho^2,\,\frac{v_E}{2v_T}\right)-A_{l+1}\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right]\sin[(2l+1)(\omega t-kz)], \tag{3.17}$$

$$E_y^{(2l+1)}=-e_yE\,\frac{(2l+1)\nu(E,l)}{4l(l+1)\omega}\left[A_l\left(\rho^2,\,\frac{v_E}{2v_T}\right)+A_{l+1}\left(\rho^2,\,\frac{v_E}{2v_T}\right)\right]\cos[(2l+1)(\omega t-kz)]. \tag{3.18}$$

Forming the ratio of the time-averaged squared electric vector given by (3.17) and (3.18) to the mean squared electric field vector of the pump (2.1) $\overline{(\mathbf{E})^2}=E^2/2$, we obtain

$$\eta^{(2l+1)}=\frac{\overline{(E^{(2l+1)})^2}}{\overline{(E)^2}}=\frac{\nu^2(E,l)}{16l^2(l+1)^2\omega^2}\,B_l\left(\rho^2,\,\frac{v_E}{2v_T}\right), \tag{3.19}$$

where

$$B_l(\rho^2,N)=A_l^2(\rho^2,N)+A_{l+1}^2(\rho^2,N)-2\rho^2A_l(\rho^2,N)A_{l+1}(\rho^2,N). \tag{3.20}$$

The expression (3.19) characterizes the harmonic generation efficiency. It must be emphasized that the right-hand side of Eq. (3.19) depends on the square of the plasma density (compare Ref. 8). The dependence of the intensity of the radiated harmonics on the polarization and intensity of the pump is determined by (3.20), which we study below.

We now present an expression for the Stokes parameters, which are given by Eqs. (3.17) and (3.18) and which characterize the polarization of the harmonics ($l>0$). Above all, we note that $\xi^{(2l+1)}=0$. This means that the harmonics, just like the pump, are completely elliptically polarized. Furthermore, we have

$$\xi_2^{(2l+1)}\equiv A^{(2l+1)}=\xi_2\frac{A_l^2(\rho^2,v_E/2v_T)-A_{l+1}^2(\rho^2,v_E/2v_T)}{B_l(\rho^2,v_E/2v_T)}, \tag{3.21}$$

$$\xi_3^{(2l+1)}=\frac{\xi_3[A_l^2(\rho^2,v_E/2v_T)+A_{l+1}^2(\rho^2,v_E/2v_T)]-2A_l(\rho^2,v_E/2v_T)A_{l+1}(\rho^2,v_E/2v_T)}{B_l(\rho^2,v_E/2v_T)}, \tag{3.22}$$

where $A^{(2l+1)}$ is the degree of circular polarization of harmonic number $(2l+1)$. The specific analysis given below will demonstrate the uniqueness of the mechanisms described by these general expressions.

4. In this section we examine the asymptotic properties of the relations obtained in the preceding section, where the inequality (2.4) holds so well that it can be assumed that $v_E/v_T\rightarrow\infty$. Neglecting the thermal motion completely, we can use the asymptotic result Eq. (A1.10). This corresponds to the fact that

$$A_l(\rho^2,\infty)\equiv A_l^{(3/2)}(\rho^2)=\frac{\Gamma(3/2-l)}{|A|^{3/2}\Gamma(3/2)}\,P_{1/2}^l\left(\frac{1}{|A|}\right), \tag{4.1}$$

where $\Gamma(z)$ is the Euler function and $P_\nu^l(z)$ is the Legendre function. Equation (4.1) corresponds to coefficients in the Fourier series expansion of the solutions of the Laplace equation (Refs. 10 and 17, §33; Ref. 18, Eq. (3.10)). The Legendre functions can be expressed in terms of the complete elliptic integrals $\mathbf{E}(k)$ and $\mathbf{K}(k)$. The corresponding expressions, obtained on the basis of Refs. 18 and 19, can be found in Appendix 2. They are helpful for describing the properties of harmonics with low numbers. For example, to characterize the intensity of the third harmonic the corresponding expression (3.20) can be represented in the form

$$\begin{aligned}B_1(\rho^2,\infty)=\frac{4(1+\rho^2)}{\pi^2\rho^8(1-\rho^4)^2}&\left\{[16-31\rho^4+15\rho^8]\mathbf{E}^2\right.\\&\times\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)+(1-\rho^2)^2(16-7\rho^4)\mathbf{K}^2\\&\times\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)-2(16-9\rho^4+3\rho^8)(1-\rho^2)\mathbf{E}\\&\left.\times\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)\mathbf{K}\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)\right\}.\end{aligned} \tag{4.2}$$

For a low degree of circular polarization A of the warming radiation (or, equivalently, for a large difference of the maximum degree of linear polarization ρ^2 from 1) the number of the generated harmonics is small. This can be easily seen from Eq. (A3.12) of Appendix 3. Conversely, the number of harmonics generated in the plasma becomes large for a small degree of circular polarization of the warming radiation

$$|A|\ll 1, \tag{4.3}$$

when the polarization is almost planar. In the limit (4.3) asymptotic formulas must be used for the Legendre functions. The corresponding expressions can be found in Appendix 3.

We turn first to the third harmonic ($l=1$) in the limit (4.3). Using Eq. (A3.4) we can write

$$A_1^{(3/2)}(\rho^2)-A_2^{(3/2)}(\rho^2)=\frac{3}{\pi\sqrt{2}}\left\{\ln\frac{64}{A^2}-\frac{14}{3}\right\}, \tag{4.4}$$

$$A_1^{(3/2)}(\rho^2)+A_2^{(3/2)}(\rho^2)=\frac{2^{5/2}}{\pi A^2}. \tag{4.5}$$

This makes it possible, specifically, to obtain for Eq. (3.20)

$$B_1(\rho^2,\infty)=8/\pi^2A^2. \tag{4.6}$$

We thereby have for the efficiency of third-harmonic generation

$$\eta^{(3)}=\frac{1}{8\pi^2A^2}\,\frac{\nu^2(E,1)}{\omega^2}. \tag{4.7}$$

The presence of the square of the low degree of circular polarization in the denominator in Eq. (4.7) attests to the important phenomenon whereby generation intensifies when the polarization of the warming pump deviates by a small amount from planar polarization.

We now consider the polarization of the third harmonic. In accordance with Eqs. (3.21) and (3.22) and the expressions (4.4) and (4.5), we obtain

$$A^{(3)}=\xi_2^{(3)}=3A\left[\ln\frac{8}{|A|}-\frac{7}{3}\right], \tag{4.8}$$

$$\xi_3^{(3)}\simeq-\left(1-\frac{1}{2}[\xi_2^{(3)}]^2\right)\approx-1. \tag{4.9}$$

Since $\xi_2^{(3)}$ is small, Eq. (4.9) indicates that the polarization of the harmonic, just like that of the polarization of the warming radiation, is almost planar. However, the overall minus sign in Eq. (4.9) means that if the pump is polarized almost along the x axis, then the harmonic is polarized in a perpendicular direction, i.e., almost along the y axis. This remarkable property contrasts sharply with the result obtained in the theory of harmonic generation by plane-polarized radiation, where the harmonics were found to be polarized in the polarization plane of the pump.[6] Another important property of the harmonic polarization is given by Eq. (4.8), according to which the degree of circular polarization of the third harmonic is logarithmically ($\ln|A|$) greater than the degree of circular polarization of the pump.

We now examine the properties of the high harmonics. Above all, we note that for the high harmonics Eq. (3.20) reduces to (4.6). Therefore we have for the generation efficiency of all harmonics, under the condition (4.3), in a cold plasma

$$\eta^{(2l+1)}=\frac{\nu^2(E,l)}{2\pi^2 l^2(l+1)^2\omega^2A^2}. \tag{4.10}$$

The general formula for the degree of circular polarization of the harmonics under the condition (4.3) is

$$A^{(2l+1)}=A\left\{(2l+1)\left[\ln\frac{2}{|A|}-\psi\left(l+\frac{3}{2}\right)-C\right]+1-\frac{2}{2l+3}\right\}. \tag{4.11}$$

Clearly the logarithmic ($\ln|A|$) increase in the circular polarization is a general property. For high harmonics, $l\gg 1$, it follows from Eq. (4.11) that

$$A^{(2l+1)}=Al\left[\ln\frac{4}{A^2l^2}-2C\right]. \tag{4.12}$$

Hence, the circular polarization of the harmonics increases as their order increases. Note that the right-hand side of Eq. (4.12) remains small compared to 1. The latter assertion is associated above all with the fact that the asymptotic representation (A3.4) is applicable only if

$$A^2l^2\ll 1. \tag{4.13}$$

The smallness of the left-hand side of the inequality (4.13) is a consequence of the anomalous increase in the generation efficiency of high harmonics compared with the case of planar polarization (compare Refs. 6 and 8).

When the inequality (4.13) is not satisfied, the coefficients $A_l^{(3/2)}(\rho^2)$ are given by the asymptotic representation (A3.13), according to which it can be asserted that if the dependence of the Coulomb logarithm on l does not lead to cutoff of the harmonic series (3.5) and (3.6), then in the cold-plasma approximation we are considering, such a cutoff arises sharply (exponentially) at $l\sim|A|^{-1}$.

5. Having established the important role of the polarization of the plasma-warming radiation, we now draw attention to the nonlinear change, following from our analysis, in this polarization accompanying absorption of the radiation. To do so, we consider the fundamental harmonic, whose current density $\mathbf{j}^{(1)}$ can be written, according to Eqs. (2.11) and (3.12),

$$\frac{\partial j_\alpha^{(1)}}{\partial t}=\frac{\omega_{Le}^2}{4\pi}E_\alpha+\sigma_{\alpha\beta}^{(1)}\frac{\partial E_\beta}{\partial t}, \tag{5.1}$$

where $\sigma_{\alpha\beta}^{(1)}$ is given by Eqs. (3.13) and (3.14). Since the electric field vector of the pump satisfies

$$\frac{\partial^2\mathbf{E}}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}+\frac{4\pi}{c^2}\frac{\partial\mathbf{j}^{(1)}}{\partial t}, \tag{5.2}$$

taking account of Eqs. (3.15) and (3.16) and assuming e_α and E to be slowly varying functions of the coordinate z, we obtain the following truncated equations for them:

$$\begin{aligned}\frac{1}{E}\frac{dE}{dz}&=-\frac{2\pi\omega}{kc^2}(\sigma_{xx}^{(1)}e_x^2+\sigma_{yy}^{(1)}e_y^2)\\&=-\frac{\omega_{Le}^2\nu(E,0)}{2kc^2\omega}[A_0^{(3/2)}(\rho^2)-\rho^2A_1^{(3/2)}(\rho^2)],\end{aligned} \tag{5.3}$$

$$\begin{aligned}\frac{d\rho^2}{dz}&=-\frac{2\pi\omega}{kc^2}(1-\rho^4)(\sigma_{xx}^{(1)}-\sigma_{yy}^{(1)})\\&=\frac{\omega_{Le}^2\nu(E,0)}{kc^2\omega}(1-\rho^4)A_1^{(3/2)(\rho^2)}.\end{aligned} \tag{5.4}$$

Making use of Eqs. (A2.3), (5.3) assumes the form

$$\frac{1}{E}\frac{dE}{dz}=-\frac{\omega_{Le}^2\nu(E,0)}{\pi kc^2\omega\sqrt{1+\rho^2}}\mathbf{K}\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right), \tag{5.5}$$

which corresponds to the law, established in Ref. 10, that the absorption of the strong polarized radiation in plasma depends nonlinearly on the polarization of the radiation. When the condition (4.3) is satisfied, it follows from Eq. (5.5) that

$$\frac{1}{E}\frac{dE}{dz}=-\frac{\omega_{Le}^2\nu(E,0)}{\omega kc^2\pi\sqrt{2}}\ln\frac{8}{|A|}\equiv\frac{1}{\zeta(E,A)}, \tag{5.6}$$

where $\zeta(E,A)$ is the characteristic absorption length. Neglecting the relatively weak logarithmic dependences, we have for the law expressing the decrease of the field

$$E^3(z)\simeq E^3(0)-3z/\zeta(E(0),A(0)). \tag{5.7}$$

Equation (5.4) describes the change in the degree of maximum linear polarization $L\equiv\rho^2$ of the warming radiation as it is absorbed while propagating into the plasma ($z>0$). The positiveness of $A_1^{(3/2)}(\rho^2)$ corresponds to the fact that L increases in the process. This property corresponds to the dependence of the polarization of the warming radiation on its intensity. This dependence is characterized by the following nonlinear equation:

$$E\frac{d\rho^2}{dE}=-\pi(1-\rho^4)\sqrt{1+\rho^2}A_1^{(3/2)}(\rho^2)\left[\mathbf{K}\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)\right]^{-1}. \tag{5.8}$$

The solution of this equation is especially simple in the case (4.3), when Eq. (5.8) reduces approximately to

$$\frac{dE}{4E}=\ln\frac{64}{A^2}d\left(\frac{A^2}{64}\right). \tag{5.9}$$

Hence follows the law expressing the decrease of the degree of circular polarization of the plasma-warming radiation in the presence of inverse bremsstrahlung absorption of this radiation:

$$\frac{A^2(z)}{64}\left[\ln\frac{64}{A^2(z)}+1\right]=\frac{A^2(0)}{64}\left[\ln\frac{64}{A^2(0)}+1\right]-\frac{1}{4}\ln\frac{E(0)}{E(z)}. \tag{5.10}$$

The physical reason for this phenomenon is the anisotropy of the dissipative conductivity tensor (3.13) and (3.14), for which in the case (4.3) we have for a cold plasma

$$\sigma_{xx}^{(1)}=\frac{e^2n_e}{m\omega^2}\nu(E,0)\frac{2^{1/2}}{\pi}\left[\ln\frac{8}{|A|}-1\right],$$

$$\sigma_{yy}^{(1)}=\frac{e^2n_e}{m\omega^2}\nu(E,0)\frac{2^{5/2}}{\pi A^2}. \tag{5.11}$$

It is obvious that the y component of the field is absorbed relatively more efficiently than the x component. This is what leads to the increase in the maximum degree of linear polarization.

6. Having established in the cold-plasma model a number of striking nonlinear polarization properties, we must now analyze the role of the thermal motion of the particles. This will make it possible to eliminate the seeming inconsistency, facilitate the transition to the case of planar polarization of the pump, and determine the magnitude of the corresponding anomalies. To this end, we turn to Eq. (3.9) for $A_l(\rho^2,N)$, where $N=v_E/2v_T\gg 1$ according to Eq. (2.4).

Equation (3.9) can be rewritten in the form

$$A_l(\rho^2,N)=A_l^{(3/2)}(\rho^2)-\delta a_l(A^2,N), \tag{6.1}$$

where

$$\delta a_l(A^2,N)=\frac{2}{\sqrt{\pi}}\int_{N^2}^{\infty}dz\,\sqrt{z}e^{-z}I_l(\sqrt{1-A^2}z). \tag{6.2}$$

For a strong plasma-warming field, by virtue of the condition (2.4), the asymptotic expansion of the function $I_l(\rho^2 z)$ (Ref. 13, Eq. 8.451.5) can be used in the integrand in Eq. (6.2). The result is

$$\delta a_l(A^2,N)=\frac{2^{3/2}}{\pi}\left\{\frac{1}{A^2}\exp\left(-\frac{A^2N^2}{2}\right)\right.$$
$$\left.-\frac{1}{4}\left(l^2-\frac{1}{4}\right)E_1\left(\frac{A^2N^2}{2}\right)\right\}, \tag{6.3}$$

where $E_1(z)$ is the exponential integral (Ref. 19, Eq. 5.1.1). Here, small terms $\sim l^2A^2v_T^2/v_E^2$ in this asymptotic expansion have been dropped.

Since $E_1(z)\sim z^{-1}\exp(-z)$ for $z\gg 1$, it is obvious that the quantity (6.3) is exponentially small when

$$N^2A^2=(Av_E/2v_T)^2\gg 1. \tag{6.4}$$

This inequality determines the region of applicability of the cold-plasma model.

It is obvious that under the condition (2.4) the inequality (6.4) can break down only if the degree of circular polarization is small, when $A_l^{(3/2)}(\rho^2)$ has the form (A3.4). This means that in the region where it is important to take account of the thermal motion Eq. (6.1) can be represented in the form

$$A_l(\rho^2,N)=\frac{2^{3/2}}{\pi}\left\{\frac{1}{A^2}\left[1-\exp\left(-\frac{A^2N^2}{2}\right)\right]\right.$$
$$+\frac{1}{2}\left(l^2-\frac{1}{4}\right)\left[\frac{1}{2}E_1\left(\frac{A^2N^2}{2}\right)-\frac{1}{2}\ln\frac{4}{A^2}\right.$$
$$\left.\left.+\psi\left(l+\frac{3}{2}\right)+C-\frac{1}{2}\right]-\frac{l}{2}\right\}. \tag{6.5}$$

This equation completely describes the effect of the thermal motion of the particles under our conditions of a strong warming field (2.4) on all anomalous polarization phenomena discussed above. Here we examine the limit of a very low degree of circular polarization

$$A\ll v_T/v_E, \tag{6.6}$$

where the transition to the previously studied case of plane-polarized radiation can be examined. Under the conditions (6.6), Eq. (6.5) assumes the form

$$A_l\left(\rho^2,\frac{v_E}{2v_T}\right)=\frac{2^{3/2}}{\pi}\left\{\frac{v_E^2}{8v_T^2}-\frac{1}{4}\left(l^2-\frac{1}{4}\right)\right.$$
$$\left.\times\left[\ln\frac{v_E^2}{2v_T^2}-2\psi\left(l+\frac{3}{2}\right)-C+1\right]-\frac{l}{2}\right\}, \tag{6.7}$$

which does not depend on the polarization. Hence we have for the high harmonics ($l\gg 1$)

$$A_l\left(\rho^2,\frac{v_E}{2v_T}\right)=\frac{2^{3/2}}{\pi}\left\{\frac{v_E^2}{8v_T^2}-\frac{l^2}{4}\right.$$
$$\left.\times\left[\ln\frac{v_E^2}{2l^2v_T^2}-C+1\right]+o(1)\right\}. \tag{6.8}$$

This equation shows, specifically, that for $l>v_E/v_T$ the harmonics series cuts off.[8] Under such conditions, we obtain for the high harmonics

$$B_l\left(\rho^2,\frac{v_E}{2v_T}\right)=\frac{2}{\pi^2}\left(l^2\left[\ln\frac{v_E^2}{2l^2v_T^2}-C\right]^2+A^2\frac{v_E^4}{16\,v_T^4}\right). \tag{6.9}$$

Accordingly, we obtain for the degree of circular polarization of the high harmonics

$$A^{(2l+1)}=\frac{A(v_E^2/2v_T^2)l[\ln(v_E^2/2l^2v_T^2)-C]}{(Av_E^2/4v_T^2)^2+l^2[\ln(v_E^2/2l^2v_T^2)-C]^2}. \tag{6.10}$$

The degree of circular polarization of the harmonics may not be small. However, in the plane polarization limit ($A=0$) the expression (6.10) vanishes. Equation (3.6) likewise vanishes, while according to the law

$$\eta^{(2l+1)}=\frac{\nu^2(E,l)}{8l^2\pi^2\omega^2}\left\{\left[\ln\frac{v_E^2}{2l^2v_T^2}-C\right]^2+\frac{A^2v_E^4}{16l^2v_T^4}\right\} \tag{6.11}$$

as $A\to 0$ the generation efficiency of the high harmonics tends to the formula corresponding to the case of a plane-polarized pump.[6,8]

7. The striking polarization effects characterizing the generation of harmonics in plasma substantially increase the efficiency of such a process. Indeed, an estimate using Eq. (6.10) at the limit of applicability of this expression, justified by Eq. (6.5), with $(A^2 v_E^2/8 v_T^2) \sim 1$ yields according to Eq. (6.9) the following estimate of the efficiency (3.19) of high-harmonic generation:

$$\eta^{(2l+1)} \sim \frac{\nu^2(E,l)}{2\pi^2\omega^2}\,\frac{v_E^2}{8l^4 v_T^2}. \tag{7.1}$$

This expression is greater than the efficiency (6.11) (in the limit $A=0$) of harmonic generation by plane-polarized plasma-warming radiation on account of the fact that the condition

$$\frac{v_E^2}{2l^2 v_T^2} \gg \left[\ln\frac{v_E^2}{2l^2 v_T^2} - C\right]^2 \tag{7.2}$$

holds for essentially all harmonics. Aside from this polarization increase in the intensity of the harmonics emitted during coherent oscillation of electrons in a laser radiation field, it is necessary to call attention to the anomaly in the polarization of the harmonics. Indeed, the Stokes parameter $\xi_3^{(2l+1)}$ of the harmonics with small but finite degree of circular polarization of the pump field turned out to be different in sign, according to Eqs. (4.9) and (4.12), from the Stokes parameter ξ_3 of the warming radiation. This means that in this case the harmonics are polarized, to a high degree of accuracy, perpendicular to the polarization plane of the pump. In the transitional region, as the degree of circular polarization of the pump decreases, $A^{(2l+1)}$ increases, reaching 1, according to Eq. (6.11), when

$$|A|\,\frac{v_E^2}{4v_T^2} = l\left[\ln\frac{v_E^2}{2l^2 v_T^2} - C\right],$$

and then decreases to zero for $A=0$. Accordingly, the Stokes parameter $\xi_3^{(2l+1)}$ changes from a value close to -1, passes through zero, and increases up to $+1$, corresponding for $A=0$ to polarization of the harmonic in the polarization plane of the pump. We have pointed out the properties of the nonlinear phenomenon whereby the polarization of the warming radiation changes as a result of inverse bremsstrahlung, having in mind the importance of the pump polarization for the effects discussed.

We present several formulas in a form convenient for making estimates. These formulas, in the first place, make it possible to see the condition of applicability of the results obtained and, in the second place, indicate the efficiency of the conversion of pump radiation into high harmonics. First, to demonstrate the obviousness of the fact that the condition (2.5) is satisfied, we write

$$\frac{v_E^2}{v_T^2} = 3.7\times 10^3\,\frac{q\lambda^2}{T}.$$

Here and below q is the flux density of the pump radiation in units of 10^{16} W/cm^2, λ is the pump wavelength in microns, and T is the electron temperature in electron volts. To see the possibility of using the nonrelativistic approximation we present the relation

$$\frac{v_E^2}{c^2} = 7.3\times 10^{-3} q\lambda^2.$$

In accordance with Eq. (6.11), the generation efficiency of the higher harmonics, which depends on the harmonic number, is given for a plane-polarized pump by

$$\frac{\nu^2(E)}{\pi^2\omega^2}\left(\ln\frac{v_E}{v_T}\right)^2 = 5\times 10^{-10}\,\frac{Z^2 n_e^2}{\lambda^4 q^3}\left(\frac{\Lambda}{10}\right)^2\left(\ln\frac{v_E}{v_T}\right)^2.$$

Here and below n_e is the electron number density in units of 10^{20} cm^{-3}.

The law (7.1), which depends on the harmonic number, of the high-harmonic generation efficiency in the presence of a low degree of circular polarization is characterized by the parameter

$$\frac{\nu^2(E) v_E^2}{(2\pi\omega)^2 v_T^2} = 4\times 10^{-7}\,\frac{Z^2 n_e}{\lambda^2 q^2 T}\left(\frac{\Lambda}{10}\right)^2.$$

The theoretical results in this paper should make it possible to formulate detailed experiments studying the phenomenon of coherent bremsstrahlung generation of harmonics in plasma. Comparing our plasma model with the atomic model of harmonic generation from Ref. 21 shows that the properties that we have established should also appear in nonionized gases if the radiation intensity is sufficiently high.

This work was performed under state support of the leading scientific schools (Project 96-15-96750), support by the Russian Fund for Fundamental Research (Project 96-02-17002), and as part of the Joint Russian–Italian project ''New Coherent Sources of Ultrashort Pulses,'' prepared in accordance with law No. 212/92 of the Italian government.

APPENDIX 1

Substituting the Fourier expansion (3.4) into Eq. (3.2) gives

$$\frac{\partial}{\partial y}\begin{Bmatrix} j_x \\ j_y \end{Bmatrix} = -E\,\frac{\omega_{Le}^2}{4\pi}\sum_{l=0}^{\infty}\frac{\nu(E,l)}{\omega} \times\begin{Bmatrix} M_l^{(+)}\left(\delta_+,\delta_-\dfrac{v_T^2}{2v_E^2}\right)\sin[(2l+1)(\omega t-\varphi_x)], \\ M_l^{(-)}\left(\delta_+,\delta_-\dfrac{v_T^2}{2v_E^2}\right)\cos[(2l+1)(\omega t-\varphi_x)], \end{Bmatrix} \tag{A1.1}$$

where $\omega_{Le}=\sqrt{4\pi e^2 n_e/m}$ is the electron Langmuir frequency, and

$$M_l^{(\pm)}(a,b;\alpha^2) = \frac{1}{\sqrt{8\pi}}\int_0^{\infty}dx\,x\int_0^{\pi}d\theta \times\exp\left(-\frac{x^2\alpha^2}{\sin^2\theta}\right)[J_{l+1}(ax)J_l(bx) \pm J_{l+1}(bx)J_l(ax)]. \tag{A1.2}$$

Using Eq. 3.363.1 of Ref. 13 (p. 329), it can be shown that

$$\int_0^\pi d\theta \exp\left(-\frac{\alpha^2x^2}{\sin^2\theta}\right)=\pi[1-\Phi(\alpha x)], \tag{A1.3}$$

where $\Phi(z)=(2/\sqrt{\pi})\int_0^z dte^{-t^2}$ is the error function. This makes it possible to represent Eq. (A1.2) in the form

$$\begin{aligned}M_l^{(\pm)}(a,b;\alpha^2)=\frac{1}{\sqrt{2\pi}}\int_\alpha^\infty d\tau\int_0^\infty dx\; x^2e^{-x^2\tau^2}\\ \times[J_{l+1}(ax)J_l(bx)\\ \pm J_{l+1}(bx)J_l(ax)].\end{aligned} \tag{A1.4}$$

Using the recurrence relations

$$J_{l+1}(z)=-\frac{d}{dz}J_l(z)+\frac{l}{z}J_l(z)$$

for the Bessel functions (Ref. 13, Eq. 8.472.2), this expression can be reduced to the well-known integral

$$\begin{aligned}M^{(\pm)}(a,b;\alpha^2)=\left[\frac{l}{a}-\frac{\partial}{\partial a}\pm\left(\frac{l}{b}-\frac{\partial}{\partial b}\right)\right]\\ \times\frac{1}{\sqrt{2\pi}}\int_\alpha^\infty d\tau\int_0^\infty dx\; xJ_l(ax)J_l(bx)e^{-x^2\tau^2}.\end{aligned} \tag{A1.5}$$

Indeed, according to Eq. 6.333.2 of Ref. 13 (p. 732), we have

$$\begin{aligned}\int_0^\infty dx\; xJ_l(ax)J_l(bx)e^{-x^2\tau^2}=\frac{1}{2\tau^2}\\ \times\exp\left(-\frac{a^2+b^2}{4\tau^2}\right)I_l\left(\frac{ab}{2\tau^2}\right),\end{aligned} \tag{A1.6}$$

where $I_l(z)$ is a modified Bessel function. Using now the recurrence relation (Ref. 13, Eq. 8.486)

$$lI_l(z)-z\frac{dI_l(z)}{dz}=-zI_{l+1}(z),$$

Eq. (A1.5) can be put into the form

$$\begin{aligned}M^{(\pm)}(a,b;\alpha^2)=\frac{(a\pm b)2}{\sqrt{\pi}}\int_0^{(1/8\alpha^2)}dz\sqrt{z}\\ \times\exp(-2[a^2+b^2]z)[I_l(4abz)\\ \mp I_{l+1}(4abz)].\end{aligned} \tag{A1.7}$$

In the cold-plasma limit, $\alpha^2\to 0$, this relation assumes the form

$$\begin{aligned}M_l^{(\pm)}(a,b;0)=\frac{(a\pm b)2}{\sqrt{\pi}}\int_0^\infty\frac{dz\sqrt{z}}{(4ab)^{3/2}}\\ \times\exp\left(-\frac{[a^2+b^2]z}{2ab}\right)[I_l(z)\mp I_{l+1}(z)].\end{aligned} \tag{A1.8}$$

Using now Eq. 6.624.5 of Ref. 13 (p. 727)

$$\begin{aligned}\int_0^\infty\exp\left\{-\frac{tz}{\sqrt{z^2-1}}\right\}I_\mu(t)t^\nu dt\\ =\Gamma(\nu+1-\mu)(z^2-1)^{(1+\nu)/2}P_\nu^\mu(z),\end{aligned} \tag{A1.9}$$

where $P_\nu^\mu(z)$ is a Legendre function, we obtain

$$M_l^{(\pm)}(a,b;0)=\frac{(a\pm b)\Gamma(3/2-l)}{|2(a^2-b^2)|^{3/2}\Gamma(3/2)}P_{1/2}^l\left(\left|\frac{a^2+b^2}{a^2-b^2}\right|\right). \tag{A1.10}$$

The latter equation makes it possible, specifically, to write an expression for the following improper integral:

$$\begin{aligned}\int_0^\infty dx\; x[J_{l+1}(ax)J_l(bx)\pm J_{l+1}(bx)J_l(ax)]\\ =\frac{(a\pm b)\Gamma(3/2-l)}{|a^2-b^2|^{3/2}\Gamma(3/2)}P_{1/2}^l\left(\left|\frac{a^2+b^2}{a^2-b^2}\right|\right).\end{aligned} \tag{A1.11}$$

APPENDIX 2

We present here several expressions for the coefficients in the Fourier expansion (4.1), which can be written in terms of complete elliptic integrals. We use, first, Eq. 8.13.5 of Ref. 19 (p. 159)

$$P_{1/2}(z)=\frac{2}{\pi}(z+\sqrt{z^2-1})^{1/2}\mathbf{E}\left(\sqrt{\frac{2(z^2-1)^{1/2}}{z+(z^2-1)^{1/2}}}\right) \tag{A2.1}$$

and, second, Eq. 3.6.1.4 of Ref. 18 (p. 149)

$$P_\nu^m(z)=(z^2-1)^{m/2}\frac{d^m}{dz^m}P_\nu(z). \tag{A2.2}$$

These relations lead to the following expressions for the coefficients (4.2):

$$\begin{aligned}A_0^{(3/2)}(\rho^2)&=\frac{2}{\pi}\frac{\sqrt{1+\rho^2}}{1-\rho^4}\mathbf{E}(k),\\ A_1^{(3/2)}(\rho^2)&=\frac{2}{\pi}\frac{\sqrt{1+\rho^2}}{1-\rho^4}\left\{\mathbf{E}(k)+\frac{1-\rho^2}{\rho^2}[\mathbf{E}(k)-\mathbf{K}(k)]\right\},\\ A_2^{(3/2)}(\rho^2)&=\frac{2}{\pi}\frac{\sqrt{1+\rho^2}}{1-\rho^4}\left\{\mathbf{E}(k)+4\frac{1-\rho^4}{\rho^4}\right.\\ &\quad\left.\times\left[\mathbf{E}(k)-\frac{1}{1+\rho^2}\mathbf{K}(k)\right]\right\},\\ A_3^{(3/2)}(\rho^2)&=\frac{2}{\pi}\frac{\sqrt{1+\rho^2}}{1-\rho^4}\left\{\mathbf{E}(k)-\frac{(1-\rho^4)(1+\rho^2)}{12\rho^4}\right.\\ &\quad\times[(27k^4-128k^2+128)\mathbf{K}(k)\\ &\quad\left.+(-3k^4+64k^2-128)\mathbf{E}(k)]\right\},\end{aligned} \tag{A2.3}$$

where

$$k=\sqrt{\frac{2\rho^2}{1+\rho^2}}\quad\text{and}\quad 1-k^2=\frac{1-\rho^2}{1+\rho^2}. \tag{A2.4}$$

APPENDIX 3

The asymptotic expansion, required for our analysis, of the Legendre functions $P_\nu^l(z)$ for half-integer ν and large arguments z can be found in Hobson's book.[17] Retaining the required higher terms of the asymptotic expansion and making convenient transformations, we can write

$$P_{1/2}^{l}(z)=\frac{1}{\sqrt{\pi}\Gamma(3/2-l)}(2z)^{1/2}\left\{1-\frac{2l}{(2z)^2}+\frac{1}{(2z)^2}\left(l^2-\frac{1}{4}\right)\times\left[-2\ln(2z)+2\psi\left(l+\frac{3}{2}\right)+2C-1\right]\right\},\tag{A3.1}$$

where $C=0.577\ldots$ is Euler's constant. This asymptotic expression makes it possible to use the following approximation for the Fourier coefficient $A_l(\rho^2,\infty)$:

$$A_l^{(3/2)}(\rho^2)=\frac{2^{3/2}}{\pi}\left\{\frac{1}{A^2}+\frac{1}{2}\left(\left[l^2-\frac{1}{4}\right]\times\left[-\ln\frac{2}{|A|}+\psi\left(l+\frac{3}{2}\right)+C-\frac{1}{2}\right]-l\right)\right\}.\tag{A3.2}$$

This expression is also applicable to low harmonic numbers. On the other hand, for high harmonics ($l\gg1$), for which $\psi(l+3/2)\approx\ln l+1/l$, we have, up to and including terms linear in l,

$$A_l^{(3/2)}(\rho^2)=\frac{2^{3/2}}{\pi}\left\{\frac{1}{A^2}+\frac{l^2}{2}\left[-\ln\frac{2}{|A|l}+C-\frac{1}{2}\right]+o(1)\right\}.\tag{A3.3}$$

The asymptotic expansion (A3.1) is inapplicable for $l^2>z^2$. We now obtain the required asymptotic representation for the Fourier coefficients $A_l^{(3/2)}(\rho^2)$ in the limit

$$l^2\gg z^2.\tag{A3.4}$$

To do so, we employ Eq. (8.702) of Ref. 13 (p. 1013)

$$P_\nu^\mu(z)=\frac{1}{\Gamma(1-\mu)}\left(\frac{z+1}{z-1}\right)^{\mu/2}F\left(-\nu,\nu+1;1-\mu;\frac{1-z}{2}\right),\tag{A3.5}$$

which expresses the Legendre function in terms of the hypergeometric function. Using also the relation

$$P_{1/2}^{-l}(z)=\frac{\Gamma(3/2-l)}{\Gamma(3/2+l)}P_{1/2}^{l}(z),\tag{A3.6}$$

we can now express the Fourier coefficients $A_l^{(3/2)}(\rho^2)$ in terms of hypergeometric series as follows:

$$A_l^{(3/2)}(\rho^2)=\frac{\Gamma(3/2+l)}{|A|^{3/2}\Gamma(3/2)\Gamma(l+1)}\left(\frac{1-|A|}{1+|A|}\right)^{l/2}\times F\left(-\frac{1}{2},\frac{3}{2};l+1;\frac{1}{2}-\frac{1}{2|A|}\right).\tag{A3.7}$$

Since the hypergeometric series $F(a,b;c;\zeta)$ for large values of c is an asymptotic expansion in c, to obtain the desired asymptotic representation of the coefficients $A_l^{(3/2)}(\rho^2)$ we employ

$$F\left(-\frac{1}{2},\frac{3}{2};l+1;\frac{1}{2}-\frac{1}{2|A|}\right)\simeq1+\frac{3(1-|A|)}{8l|A|}+o\left(\frac{1}{l^2A^2}\right).\tag{A3.8}$$

Since, further, according to Ref. 20 (p. 87)

$$\frac{\Gamma(3/2+l)}{\Gamma(l+1)}\simeq\sqrt{l}\left(1+\frac{3}{8l}+\ldots\right),\tag{A3.9}$$

we obtain

$$A_l^{(3/2)}(\rho^2)\simeq\frac{2\sqrt{l}}{\sqrt{\pi}|A|^{3/2}}\left(1+\frac{3}{8l|A|}+\ldots\right)\times\exp\left\{-\frac{l}{2}\ln\frac{1+|A|}{1-|A|}\right\}.\tag{A3.10}$$

This asymptotic formula can also be used for degrees of circular polarization that are not small. Assuming now that $|A|\ll1$, it follows from Eq. (A3.10) that

$$A_l^{(3/2)}(\rho^2)\simeq\frac{2\sqrt{l}}{\sqrt{\pi}|A|^{3/2}}\exp(-l|A|).\tag{A3.11}$$

Finally, we present an interpolation formula describing the leading terms of the asymptotic expansions of Eqs. (A3.2) and (A3.11) for large harmonic numbers:

$$A_l^{(3/2)}(\rho^2)=\frac{2^{3/2}l}{\pi|A|}K_1(l|A|),\tag{A3.12}$$

where $K_1(z)$ is a modified Bessel function.

---

[*] E-mail: silin@sci.lpi.ac.ru

---

[1] W. H. Knox, IEEE J. Quantum Electron. **24**, 388 (1988).

[2] H. Schulz and D. von der Linde, Proc. SPIE **1268**, 30 (1990).

[3] S. Svanberg, J. Larsson, A. Persson, and C.-G. Wahlström, Phys. Scr. **49**, 187 (1994).

[4] S. M. Gladkov and N. I. Koroteev, Usp. Fiz. Nauk **160**, 105 (1990) [Sov. Phys. Usp. **33**, 554 (1990)].

[5] C.-G. Wahlström, Phys. Scr. **49**, 201 (1994).

[6] V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1965)].

[7] F. Giammanco, P. Ceccherini, C. Tagliavini, M. Malvezzi, P. Villoresi, and G. Tondello, Laser Phys. **7**, 22 (1997).

[8] G. Ferrante, S. A. Uryupin, M. Zarcone, and P. I. Porshnev, J. Opt. Soc. Am. B **14**, 1716 (1997).

[9] F. Giammanco and N. Spinelli (Eds.), *Plasma Collective Effects in Atomic Physics*, Edizioni ETS, Pisa (1996).

[10] V. P. Silin, Zh. Éksp. Teor. Fiz. **111**, 478 (1997) [JETP **84**, 262 (1997)].

[11] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, New York (1975).

[12] V. P. Silin, *Introduction to the Kinetic Theory of Gases* [in Russian], Nauka, Moscow (1971).

[13] I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York (1965).

[14] H. A. Kramers, Philos. Mag. **46**, 836 (1923).

[15] V. L. Ginzburg, *Propagation of Electromagnetic Waves in Plasma*, 2nd ed., Pergamon Press, New York (1970).

[16] V. P. Silin, Zh. Éksp. Teor. Fiz. **38**, 1771 (1960) [Sov. Phys. JETP **11**, 1277 (1960)].

[17] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics*, Cambridge University Press, Cambridge (1931).

[18] A. Erdélyi (Ed.), *Higher Transcendental Functions (Bateman Manuscript Project)*, Vols. 1–3, McGraw-Hill, New York (1953–1955).

[19] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington (1965).
[20] E. T. Copson, *Asymptotic Expansions*, Cambridge University Press, Cambridge (1965).
[21] R. V. Karapetyan and V. B. Fedorov, Kratkie Soobshcheniya po fizike, FIAN, No.7–8, 76 (1995).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Structure and evolution of shock waves in relativistic magnetohydrodynamics

V. I. Zhdanov*) and P. V. Titarenko

Astronomical Observatory, T. G. Shevchenko Kiev University, 254053 Kiev, Ukraine

(Submitted 1 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 881–891 (September 1998)

We derive the existence conditions for relativistic shock waves propagating in a perfectly conducting fluid with a general equation of state that guarantees that the stationary wave has a continuous profile in the presence of weak viscosity. To this end we study the one-dimensional solutions of the magnetohydrodynamic equations with a relativistic viscosity tensor. We allow for anomalous regions of thermodynamic variables and do not use the well-known condition for the convexity of Poisson adiabats. The results lead to relationships among the velocities of magnetoacoustic, Alfvén, and shock waves in front of and behind the discontinuity that prove to be more stringent than the corollaries of the evolution conditions. In the nonrelativistic case and in parallel and perpendicular shock waves, any difference between the two conditions disappears. © *1998 American Institute of Physics.* [S1063-7761(98)00909-3]

1. INTRODUCTION

The physical criteria determining the admissibility of discontinuities in the solutions of hydrodynamic equations play an important role in the theory of shock waves. This role is especially prominent in the complicated equations of state of superdense matter used, for instance, in relativistic astrophysics. Qualitatively, the derivation of these criteria in relativistic hydrodynamics is no different from the derivation of such criteria in the nonrelativistic theory. However, our results illustrate the specific features of the relativistic treatment that emerge when one allows for the additional degrees of freedom related to a magnetic field. It turns out that in magnetohydrodynamics (MHD) the derivation of the two fundamental existence criteria for shock waves, the evolution condition and the structure condition, i.e., the existence of a continuous profile in the shock wave in the presence of arbitrarily weak viscosity, differ in the relativistic case, although they coincide in the nonrelativistic case.

It is well known that the conservation laws that relate thermodynamic quantities on the two sides of the shock front are insufficient to define unambiguously a physically admissible shock transition. This forces one to introduce additional criteria that ensure the uniqueness of the solutions of the hydrodynamic equations near the discontinuity. Among these additional criteria is the most general evolution condition for shock waves.[1] In classical MHD, the evolution criterion leads to relationships between the velocities of magnetoacoustic and Alfvén waves, and the velocity of the shock wave with respect to the medium in front of and behind the discontinuity,[2,3] which bounds the domain of existence of various types of shock wave.

It is interesting to compare the evolution conditions with other constraints on the parameters of a shock wave that emerge in the weak-viscosity method. Introduction of viscosity is one of the most effective and physically justifiable tools that make it possible to define discontinuous flow as the limit of continuous solutions (see, e.g., Ref. 4). Here the admissibility condition for a shock transition from a state 0 of the medium in front of the shock front to a state 1 behind the front is obtained as the existence condition for a viscous profile of the shock wave, stationary viscous flow whose parameters change continuously from state 0 to state 1.

In the present paper, to obtain this condition we introduce into the equations of relativistic MHD[5,6] a relativistic viscosity tensor[1] and then pass to the zero-viscosity limit. We found that it is possible to examine the general equation of state without limiting ourselves to the requirement that the Poisson adiabats be convex, a requirement that plays an important role in shock wave theory as one of the Bethe–Weyl conditions for a normal medium.[4] Classical hydrodynamics states[4] that if the convexity condition is violated, the entropy criterion does not eliminate all nonphysical solutions and, therefore, is not sufficient for selecting the admissible solutions. The results of shock-wave studies in relativistic MHD[5,6] also rely on the convexity property, which in the relativistic range has the form $(\partial^2 p/\partial X^2)_S > 0$, where $X = (\varepsilon + p)V^2$, p is the pressure, ε is the entropy density, V is the specific volume (per baryon), and S is the specific entropy.

The convexity property is not a thermodynamic requirement,[1] and it can break down in complicated equations of state. In the past this fact has drawn much attention to the study of relativistic shock waves propagating in media with anomalous properties in connection with hydrodynamic models of a quark-to-hadron phase transition (see, e.g., Refs. 7–12 and the literature cited therein). The general admissibility criteria for relativistic shock waves with a viscous profile were obtained by Bugaev *et al.*[11,12] One result of this work was a constraint on the shock wave velocities[12] that coincides with the conclusions drawn from the evolution conditions (just as in nonrelativistic hydrodynamics[1]).

In the present paper we generalize the results of Bugaev *et al.*[11,12] to the case of MHD for a fluid with perfect conductivity.[5,6] The existence condition for a viscous profile in a relativistic shock wave makes it possible to formulate the admissibility criterion for such a wave in terms of inequalities that contain the equation of state. This leads to restrictions on the shock wave velocities, which in our case differ from the corollaries of the evolution conditions. This difference disappears, however, in the nonrelativistic limit.

2. BASIC RELATIONSHIPS

The equations of motion of an ideal relativistic fluid with infinite conductivity and unit permeability in a magnetic field are derivable from the conservation laws for the magnetohydrodynamic energy–momentum tensor[5,6]

$$T^{\mu\nu}=(p^*+\varepsilon^*)u^\mu u^\nu-p^*g^{\mu\nu}-\frac{1}{4\pi}h^\mu h^\nu, \tag{1}$$

where u^μ is the 4-velocity of the fluid, $g^{\mu\nu}=g_{\mu\nu}=\mathrm{diag}(1,-1,-1,-1)$, $h^\mu=-(1/2)e^{\mu\alpha\beta\gamma}F_{\alpha\beta}u_\gamma$ is the magnetic field, $e^{\alpha\beta\gamma\delta}$ is the Levi-Civita symbol, $F_{\alpha\beta}$ is the electromagnetic field tensor, and

$$p^*=p+\frac{1}{8\pi}|h|^2, \quad \varepsilon^*=\varepsilon+\frac{1}{8\pi}|h|^2, \quad |h|^2=-h^\alpha h_\alpha>0.$$

We assume that the pressure p is related to the energy density ε and the baryonic number density n (or the density of any other conserved charge) via a sufficiently smooth equation of state $p=p(\varepsilon,n)$.

Energy–momentum conservation must be supplemented by baryonic charge conservation and the Maxwell equations. Continuous fluid flow is described by the equations of motion

$$\partial_\mu T^{\mu\nu}=0, \tag{2}$$

$$\partial_\mu(nu^\mu)=0, \tag{3}$$

$$\partial_\mu(u^\mu h^\nu-u^\nu h^\mu)=0. \tag{4}$$

If there is a discontinuity in the flow, the conservation laws relate the hydrodynamic and electrodynamic properties on the two sides of the shock wave:[1]

$$D_{10}[T^{\mu\nu}l_\mu]=0, \tag{5}$$

$$D_{10}[nu^\mu l_\mu]=0, \quad D_{10}[(u^\mu h^\nu-h^\mu u^\nu)l_\mu]=0, \tag{6}$$

where l_μ is the normal to the discontinuity's hypersurface, $l_\mu l^\mu=-1$, and $D_{10}(F)=F_1-F_0$, with F_0 and F_1 the values of the quantity F in front of and behind the discontinuity. For a given state of the medium in front of the shock front, Eqs. (5) and (6) determine the curve of the shock transitions, or the Hugoniot–Taub–Lichnerowicz shock adiabat.[5,6] However, not all such transitions are admissible.

To analyze the admissible shock waves that satisfy Eqs. (5) and (6), we ''smear'' the discontinuity by introducing viscosity effects, i.e., we replace Eq. (2) with

$$\partial_\mu(T^{\mu\nu}+\tau^{\mu\nu})=0, \tag{7}$$

where

$$\tau_{\mu\nu}=\eta(u_{\mu,\nu}+u_{\nu,\mu}-u_\mu u^\alpha u_{\nu,\alpha}-u_\nu u^\alpha u_{\mu,\alpha})+\left(\zeta-\frac{2}{3}\eta\right)\frac{\partial u^\alpha}{\partial x^\alpha}(g_{\mu\nu}-u_\mu u_\nu) \tag{8}$$

is the relativistic viscosity tensor.[1] We then consider Eqs. (3) and (4) together with Eq. (7) under the assumption that there exists a solution that tends to a limit as $\eta\to 0$ and $\zeta\to 0$ and describes discontinuous flow of an ideal fluid.

3. EXISTENCE CONDITIONS FOR A SHOCK TRANSITION

Strictly speaking, a thorough study of such phenomena presupposes a detailed analysis of the physical processes at the shock front that takes into account the extent of the shock front. However, when studying superdense matter, the researcher is limited to phenomenological models, which do not allow for a description of the microstructure of the shock wave. Hence it would be interesting to analyze the various methods of regularizing discontinuous flow and to compare the different requirements that limit the class of admissible relativistic shock waves. Such studies have been done in classical MHD (see, e.g., Ref. 13 and the literature cited therein). The conclusions drawn in these studies show that, generally speaking, the criteria for the existence of shock waves depend on the way in which the discontinuity is smeared. For instance, Kulikovskiĭ and Lyubimov[13] found that in the presence of magnetic viscosity, only evolutionary shock waves possess a continuous profile for any relationship between the dissipation coefficients. A different introduction of dissipation effects presupposes allowance for finite electrical or thermal conductivity. The method of smearing via the tensor (8), which is used in this paper, has been chosen because of its simplicity and the possibility of comparing the results with those of a nonmagnetic case.

Below we consider shock waves that are obtained irrespective of the way in which the limit of $\eta\to 0$ and $\zeta\to 0$ is attained. This makes it possible to limit ourselves to the case where $\eta=0$, and one viscosity coefficient $\zeta\neq 0$ is sufficient for the shock wave to have a continuous profile and to block the evolution conditions.

A stationary shock wave propagating in the spacelike direction l_μ can be locally represented in an appropriate reference frame by a stationary viscous flow that depends only on one variable, $x=x^\mu l_\mu$. We select a system of coordinates such that $\{l_\mu\}=\{0,1,0,0\}$ and $x=x^1$. Then, as $x\to-\infty$, all viscous-flow parameters tend to constant values, to which we attach the index ''0,'' corresponding to the state in front of the shock wave; as $x\to\infty$, they tend to values to which we attach the index ''1,'' corresponding to the state behind the shock wave. Accordingly, $u^1>0$.

Equations (7), (3), and (4) yield the following constants of the motion:

$$T^{1\nu}+\tau^{1\nu}=\mathrm{const}, \tag{9}$$

$$u^1h^\nu-h^1u^\nu\equiv H^\nu=\mathrm{const}, \tag{10}$$

$$nu^1\equiv j=\mathrm{const}. \tag{11}$$

Since $\tau^{\mu\nu}\to 0$ as $x\to\pm\infty$, Eqs. (5) and (6) are valid for the corresponding asymptotic values of $T^{\mu\nu}$, n, and h^{μ} obtained from the continuous solutions of the system (9)–(11). Reasoning by analogy with classical hydrodynamics,[4] we find that the existence conditions for the solutions of this boundary value problem can be interpreted as the admissibility conditions for the corresponding shock transition $u^{\mu}_{(0)}, h^{\mu}_{(0)}, n_0, p_0 \to u^{\mu}_{(1)}, h^{\mu}_{(1)}, n_1, p_1$, and these states must satisfy Eqs. (5) and (6) for a shock wave.

Next, without loss of generality, we can put $u^3\equiv 0$ and $h^3\equiv 0$; the components u^1 and h^1 are normal to the surface $x=\text{const}$ parallel to the shock front, and u^2 and h^2 are the tangential components. We are still free to select the reference frame, so we choose one in which $u^2_{(0)}=0$.

Equation (10) yields

$$h^{\mu}=\frac{1}{u^1}[H^{\mu}-u^{\mu}H^{\alpha}u_{\alpha}]. \tag{12}$$

Multiplying (9) by u_{ν} and allowing for the fact that $\tau^{\mu\nu}u_{\nu}=0$, we find that

$$\varepsilon^{*}u^1=T^{1\mu}_{(0)}u_{\mu}, \tag{13}$$

which can be used to express ε and h^{μ} in terms of u^1 and u^2.

Eliminating the single derivative du^1/dx from (9) at $\eta=0$ for $\mu=1,2$, we arrive at a relationship that links u^1 and u^2. Combining this with (12) and (13), we obtain

$$(H^2u^2-H^0u^0)(H^2u^0-H^0u^2)=4\pi u^1(T^{10}_{(0)}u^2-T^{12}_{(0)}u^0).$$

This leads to a relationship between the components of the three-dimensional velocity, $v_1=u^1/u^0$ and $v_2=u^2/u^0$:

$$v_1=-A\frac{(v_2-a)(v_2-b)}{v_2-c}, \tag{14}$$

where

$$A=\frac{H^0H^2}{4\pi T^{10}_{(0)}},\quad a=\frac{H^2}{H^0},\quad b=\frac{1}{a},\quad c=\frac{T^{12}_{(0)}}{T^{10}_{(0)}},$$

which is linear in v_1 and quadratic in v_2. Here we are interested in the connected part of the curve (14), which contains the point $v_2=0$. At $\mu=1$ Eq. (9) yields

$$\zeta[1+(u^1)^2]\frac{du^1}{dx}=T^{11}-T^{11}_{(0)}. \tag{15}$$

Since velocities can be expressed in terms of the density n (see Eq. (11)),

$$u^1=\frac{u^1_{(0)}n_0}{n}, \tag{16}$$

the problem reduces to a study of a first-order ordinary differential equation in $n(x)$ or $u^1(x)$. Equations (12)–(16) fully determine the structure of the shock transition.

Using Eqs. (12)–(14) and (16), we can rewrite Eq. (15) in the form

$$\zeta\frac{n_0u^1_0}{n^2}\frac{dn}{dx}=\tilde{p}(n)-p(\tilde{\varepsilon}(n),n), \tag{17}$$

where we have introduced the notation

$$\tilde{p}(n)=\{1+(u^1)^2\}^{-1}\left\{T^{11}_{(0)}+\frac{1}{4\pi}(H^{\alpha}u_{\alpha})^2-T^{1\mu}_{(0)}u_{\mu}u^1\right\}-\frac{1}{8\pi}(u^1)^{-2}\{(H^{\alpha}u_{\alpha})^2-H^{\alpha}H_{\alpha}\},$$

$$\tilde{\varepsilon}(n)=\frac{T^{1\mu}_{(0)}u_{\mu}}{u^1}-\frac{1}{8\pi}(u^1)^{-2}\{(H^{\alpha}u_{\alpha})^2-H^{\alpha}H_{\alpha}\}.$$

The requirement that the right-hand side of Eq. (17) have a constant sign inside the interval (n_0,n_1) (the right-hand side vanishes at the endpoints of the interval) ensures the existence of a continuous solution whose asymptotic behavior is $n(x)\to n_0$ as $x\to-\infty$ and $n(x)\to n_1$ as $x\to\infty$. If, for instance, $n_1>n_0$ and the right-hand side of Eq. (17) is a positive, continuously differentiable function of $n\in(n_1,n_0)$, a solution with such properties does indeed exist. But if the right-hand side of Eq. (17) changes sign at a certain point between n_0 and n_1, there can be no such solution. One must make sure that there is no transition through a branch point of the function $v_2(u^1)$, where the solution can cease to be differentiable. This point corresponds to the extremum $v_2=v_2^{*}$ of the inverse function $u^1(v_2)$. To prove this we need only write Eq. (17) in the form

$$\zeta\frac{du^1}{dv_2}\frac{dv_2}{dx}=p(\tilde{\varepsilon}(n),n)-\tilde{p}(n),$$

in which $n=n(v_2)$ by virtue of (11) and (14), and the right-hand side is a single-valued function of v_2, and study it in the neighborhood of this point, where $du^1/dv_2\approx C(v_2-v_2^{*})$, $C\neq 0$. Assuming, *reductio ad absurdum*, that $v_2(x^{*})=v_2^{*}$ at a point x^{*}, we can easily see that the solution can be continued through this point only if the sign of the right-hand side of Eq. (17) also changes at the same point. But even if this coincidence occurs, it will not be preserved under small variations of the equation of state in the neighborhood of the point $\{\varepsilon(v_2^{*}),p(v_2^{*})\}$. Such variations would change the position of the root of the right-hand side of Eq. (17) but not the position of the point v_2^{*}, which depends on the equation of state only at points 0 and 1 (see Eq. (14)). Hence this case must be ruled out, and we must assume that u^1 and u^2 vary monotonically along the solutions.

Thus, we have arrived at the admissibility criterion for a stationary shock transition,

$$(n_0-n_1)(p(\tilde{\varepsilon}(n),n)-\tilde{p}(n))\geqslant 0, \tag{18}$$

for all n between n_0 and n_1, where it is assumed that Eq. (14) has in this interval a unique regular solution with respect to v_2. Here we cannot rule out the possibility that the curves p and $\tilde{p}$ touch (at the Chapman–Jouguet point), which is allowed in the weak-viscosity limit, when the left-hand side of Eq. (18) can vanish at a point between n_0 and n_1 but does not change sign. This criterion is considered a necessary one. It is not, however, a sufficient criterion, e.g., for an equation of state that leads to multivalued shock adiabats: from classical hydrodynamics we know that here there can be more than one discontinuous solution even if the existence conditions for a viscous profile of the shock wave are met, and removing this ambiguity requires imposing additional conditions of a nonhydrodynamic nature.[14,15]

The same criterion can be formulated in terms of the Hugoniot–Taub–Lichnerowicz shock adiabat[4,5] (see Appendix A), with the two forms being equivalent.

4. COROLLARIES OF CRITERION (18)

Below we will need one more relationship that is satisfied identically along the solutions of Eqs. (9)–(11). This relationship can be obtained if we allow for the explicit form of the energy–momentum tensor (1):

$$T^{12}_{(0)}H^0 - T^{10}_{(0)}H^2 = (T^{12}+\tau^{12})H^0 - (T^{10}+\tau^{10})H^2$$

$$= -\sqrt{1+(u^1)^2}\,\tilde{h}^2\left[R - \zeta (u^1)^2 \frac{du^1}{dx}\right]. \qquad (19)$$

Here $R=(p^*+\varepsilon^*)(u^1)^2-(h^1)^2/4\pi$, and $\tilde{h}^2$ is the transverse component of the magnetic field in the reference frame in which $u^2=0$:

$$\tilde{h}^2 = \frac{h^2(1+(u^1)^2)-u^1u^2h^1}{u^0\sqrt{1+(u^1)^2}} = u^0\frac{h^2-v_2h^0}{\sqrt{1+(u^1)^2}}.$$

Since we are interested in a continuous solution in the interval $(-\infty,\infty)$, from (19) it follows that $\tilde{h}^2$ does not change sign (or $\tilde{h}^2\equiv 0$ in front of and behind the shock wave). This implies that in relativistic MHD there can be no turning on or turning off of shock waves (cf. Ref. 5). Recall that the tangential magnetic field in front of a shock wave that is suddenly turned on is zero, and behind it the field is nonzero; such shock waves are admissible in classical MHD.[3] The fact that there can be no relativistic turning on or turning off of a shock wave has also been analyzed in detail using another method[5]

Since $du/dx\rightarrow 0$ as $x\rightarrow\pm\infty$ and the right-hand side of Eq. (19) at these limits is proportional to R, Eq. (19) implies that in a nontrivial case R does not change sign either. The relationship

$$(u_A)^2 = \frac{(h^1)^2}{4\pi(p^*+\varepsilon^*)}$$

determines the component of the 4-velocity of Alfvén waves propagating along the x axis, i.e., the sign of R is the same as that of the difference $(u^1)^2-(u^1_A)^2$, which, consequently, is conserved in the transition through the shock wave. Accordingly, if $R>0$, the equations describe in the weak-viscosity limit a fast shock wave, while if $R<0$, they describe a slow shock wave.

An additional condition imposed on velocities follows from the result that du^1/dv_2 must not change sign along an admissible path. Using Eq. ((14), we calculate this quantity in front of the shock front ($x\rightarrow-\infty$) in a reference frame in which $u^2_{(0)}=0$:

$$\left.\frac{du^1}{dv_2}\right|_{(0)} = -\frac{4\pi u^0}{h^1\tilde{h}^2}R^*\Big|_{(0)}, \qquad (20)$$

where

$$R^* = (p+\varepsilon)(u^1)^2[1+(u^1)^2] - \frac{1}{4\pi}(h^1)^2$$

contains components of velocity and magnetic field that are normal to the plane $x=\text{const}$, and has the same form in inertial reference frames moving parallel to that plane. Since the sign of du^1/dv_2 does not change in such motion, R^* determines this sign in an arbitrary reference frame of this type. And since du^1/dv_2 also retains its sign along a solution, the sign of $R^*_{(0)}$ in front of the shock front is the same as that of $R^*_{(1)}$ behind the shock front. Here we must use the fact that $\tilde{h}^2$ and h^1 retain their signs (see below) along an admissible path. Note that in the nonrelativistic limit R^* and R coincide.

We thus obtain a relationship similar to (19),

$$T^{12}_{(0)}H^2 - T^{10}_{(0)}H^0 = u^1h^1\left(p+\varepsilon-\zeta\frac{du^1}{dx}\right), \qquad (21)$$

which states that the sign of the longitudinal component h^1 of the magnetic field is preserved.

Now we analyze formula (14) in a reference frame in which $u^2_{(0)}=0$. We assume that $h^1_{(0)}\neq 0$ and $h^2_{(0)}\neq 0$. (Note that the cases of perpendicular ($h^1_{(0)}=0$) and parallel ($h^2_{(0)}=0$) shock waves are much simpler.[16]) In such a reference frame the expressions for the constants in (14) are

$$A=\frac{u^1_{(0)}c}{u^0_{(0)}}, \quad a=-\frac{u^1_{(0)}u^0_{(0)}h^2_{(0)}}{h^1_{(0)}}, \quad b=\frac{1}{a},$$

$$c=-h^1_{(0)}h^2_{(0)}\left[4\pi\left(p_0+\varepsilon_0+\frac{(h^2_{(0)})^2}{4\pi}\right)u^1_{(0)}u^0_{(0)}\right]^{-1}.$$

Using Eqs. (19) and (20), one can easily see that the case where c is between a and b corresponds to a negative R and describes a slow shock wave; here the branch of the dependence (14) containing the initial point $v_2=0$ is monotonic. But if c is outside the interval $[a,b]$, then $R>0$, with the result that the shock wave is fast; in this case the branch has an extremum.

Now we establish a relationship between the velocity of a shock wave and the velocities of magnetoacoustic waves in front of and behind the discontinuity. We expand (18) at 0 and 1. Allowing for (12)–(14) and (16), we see that direct calculations in the neighborhood of 0 yield

$$p-\tilde{p} = \left.\frac{d(p-\tilde{p})}{du^1}\right|_{u^1=u^1_{(0)}}(u^1-u^1_{(0)}) = \frac{(p_0+\varepsilon_0)^2}{u^1_{(0)}}D(u^1_{(0)})$$

$$\times(u^1-u^1_{(0)}),$$

where

$$D(u^1)=\frac{Q(u^1)}{R^*(u^1)},$$

$$Q(\xi)=\xi^4(1-c_s^2)-\xi^2\left(c_s^2+\frac{|h|^2}{4\pi(p+\varepsilon)}\right)+\frac{(h^1)^2c_s^2}{4\pi(p+\varepsilon)},$$

with $c_s^2=(\partial p/\partial\varepsilon)_S$ the speed of sound; all quantities were calculated in the neighborhood of 0.

Note that $Q(\xi)=0$ coincides with the equation for the normal components of the 4-velocityof fast and slow magnetoacoustic waves, ξ_f and $\xi_{\rm sl}$ (see Refs. 5 and 6), which are the roots of the function

$$Q(\xi)\equiv(1-c_s^2)(\xi^2-\xi_f^2)(\xi^2-\xi_{\rm sl}^2).$$

If we now use the condition (18) in the neighborhood of 0, we obtain $D(\xi)|_{\xi=u^1_{(0)}}>0$.

Similar reasoning for the final point 1 (with allowance for appropriate notation; now ξ_f and $\xi_{\rm sl}$ are calculated at this point) yields $D(\xi)|_{\xi=u^1_{(1)}}<0$. We see that if, for instance, in the initial state $(u^1_{(0)})^2>\xi^2_{f(0)}$, in the final state (with allowance for the fact that R does not change sign) we obtain $\xi^2_{f(1)}>(u^2_{(1)})^2>(u_{A(1)})^2$.

An additional lower bound on the velocity of fast shock waves with respect to the medium is obtained if we note that R^* has the same sign at 0 and 1. We denote the components of the 4-velocity at which R^* vanishes by

$$u_A^*=\sqrt{\frac{1}{2}\left(\sqrt{1+\frac{(h^1)^2}{\pi(p+\varepsilon)}}-1\right)}.$$

Allowing for the relationship between u_A and u_A^* (see Appendix B), we arrive at a relationship for the parameters in front of the shock front: $\xi^2_{f(1)}>(u^2_{(1)})^2>(u^*_{A(1)})^2\geqslant(u_{A(1)})^2$.

Examining the feasible alternatives with these inequalities in mind, we obtain relationships for the following three-dimensional velocities in front of and behind the shock front: the shock-wave velocity $v_{\rm sh}$, the fast and slow magnetoacoustic-wave velocities v_f and $v_{\rm sl}$, the Alfvén-wave velocity v_A, and v_A^*, which are all related to the components of the 4-velocities $u_{\rm sh}$, ξ_f, $\xi_{\rm sl}$, u_A, and u_A^* in the usual manner. For fast shock waves the characteristic velocities obey

$$u_{\rm sh(0)}>v_{f(0)}>v^*_{A(0)}>v_{A(0)}>v_{\rm sl(0)} \tag{22}$$

in front of and

$$u_{f(1)}>v_{\rm sh(1)}>v^*_{A(1)}>v_{A(1)}>v_{\rm sl(1)} \tag{23}$$

behind the shock front. For slow shock waves we have, respectively,

$$v_{A(0)}>v_{\rm sh(0)}>v_{\rm cl(0)}, \quad v_{\rm sl(1)}>v_{\rm sh(1)}. \tag{24}$$

Here allowance for the monotonic variation of velocities in a fast shock wave leads to the emergence of a forbidden region $[v_{A(1)},v^*_{A(1)}]$. Such a region is a hallmark of relativistic MHD: in the classical limit $v_{A(1)}\approx v^*_{A(1)}$, and the region disappears. We also note that for a parallel shock wave $v_A=v_A^*$; for a perpendicular shock wave both velocities vanish.

5. DISCUSSION

The necessary condition (18) for the existence of a continuous viscous profile of a shock wave in relativistic MHD also guarantees the growth of entropy, since it originates in the equations with viscosity. This criterion and its alternate form in terms of shock adiabats (see Appendix A) agree in the case of a zero magnetic field with the results of Bugaev *et al.*,[11,12] where a similar condition proves sufficient for the realization of a continuous profile of a shock and its evolution. When the magnetic field is taken into account, the relationships (24) for slow shock waves coincide with the well-known restrictions on shock-wave velocities,[5,6] which can be derived from the evolution conditions, while the inequalities (23) for fast shock waves are more stringent.

Note that condition (18) must be met over the entire interval $[n_0,n_1]$ and not only at the endpoints; if only for this reason it is more stringent that the evolution conditions. However, in a normal nonconducting medium (in nonmagnetic relativistic hydrodynamics), where the Poisson adiabats and hence the shock adiabats do not change the sign of convexity, the principal constraint is on the velocities of the shock and acoustic waves. Here, the result is the same, whether we start with the evolution conditions or with the assumption that there is a viscous profile,[11,12] as in the nonrelativistic case.[4] But when we consider the continuous structure of a shock wave in relativistic MHD, a new characteristic velocity v_A^* appears, and limits the velocity of fast shock waves with respect to the medium behind the shock front. In the nonrelativistic limit, this parameter coincides with the ordinary Alfvén velocity, and so do the admissible intervals of shock-wave velocities. Thus, the more stringent constraint on velocities is a reflection of the relativistic nature of the process; however, this constraint is lifted for perpendicular and parallel relativistic shock waves.

In conclusion we reiterate that here we have used a special model for smearing the discontinuity by introducing weak viscosity. This is not the only possibility, and we believe it would be interesting to examine other methods of regularization, in particular, those that allow for other dissipation effects.

APPENDIX A: ANOTHER FORM OF THE EXISTENCE CONDITIONS FOR A SHOCK TRANSITION

For a given initial state $u^\mu_{(0)}$, $h^\mu_{(0)}$, V_0, p_0, where $V=1/n$ is the specific volume, the shock adiabat $p_H(V)$ (see Refs. 5 and 6) determines the possible shock transitions via the conditions (5) and (6). The p_H vs. V dependence might not be single-valued. By their very definition, the curves $p(\tilde{\varepsilon}(V),V)$, $\tilde{p}(V)$, and $p_H(V)$ have common intersection points only: if at a point V^* we have $p(\tilde{\varepsilon}(V^*),V^*)=\tilde{p}(n^*)$, then (see the right-hand sides of (15) and (17)) that $T^{1\nu}_{(*)}=T^{1\nu}_{(0)}$, i.e., the point V^* belongs to the shock adiabat and $p(\tilde{\varepsilon}(V^*),V^*)=p_H(V^*)$. If there is a viscous profile, then by virtue of (15), these curves intersect only at the initial and final points and have no common points in between (with the possible exception of points of tangency).

Let us examine the relative position of these curves. By H we denote the function that defines a shock adiabat:[5]

$$H\equiv w^2V^2-w_0^2V_0^2-(wV^2+w_0V_0^2)(p-p_0)$$
$$+\frac{1}{2}(wV^2-w_0V_0^2)(\psi-\psi_0)^2=0,$$

where

$$w=\varepsilon+p, \quad \psi^2=h_n^2+|h|^2(1-u_n^2).$$

At the initial point of the adiabat,

$$\left.\frac{\partial H}{\partial p}\right|_{V=V_0}=\frac{2\tau_0 T_0}{V_0}\left.\frac{\partial S}{\partial p}\right|_{V=V_0}>0, \quad \tau=wV^2,$$

which means that H is positive above the initial point and negative below. Now we calculate the value of H on the curve $p(\tilde{\varepsilon}(V),V)$, i.e., $H(p(\tilde{\varepsilon}(V),V),V)$ in the neighborhood of the initial point. Some tedious but straightforward calculations yield

$$H(p(\tilde{\varepsilon}(V),V),V)$$

$$\approx\frac{\tau_0}{V_0}\left(\frac{dp(\tilde{\varepsilon}(V),V)}{dV}-\frac{d\tilde{p}(V)}{dV}\right)_0(V-V_0)^2. \tag{A1}$$

Bearing in mind that the sign of $p(\tilde{\varepsilon}(V),V)-\tilde{p}(V)$ in the neighborhood of the initial point determines the sign of (A1), we find that

$$(V_1-V_0)(p_H(V)-\tilde{p}(V))\geqslant 0 \tag{A2}$$

in the neighborhood of the initial point. This condition, when considered in conjunction with the fact that the shock adiabat does not intersect the template curve, is an equivalent form of the existence criterion for a shock wave. If, in addition, $p_H(V)$ is a single-valued function, then since these curves do not intersect between the initial and final points, the sign of the left-hand side of (A2) remains the same everywhere between these points. Condition (A2) must then hold over the entire interval between V_0 and V_1, and is equivalent to (18). This result agrees with the conclusions drawn in Refs. 11 and 12 for the case of zero magnetic field.

APPENDIX B: RELATIONSHIPS BETWEEN THE CHARACTERISTIC VELOCITIES

In both relativistic and classical MHD we have $v_f>v_A>v_{\rm sl}$ in front of and behind the shock front.

To compare the quantities u_A and u_A^*, which respectively make R and R^* vanish, we note that they do not depend on motion of the reference frame parallel to the shock front. Then, using a reference frame in which $u^2=0$ and allowing for the fact that the 4-vectors u^μ and h^μ are mutually orthogonal, we obtain

$$R=\left[p+\varepsilon+\frac{(h^2)^2}{4\pi}\right](u^1)^2-\frac{(h^1)^2}{4\pi[1+(u^1)^2]}.$$

Comparing the expressions for R^* and R, we find that $(u_A)^2\leqslant(u_A^*)^2$. Clearly, for a parallel shock wave we have $u_A=u_A^*$. For a perpendicular shock wave both velocities vanish.

In the nonrelativistic limit $p^*+\varepsilon^*\approx p+\varepsilon$, so that

$$(u_A^*)^2\approx\frac{(h^1)^2}{4\pi(p+\varepsilon)}\approx(u_A^1)^2,$$

and condition (23) coincides in this limit with the evolution conditions in classical MHD.

Substituting u_A^* into the explicit expression for $Q(\xi)$, we obtain

$$Q(u_A^*)\equiv(u_A^*)^4[(u_A^*)^2+1](1-c_S^2)-(u_A^*)^2\frac{p^*+\varepsilon^*}{p+\varepsilon}$$

$$+\frac{(h^1)^2c_S^2}{4\pi(p+\varepsilon)}=\frac{p^*+\varepsilon^*}{p+\varepsilon}[(u_A)^2-(u_A^*)^2]<0.$$

Since ξ_f^2 and $\xi_{\rm sl}^2$ are the roots of the quadratic trinomial $Q(\xi)$, we thereby obtain $\xi_{\rm sl}^2<(u_A)^2<\xi_f^2$.

*)E-mail: zhdan@aoku.freenet.kiev.ua

---

[1] L. D. Landau and E. M. Lifshitz, *Hydrodynamics* [in Russian], Pergamon Press, Oxford (1986).

[2] A. I. Akhiezer, G. Ya. Lyubarskiĭ, and R. V. Polovin, Zh. Éksp. Teor. Fiz. **35**, 731 (1958) [Sov. Phys. JETP **8**, 507 (1959)].

[3] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).

[4] B. L. Rozhdestvenskiĭ and N. N. Yanenko, *Systems of Quasilinear Equations* [in Russian], Nauka, Moscow (1978).

[5] A. A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics*, W. A. Benjamin, New York (1967).

[6] N. R. Sibgatullin, *Oscillations and Waves in Strong Gravitational and Electromagnetic Fields* [in Russian], Nauka, Moscow (1984).

[7] L. Van Hove, Z. Phys. C **21**, 93 (1983).

[8] H. W. Barz, L. P. Csernai, B. Kampfer, and B. Lukács, Phys. Rev. D **32**, 115 (1985).

[9] P. Danielewicz and P. V. Ruuskanen, Phys. Rev. D **35**, 344 (1987).

[10] J. P. Blaizot and J. Y. Ollitrault, Phys. Rev. D **36**, 916 (1987).

[11] K. A. Bugaev, M. I. Gorenstein, and V. I. Zhdanov, Z. Phys. C **39**, 365 (1988).

[12] K. A. Bugaev, M. I. Gorenshteĭn, and V. I. Zhdanov, Teor. Mat. Fiz. **80**, 138 (1989).

[13] A. G. Kulikovskiĭ and G. A. Lyubimov, Prikl. Mat. Mekh. **25**, 125 (1961).

[14] R. Menikoff and B. J. Plohr, Rev. Mod. Phys. **61**, 75 (1989).

[15] N. N. Kuznetsov, Zh. Éksp. Teor. Fiz. **88**, 470 (1985) [Sov. Phys. JETP **61**, 275 (1985)].

[16] V. I. Zhdanov and P. V. Titarenko, Nonlin. Math. Phys. **4**, 214 (1997).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Dynamic properties of moderately concentrated magnetic liquids

A. Yu. Zubarev*) and A. V. Yushkov

Ural State University, 620083 Ekaterinburg, Russia

(Submitted 18 November 1997)
Zh. Éksp. Teor. Fiz. **114**, 892–909 (September 1998)

On the basis of statistical analysis, we derive expressions for the dynamic susceptibility, magnetization relaxation times, and the effective rheological characteristics of a moderately concentrated homogeneous ferrocolloid consisting of identical spherical ferroparticles suspended in a Newtonian liquid. The magnetic moment of a particle is assumed constant and rigidly ''frozen'' into the body of the particle. We also estimate how the magnetodipole and hydrodynamic interactions of the particles influence the effective dynamic properties of the ferrocolloid. © *1998 American Institute of Physics.* [S1063-7761(98)01009-9]

1. INTRODUCTION

One of the central goals of the physics of magnetic liquids (ferrocolloids) is to establish the relationship that exists between the macroscopic dynamic characteristics describing the properties of the colloid in the continuum approximation and the parameters characterizing it on the microscopic level, such as the size, shape, properties, and concentration of the particles and the properties of the liquid carrier medium. Despite a large body of data (see, e.g., Refs. 1–6), a theory that adequately describes the macroscopic dynamic properties of magnetic liquids has yet to be developed. This leads to a situation in which the results of many experiments are interpreted by the researchers differently.

The existing consistent theoretical models of the dynamic properties of magnetic liquids deal primarily with extremely dilute systems, in which interparticle interactions can be completely ignored (see, e.g., Refs. 7–9). However, in various ferrocolloids such interactions play an important, and sometimes crucial, role: they lead not only to correlations between particles but also to the emergence of drop,[10–12] chain,[13–15] and other heteroaggregates capable of radically altering the effective characteristics of the colloid. The effect of chain aggregates on the macroscopic properties of magnetic liquids was studied theoretically in Refs. 16–18. The analysis of the macroscopic properties of magnetic liquids is complicated by the fact that the ferrocolloids commonly used in experiments and applications are polydisperse. Since particles belonging to different fractions are involved in different mechanisms of magnetic moment relaxation (Brownian, Néel, etc.) and the contributions of the different fractions to the overall response are not additive in systems with interaction, the results of experiments that establish the properties of polydisperse magnetic liquids can hardly be expected to be interpreted correctly, even qualitatively.

Under such circumstances it is advisable to study monodisperse systems, both theoretically and experimentally, so that we can separately analyze the effect of each factor on the macroscopic properties of dense magnetic liquids. We believe that only this line of research will provide a coherent picture of the behavior and properties of these systems.

This paper is a theoretical study of the dynamic susceptibility and the rheological characteristics of a moderately concentrated ferrocolloid consisting of identical spherical particles with their magnetic moments being constant in magnitude and ''frozen'' into their bodies. We ignore the possibility of formation of drop, chain, and other heteroaggregates. The problem of the conditions for the emergence of such aggregates requires a special study.

Attempts to calculate the dynamic response functions for dense homogeneous ferrocolloids were made by Tsebers[19] and Kasherskiĭ.[20] However, they took the magnetodipole interaction of the particles into account by employing Weiss's theory of a self-consistent field. Such a model suggests that the interaction of the particles effectively manifests itself only after the ferrocolloid has gone into the ferromagnetic phase. However, there are no experimental indications that ferromagnetism is present in magnetic liquids or, to that matter, in other similar polar systems. Hence one should be careful when using the results of Refs. 19 and 20.

In Ref. 21 a consistent thermodynamic perturbation theory (a variant of the second virial coefficient approximation) was used to account for the magnetic interaction of particles in the analysis of the kinetics of alternating magnetization of moderately concentrated homogeneous ferrocolloids. Byevich and Ivanov[22] and Pshenichnikov[23] demonstrated that the theory could be used to describe the equilibrium properties of magnetic liquids. In Ref. 21 the magnetization relaxation equation was derived from the Fokker–Planck equation for the many-particle distribution function of the positions and orientations of the particles. The effective-field method[7] was employed to reduce this equation. (The high accuracy of this method as applied to dilute ferrocolloids has been demonstrated in the numerical experiments of Tsebers.[24]) However, in Ref. 21 the dynamics of magnetization variations was analyzed only for weakly nonequilibrium systems, which means that we still do not know the nature of the response of magnetic liquids to rapidly varying fields and the rheological properties of such liquids.

In the present paper we study the dynamic response of a moderately concentrated ferrocolloid that is arbitrarily far from equilibrium, and the rheological characteristics of such a system. Methodologically, the work is close to Ref. 21, and we also correct the errors discovered in that paper.

2. MATHEMATICAL MODEL AND BASIC EQUATIONS

Consider a ferrocolloid occupying a volume V and consisting of N identical spherical particles of radius a. The absolute value m of the magnetic moment $\mathbf{m}$ of a particle is constant: the moment is ''frozen'' into the body of the particle (the magnetic anisotropy energy is high). We denote the unit vector directed along the magnetic moment of the ith particle by $\mathbf{e}_i=\mathbf{m}_i/m$ and the radius vector of this particle by $\mathbf{r}_i$.

To calculate the macroscopic characteristics of the system, we must determine the N-particle distribution $P_N(\mathbf{e}_1,\dots,\mathbf{e}_N,\mathbf{r}_1,\dots,\mathbf{r}_N)$, which can be found by solving the appropriate Fokker–Planck equation, in which we must allow for the magnetic and hydrodynamic interactions of the particles with each other. The fact that this equation contains terms corresponding to nonpotential hydrodynamic interparticle interactions complicates the mathematics significantly. To simplify analysis and obtain results in a tangible form that can be analyzed from the standpoint of physics, instead of explicitly writing the appropriate terms in the Fokker–Planck we effectively allow for the hydrodynamic effect by using renormalized particle diffusion coefficients in that equation. These coefficients already allow for the hydrodynamic interparticle interactions. Such a model of an effective medium has been repeatedly employed with success in the theory of the macroscopic properties of suspensions and composite materials.

If we adopt a system of coordinates in which the average velocity $\mathbf{u}$ of the colloid as a whole is zero, the Fokker–Planck equation reduced in this manner has the form

$$\frac{\partial P_N}{\partial t}=\sum_i \hat{\mathbf{I}}_i\cdot\left[\left(\frac{D_r}{T}\hat{\mathbf{I}}_i U-\mathbf{\Omega}\right)P_N\right]+\sum_i \nabla_i\cdot\left[\left(\frac{D_t}{T}\nabla_i U\right)P_N\right]$$

$$+D_r\sum_i \hat{\mathbf{I}}_i^2 P_N+D_t\sum_i \nabla_i^2 P_N,$$

$$\hat{\mathbf{I}}_i=\mathbf{e}_i\times\frac{\partial}{\partial \mathbf{e}_i},\quad \nabla_i=\frac{\partial}{\partial \mathbf{r}_i},\quad U=-T\sum_i \boldsymbol{\alpha}\cdot\mathbf{e}_i+\frac{1}{2}\sum_{i\neq j} w_{ij},$$

$$\boldsymbol{\alpha}=\frac{m\mathbf{H}}{T},\quad \mathbf{\Omega}=\frac{1}{2}\,\mathrm{curl}\,\mathbf{u},$$

$$w_{ij}=m^2\frac{(\mathbf{e}_i\cdot\mathbf{e}_j)r_{ij}^2-3(\mathbf{e}_i\cdot\mathbf{r}_{ij})(\mathbf{e}_j\cdot\mathbf{r}_{ij})}{r_{ij}^5}. \tag{1}$$

Here T is the absolute temperature in energy units, D_r and D_t are the effective coefficients of rotational and translational diffusion of the particles calculated with allowance for the hydrodynamic and steric interactions of the particles, U is the total potential energy of the particle system, and w_{ij} is the energy of the dipole–dipole interactions of the ith and jth particles. Summation in (1) is over particle numbers. In solving Eq. (1) we take into account the condition that the particles do not overlap: $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|\geqslant 2a$.

Even the reduced Fokker–Planck equation (1) cannot be solved exactly. Here, as in Ref. 21, we use the effective-field approximation,[7] a variant of the trial-function method. According to Ref. 7, the solution of Eq. (1) must be sought in the form of an equilibrium function with respect to an effective potential energy U_e, which differs from U in Eq. (1) in that instead of the true magnetic field $\mathbf{H}$ it contains an effective field $\mathbf{H}_e$, which must be found. Thus, we postulate the validity of the equation

$$\sum_i \hat{\mathbf{I}}_i\cdot\left[\left(\frac{D_r}{T}\hat{\mathbf{I}}_i U_e\right)P_N\right]+\sum_i \nabla_i\cdot\left[\left(\frac{D_t}{T}\nabla_i U_e\right)P_N\right]$$

$$+D_r\sum_i \hat{\mathbf{I}}_i^2 P_N+D_t\sum_i \nabla_i^2 P_N=0, \tag{2}$$

$$U_e=-T\sum_i \boldsymbol{\alpha}_e\cdot\mathbf{e}_i+\frac{1}{2}\sum_{i\neq j} w_{ij},\quad \boldsymbol{\alpha}_e=\frac{m\mathbf{H}_e}{T}.$$

Subtracting (2) from (1), we get

$$\frac{\partial P_N}{\partial t}=-\sum_i \hat{\mathbf{I}}_i\cdot\left[\left(\frac{D_r}{T}\hat{\mathbf{I}}_i[\delta\boldsymbol{\alpha}\cdot\mathbf{e}_i]+\mathbf{\Omega}\right)P_N\right], \tag{3}$$

where $\delta\boldsymbol{\alpha}=\boldsymbol{\alpha}_e-\boldsymbol{\alpha}$.

Up to this point all the transformations were exact: simply put, instead of the unknown function P_N we introduced a new unknown $\boldsymbol{\alpha}_e$ linked to P_N by the Gibbs formula

$$P_N=Z^{-1}\exp\left(-\frac{U_e}{T}\right),$$

$$Z=\int \exp\left(-\frac{U_e}{T}\right)d\mathbf{e}_1\cdots d\mathbf{e}_N\,d\mathbf{r}_1\cdots d\mathbf{r}_N. \tag{4}$$

The crucial assumption of the method developed in Ref. 7 is that the effective field $\mathbf{H}_e$ is independent of the vectors $\mathbf{e}_i$ and $\mathbf{r}_i$ and that its components can be found from the equation for the first moment of $\mathbf{e}_1$. Obviously, after this assumption is made, Eq. (3) ceases to be equivalent to Eq. (1) and the function P_N that this new equation yields is only an approximation to the true function satisfying Eq. (1). As noted earlier, Tsebers[24] and Pokrovskiĭ[25] showed that this approximation is extremely accurate when applied to dilute ferrocolloids. Note that similar ideas have been successfully used in analyzing the dynamic properties of polar liquid crystal systems.[26,27]

As is common in statistical physics, interparticle interactions do not make it possible to exactly calculate the average values of quantities via (4). From now on we assume that the concentration of the colloid is low or moderate and use the second virial coefficient approximation. As noted earlier, the fact that this approximation produces adequate results in describing the properties of many real ferrocolloids has been demonstrated in Refs. 22 and 23.

First, we will find it convenient if P_N is averaged over the particle coordinates. Introducing the Mayer function $f_{ij}=\exp(-w_{ij}/T)-1$ and averaging (4) over all the $\mathbf{r}_k$, we obtain

$$p_N=\int P_N\prod_k d\mathbf{r}_k=Z^{-1}\exp\left(\boldsymbol{\alpha}_e\cdot\sum_i \mathbf{e}_i\right)$$
$$\times\int\prod_{i>j}(1+f_{ij})\prod_k d\mathbf{r}_k. \tag{5}$$

Expanding (5) in a power series in f_{ij}, keeping only the first two terms, and performing standard transformations, we get

$$p_N=\left(\prod_k\varphi_k\right)\left(1-\rho(N-1)G_e+\frac{1}{V}\sum_{i>j}Q_{ij}\right), \tag{6}$$
$$\rho=\frac{N}{V}v,\quad v=\frac{4}{3}\pi a^3,\quad \varphi_i=\varphi(\mathbf{e}_i)=\frac{\exp(\boldsymbol{\alpha}_e\cdot\mathbf{e}_i)}{z_1},$$
$$z_1=\int\exp(\boldsymbol{\alpha}_e\cdot\mathbf{e})\,d^3e=4\pi\frac{\sinh\alpha_e}{\alpha_e},$$
$$Q_{ij}=\int_{r_{ij}>2a}f_{ij}\,d\mathbf{r}_{ij},\quad \mathbf{r}_{ij}=\mathbf{r}_i-\mathbf{r}_j,$$
$$G_e=\frac{1}{2v}\langle\varphi_1\varphi_2Q_{12}\rangle_{12},$$
$$\langle\cdots\rangle_{i_1\cdots i_n}=\int\cdots\prod_{k=1}^{n}d\mathbf{e}_{i_k},\quad i_k=1,\ldots,N.$$

Here and in what follows, ρ is the volume concentration of the particles.

When calculating the integral in the expression for Q_{ij}, we must bear in mind that the result depends on the shape of the infinite volume over which the integration is done. Proper selection of this volume was done by Byevich and Ivanov.[22] Unfortunately, the complicated form of the Mayer function makes an analytic calculation of Q_{ij} impossible. Here, as in Ref. 22, we limit ourselves to an analysis of a system in which w_{ij} is small in comparison to the thermal energy of the system. Note that this assumption means that there are no heteroaggregates in the ferrocolloid.

Expanding the Mayer function in a power series in w_{ij}, keeping only linear terms, and using the method of Ref. 22 to calculate the integral in Q_{ij}, we obtain

$$Q_{ij}=8\gamma v\,\mathbf{e}_i\cdot\mathbf{e}_j,\quad G(x)=4\gamma L^2(x),\quad L(x)=\coth x-\frac{1}{x},$$
$$G=G(\alpha),\quad G_e=G(\alpha_e),\quad \gamma=\frac{m^2}{8a^3T}. \tag{7}$$

The parameter γ characterizes the ratio of the dipole–dipole interaction of two closely located particles to the thermal energy of the system.

Averaging (3) over the radius vectors of the particles, we arrive at an equation that is identical to Eq. (3), with P_N replaced by p_N. Multiplying the resulting equation by $\mathbf{e}_1$ and averaging over all $\mathbf{e}_i$, we obtain

$$\frac{\partial\boldsymbol{\mu}}{\partial t}=-D_r\left\langle\mathbf{e}_1\sum_i\hat{\mathbf{I}}_i([\delta\boldsymbol{\alpha}\cdot\mathbf{e}_i]p_N)\right\rangle-\left\langle\mathbf{e}_1\boldsymbol{\Omega}\sum_i\hat{\mathbf{I}}_ip_N\right\rangle,$$
$$\langle\cdots\rangle=\langle\cdots\rangle_{1\cdots N},\quad \boldsymbol{\mu}=\langle\mathbf{e}_1p_N\rangle, \tag{8}$$

where $\boldsymbol{\mu}$ is the average value of the vector $\mathbf{e}_1$ of a particle. Obviously, the colloid's magnetization is

$$\mathbf{M}=mn\boldsymbol{\mu},\quad n=\frac{N}{V}=\frac{\rho}{v}. \tag{9}$$

Using (6) and (7), we easily find that

$$\boldsymbol{\mu}=\mu_e\mathbf{h}_e,\quad \mu_e=L_e+\frac{N-1}{V}v\frac{dG_e}{d\alpha_e}, \tag{10}$$
$$L_e=L(\alpha_e),\quad \mathbf{h}_e=\frac{\mathbf{H}_e}{H_e}.$$

In the thermodynamic limit,

$$\mu_e=L_e+\rho\frac{dG_e}{d\alpha_e}. \tag{11}$$

If we employ the fact that the angular momentum operator $\hat{\mathbf{I}}_i$ is antihermitian and use the approximation (6) for p_N, we get

$$\left\langle\mathbf{e}_1\sum_i\hat{\mathbf{I}}_i([\delta\boldsymbol{\alpha}\cdot\mathbf{e}_i]p_N)\right\rangle=\langle\boldsymbol{\xi}p_N\rangle=\langle\boldsymbol{\xi}\varphi_1\rangle_1$$
$$+\frac{N-1}{V}v[\langle\boldsymbol{\xi}b\varphi_1\rangle_1$$
$$-2G_e\langle\boldsymbol{\xi}\varphi_1\rangle_1], \tag{12}$$

where $\boldsymbol{\xi}=\mathbf{e}_1\times(\delta\boldsymbol{\alpha}\times\mathbf{e}_i)$ and $b=1/v\,\langle\varphi_2Q_{12}\rangle_2$.

In the thermodynamic limit,

$$\left\langle\mathbf{e}_1\sum_i\hat{\mathbf{I}}_i([\delta\boldsymbol{\alpha}\cdot\mathbf{e}_i]p_N)\right\rangle=\langle\boldsymbol{\xi}\varphi_1\rangle_1+\rho[\langle\boldsymbol{\xi}b\varphi_1\rangle_1$$
$$-2G_e\langle\boldsymbol{\xi}\varphi_1\rangle_1]. \tag{13}$$

Allowing for the antihermitian nature of $\hat{\mathbf{I}}_i$ and using (6) and (10), we easily find that

$$\left\langle\mathbf{e}_1\boldsymbol{\Omega}\sum_i\hat{\mathbf{I}}_ip_N\right\rangle=\langle\mathbf{e}_1\times\boldsymbol{\Omega}p_N\rangle=\langle\mathbf{e}_1p_N\rangle\times\boldsymbol{\Omega}$$
$$=-\mu_e\boldsymbol{\Omega}\times\mathbf{h}_e. \tag{14}$$

If we now combine (6), (7), and (13) and perform simple transformations, we obtain

$$\left\langle\mathbf{e}_1\sum_i\hat{\mathbf{I}}_i([\delta\boldsymbol{\alpha}\cdot\mathbf{e}_i]p_N)\right\rangle=A_e\delta\boldsymbol{\alpha}-B_e(\mathbf{h}_e\cdot\delta\boldsymbol{\alpha})\mathbf{h}_e, \tag{15}$$
$$A_e=A(\alpha_e),\quad B_e=B(\alpha_e),$$
$$A(x)=1-\frac{L(x)}{x}+8\rho\gamma\frac{L(x)}{x}(L^2(x)-C(x)),$$
$$B(x)=C(x)+24\rho\gamma(L^2(x)-C(x)),$$
$$C(x)=1-3\frac{L(x)}{x}.$$

Allowing for (14) and (15), we can write Eqs. (8) and (10) as

$$\frac{\partial \boldsymbol{\mu}}{\partial t} = -D_r[A_e\delta\boldsymbol{\alpha} - B_e(\mathbf{h}_e\cdot\delta\boldsymbol{\alpha})\mathbf{h}_e] + \mu_e\boldsymbol{\Omega}\times\mathbf{h}_e,$$

$$\boldsymbol{\mu}=\mu_e\mathbf{h}_e, \quad \mu_e = L_e + \rho\frac{dG_e}{d\alpha_e},$$

$$\delta\boldsymbol{\alpha}=\boldsymbol{\alpha}_e-\boldsymbol{\alpha}, \quad \boldsymbol{\alpha}_e=\alpha_e\mathbf{h}_e. \tag{16}$$

We have arrived at a system of equations for the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}_e$. For subsequent calculations it is convenient to write this system in the form of a single equation with respect to $\boldsymbol{\alpha}_e$. To this end we employ the fact that $\boldsymbol{\mu}=\mu_e\mathbf{h}_e$ and write

$$\frac{d\boldsymbol{\mu}}{dt} = J_e\dot{\alpha}_e\mathbf{h}_e + \mu_e\dot{\mathbf{h}}_e, \quad J_e = \frac{d\mu_e}{d\alpha_e} \tag{17}$$

(as usual, the dot indicates a time derivative).

Allowing for (7) and (16), in the linear approximation in $\rho\gamma$ we get

$$J_e = \frac{dL_e}{d\alpha_e} + 8\rho\gamma\left(\left(\frac{dL_e}{d\alpha_e}\right)^2 + L_e\frac{d^2L_e}{d\alpha_e^2}\right). \tag{18}$$

Plugging (17) into the first equation in (16) and writing the scalar product of the result and the vector $\mathbf{h}_e$, we get

$$\dot{\alpha}_e = \frac{D_r}{J_e}(B_e - A_e)\mathbf{h}_e\cdot\delta\boldsymbol{\alpha}. \tag{19}$$

If we now plug (19) into (17) and the result into (16), we arrive at the equation

$$\dot{\mathbf{h}}_e = -D_r\frac{A_e}{\mu_e}(\delta\boldsymbol{\alpha} - (\mathbf{h}_e\cdot\delta\boldsymbol{\alpha})\mathbf{h}_e) + \boldsymbol{\Omega}\times\mathbf{h}_e, \quad |\mathbf{h}_e| = 1. \tag{20}$$

Equations (19) and (20) form a system of equations for finding α_e and $\mathbf{h}_e$, which can easily be reduced to a single equation:

$$\frac{d\boldsymbol{\alpha}_e}{dt} = -D_r\left[\frac{A_e - B_e}{\alpha_e^2 J_e}(\boldsymbol{\alpha}_e\cdot\delta\boldsymbol{\alpha})\boldsymbol{\alpha}_e + \frac{A_e}{\mu_e}\alpha_e\left(\delta\boldsymbol{\alpha} - \frac{\boldsymbol{\alpha}_e\cdot\delta\boldsymbol{\alpha}}{\alpha_e^2}\boldsymbol{\alpha}_e\right)\right] + \boldsymbol{\Omega}\times\boldsymbol{\alpha}_e. \tag{21}$$

Finding $\boldsymbol{\alpha}_e$ from (21) or (19) and (20) and plugging the result into (10) and (11) or into the second and third relations in (16), we arrive at an expression for the average vector $\boldsymbol{\mu}$. If we then plug the expression for $\boldsymbol{\mu}$ into (9), we obtain the nonequilibrium magnetization of the ferrocolloid. As $\rho\to 0$, Eq. (21) becomes the equation for the effective field derived in Ref. 7. Note that in deriving (21) we corrected the errors that were made in deriving a similar expression in Ref. 21.

3. DYNAMIC SUSCEPTIBILITY

Equation (21) is nonlinear and, generally speaking, can be solved only numerically. In this section we study the response of a ferrocolloid that macroscopically is at rest ($\boldsymbol{\Omega} = 0$) to a linearly polarized oscillating magnetic field. Let

$$H_x = H_0\cos\omega t, \quad H_y = H_z = 0. \tag{22}$$

We write the magnetization, defined in (9), in the form

$$M(t) = \sum_{n=0}^{\infty} M_1(n\omega)\cos n\omega t + \sum_{n=1}^{\infty} M_2(n\omega)\sin n\omega t,$$

$$M_1(n\omega) = \frac{1}{T}\int_T M(t)\cos n\omega t\,dt,$$

$$M_2(n\omega) = \frac{1}{T}\int_T M(t)\sin n\omega t\,dt, \quad T = \frac{2\pi}{\omega}. \tag{23}$$

Now we define the real χ'_ω and imaginary χ''_ω parts of the magnetic susceptibility in the following manner:

$$\chi'_\omega = \frac{M_1(\omega)}{H_0}, \quad \chi''_\omega = \frac{M_2(\omega)}{H_0}. \tag{24}$$

For H_0 small and the dependence of M on H linear, the definitions (24) coincide with the ordinary definitions for the components of the complex-valued dynamic susceptibility.

The system of equations (19) and (20) was solved numerically with respect to α_e and $\mathbf{h}_e$ for

$$\boldsymbol{\Omega} = 0, \quad \delta\boldsymbol{\alpha} = \alpha_e\mathbf{h}_r - \boldsymbol{\alpha},$$

$$\boldsymbol{\alpha} = (\alpha_0\cos\omega t, 0, 0), \quad \alpha_0 = \frac{mH_0}{T}. \tag{25}$$

The result was then plugged into the second and third relationships in (16), which together with (9) yielded $M(t)$. After numerical Fourier transformations were carried out, the Fourier transforms of $\boldsymbol{\mu}$ were calculated,

$$\mu_1(\omega) = \frac{M_1(\omega)}{mn}, \quad \mu_2(\omega) = \frac{M_2(\omega)}{mn},$$

and so were the quantities

$$\kappa'(\omega) = \frac{\mu_1(\omega)}{\alpha_0}, \quad \kappa''(\omega) = \frac{\mu_2(\omega)}{\alpha_0}, \tag{26}$$

which correspond to the nonlinear response of a particle at the external field frequency.

Equations (19) and (20) contain the effective rotational diffusion coefficient D_r. In the approximation of two hydrodynamically interacting particles (similar to the second virial coefficient approximation) this diffusion coefficient was estimated by Perez-Madrid and Pubi:[28]

$$D_r = D_0\left(1 - \frac{14}{5}\rho\right), \quad D_0 = \frac{T}{6\eta_0 v}, \tag{27}$$

where η_0 is the viscosity coefficient of the liquid carrier medium, and D_0 is the rotational diffusion coefficient for a single particle. Note that the value of D_r used in Ref. 21, chosen intuitively, differs somewhat from the value given by (27), which we use below.

The results of calculating κ' and κ'' are depicted in Figs. 1–3. Figure 1 shows that at low frequencies the real part κ' of the specific susceptibility increases with the parameter γ of the magnetodipole interaction of the particles, while at high frequencies it decreases. This agrees with the well-known fact that the magnetodipole interparticle interaction

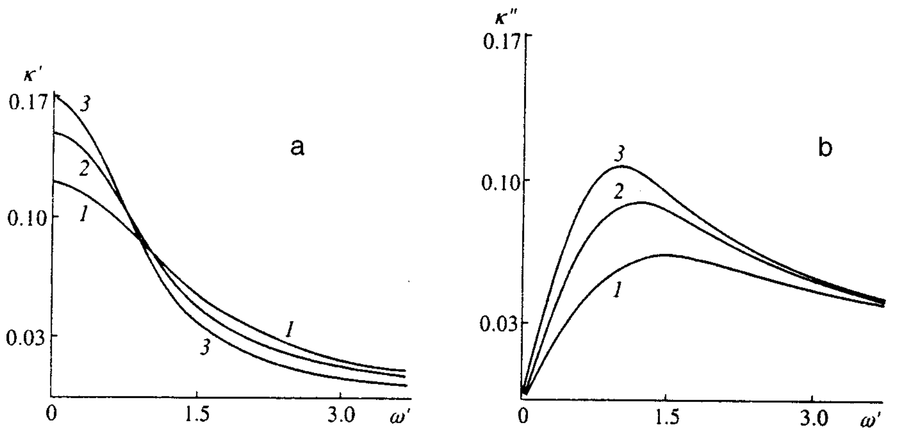


FIG. 1. Plots of κ' (a) and κ'' (b) vs. the dimensionless frequency $\omega'=\omega/D_0$ of an alternating magnetic field at $\alpha_0=3$ and $\rho=0.15$. Curves *1* correspond to $\gamma=0$, curves *2* to $\gamma=1.5$, and curves *3* to $\gamma=2.5$.

enhances the dc susceptibility of magnetic liquids (see, e.g., Ref. 23). At the same time, at high frequencies ω, collective effects hinder the rotation of particles, which results in κ' decreasing with increasing γ. The same physical reasons lead to an increase in κ'' and to a shift of its maximum to the left along the frequency axis as γ increases.

Figure 2 shows that as the volume concentration ρ increases, the parameter κ' at low frequencies also increases, which again agrees with the well-known properties of dc susceptibility. At high frequencies, κ' decreases with increasing ρ, since ‘‘braking’’ effects of the magnetodipole interaction of the particles begin to manifest themselves and the decrease of the effective rotational mobility D_r/T of the particles comes into play. A comparison of Figs. 1 and 2 reveals that hydrodynamic interparticle interactions have a stronger effect on the dynamic susceptibility (especially on its imaginary part) than dipole–dipole interactions.

Figure 3 shows that an increase in the magnetic field strength reduces, on the whole, both κ' and κ'' and shifts the maximum in κ'' to the left along the frequency axis. This is understandable since an increase in field strength without limit cannot lead to a similar increase in magnetization—it reaches a plateau and becomes saturated; the characteristic time of response of the system to variations in field strength decrease in the process.

Our approach can be used to calculate response functions at frequencies that are integral multiples of the frequency ω of the external field. Such calculations were done, and we found that the corresponding susceptibilities were smaller by several orders of magnitude than the susceptibilities κ' and κ'' at the signal frequency.

4. RELAXATION TIMES

Analytic treatment of (16) is possible only for weakly nonequilibrium processes, in which $\boldsymbol{\mu}(t)$ differs only slightly from its equilibrium value corresponding to the current magnetic field $\mathbf{H}(t)$. Formally this means that $\delta\alpha(t)$ in (16) is small. In the approximation linear in $\delta\alpha(t)$, instead of (16) we have

$$\frac{\partial\boldsymbol{\mu}}{\partial t}=-D_r[A\,\delta\boldsymbol{\alpha}-B(\mathbf{h}\cdot\delta\boldsymbol{\alpha})\mathbf{h}],$$

$$\boldsymbol{\mu}=\boldsymbol{\mu}(\boldsymbol{\alpha}+\delta\boldsymbol{\alpha})=\boldsymbol{\mu}_0(\alpha)+\frac{\mu_0}{\alpha}\,\delta\boldsymbol{\alpha}+\left(J-\frac{\mu_0}{\alpha}\right)(\mathbf{h}\cdot\delta\boldsymbol{\alpha})\mathbf{h},$$

$$\boldsymbol{\mu}_0=\mu_0\mathbf{h},\quad \mu_0=L(\alpha)+\rho\frac{dG(\alpha)}{d\alpha},$$

$$A=A(\alpha),\quad B=B(\alpha),\quad J=J(\alpha),\quad \mathbf{h}=\frac{\mathbf{H}}{H},\tag{28}$$

where $\boldsymbol{\mu}_0$ is the equilibrium value of the vector $\boldsymbol{\mu}$ corresponding to $\mathbf{H}(t)$. In deriving the first relationship in (28) we again assumed that macroscopically the colloid is at rest $(\Omega=0)$. The second relationship in (28) is the expansion of $\boldsymbol{\mu}(\boldsymbol{\alpha}+\delta\boldsymbol{\alpha})$ in a Taylor series and is independent of Ω.

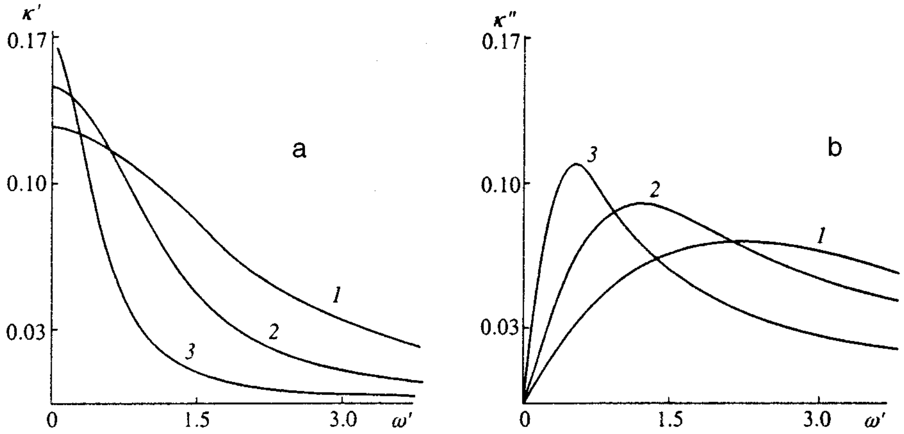


FIG. 2. Plots of κ' (a) and κ'' (b) vs. the dimensionless frequency $\omega'=\omega/D_0$ of an alternating magnetic field at $\alpha_0=3$ and $\gamma=1.5$. Curves *1* correspond to $\rho=0.05$, curves *2* to $\rho=0.15$, and curves *3* to $\rho=0.25$.

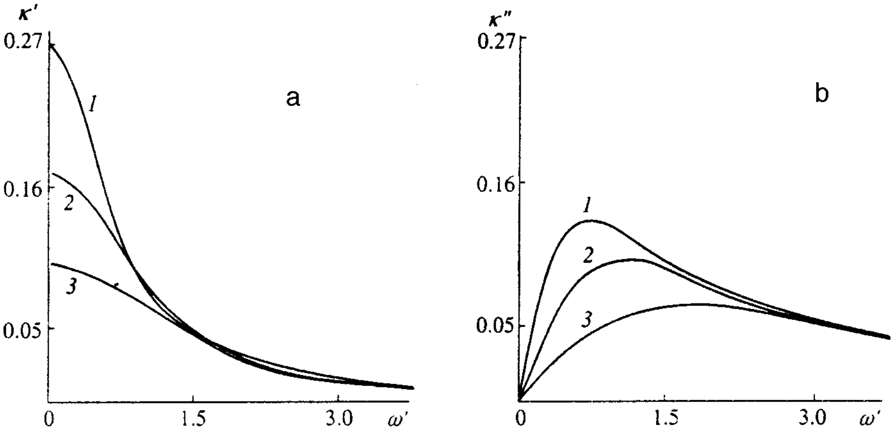


FIG. 3. Plots of κ' (a) and κ'' (b) vs. the dimensionless frequency $\omega' = \omega/D_0$ of an alternating magnetic field at $\gamma = 1.5$ and $\rho = 0.15$. Curves *1* correspond to $\alpha_0 = 0.01$, curves *2* to $\alpha_0 = 2.5$, and curves *3* to $\alpha_0 = 5$.

We introduce a local system of coordinates whose z axis is directed along $\mathbf{H}$. This means that we can write (28) as follows:

$$\frac{d\mu_{x,y}}{dt} = -D_r A\,\delta\alpha_{x,y}, \quad \frac{d\mu_z}{dt} = -D_r(A-B)\,\delta\alpha_z,$$

$$\mu_{x,y} = \mu_0 \frac{\delta\alpha_{x,y}}{\alpha}, \quad \mu_z = \mu_0 + J\,\delta\alpha_z. \tag{29}$$

Eliminating the components of $\delta\boldsymbol{\alpha}$ and using (9), we arrive at the magnetization relaxation equations

$$\frac{dM_{x,y}}{dt} = -\frac{1}{\tau_\perp} M_{x,y}, \quad \frac{dM_z}{dt} = -\frac{1}{\tau_\parallel}(M_z - M_0), \tag{30}$$

with $M_0 = mn\mu_0$ the equilibrium magnetization in the field H, and

$$\tau_\parallel = \left(D_r \frac{A-B}{J} \right)^{-1} = \tau_{0\parallel} \frac{1 + 8\rho\gamma(W + LYW^{-1})}{1 + 8\rho\gamma(C - L^2)},$$

$$\tau_\perp = \left(D_r \frac{\alpha A}{\mu_0} \right)^{-1} = \tau_{0\perp} \frac{1 + 8\rho\gamma W}{\alpha - L + 8\rho\gamma(L^2 - C)} (\alpha - L),$$

$$\tau_{0\parallel} = \frac{\alpha}{2D_r L} \frac{\partial L}{\partial \alpha}, \quad \tau_{0\perp} = \frac{1}{D_r} \frac{L}{\alpha - L},$$

$$W = \frac{dL}{d\alpha}, \quad Y = \frac{d^2 L}{d\alpha^2}. \tag{31}$$

The parameters $\tau_\parallel$ and $\tau_\perp$ are the times it takes the components of magnetizations parallel and perpendicular to the field to relax to their equilibrium values, and $\tau_{0\parallel}$ and $\tau_{0\perp}$ are the values of these times at $\gamma = 0$. The results of calculating $\tau_\parallel$ and $\tau_\perp$ are depicted in Figs. 4–8. A combination of the following factors determines the above dependence:

1. An increase in the strength of the magnetodipole interaction of the particles (both p and γ increase) leads to an increase in μ_0 and hence to an increase in the strength with which the particles are coupled to the field and a decrease in the relaxation times τ.

2. The magnetodipole interaction of the particles makes the configuration of their moments more stable, which hinders particle rotation in the field and hence enhances τ. Thus, the magnetic interaction of the particles affects τ in two ways, and the predominant behavior of τ depends on the conditions. In Ref. 21 it was shown that at large values of α, an increase in γ leads to a decrease in τ, while at small values of α, an increase in γ leads to an increase in τ.

3. The hydrodynamic interaction of the particles, which becomes stronger as ρ increases, leads to an increase in τ.

4. As the magnetic field strength grows, the strength with which the particle moment is coupled to the field increases, which leads to a decrease in τ.

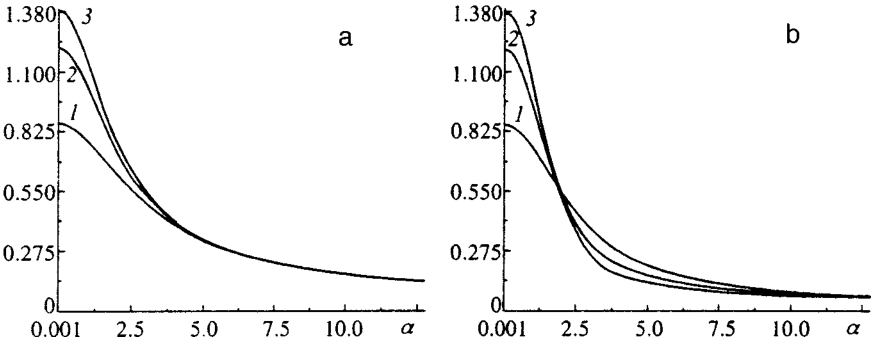


FIG. 4. Plots of $\tau_\perp$ (a) and $\tau_\parallel$ (b) vs. the constant external magnetic field α at $\rho = 0.15$. Curves *1* correspond to $\gamma = 0$, curves *2* to $\gamma = 1$, and curves *3* to $\gamma = 1.5$.

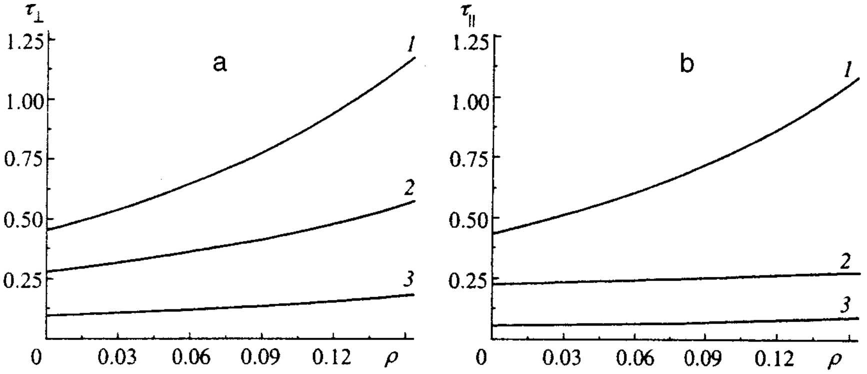


FIG. 5. Plots of $\tau_\perp$ (a) and $\tau_\parallel$ (b) vs. the volume concentration ρ at $\gamma=1.5$. Curves *1* correspond to $\alpha_0=1$, curves *2* to $\alpha_0=3$, and curves *3* to $\alpha_0=10$.

5. EFFECTIVE COLLOID VISCOSITY

When there is a magnetic field, the average hydrodynamic stress tensor σ in a ferrocolloid becomes asymmetric. We write this tensor as

$$\sigma=\sigma^S+\sigma^A, \tag{32}$$

where the superscripts S and A designate the symmetric and antisymmetric parts of σ. The literature devoted to calculations of σ^S in dense suspensions of neutral spheres is vast. The most rigorous result in the approximation linear in the gradient of the average velocity was obtained by Batchelor and Green[29] in the approximation of two interacting particles:

$$\sigma_{ij}=2\eta^S\Gamma_{ij}, \quad \eta^S=\eta_0(1+2.5\rho+7.6\rho^2),$$
$$\Gamma_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right), \quad i,j=x,y,z, \tag{33}$$

where η^S is the effective viscosity of a suspension of neutral hard spheres. This result agrees well with the experimental data if $\rho\leqslant 0.1$; at higher concentrations it leads to underestimated results.

Our goal in this section and in Sec. 6 is to calculate the components of the antisymmetric part σ^A of the average stress tensor in an approximation linear in the components of the tensor of the ferrocolloid's average velocity gradient. Here we examine steady flow and estimate the effective viscosity of the ferrocolloid.

The components of the antisymmetric part of the stress tensor of a magnetic liquid can be written as follows (see, e.g., Refs. 25 and 30):

$$\sigma^A_{ij}=\frac{1}{2}(M_iH_j-M_jH_i)=\frac{1}{2}n\alpha T(\mu_ih_j-\mu_jh_i),$$
$$i,j=x,y,z; \quad n=\frac{N}{V}. \tag{34}$$

Thus, to calculate the components of σ^A, we must find the components of the nonequilibrium magnetization $\mathbf{M}$ and plug them into (34). To this end we go back to (16), assuming that the flow and external field are steady-state.

In the approximation linear in Ω and $\delta\alpha$, Eqs. (19) and (20) yield a system of time-independent equations,

$$\frac{B-A}{J}(\mathbf{h}\cdot\delta\boldsymbol{\alpha})=0, \tag{35}$$
$$D_r\frac{A}{\mu_0}(\delta\boldsymbol{\alpha}-(\mathbf{h}\cdot\delta\boldsymbol{\alpha})\mathbf{h})+\boldsymbol{\Omega}\times\mathbf{h}=0.$$

From the first equation in (35) it immediately follows that $\mathbf{h}\cdot\delta\boldsymbol{\alpha}=0$. Using (28), we obtain

$$\boldsymbol{\mu}=\boldsymbol{\mu}_0+\frac{\mu_0}{\alpha}\delta\boldsymbol{\alpha}. \tag{36}$$

Finding $\delta\boldsymbol{\alpha}$ from the second equation in (35) and plugging the result into (36), we get

$$\boldsymbol{\mu}=\boldsymbol{\mu}_0-\frac{1}{D_r}\frac{\mu_0^2}{\alpha A}(\boldsymbol{\Omega}\times\mathbf{h}). \tag{37}$$

Plugging the components of $\boldsymbol{\mu}$ from (37) into (34), we find that

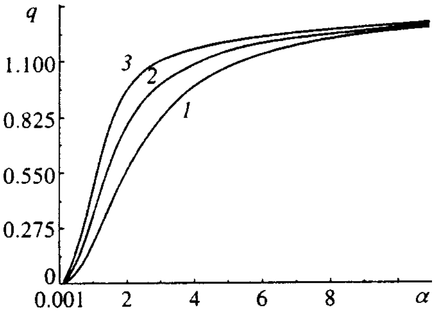


FIG. 6. Plots of q vs. the constant external magnetic field α at $\rho=0.1$. Curve *1* corresponds to $\gamma=0$, curve *2* to $\gamma=1.5$, and curve *3* to $\gamma=2.5$.

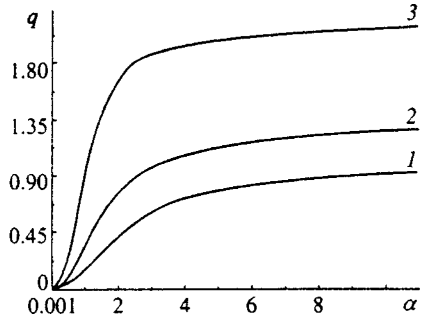


FIG. 7. Plots of q vs. the constant external magnetic field α at $\gamma=1.5$. Curve *1* corresponds to $\rho=0.01$, curve *2* to $\rho=0.1$, and curve *3* to $\rho=0.2$.

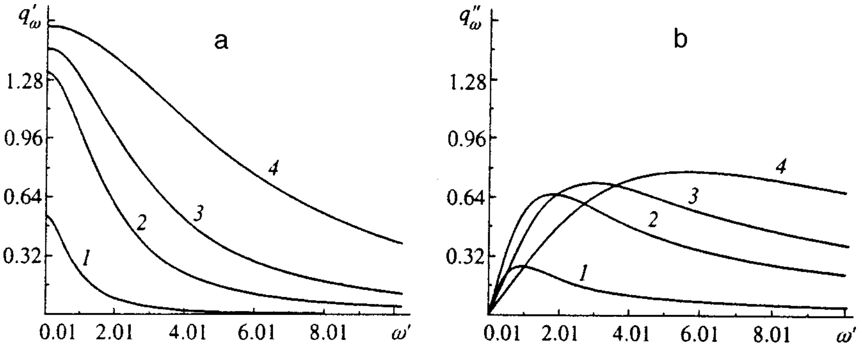


FIG. 8. Plots of q'_ω (a) and q''_ω (b) vs. the dimensionless frequency $\omega'=\omega/D_0$ of an external perturbation at $\gamma=1.5$ and $\rho=0.15$. Curves *1* correspond to $\alpha=1$, curves *2* to $\alpha=3$, curves *3* to $\alpha=5$, and curves *4* to $\alpha=10$.

$$\sigma^A_{ij}=3\frac{D_0}{D_r}\eta_0\frac{\mu_0^2}{A}T_{ijkm}\Omega_{km},$$

$$T_{ijkm}=h_m(\delta_{ik}h_j-\delta_{jk}h_i),\qquad \Omega_{km}=\frac{1}{2}\left(\frac{\partial u_k}{\partial x_m}-\frac{\partial u_m}{\partial x_k}\right), \tag{38}$$

which can be written

$$\sigma^A_{ij}=2\eta^A\sin^2\theta\,\Omega_{ij},\qquad \eta^A=\eta_0\frac{3}{2}\frac{D_0}{D_r}\frac{\mu_0^2}{A}, \tag{39}$$

with θ the angle between the vectors **H** and $\mathbf{\Omega}$.

The quantity $\eta^A\sin^2\theta$, which acts as the effective viscosity coefficient for the antisymmetric part of the stress tensor, is sometimes called the rotational viscosity coefficient, with amplitude η^A. Figures 6 and 7 depicts the results of calculating the dimensionless quantity

$$q=\frac{\eta^A}{(3/2)}\eta_0\rho \tag{40}$$

($3\eta_0\rho/2$ is the limit of the rotational viscosity coefficient of a dilute ferrocolloid placed in an infinitely strong magnetic field). The increase in q with γ in Fig. 6 can be explained by the fact that the magnetodipole interaction of the particles enhances their coupling to the external field, and hinders rotation in the field of a hydrodynamic vortex. As a result, the perturbation introduced by the particles into the suspension flux grows, which manifests itself in an increase in the suspension's viscosity. The increase in q with ρ in Fig. 7 is related to the magnetodipole and hydrodynamic interaction of the particles. A comparison of Figs. 6 and 7 reveals that hydrodynamic effects have a stronger influence on rotational viscosity than dipole–dipole effects.

The total effective viscosity coefficient of a ferrocolloid is, of course, the sum of η^S and η^A. Although the problem of the effect of the magnetodipole interaction of the particles on η^S remains unresolved, the results of some experiments (see, e.g., Refs. 31 and 32) suggest that if γ is not too large and no heteroaggregates are formed in the magnetic liquid, η^S in the first approximation can be estimated by formulas valid for suspensions of neutral spheres. For instance, for systems with a low concentration we can use (33).

6. VISCOELASTIC PROPERTIES

As is known, suspensions of neutral particles may exhibit viscoelastic properties.[33,34] The reason is that in a moving suspension the two-point correlation function of the particles changes, which generates additional specific stresses.[35] If the flow velocity gradient changes in time, the correlation function relaxes to it in the course of a finite time interval, and this is the cause of macroscopic viscoelastic effects. Since these effects are due to correlations, in low and moderately concentrated suspensions such effects are not large and manifest themselves only in subtle experiments.

In ferrocolloids, even extremely dilute ones, the viscoelastic nature of the relationship between the tensors σ^A_{ij} and Ω_{ij} is determined by the finite rate of tuning of the average moments of separate particles to the steady-state values corresponding to the fixed Ω_{ij}. Since interparticle interactions have a profound effect on the kinetics of this process (see Secs. 3 and 4), they should also have a profound effect on the viscoelastic behavior of magnetic liquids. In this section we study the nature of the relationship between σ^A_{ij} and Ω_{ij} in a time-dependent flow in a constant magnetic field.

In the approximation linear in Ω and $\delta\alpha$, the time-dependent equations (19) and (20) can be written

$$\frac{d\boldsymbol{\alpha}_e}{dt}=-D_r\left[\frac{A-B}{J}(\mathbf{h}\cdot\delta\boldsymbol{\alpha})\mathbf{h}+\frac{\alpha A}{\mu_0}(\delta\boldsymbol{\alpha}-(\mathbf{h}\cdot\delta\boldsymbol{\alpha})\mathbf{h})\right]+\alpha(\mathbf{\Omega}\times\mathbf{h}),$$

$$\boldsymbol{\alpha}_e=\boldsymbol{\alpha}+\delta\boldsymbol{\alpha},\qquad \boldsymbol{\alpha}=\text{const.} \tag{41}$$

After a Fourier transformation of (41) with respect to time is carried out, we get

$$i\omega\delta\boldsymbol{\alpha}_\omega=-D_r\left[\frac{A-B}{J}(\mathbf{h}\cdot\delta\boldsymbol{\alpha}_\omega)\mathbf{h}+\frac{\alpha A}{\mu_0}(\delta\boldsymbol{\alpha}_\omega-(\mathbf{h}\cdot\delta\boldsymbol{\alpha}_\omega)\mathbf{h})\right]+\alpha(\mathbf{\Omega}\times\mathbf{h}), \tag{42}$$

where ω is the Fourier frequency, and the corresponding subscripts indicate that we are dealing with the amplitudes of the harmonics of the physical quantities involved.

From (42) we find

$$\delta\boldsymbol{\alpha}_\omega=\frac{\alpha\mu_0}{D_r\alpha A+\mu_0 i\omega}(\mathbf{\Omega}\times\mathbf{h}). \tag{43}$$

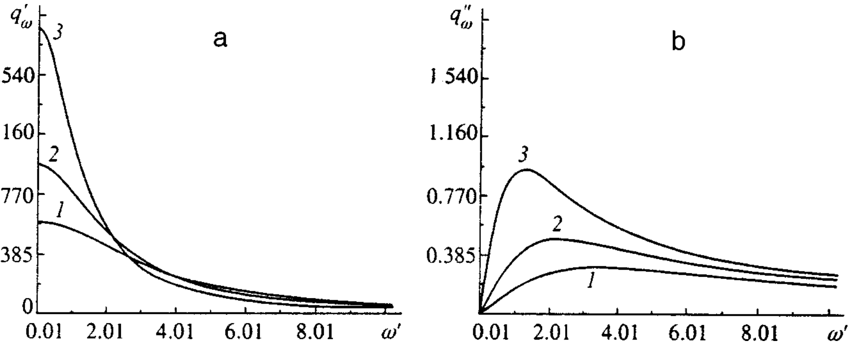


FIG. 9. Plots of q'_ω (a) and q''_ω (b) vs. the dimensionless frequency $\omega' = \omega/D_0$ of an external perturbation at $\alpha = 3$ and $\gamma = 1.5$. Curves *1* correspond to $\rho = 0.01$, curves *2* to $\rho = 0.1$, and curves *3* to $\rho = 0.2$.

In the first approximation in the small quantity $\delta\alpha$, instead of (36) we have

$$\boldsymbol{\mu}_\omega = \boldsymbol{\mu}_0 \delta(\omega) + \frac{\mu_0}{\alpha} \delta \boldsymbol{\alpha}_\omega . \tag{44}$$

Plugging (44) into (34), we obtain

$$\sigma^A_{\omega ij} = 2\,\eta^A_\omega T_{ijkm} \Omega_{km} = 2\,\eta^A_\omega \sin^2\theta\, \Omega_{ij} ,$$

$$\eta^A_\omega = \frac{3}{2} \rho \eta_0 \frac{D_0}{D_r} \frac{\mu_0^2}{A} \frac{1 - i\omega\tau^A}{1 + (\omega\tau^A)^2}, \qquad \tau^A = \frac{\mu_0}{D_r \alpha A} . \tag{45}$$

The quantity τ^A acts as the characteristic relaxation time for the antisymmetric part of the hydrodynamic stress tensor.

We introduce the complex-valued dimensionless quantity

$$q_\omega = q'_\omega + i q''_\omega ,$$

$$q'_\omega = \frac{1}{(3/2)\eta_0 \rho} \operatorname{Re} \eta^A_\omega , \quad q''_\omega = \frac{1}{(3/2)\eta_0 \rho} \operatorname{Im} \eta^A_\omega . \tag{46}$$

The results of several calculations of q'_ω and q''_ω are depicted in Figs. 8–10. Clearly, the hydrodynamic interaction of the particles has a more profound effect on the viscoelastic properties of a homogeneous ferrocolloid than the dipole–dipole interaction.

7. DISCUSSION AND CONCLUSIONS

In this paper we have studied the macroscopic dynamic characteristics of a moderately concentrated homogeneous magnetic liquid with the Brownian mechanism of relaxation of the moments of the particles. From our analysis we draw the general conclusion that even if no heteroaggregates are formed in the ferrocolloid, the dipole–dipole interparticle interaction has a substantial effect on the ferrocolloid's properties, but the effect of hydrodynamic interparticle interaction is somewhat stronger. In principle, this result coincides with the one obtained in experiments.[36] It must be noted, however, that it is the magnetodipole interaction of the particles that is responsible for the formation of drop, chain, and other heteroaggregates in many real magnetic liquids. The emergence of such aggregates is capable of strongly affecting the macroscopic characteristics of ferrocolloids.[16–18]

Unfortunately, it is extremely difficult to compare our results with those of well-known experiments, since such experiments ordinarily use polydisperse ferrocolloids, and particles belonging to different fractions in such colloids are involved in different mechanisms of magnetic moment relaxation. We believe that developing an coherent picture of the dynamic properties of magnetic liquids is possible only when model single-fraction systems have been studied. The present paper analyzes such a system theoretically. Note that all the adopted approximations are either mathematically consistent (allowance for the dipole–dipole interaction within the context of the second virial coefficient approximation) or have been thoroughly checked in previous studies (the effective-medium method in reducing Eq. (1) to (2) and the effective-field method[7]).

ACKNOWLEDGMENTS

This work was made possible by a grant from the Russian Fund for Fundamental Research (Projects 96-15-96904 and 98-01-00031).

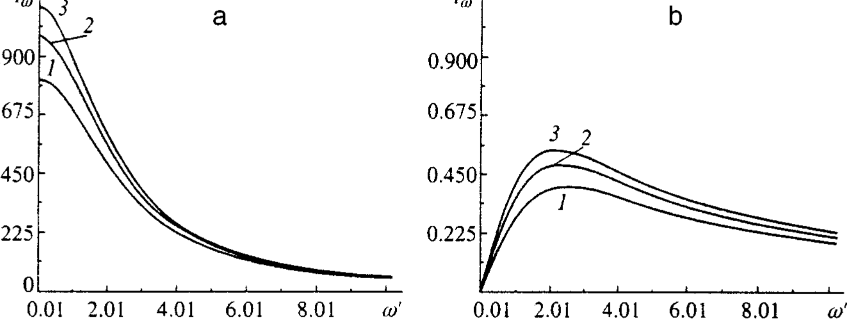


FIG. 10. Plots of q'_ω (a) and q''_ω (b) vs. the dimensionless frequency $\omega' = \omega/D_0$ of an external perturbation at $\alpha_0 = 3$ and $\rho = 0.1$. Curves *1* correspond to $\gamma = 0$, curves *2* to $\gamma = 1.5$, and curves *3* to $\gamma = 2.5$.

*)E-mail: andrey.zubarev@usu.ru

---

[1] É. Ya. Blum, M. M. Maĭorov, B. L. Nikoaru, and A. O. Tsebers, Magn. Gidrodinam. No. 1, 53 (1987).

[2] A. F. Pshenichnikov and A. V. Lebedev, Zh. Éksp. Teor. Fiz **95**, 869 (1989) [Sov. Phys. JETP **68**, 498 (1989)].

[3] Yu. L. Raĭkher and A. F. Pshenichnikov, JETP Lett. **41**, 132 (1985).

[4] A. Tari, J. Popplewell, and S. W. Charles, J. Magn. Magn. Mater. **15**, 1125 (1980).

[5] K. O'Gady, J. Popplewell, and S. W. Charles, J. Magn. Magn. Mater. **39**, 56 (1983).

[6] A. A. Minakov, I. A. Zaitsev, and V. I. Lesnin, J. Magn. Magn. Mater. **85**, 60 (1990).

[7] M. A. Martsenyuk, Yu. L. Raĭkher, and M. I. Shliomis, Zh. Éksp. Teor. Fiz **65**, 834 (1973) [Sov. Phys. JETP **38**, 413 (1974)].

[8] M. C. Miguel and J. M. Rubi, Physica A **231**, 288 (1996).

[9] J. P. Shen and M. Doi, J. Phys. Soc. Jpn. **59**, 111 (1990).

[10] C. F. Hayes, J. Colloid Interface Sci. **52**, 239 (1975).

[11] E. A. Peterson and A. A. Kruger, J. Colloid Interface Sci. **62**, 24 (1977).

[12] A. F. Pshenichnikov and I. Yu. Shurtsbor, Magn. Gidrodinam. No. 2, 137 (1986).

[13] P. H. Hess and P. H. Parker, J. Appl. Polym. Sci. **10**, 1915 (1966).

[14] P. Goldberg, J. Hansford, and P. J. van Heerden, J. Appl. Phys. **42**, 3874 (1971).

[15] N. A. Yusuf, Physica D **22**, 1916 (1989).

[16] S. Kamiyama and A. Satoh, J. Colloid Interface Sci. **127**, 173 (1989).

[17] M. Doi and H. See, J. Phys. Soc. Jpn. **61**, 2090 (1992).

[18] A. Yu. Zubarev and L. Yu. Iskakova, Zh. Éksp. Teor. Fiz **107**, 1534 (1995) [JETP **80**, 857 (1995)].

[19] A. O. Tsebers, Magn. Gidrodinam. No. 1, 3 (1983).

[20] B. É. Kashevskiĭ, Magn. Gidrodinam. No. 3, 21 (1989).

[21] A. Yu. Zubarev, Kolloidn. Zh. **57**, 335 (1995).

[22] Yu. A. Byevich and A. O. Ivanov, Physica A **190**, 276 (1992).

[23] A. F. Pshenichnikov, J. Magn. Magn. Mater. **145**, 319 (1995).

[24] A. O. Tsebers, Magn. Gidrodinam. No. 4, 17 (1984).

[25] V. N. Pokrovskiĭ, *Statistical Mechanics of Dilute Suspensions* [in Russian], Nauka, Moscow (1978), p. 88.

[26] V. N. Pokrovskiĭ, Zh. Éksp. Teor. Fiz **71**, 1880 (1976) [Sov. Phys. JETP **44**, 987 (1976)].

[27] A. Yu. Zubarev and L. Yu. Iskakova, Physica A **229**, 188 (1996).

[28] A. Perez-Madrid and J. M. Pubi, Physica A **132**, 438 (1985).

[29] G. K. Batchelor and J. T. Green, J. Fluid Mech. **56**, 375 (1972).

[30] R. Rosensweig, *Ferrohydrodynamics*, Cambridge University Press, London (1985).

[31] Yu. D. Varlamov and A. B. Kaplun, Magn. Gidrodinam. No. 3, 43 (1986).

[32] V. M. Buzmakov and A. F. Pshenichnikov, Magn. Gidrodinam. No. 1, 18 (1991).

[33] A. Yu. Zubarev, Inzh.-Fiz. Zh. **59**, 41 (1990).

[34] B. Cichocki and B. U. Felderhof, J. Chem. Phys. **101**, 7850 (1994).

[35] G. K. Batchelor, in *Hydrodynamic Interaction of Particles in Suspensions* [in Russian], Mir Publishers, Moscow (1980), p. 46.

[36] A. F. Pshenichnikov, Dr. Sci. (Phys.–Math.) thesis, Institute of Physics of Metals, Ural Branch of the Russian Academy of Sciences, Ekaterinburg (1992).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Contribution to the statistical theory of solutions of polymers in a critical solvent

I. Ya. Erukhimovich*)

A. M. Nesmeyanov Institute of Heteroorganic Compounds, Russian Academy of Sciences, 117813 Moscow, Russia

(Submitted 6 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 910–929 (September 1998)

The thermodynamics of high polymers in equilibrium with a low-molecular solvent with a large correlation radius (super- and near-critical solvent) is studied. Special attention is devoted to the analysis of typical phase diagrams describing the conditions of solubility of a polymer in such a solvent. The nature of these diagrams is determined by the existence of long-range multiparticle attraction between the monomers, which increases as the critical point of the solvent is approached. At the critical point the contribution of this attraction to the free energy of the system is nonanalytic with respect to the polymer concentration. It is shown that the nontrivial dependences of the polymer–polymer and polymer–solvent coupling constants, which appear in the phenomenological analysis, on the pressure and temperature of the solvent play an important role in the quantitative analysis of the phase diagrams of the solubility of the polymer. These dependences are found in explicit form under the assumption that in the absence of intermonomer bonds the system can be described as a compressible two-component lattice alloy. The partition function of the system under study is represented as a functional integral over two coupled, strongly fluctuating fields, one of which, describing the fluctuations of the polymer density, is the 0 component. By virtue of the specific nature of the problem, the effective temperature corresponding to the 0-component of the field cannot be specified independently, but can be determined by minimizing the total free energy of the system. © *1998 American Institute of Physics.* [S1063-7761(98)01109-3]

1. INTRODUCTION

The behavior of polymers in super- and near-critical solvents has been of great interest in recent years. On the one hand, this interest is due to the myriad applications of such solvents on industrial scales for the most diverse purposes: extraction, chromatography, purification, dissolution, and so on (see, for example, Ref. 1). These applications are all based on the feasibility of closely regulating the solubility of polymers in supercritical liquids by combining the fast mass-transfer processes that they (and gases) typically display with the strong temperature and pressure dependence of the dissolution power typical of critical solvents. The direct relationship of the latter factor to high susceptibility, and thereby to the existence of anomalously large fluctuations typical of near-critical systems, has also focused the attention of theoreticians on the industrial properties of critical solvents. Thus, the characteristic features of the solubility of impurities near a critical point of a pure solvent have been analyzed in Refs. 2 and 3 from the phenomenological standpoint, neglecting the microscopic structure of the impurities.

On the other hand, the behavior of polymers in critical solvents is also of special interest. Indeed, because of the anomalously high susceptibility and large correlation radius of polymer solutions (as compared with low-molecular substances), they exhibit a well-known analogy with an n-component magnetic material, as first pointed out by de Gennes[4,5] and des Cloizeaux[6] for the case $n=0$, and by the present author[7,8] for the continuous range $n>0$ (see also Refs. 9–14). For solutions of polymers in critical solvents, an analogy is therefore to be expected with a system of two coupled fluctuating fields described by differing effective temperatures and coupling constants.[15,16] The development of a theory of solutions of polymers in critical solvents might thereby make it possible to derive new consequences from the general fluctuation theory of phase transitions, and to test them under laboratory and industrial conditions. The quantitative microscopic description of the behavior of polymers in critical and supercritical solvents and the analysis of some of the characteristic features of such a description are the aims of the present paper.

We proceed as follows. In Sec. 2 we review the basic results of the theory of semidilute polymer solutions, and we introduce the concept of the effective free energy $F_{\mathrm{eff}}(\rho,\mu,T)$ for polymers in critical solvents, where μ and T are the chemical potential of the solvent and the temperature of the system, and ρ is the density of monomers. We also describe the renormalization of the two-body interaction and the second virial coefficient of the monomers as a result of screening by particles of the critical solvent.[17–22]

In Sec. 3 the renormalization of the multiparticle interaction of the monomers is taken into account for the first time and an expression is obtained for the nonpolynomial function $F_{\mathrm{eff}}(\rho,\mu,T)$ of ρ, which is used to study a number of properties of the condensed polymer phase.

We analyze and discuss the characteristic features of the phase diagrams for a simple model of a polymer solution in a critical solvent in Sec. 4, where for this model we derive the explicit dependence of the polymer–polymer and polymer–solvent coupling constants on the temperature T and pressure P. In the preceding sections these were introduced only phenomenologically.

Finally, we derive the basic equations of the theory of the coil–globule transition for a polymer chain in a critical solvent are derived in Sec. 5 by representing the partition function as a functional integral over two coupled fields. The analysis performed here leads to no additional results, but it does place the system of interest in the general context of the theory of phase transitions,[16,23] and makes it possible to study the system in the Landau approximation. We briefly summarize our results in Sec. 6.

2. PHYSICS OF SEMIDILUTE SOLUTIONS OF POLYMERS IN SIMPLE AND CRITICAL SOLVENTS

Consider a system whose volume V contains n solvent particles and M identical linear polymer chains, each of which consists of N monomers. The characteristic parameters of this system are the concentration of solvent particles $\rho_s = n/V$, chains $\nu = M/V$, and monomers $\overline{\rho} = \nu N$, the rms radius of gyration of a chain as a whole $R_G \sim a\sqrt{N}$ (a is the average distance between neighboring monomers (bond length)), and the range r_0 of the interaction potential between monomers belonging to different chains. For

$$\overline{\rho} r_0^3 \ll 1, \tag{2.1}$$

the collision probability between monomers belonging to different chains is low. To describe their interaction, it is therefore sufficient to use the virial expansions.[23,24]

If

$$\nu R_G^3 \sim \overline{\rho} a^3 \sqrt{N} \gg 1, \tag{2.2}$$

which always holds for sufficiently high degrees of polymerization N, then the volume R_G^3 of each chain will contain many other chains simultaneously. Such solutions are said to be semidilute,[25] and the dominant contribution to the interaction energy in them comes from collisions of monomers that belong to different polymer chains. Averaging of the contribution of these collisions when calculating the partition function of such solutions is performed on low-molecular scales $\sim r_0$ and is therefore statistically independent of the averaging of the contribution of the conformational collection of chains that occurs on macromolecular scales $\sim R_G$. In other words, the partition function of semidilute polymer solutions with a prescribed (generally speaking, nonuniform and nonequilibrium) spatial distribution $\{\rho(\mathbf{r})\}$ of their monomer densities factorizes into a product, while the free energy can be expanded in a sum of the corresponding contributions:

$$F_{\mathrm{poly}}(\{\rho(\mathbf{r})\}) = F^*(\{\rho(\mathbf{r})\}) + F_{\mathrm{str}}(\{\rho(\mathbf{r})\}). \tag{2.3}$$

The first term of Eq. (2.3) is therefore independent of the degree of polymerization of the chains and can be identified with the energy contribution of disconnected monomers to the free energy of the system (low-molecular system whose particles have the same interaction energy as the monomers). This term is also called the contribution of the ‘‘volume interaction’’ of monomers and is written in the form

$$\begin{aligned} F^*(\{\rho(\mathbf{r})\}) &= F_{\mathrm{bls}}(\{\rho(\mathbf{r})\})T - T\int d\mathbf{r}\ \rho(\mathbf{r})\ln(\rho(\mathbf{r})/e) \\ &= \int d\mathbf{r}\ f^*(\rho(\mathbf{r})), \\ f^*(\rho) &= T\sum_{i=2}^{i=\infty} \frac{a_i\rho^i}{i-1}, \end{aligned} \tag{2.4}$$

where the a_i are the virial coefficients that appear in the expansion of the equation of state of a system of disconnected monomers in powers of the density:[23,24]

$$P_{\mathrm{bls}} = T\rho + P^*(\rho) = T(\rho + a_2\rho^2 + a_3\rho^3 + \dots). \tag{2.5}$$

The second term in Eq. (23) is the free energy of a system of noninteracting N-mers with distribution $\{\rho(\mathbf{r})\}$, and it is ordinarily called the structural–entropy term. The first equality in Eq. (2.4) follows from the fact that for a system of disconnected monomers this term is the entropy of an ideal gas. The necessity of averaging over different scales in order for the additive decomposition (2.3) of the free energy of polymers to be valid and the idea of a system of disconnected monomers were first formulated by I. M. Lifshitz[26] (see also Refs. 25 and 27).

Neglecting fluctuation effects, Eq. (2.3) assumes the simple form[25]

$$F_{\mathrm{poly}}(\{\rho(\mathbf{r})\}) = F^*(\{\overline{\rho}\}) + VT\nu\ \ln(\nu/e), \tag{2.6}$$

whence follows the virial equation of state of semidilute polymer solutions in the mean-field approximation[5,25,28]

$$\begin{aligned} P_{\mathrm{poly}} &= -(\partial F/\partial V)_{\upsilon,T} = (T\rho/N) + P^*(\rho) \\ &= T((\rho/N) + a_2\rho^2 + a_3\rho^3 + \dots), \end{aligned} \tag{2.7}$$

where $P^*(\rho)$ (the energy contribution to the pressure, identical for polymers and disconnected monomers), which also appears in Eq. (2.5), is given by

$$P^*(\rho) = \rho(\partial f^*/\partial\rho)_T - f^*(\rho, T). \tag{2.8}$$

Thus, the equation of state (2.7) of a semidilute polymer solution and the equation of state (2.5) of a system of disconnected monomers have identical virial coefficients, but the linear term of the equation of state for high-polymer systems ($N \to \infty$) is negligible because of the anomalously low translational entropy of large molecules. The behavior of polymer solutions therefore depends heavily on the sign of the second virial coefficient a_2.

Specifically, for $a_2 < 0$ the range of low monomer densities bounded by the inequality

$$(\partial P_{\mathrm{poly}}/\partial\rho)_T = T(2a_2 + 3a_3\rho + \dots)\rho < 0, \tag{2.9}$$

is unstable against condensation of the polymer into a phase whose density ρ_g can be determined by requiring that the (osmotic) pressure of the polymer vanish:

$$P^*(\rho_g) = a_2\rho_g^2 + a_3\rho_g^3 + \dots = 0. \tag{2.10}$$

The notation for ρ_g derives from the fact that the solution of Eq. (2.10) also determines the density of a globule formed by a single infinitely long polymer chain.[25,26] If in addition to the inequality (2.8) we also have

$$\varepsilon=|a_2a_4|/a_3^2\ll 1, \tag{2.11}$$

then Eq. (2.10) yields

$$\rho_g=-a_2/a_3 \tag{2.12}$$

for the density of the condensed phase. This quantity vanishes at $a_2=0$, and plays the role of the triple point in semidilute polymer solutions.[29] (We have ignored a small correction $\sim O(\varepsilon)$ to the density of the condensed phase, and vanishingly small corrections $\sim O(N^{-1})$ to the density of the dilute phase and the position of the critical point, which in the theory of polymer solutions is usually called a θ point.)

Thus, the sign of the second virial coefficient of the system of disconnected monomers qualitatively determines the behavior of semidilute polymer solutions. To investigate the dependence of a_2 on the state of the solvent, we examine the change in the free energy of the solvent when monomers with a distribution $\{\rho(\mathbf{r})\}$ are added to it:

$$\Delta F(\{\rho(\mathbf{r})\},T)=-T\ln\frac{\int\delta\rho_s(\mathbf{r})\exp(-F(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T)/T)}{\int\delta\rho_s(\mathbf{r})\exp(-F_{\text{solv}}(\{\rho_s(\mathbf{r})\},T)/T)}, \tag{2.13}$$

where the functional integral extends over all density distributions $\{\rho_s(\mathbf{r})\}$ of the solvent particles, $F_{\text{solv}}(\{\rho_s(\mathbf{r})\},T)$ is the free energy of the solvent in the absence of a polymer, and $F_{\text{solv}}(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T)$ is the free energy of the system under study as a functional of both density distributions. Repeating the arguments used in the derivation of the expression (2.3), the free energy of the system can be represented in the form

$$\begin{aligned}F(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T)=F_{\text{solv}}(\{\rho_s(\mathbf{r})\},T)+F_0^*(\{\rho(\mathbf{r})\},T)\\+F_1^*(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T)\\+F_{\text{str}}(\{\rho(\mathbf{r})\},T),\end{aligned} \tag{2.14}$$

where the terms $F_1^*(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T)$, and $F_0^*(\{\rho(\mathbf{r})\},T)$ describe the respective contributions of monomer interactions with solvent particles and one another. Substituting (2.14) into Eq. (2.13), we obtain

$$\Delta F=F_{\text{eff}}^*(\{\rho(\mathbf{r})\},T)+F_{\text{str}}(\{\rho(\mathbf{r})\},T), \tag{2.15}$$

where the contribution of the effective volume interaction is given by

$$F_{\text{eff}}^*(\{\rho(\mathbf{r})\},T)=F_0^*(\{\rho(\mathbf{r})\},T)+\Delta F^*(\{\rho(\mathbf{r})\},T), \tag{2.16}$$

i.e., the sum of the initial volume interaction $F_0^*(\{\rho(\mathbf{r})\},T)$ and the renormalization $\Delta F^*(\{\rho(\mathbf{r})\},T)$ due to the change in solvent free energy as a result of solvent particle redistribution in the field produced by the prescribed monomer distribution $\{\rho(\mathbf{r})\}$:

$$\begin{aligned}\frac{\Delta F^*(\{\rho(\mathbf{r})\},T)}{T}&=-\ln\frac{\int\delta\rho_s(\mathbf{r})\exp(-(\tilde{F}_{\text{solv}}(\{\rho_s(\mathbf{r})\},\{\rho(\mathbf{r})\},T))/T)}{\int\delta\rho_s(\mathbf{r})\exp(-F_{\text{solv}}(\{\rho_s(\mathbf{r})\},T)/T)},\\ \tilde{F}_{\text{solv}}(\{\rho_s(\mathbf{r})\},\{\rho(\mathbf{r})\},T)&=F_{\text{solv}}(\{\rho_s(\mathbf{r})\},T)\\&\quad+F_1^*(\{\rho(\mathbf{r})\},\{\rho_s(\mathbf{r})\},T).\end{aligned} \tag{2.17}$$

This renormalization grows in an obvious manner with the susceptibility of the solvent. Indeed, let us expand $\tilde{F}_{\text{solv}}(\{\rho_s(\mathbf{r})\},\{\rho(\mathbf{r})\},T)$, which appears in Eq. (2.17), in powers of the density fluctuations of the solvent particles $\psi(\mathbf{r})=\rho_s(\mathbf{r})-\overline{\rho_s}$ ($\overline{\rho_s}$ is the mean density of the latter) and the average monomer density $\overline{\rho}$, limiting the expansion to terms no higher than second order in ψ and first order in $\overline{\rho}$:

$$\begin{aligned}\tilde{F}_{\text{solv}}(\overline{\rho},\{\rho_s(\mathbf{r})\},T)=F_{\text{solv}}(\overline{\rho_s},T)+F_1^*(\overline{\rho},\overline{\rho_s},T)\\+\mu_s(\overline{\rho_s}),T)\int\psi(\mathbf{r})dV\\+\frac{T}{2}\int dV(\psi\hat{G}^{-1}\psi+v_0b_2\overline{\rho}\psi^2(\mathbf{r}))\\+Tb_1\overline{\rho}\int\psi(\mathbf{r})dV.\end{aligned} \tag{2.18}$$

Here the operator $\hat{G}^{-1}$ is the inverse of the integral operator

$$\hat{G}f=\int d\mathbf{r}'G(\mathbf{r}-\mathbf{r}')f(\mathbf{r}'),$$

whose kernel is the density–density correlation function of particles of the pure solvent. Limiting attention here for simplicity to a description of the properties of a critical solvent in the Landau approximation, we can represent the Fourier transform of this function in the standard Ornstein–Zernike form[23]

$$\begin{aligned}G(\mathbf{r})&=\langle\psi(0)\psi(\mathbf{r})\rangle,\\ \tilde{G}(\mathbf{q})&=\int d\mathbf{r}\,G(\mathbf{r})\exp(i\mathbf{qr})=(v_0(\tau_{\text{eff}}+d^2q^2))^{-1}.\end{aligned} \tag{2.19}$$

In (2.18) and (2.19) the scale $d\sim r_0$ and the parameters b_1, b_2, and $v_0=\rho_{\max}^{-1}$ ($\rho_{\max}$ is the maximum packing density of the solvent particles), which have dimensions of volume, are phenomenological constants, while the parameter $\tau_{\text{eff}}=(\partial\mu_s/\partial\rho_s)_T/v_0T$ vanishes at the critical point of the solvent.

Substituting (2.18) into Eq. (2.17), taking account of the fact that the term in Eq. (2.18) that is proportional to the chemical potential μ_s of the solvent and linear in ψ vanishes (since the number of solvent particles is conserved), and calculating the resulting Gaussian integral, we have

$$\Delta F^*(\{\rho(\mathbf{r})\},T)=-\frac{T}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\frac{b_1^2|\rho_{\mathbf{q}}|^2/v_0}{\tau_{\text{eff}}+b_2\rho+d^2q^2} \tag{2.20}$$

whence the final expression for the first nonvanishing term in the free energy of a semidilute polymer solution is, to terms $\sim O(N^{-1})$,

$$\frac{F^*_{\mathrm{eff}}(\{\rho(\mathbf{r})\},T)}{T}=\int\frac{d^3\mathbf{q}}{(2\pi)^3}\left(a_2-\frac{1}{2}\frac{b_1^2/v_0}{\tau_{\mathrm{eff}}+d^2q^2}\right)|\rho_{\mathbf{q}}|^2. \tag{2.21}$$

As we can see from Eq. (2.21), the redistribution of the solvent particles around the particles of the dissolved substance (screening) introduced into the solvent always behaves like an attraction, whose strength and range increase as τ decreases, i.e., as the critical point is approached. This effect, a special case of which is the familiar Debye–Hückel screening in the theory of electrolytes,[23] was first studied in the theory of polymers by Edwards[17] (see also Refs. 8, 14, and 18–22).

Since in the present work we are interested primarily in phase equilibrium, we neglect the existence of polymer density fluctuations and we rewrite Eq. (2.21) in the form

$$F^*_{\mathrm{eff}}(\rho,T)=VTa_2^{\mathrm{eff}}\rho^2,\quad a_2^{\mathrm{eff}}=a_2-(b_1^2/2v_0)\tau_{\mathrm{eff}}^{-1}. \tag{2.22}$$

Thus the condition $a_2^{\mathrm{eff}}<0$, which ensures the instability of a semidilute polymer solution in critical solvents against formation of a condensed phase, holds *a fortiori* near the critical point of the solvent. A coil–globule transition is therefore to be expected here for a long single chain, while for a system of many chains a condensed phase where the polymer concentration ρ_g depends on proximity to the critical point of the solvent can be expected to precipitate from the solution.

At first glance, ρ_g is determined, in accordance with (2.12), by the balance of the attraction due to the negative renormalized value of the second virial coefficient a_2^{eff} and the repulsion due to the presence solid cores in the monomers, which is described by the third virial coefficient $a_3>0$:

$$\rho_g=-a_2^{\mathrm{eff}}/a_3. \tag{2.23}$$

(A similar result was obtained in Ref. 30.) Since, however, the above-noted long-range nature of the interaction of the particles can lead to violation of (2.1), it is not sufficient to calculate only the first few terms of the virial expansion to determine the properties of the condensed phase. We therefore carry out a more detailed analysis of the contributions of the effective volume interaction to the free energy $F^*_{\mathrm{eff}}(\rho,T)$ and the pressure $P^*_{\mathrm{eff}}(\rho,T)$, which are related by (2.8).

3. FREE ENERGY OF THE EFFECTIVE VOLUME INTERACTION OF MONOMERS IN A CRITICAL SOLVENT

Let us consider as the simplest microscopic model of a critical solvent, where it is possible to express the vertices of the phenomenological Ginzburg–Landau Hamiltonian in terms of the microscopic parameters of the solvent particles, a lattice gas (liquid) whose free energy has the form[31]

$$F=\int\frac{dV}{v_0}\left(f(\phi(\mathbf{r}),T)+2Td^2(\nabla\phi)^2\right), \tag{3.1}$$

$$f(\phi,T)=T(\phi\ln\phi+(1-\phi)\ln(1-\phi))-2\phi^2, \tag{3.2}$$

where v_0 is the volume of a unit cell of the lattice gas, $\phi(\mathbf{r})$ is the local value of the volume fraction of cells occupied by solvent particles, d is a phenomenological scale, and the temperature, free energy, and chemical potential are expressed in units such that the temperature at the critical point equals 1.

In the mean-field approximation the pressure and chemical potential as functions of T and the average value of ϕ have the form

$$\tilde{P}=v_0P=-T\ln(1-\phi)-2\phi^2,$$

$$\mu=T\ln(\phi/(1-\phi))-4\phi. \tag{3.3}$$

The condition $\partial^2 f/\partial\phi^2=0$ determines the chemical potential $\mu_c=-2$, the volume fraction of the solvent $\phi_c=0.5$, and the pressure $v_0P_c=p_c=\ln 2-0.5$ at the critical point (the condition $\partial^3 f/\partial\phi^3=0$ holds automatically at $\phi_c=0.5$).

The Gibbs free energy of the grand canonical ensemble, convenient for studying phase equilibria, can be written near the critical point as

$$\begin{aligned}\Omega(\mu,T,V)&=-T\ln\int\delta\phi(\mathbf{r})\\&\quad\times\exp\left(\frac{\mu\int(dV/v_0)\phi(\mathbf{r})-F(\{\phi(\mathbf{r})\},T)}{T}\right)\\&=F(\phi_c,T,V)-(V\mu/2v_0)+\Omega_c(h,\tau,V),\end{aligned} \tag{3.4}$$

where $\tau=1-T^{-1}$, $h=(\mu_c-\mu)/T\mu_c$ and the function

$$\begin{aligned}\Omega_c(h,\tau,V)&=-T\ln\int\delta\Psi(\mathbf{r})\exp\left\{\int\frac{dV}{v_0}\right.\\&\quad\left.\times\left(h\Psi-\frac{\tau\Psi^2+d^2(\nabla\Psi)^2}{2}-\frac{\Psi^4}{12}\right)\right\},\end{aligned} \tag{3.5}$$

describes the thermodynamics of the solvent near the critical point.

To give the functional integrals in Eqs. (3.4) and (3.5) a more precise meaning, they can be defined by subtracting out the fluctuation contribution to the free energy of the lattice gas (in the limit $T\to\infty$):

$$\begin{aligned}\frac{\Omega_c(h,\tau,V)}{T}&=-\ln\\&\times\frac{\int\delta\Psi(\mathbf{r})\exp\int(dV/v_0)(h\Psi-(\tau\Psi^2+d^2(\nabla\Psi)^2)/2-\Psi^4/12)}{\int\delta\Psi(\mathbf{r})\exp(-\int dV(\Psi^2+d^2(\nabla\Psi)^2)/2v_0)}.\end{aligned} \tag{3.6}$$

This representation, somewhat different from that obtained in Ref. 23 on the basis of a different (isothermally isobaric) potential, is valid so long as the values of the order parameter $\Psi(\mathbf{r})=2\phi(\mathbf{r})-1$ that make the principal contribution to the integrals (3.5) and (3.6) remain small compared to 1.

For simplicity, we confine our attention in the present paper to the Landau approximation,1) in which there is no difference between (3.5) and (3.6); the latter then lead to the simple result

$$\Omega=-P_{\mathrm{solv}}V=-\frac{V}{v_0}\left(T(\ln 2-0.5)+\frac{\mu-\mu_c}{2}-TR(h,\tau)\right), \tag{3.7}$$

where the function $R(h,\tau)$ is the absolute minimum of the function

$$H(h,\tau,\Psi)=\frac{\Psi^4}{12}+\tau\frac{\Psi^2}{2}-h\Psi. \tag{3.8}$$

Using Eq. (3.7), the pressure of the pure solvent near the critical point can be written conveniently in terms of the reduced pressure Π:

$$v_0P_{\mathrm{solv}}=T(p_c+\Pi), \tag{3.9}$$

where Π can be written as a function of the chemical potential and temperature:

$$\Pi=h-R(h,\tau). \tag{3.10}$$

Similarly, generalizing the arguments of the preceding section that led to (2.18) and (2.22), it is easy to obtain the Gibbs free energy of a polymer solution with prescribed values of the chemical potential μ of the solvent and the polymer density:

$$\Phi(\mu,T,\rho,V)=V\Bigg(f^*(\rho,T,\phi_c)+Tv_0^{-1}\times\left(R\left(h+\frac{b_1\rho}{2},\tau+\frac{b_2\rho}{4}\right)-h-p_c\right)\Bigg), \tag{3.11}$$

where b_1 and b_2 are coefficients in the expansion of the free energy (2.18). An expression for the pressure of such a solution follows from Eq. (3.11):

$$P=-\left(\frac{\partial\Phi}{\partial V}\right)_{T,\mu}=P_{\mathrm{solv}}+P^*+\Delta P, \tag{3.12}$$

where the contributions P_{solv} and P^* are given by (3.9) and (2.8), respectively, while the correction to the pressure due to long-range screening by solvent particles of the interactions between monomers is

$$v_0\Delta P=\rho\left(\frac{\partial\Delta f}{\partial\rho}\right)_T-\Delta f, \tag{3.13}$$

where

$$\Delta f(\rho,T,h,\tau)/T=R(h+b_1\rho/2,\tau+b_2\rho/4)-R(h,\tau).$$

Carrying out the differentiation in Eq. (3.13), we obtain

$$\Delta P=\frac{T}{v_0}\left(\frac{\Psi_0^4-\Psi^4}{12}+\tau\frac{\Psi_0^2-\Psi^2}{2}-h(\Psi_0-\Psi)\right), \tag{3.14}$$

where Ψ and Ψ_0 are the respective positions of the absolute minima, as given by Eq. (3.8), of the functions $H(h+b_1\rho/2,\tau+b_2\rho/4,\Psi_0)$ and $H(h,\tau,\Psi_0)$.

It is convenient to rewrite the correction (3.14) as an expansion

$$\Delta P=-\frac{T}{v_0}(\Psi-\Psi_0)^2\left(\frac{\tau_{\mathrm{eff}}}{8}+\Psi_0\frac{\Psi-\Psi_0}{3}+\frac{(\Psi-\Psi_0)^2}{12}\right), \tag{3.15}$$

where $\tau_{\mathrm{eff}}=4(\tau+\Psi_0^2)$, which has the same meaning as in the preceding section, is everywhere positive (except at the critical point). Substituting the limit $\Psi-\Psi_0=2b_1\rho/\tau_{\mathrm{eff}}$ (as $\rho\to 0$) into Eq. (3.15), we again arrive at (2.22).

At the critical point itself $h=\tau=\tau_{\mathrm{eff}}=0$, the long-range screening of the monomer-monomer interaction by solvent particles results in the contributions to the free energy and pressure which have (similarly to the case of electrolytes[14]) a nonanalytical form

$$\Delta P(\rho,T,0,0)=-\frac{T}{12v_0}\Psi^4\approx-\frac{T}{12v_0}\begin{cases}(3b_1\rho/2)^{4/3}, & b_1\rho\ll b_1^3/b_2^3,\\ (2b_1/b_2)^4, & b_1\rho\gg b_1^3/b_2^3.\end{cases} \tag{3.16}$$

The contribution (3.16) decreases with decreasing polymer density more slowly than the renormalized contribution of two-body collisions (2.22) obtained above. This means that as the critical point is approached, all terms of the virial expansion become important, and in this sense even a very dilute solution can be considered to be concentrated.[13] However, as before, the sum (3.13) of all these terms corresponds to attraction, determining the existence condition and density of the condensed phase (globular state) of polymers near the critical point of the solvent.

Indeed, the polymer density ρ_g in the condensed phase can be determined by requiring that the pressure (3.11) of this phase equal the pressure (3.8) of the pure solvent, i.e., by requiring that the osmotic pressure of the polymer vanish:

$$v_0\frac{P^*+\Delta P}{T}=a_2v_0\rho_g^2-\left(\frac{\Psi^4(\rho_g)-\Psi_0^4}{12}+\tau\frac{\Psi^2(\rho_g)-\Psi_0^2}{2}-h(\Psi(\rho_g)-\Psi_0)\right)=0. \tag{3.17}$$

Specifically, at the critical point itself, substituting (3.16) into Eq. (3.17) yields

$$\phi_g=v_0\rho_g=\sqrt{\frac{v_0}{3a_2}}\,\frac{\Psi^2}{2}\approx\begin{cases}\sqrt{\dfrac{v_0}{3a_2}}\,\dfrac{3b_1^2}{32a_2v_0}, & b_2^2\ll a_2v_0,\\ \sqrt{\dfrac{v_0}{3a_2}}\,\dfrac{2b_1^2}{b_2^2}, & b_2^2\gg a_2v_0.\end{cases} \tag{3.18}$$

Here, clearly, the equilibrium volume fraction ρ_g is completely independent of a_3, contrary to the result (2.23) in Ref. 30. This is due to the fact that the repulsive component of the two-body interactions of monomers, described by the value $a_2>0$ of the initial second virial coefficient, is in fact sufficient to balance the nonpolynomial attraction (3.16) of the monomers at the critical point. However, it must be borne in mind that as ϕ_g increases, the quantitative contribution of three-body and higher-order collisions can again become critically important. Specifically, higher-order collisions ensure that $\phi_g<1$. The applicability of (3.18) is thus constrained by

$$\phi_g \ll 1. \tag{3.19}$$

The requirement that the density ρ_g vanish,

$$\rho_g(h,\tau)=0, \tag{3.20}$$

or equivalently that the effective second virial coefficient (2.22) vanish

$$\tau_{\text{eff}}=\tau_0\equiv\frac{b_1^2}{2a_2v_0}, \tag{3.21}$$

determines the locus of critical points of the polymer solution in the PT plane. The second equality in Eq. (3.21) is simply a definition of the parameter τ_0 that characterizes the relative interaction force between the monomers and solvent, compared with their repulsion of one another, in terms of the microscopic parameters b_1 and a_2.

4. PHASE DIAGRAMS OF POLYMER SOLUTIONS IN A CRITICAL SOLVENT AND *PT* DEPENDENCE OF THE COUPLING CONSTANTS b_1 AND a_2

At first glance, it follows from Eq. (3.21) that all possible partitions of the PT plane into one- and two-phase regions (where one phase is always the pure solvent phase) in the mean-field approximation can be specified by the contour lines of the function

$$\tau_{\text{eff}}(P,T)=T^{-1}\left(\frac{\partial\mu_s}{\partial\rho_s}\right)_T=\frac{1}{\phi(1-\phi)}-\frac{4}{T}=\tau_0, \tag{4.1}$$

where the volume fraction ϕ of the solvent is a single-valued function of P and T that can be derived from the first of Eqs. (3.3) for the thermodynamic equilibrium phase at given P and T. As shown in Fig. 1, these lines are closed curves that encircle the critical point, and the region bounded by the latter, in which a condensed phase forms, grows as τ_0 (i.e., the interaction of the polymer with the critical solvent) increases.

However, these curves are not true phase transition lines, as the microscopic parameter τ_0 itself depends on P and T. To take account of this dependence explicitly (albeit phenomenologically), we examine the polymer solution in the limit $N\to\infty$. In this limit, according to Eqs. (2.4) and (2.6), its free energy takes the form

$$F=(V/v)f^*(\phi_s,\phi,T), \tag{4.2}$$

where ϕ and ϕ_s are the respective volume fractions of the monomers and solvent, and

$$f^*(\phi_s,\phi,T)=f_{\text{bls}}(\phi_s,\phi,T)-\phi\ \ln(\phi/e). \tag{4.3}$$

Thus, in this limit the description of a polymer solution reduces to choosing a free energy that corresponds to a two-component system of disconnected monomers. It is natural to use as the latter the following generalization of the free energy (3.2) of a one-component lattice gas:

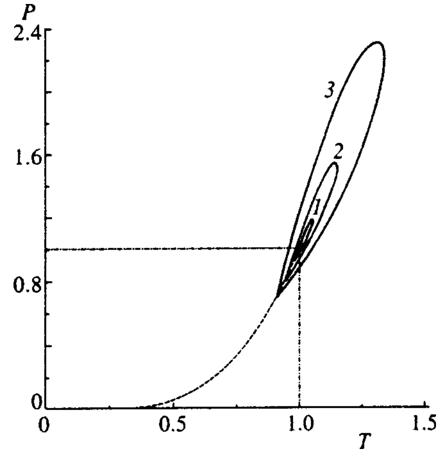


FIG. 1. PT phase diagrams of the solubility of a polymer, neglecting the dependence of the phenomenological constants a_2 and b_1 on the pressure and temperature of the solvent. The contour lines of the function (4.1) with $\tau_0=0.2$, 0.5, and 1.0 are shown by the solid curves *1, 2*, and *3*, respectively; the dashed line shows the liquid–gas phase transition for the pure solvent; the dot-dashed lines show the position of the critical point of the pure solvent.

$$f_{\text{bls}}(\phi_s,\phi,T)=T(\phi\ \ln\ \phi+\phi_s\ \ln\ \phi_s+(1-\phi-\phi_s)\times\ln(1-\phi-\phi_s))-2(\phi_s+\delta\phi)^2. \tag{4.4}$$

Here we again take the temperature of the critical point of the pure solvent to be the unit of temperature, while δ is a parameter that characterizes the attraction between dissolved monomers and solvent particles (and one another); it is related to the critical temperature of a pure system of monomers:

$$T_c^{(1)}=\delta^2. \tag{4.5}$$

An expression for the pressure of a polymer solution with an infinite degree of polymerization (or, equivalently, globules) follows from Eqs. (4.2)–(4.4):

$$\tilde{P}=v_0P=-T(\ln(1-\phi_s-\phi)+\phi)-2(\phi_s+\phi\delta)^2, \tag{4.6}$$

which in the absence of the solvent ($\phi_s=0$) assumes the form

$$\tilde{P}=-T(\ln(1-\phi)+\phi)-2\delta^2\phi^2. \tag{4.7}$$

At temperatures $T>4\delta^2$, the isotherm $P(\phi)$ given by Eq. (4.7) is positive, and increases monotomically over the entire range $0<\phi<1$, while for $T<4\delta^2$ it decreases on the interval $0<\phi<\phi_c$ and increases on the interval $\phi_c<\phi<1$ ($\phi_c=1-T/4\delta^2$), remaining negative for $0<\phi<\phi_g$. Here ϕ_g is a root of the equation

$$\tilde{P}(\phi_g)=-T(\ln(1-\phi_g)+\phi_g)-2\delta^2\phi_g^2=0 \tag{4.8}$$

and represents the volume fraction of the pure polymer for $T<4\delta^2$ in the absence of external pressure. As shown in Fig. 2, ϕ_g decreases monotonically with increasing temperature and vanishes at $T=4\delta^2$, while the volume of a polymer

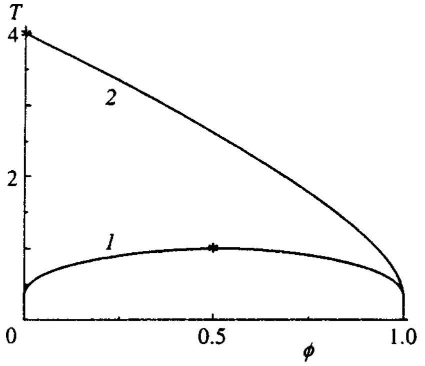


FIG. 2. Monomer volume fraction phase diagrams (ϕ vs. temperature T) for a system of disconnected monomers (*1*) and a system consisting of the same monomers of high polymer (in the limit of an infinite degree of polymerization) (*2*). Asterisks mark the critical points.

sample with an infinite degree of polymerization increases continuously (the polymer swells), until at $T=4\delta^2$ it fills the entire accessible volume (no matter how large). The unfilled part of the volume can be formally interpreted as a coexisting phase with zero polymer concentration (void).

In other words, in the limit $N\rightarrow\infty$ the temperature and volume fraction of the pure polymer at its critical point become[19]

$$\phi_c^{(\infty)}=0, \quad T_c^{(\infty)}=4\delta^2=4T_c^{(1)}. \tag{4.9}$$

(At finite N these values acquire corrections $\sim O(N^{-1/2})$, and accordingly the increase in volume as the polymer approaches its critical point is governed by the large parameter $N^{1/2}$; henceforth we neglect such refinements.)

If two-phase separation occurs in a solvent, then the solvent volume fraction ϕ_s, in Eq. (4.6) is not an independent quantity in the polymer phase, but is instead determined by the chemical potential μ of the solvent in the pure solvent phase (which can be both liquid and vapor) that coexists with it:

$$\mu_s(T,\phi_s,\phi)=T\ \ln(\phi_s/(1-\phi_s-\phi)) \\ -4(\phi_s+\phi\delta)=\mu_s(T,\phi_s^0,0)=\mu. \tag{4.10}$$

Here $\phi_s^0(P,T)$ and $\phi_s(P,T,\phi)$ are the solvent volume fractions in the coexisting phases, while the explicit dependence $\mu_s(T,\phi_s,\phi)$ given by the first equality in Eq. (4.10) follows from Eqs. (4.2)–(4.4). It is convenient to give the pressure of the polymer phase parametrically, supplementing the equation of state (4.6) with equations for the volume fractions of the components:

$$\tilde{P}=\tilde{P}(u)=-T(\ln((1-u)(1-\phi(u))) \\ +\phi(u))-2(\phi_s(u)+\phi(u)\delta)^2, \tag{4.11}$$

$$\phi_s=\phi_s(u)=u(1-\phi(u)), \tag{4.12}$$

$$\phi=\phi(u)=\frac{T\ \ln(u/(1-u))-4u-\mu}{4(\delta-u)}, \tag{4.13}$$

where as one can see from Eq. (4.12), the parameter u is just the reduced volume fraction of the solvent, i.e., the fraction of locations not occupied by monomers that are nevertheless occupied by solvent particles.

Direct calculation easily yields

$$\frac{\partial(\tilde{P}/T)}{\partial\phi}=\frac{\phi T}{1-\phi}\left(1-4\frac{(\delta-u)^2}{\tau u(1-u)}\frac{1-\phi}{T}\right), \tag{4.14}$$

$$\left(\frac{\partial^2(\tilde{P}/T)}{\partial\phi^2}\right)\Bigg|_{\phi=0}=\frac{2a_2^{\text{eff}}}{v_0}=1-4\frac{(\delta-\phi_s^0)^2}{T} \\ -\frac{(4(\delta-\phi_s^0)/T)^2}{\tau_{\text{eff}}(P,T)}, \tag{4.15}$$

where $\tau_{\text{eff}}(P,T)$ is given by (4.1). It follows from Eqs. (4.14) and (4.15) that for a negative right-hand side of (4.15),

$$1-4\frac{(\delta-\phi_s^0)^2}{T}-\frac{(4(\delta-\phi_s^0)/T)^2}{\tau_{\text{eff}}(P,T)}<0, \tag{4.16}$$

the addition of a small amount of polymer to a pure solvent with chemical potential μ at temperature T and pressure P reduces the pressure of the polymer phase (the corresponding isotherm is shown in Fig. 3a). In this case, the system is always unstable against separation into two phases, one of which is pure solvent. The second phase contains both polymer and solvent, whose volume fractions can be determined by substituting the function $u(T,\mu)$ given implicitly by the equation

$$\Delta\tilde{P}=\tilde{P}(u,T,\mu)-\tilde{P}_s(\mu,T)=0 \tag{4.17}$$

into Eqs. (4.12) and (4.13), where $\tilde{P}_s(\mu,T)$ is the pressure of the pure solvent given by Eqs. (3.3). Since the pure solvent volume fraction ϕ_s^0 in the condition (4.16) depends on P and

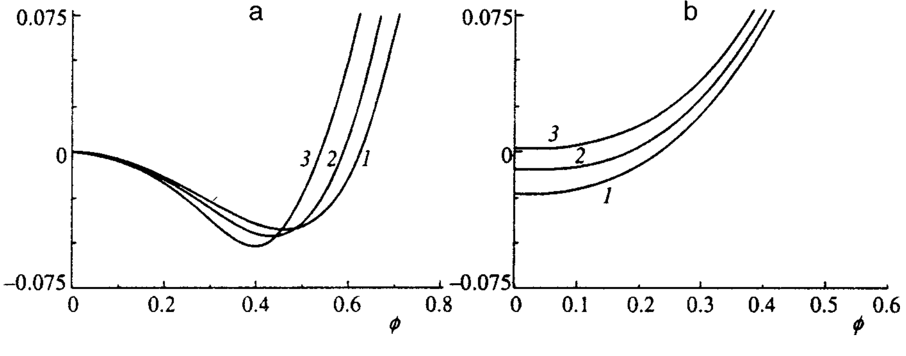


FIG. 3. Typical curves of the reduced pressure of the polymer phase vs. the volume fraction of that phase (for definiteness, $\delta>0.5$): a) outside the region T_1T_2 below the line (4.18); b) inside the region T_1T_2. The lines are numbered in order of increasing pressure of the pure solvent.

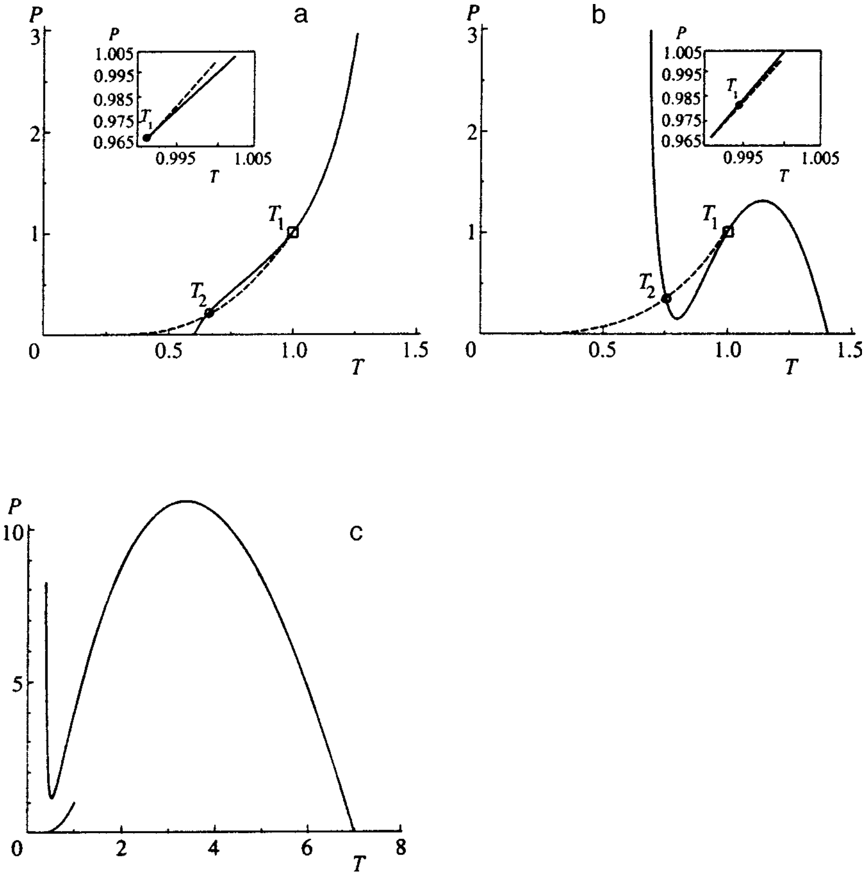


FIG. 4. Reduced solubility diagrams (pressure P vs. temperature T) for a high polymer in a low-molecular solvent in the limit of an infinite degree of polymerization: a) $\delta<0.5$; b) $0.5<\delta<6.79$; c) $\delta>6.79$. The liquid–vapor transition of the pure solvent and solubility of the polymer (4.18) are shown as dashed and solid lines, respectively. Near a critical point of the pure solvent (small rectangles in Figs. 4a and 4b), these essentially merge, so that they can only be distinguished in the much more detailed view in the insets.

T, it determines the boundary of the region of solubility of the polymer in the PT plane. This boundary can be represented in the form

$$4\phi_s^0(P,T)=\frac{T-4\delta^2}{1-2\delta}. \tag{4.18}$$

The shape of the curve (4.18) depends on the value of δ in Eq. (4.4) and in subsequent expressions, which represents the relative cohesion energy of the polymer (as compared with the solvent). Specifically, at the ends of this line

$$\phi_s^0=0, \quad P=0, \quad T=4\delta^2,$$

$$\phi_s^0=1, \quad P=\infty, \quad T=4(1-\delta)^2, \tag{4.19}$$

As one can see from Fig. 4, for $\delta<\delta_c=6.79$ it crosses the curve describing the liquid–vapor phase transition of the pure solvent at the points T_1 and T_2, taking a detour above the critical point of the latter for $\delta>0.5$, and below it for $\delta<0.5$. In the first instance, all points on the line (4.18) to the left of the critical point of the pure solvent correspond to the liquid phase (for these points, Eq. (4.18) yields $\phi_s^0>0.5$), and the part below the line of the phase transition of the pure solvent corresponds to a metastable liquid phase. The condensed phase of the polymer is therefore also stable inside the region bounded by the segments T_1T_2 of the lines (4.18) and the phase transition of the pure solvent, even though the pressure increases monotonically with polymer concentration (see Fig. 3b), on account of the negative osmotic pressure $\Delta\tilde{P}$ due to the metastability of the liquid phase of the pure solvent in the indicated region.

Summarizing the foregoing analysis, which can be directly extended to situations in which $\delta<0.5$, we conclude that the density of the condensed polymer phase (globules) vanishes on that line of the two indicated above (liquid–vapor transition of the pure solvent or (4.18)) that corresponds to the higher or lower pressure (for $\delta>0.5$). However, upon crossing line segments of the liquid–vapor transition of the pure solvent that lie below (for $\delta<0.5$) or above (for $\delta<0.5$) the line (4.18), the density of a globule remains constant but its pressure and temperature derivatives are discontinuous (together with the derivatives of the chemical potential of the pure solvent). We also note that for $\delta>\delta_c=6.79$ the region T_1T_2 vanishes, and the solubility diagram assumes the form shown in Fig. 4c.

In concluding this section we note that both the model phase diagrams in Fig. 1 and the diagrams in Fig. 4 have a loop around the critical point. The existence of this loop indicates that the effective attraction of monomers in a critical solvent plays a decisive role in the formation of the region of the swelled undissolved polymer. Differences among other details of the phase diagrams in Figs. 1 and 4 can be explained by the existence of a strong PT dependence of the previously introduced phenomenological parameters. Indeed, direct comparison of the right-hand side of Eq. (4.15) with Eq. (2.22) yields

$$a_2(P,T)=v_0\left(\frac{1}{2}-2\,\frac{(\delta-\phi_s^0(P,T))^2}{T}\right),$$

$$b_1(P,T)=4v_0\frac{\delta-\phi_s^0(P,T)}{T},$$

$$\tau_0=\frac{b_1^2}{2a_2v_0}=\frac{16(\delta-\phi_s^0(P,T))^2}{T(T-4(\delta-\phi_s^0(P,T))^2)}. \tag{4.20}$$

The similarity of the experimentally observed phase diagram[1] and the diagrams in Fig. 4 underscores the important role of factors taken into account by our analysis: long-range attraction of monomers which increases as the critical point of the solvent is approached; a decrease in the polymer–critical-solvent coupling constant; and an increase in repulsion due to direct two-body collisions of monomers as the volume fraction of the solvent increases.

5. TWO-FIELD MODEL OF A SOLUTION OF A POLYMER IN A CRITICAL SOLVENT

The analysis in the preceding sections was conducted in the Landau (mean-field) approximation. Although numerical calculations of phase diagrams outside this approximation are not studied in the present paper, it would be useful here to suggest a general technique that makes it possible to take account of the fluctuation corrections in the general case. To do so, we recall that in the absence of a solvent (or taking account of its structure explicitly), the partition function of a polymer chain consisting of N monomers and filling a volume V can be represented in the form[8–13,32]

$$Z(V,T,N)=\frac{1}{2\pi i}\oint Z(V,T,z)\frac{dz}{z^{N+1}}, \tag{5.1}$$

where the integral is taken over a small contour around the origin in the complex z plane, while the thermodynamic potential of the grand canonical ensemble can be represented in the form

$$Z(V,T,z)=\int d\mathbf{R}\times\frac{\int\delta\varphi(\mathbf{r})\varphi(\mathbf{R})\varphi(0)\exp(\int dV\chi(\tau(\varphi(\mathbf{r})))-L(\{\varphi(\mathbf{r})\}))}{\int\delta\varphi(\mathbf{r})\exp(-L(\{\varphi(\mathbf{r})\}))}, \tag{5.2}$$

$$\tau(\varphi(\mathbf{r}))=z\varphi^2(\mathbf{r})/2,\quad L(\{\varphi(\mathbf{r})\})=\int\varphi g^{-1}\varphi dV/2. \tag{5.3}$$

Here $\hat{g}^{-1}$ is the operator inverse of the integral operator

$$\hat{g}\varphi=\int g(\mathbf{r}-\mathbf{r}')\varphi(\mathbf{r}')dV',$$

where the nearest-neighbor correlation function $g(\mathbf{r})$ is usually given by its Fourier expansion in the form

$$g_k=\int dV\, g(\mathbf{r})\exp(i\mathbf{kr})=1+(ka)^2/6+\dots \tag{5.4}$$

(the phenomenological parameter a^2 represents the rms distance between neighboring monomers in a polymer chain), and $\chi(\tau)$ is the generating function, familiar from the theory of simple liquids, of all connected Mayer diagrams[24] for the corresponding system of disconnected monomers:

$$\chi(\tau)=\tau-a_2\tau^2+\left(2a_2^2-\frac{a_3}{2}\right)\tau^3+\dots \tag{5.5}$$

The function $\chi(t)$ makes it possible to write the equation of state of a system of disconnected monomers in the parametric form[24]

$$P/T=\chi(\tau),\quad \rho=\tau\chi(\tau). \tag{5.6}$$

In the mean-field approximation the partition function of a polymer chain is the saddle point value of the integral (5.1), obtained after (5.2)–(5.4) are substituted into this integral. The condensed (globular) state of the chain corresponds to the solution characterized by certain optimal values of the complex activity z, globule volume V occupied by a chain of N monomers, and the constant value of the field φ_0 in this volume (with the exception of the surface layer). The corresponding equations for the extremum have the form

$$\tau\partial\chi/\partial\tau=N/V,\quad \chi(\tau_0)=\varphi_0^2/2,$$

$$\varphi_0^2=\varphi_0\partial\chi/\partial\varphi_0=2\tau\partial\chi/\partial\tau. \tag{5.7}$$

Taking account of Eqs. (5.6) and the fact that N/V is the average density ρ_g of a globule, Eqs. (5.7) can be reduced to the classical equations for a ‘‘large globule’’ [26,33]

$$P^*(\rho_g)=0,\quad F(T,N)=-T\ln Z(N,T)=N\lambda,$$

$$\lambda=T\ln z_0=\mu_{\text{bls}}(\rho_g)-T\ln\rho_g,\quad V=N/\rho_g. \tag{5.8}$$

Equations (5.7) and (5.8) always have a trivial solution corresponding to a coiled state of the chain:

$$\rho_g=\varphi_0=0,\quad \lambda=0. \tag{5.9}$$

The nontrivial thermodynamic equilibrium value of ρ_g can be easily found if it is small, whereupon only the first few nontrivial terms of the virial expansion of the function χ need be retained in the integral (5.2):

$$Z(V,T,N)=\frac{1}{2\pi i}\oint\frac{dz}{z^{N+1}}\int d\mathbf{R}\times\frac{\int\delta\varphi(\mathbf{r})\varphi(\mathbf{R})\varphi(0)\exp(-H(\{\varphi(\mathbf{r})\})/T)}{\int\delta\varphi(\mathbf{r})\exp\{-\int dV(c^2(\nabla\varphi)^2+\varphi^2)/2\}},$$

$$H(\{\varphi(\mathbf{r})\})=T\int dV\left\{\frac{c^2(\nabla\varphi)^2+\varepsilon\varphi^2}{2}+a_2\left(\frac{z\varphi^2}{2}\right)^2+\frac{a_3-4a_2^2}{2}\left(\frac{z\varphi^2}{2}\right)^3\right\}, \tag{5.10}$$

where we have used the expansions (5.4) and (5.5) and introduced the parameter $c^2=a^2/6$ and the effective temperature $\varepsilon=1-z$. If the inequality

$$a_3\gg a_2^2, \tag{5.11}$$

holds, as it usually does for real systems, then for $a_2>0$ the trivial solution of Eqs. (5.8) remains the only one, while for $a_2<0$ it becomes unstable and can be replaced by a stable nontrivial solution describing the condensed (globular) state of the chain:

$$\rho_g=-a_2/a_3,\quad \varphi_0^2=2\rho_g,\quad \lambda=a_2\rho_g/2<0. \tag{5.12}$$

Thus, as first reported by de Gennes,[29] the triple point $a_2=\varepsilon=0$ of the Landau Hamiltonian in Eq. (5.10) corresponds to the coil–globule transition point in the mean-field approximation. In the condensed state itself, i.e., for $\varepsilon<0$, $a_2<0$, the effective temperature ε is not an independent variable, but is instead determined by the conditions (5.6) for the maximum of the functional integrals (5.2) and (5.8). (This specific feature of the theory of critical phenomena applied to the physics of the condensed state of polymers, which follows from the derivation above, was first pointed out by the present author in Ref. 8.)

To extend the above analysis to the case of a polymer solution in a critical solvent, we note that the partition function $Z(V,T,z,\mu_s)$ of a solution of disconnected monomers in the latter, where $z=\exp(\mu/T)$ and μ and μ_s are the chemical potentials of the disconnected monomers and solvent, respectively, can be written, using the results obtained in Secs. 2 and 3, in the form

$$
\begin{aligned}
Z(V,T,z,\mu_s)=&\int\delta\Psi(\mathbf{r})\exp\int\frac{dV}{v}\biggl\{(\ln 2-0.5)\\
&+h+\chi_0(\tilde z)+\left(h+\frac{b_1\tilde z}{2}\right)\Psi\\
&-\frac{1}{2}\left[\left(\tau+\frac{b_2\tilde z}{4}\right)\Psi^2+d^2(\nabla\Psi)^2\right]-\frac{\Psi^4}{12}\biggr\}\\
&\times\biggl\{\int\delta\Psi(\mathbf{r})\exp\biggl\{-\int dV\\
&\times\frac{\Psi^2+d^2(\nabla\Psi)^2}{2v}\biggr\}\biggr\}^{-1}, \qquad (5.13)
\end{aligned}
$$

where $\tilde z=z\,\exp(-\alpha(T)/T)$ and $\alpha(T)$ is the change that would be introduced in the thermodynamic potential of the solvent with $\phi=0.5$ when a single monomer is introduced into it.

Substituting (5.13) into (5.1) and (5.2), we obtain the final expression for the partition function of a solution of a polymer in a critical solvent:

$$
\begin{aligned}
Z(V,T,N,\mu_s)=&\frac{\exp\{(V/v_0)(\ln 2-0.5+h)\}}{2\pi i}\\
&\times\oint Z(V,T,z,\mu_s)\,\frac{dz}{z^{N+1}}, \qquad (5.1\text{a})
\end{aligned}
$$

$$
\begin{aligned}
Z(V,T,z,\mu_s)=&\int\delta\Psi(\mathbf{r})\,\delta\varphi(\mathbf{r})\exp\biggl\{\int\frac{dV}{v}\biggl[\biggl(h+\frac{b_1\tilde z}{2}\\
&\times\frac{\varphi^2(\mathbf{r})}{2}\biggr)\Psi-\frac{1}{2}\biggl[\biggl(\tau+\frac{b_2\tilde z}{4}\,\frac{\varphi^2(\mathbf{r})}{2}\biggr)\Psi^2\\
&+d^2(\nabla\Psi)^2\biggr]-\frac{\Psi^4}{12}\biggr]-\frac{H(\{\varphi(\mathbf{r})\})}{T}\biggr\}\\
&\times\biggl\{\int\delta\Psi(\mathbf{r})\,\delta\varphi(\mathbf{r})\exp\biggl\{-\int dV\\
&\times\left[\frac{\Psi^2+d^2(\nabla\Psi)^2}{2v}+\frac{c^2(\nabla\varphi)^2+\varphi^2}{2}\right]\biggr\}\biggr\}^{-1}, \qquad (5.13\text{a})
\end{aligned}
$$

where the functional $H(\{\varphi(\mathbf{r})\})$ is given by (5.10) with the substitution $\varepsilon=1-\tilde z$.

The representation (5.1a) reduces the description of a solution of a polymer in a critical solvent to analysis of a system of two coupled fluctuating fields,[15,16] with the aforementioned specific additional condition that the effective temperature ε of the field $\varphi(\mathbf{r})$ describing the polymer is dictated by the conditions that maximize the total partition function (5.1a). In the Landau approximation these conditions are similar to (5.7) and equivalent to Eq. (3.17).

Beyond this approximation (in the scaling region), when calculating the fluctuation renormalizations of the free energy, the fields $\varphi(\mathbf{r})$ in the integrals (5.2) and (5.13) must be interpreted as $(n\rightarrow 0)$-component vector fields[5–14] with a fixed temperature ε, and the expression for the total free energy of the system should once again be minimized with respect to ε and the volume occupied by the condensed polymer phase. Although the corresponding analysis is beyond the scope of the present paper, it makes sense to note that the derivation in the preceding section of the explicit PT dependences (4.20) for the phenomenological parameters a_2, b_1, and τ_0 of the field description under study opens up the possibility of explicitly calculating the lines of the fluctuation-induced first-order swelled–dissolved polymer phase transitions[15,16] in a critical solvent.

6. CONCLUSIONS

I now summarize the basic results of the foregoing analysis.

1. A long polymer chain located in a low-molecular solvent near the critical point of the solvent always forms a globule, while a solution of many high-polymer chains forms a condensed phase. The monomer density ρ_g in such a phase is determined by the balance between multiparticle attraction among the monomers, which is associated with redistribution of the solvent particles around them and increases as the critical point of the solvent is approached, and repulsion due to hard-core effects in monomers undergoing two-body collisions (see Sec. 3 and Eq. (3.18)). Attempts to restrict this aspect of the analysis to just an effective two-body interaction[30] lead to much too high a value of ρ_g.

2. At the same time, the global nature of the phase diagrams describing solubility conditions of a polymer in such a solvent is determined in many respects by the actual solvent pressure and temperature dependence of the coupling constants describing the indicated attraction. As shown for a simple model in Sec. 4, the important molecular parameter that governs these dependences is the ratio δ of the critical temperature of the system of disconnected monomers to that of the solvent. The solubility of the polymer can increase as the pressure increases (for $\delta>0.5$) or decreases (for $\delta<0.5$).

3. There exists a mathematical isomorphism between models of a high polymer in thermodynamic equilibrium with a critical solvent and a system of two strongly fluctuating fields (0-component field for the polymer and scalar for the solvent) with differing effective temperatures. One typical feature of polymer systems is that the effective tempera-

ture for the field, which describes the density fluctuations of the polymer, cannot be chosen independently, but is instead dictated (along with the density of the condensed phase of the polymer) by the minimization of the total free energy of the system.

I thank E. E. Said-Galiev and A. R. Khokhlov for stimulating and useful discussions.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-03-33521).

[*] E-Mail: ierukhs@ierukh.phys.msu.su

[1] The condition of applicability (the first correction to the renormalized fourth vertex of the Hamiltonian (3.5) must be small compared with the initial value of this vertex[16]) of the Landau approximation in our case assumes the form $|\tau| \gg ((3/16\pi)(v_0/d^3))^2$ and is not too stringent, since the ratio v_0/d^3 is at most of order 1.

---

[1] M. A. McHugh and V. J. Krukonis, in *Supercritical Fluids Extraction, Principles and Practice*, 2nd ed., Butterworth-Heinemann Series in Chemical Engineering, Butterworth-Heinemann, Boston (1994).
[2] M. Gitterman and I. Procaccia, J. Chem. Phys. **78**, 2848 (1983).
[3] M. Gitterman, Am. J. Phys. **56**, 1000 (1988).
[4] P. G. de Gennes, Phys. Lett. A **38**, 339 (1972).
[5] P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, New York (1979).
[6] J. des Cloizeaux, J. de Phys. **36**, 281 (1975).
[7] I. Ya. Erukhimovich, Vysokomol. Soedin., Ser. B **20**, 437 (1978).
[8] I. Ya. Erukhimovich, Candidate's Dissertation in Mathematical Physics, Moscow (1979).
[9] J. C. Wheeler, S. J. Kennedy, and P. Pfeuty, Phys. Rev. Lett. **45**, 1748 (1980).
[10] E. S. Nikomarov and S. P. Obukhob, Zh. Éksp. Teor. Fiz. **80**, 650 (1981) [Sov. Phys. JETP **53**, 328 (1981)].
[11] P. Pfeuty and J. C. Wheeler, Phys. Lett. A **85**, 493 (1981).
[12] S. Stepanow, Acta Polym. **32**, 98 (1981).
[13] R. Cordery, Phys. Rev. Lett. **47**, 457 (1982).
[14] I. Ya. Erukhimovich, Doctoral Dissertation in Mathematical Physics, Moscow (1994).
[15] I. F. Lyuksyutov, V. L. Pokrovskiĭ, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **69**, 1817 (1975) [Sov. Phys. JETP **42**, 923 (1975)].
[16] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* [in Russian], Nauka, Moscow (1982) [translation of 1st ed., Pergamon Press, New York (1979)].
[17] S. F. Edwards, Proc. Phys. Soc. **88**, 265 (1966).
[18] I. Ya. Erukhimovich, V. I. Irzhak, and V. G. Rostiashvili, Vysokomol. Soedin., Ser. A **18**, 1470 (1976).
[19] P. G. de Gennes, J. de Phys. **37**, L59 (1976).
[20] A. Yu. Grosberg, I. Ya. Erukhimovich, and E. I. Shakhnovich, Biopolymers **21**, 2413 (1982).
[21] J. J. Magda, G. H. Fredrickson, R. G. Larson, and E. Helfand, Macromolecules **21**, 726 (1988).
[22] M. G. Brereton and T. A. Vilgis, J. Phys. France **50**, 245 (1989).
[23] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Pt. 1* [in Russian], Nauka, Moscow (1976).
[24] G. E. Uhlenbeck, G. W. Ford, and E. W. Montroll, *Lectures in Statistical Mechanics*, Amer. Math. Soc., Providence (1963).
[25] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* [in Russian], Nauka, Moscow (1989).
[26] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **55**, 2408 (1968) [Sov. Phys. JETP **28**, 1280 (1969)].
[27] A. R. Khokhlov, Polymer **19**, 1387 (1978).
[28] P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York (1953).
[29] P. G. de Gennes, J. de Phys. Lett. **36**, L55 (1975).
[30] T. A. Vilgis, A. Sans, and G. Jannink, J. de Phys. II **3**, 1779 (1993).
[31] R. Brout, *Phase Transitions* [Russian translation], Mir, Moscow (1965).
[32] S. I. Kuchanov, S. V. Korolev, and S. V. Panyukov, in *Applications of Graph Theory in Chemistry* [in Russian], N. S. Zefirov and S. I. Kuchanov (eds.), Nauka, Nobosibirsk (1988).
[33] I. M. Lifshits, A. Yu. Grosberg, and A. R. Khokhlov, Usp. Fiz. Nauk **127**, 353 (1979) [Sov. Phys. Usp. **22**, 123 (1979)].

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Turbulent diffusion in a compressible medium

N. A. Silant'ev*)

Instituto Nacional de Astrofisica, Optica y Electronica, Apartado Postal 51 y 216 C.P. 72000 Puebla, Pue. Mexico; Central Astronomical Observatory, Russian Academy of Sciences, 196140 St. Petersburg, Russia

(Submitted 23 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 930–945 (September 1998)

This paper examines the diffusion of impurity particles in a compressible turbulent medium and compares it to diffusion in an incompressible medium. The turbulent diffusion coefficients are calculated using exact formulas expressed in terms of the Green's function describing impurity transport in an infinite homogeneous, isotropic, stationary turbulent medium. To obtain an approximate expression for the Green's function, numerical solutions of the nonlinear DIA (direct interaction approximation) equation (which in this paper are obtained for the first time for the case of compressible turbulence) are employed. Two types of turbulence are examined, acoustic and a mixture of shock waves. These are described by different generalized spectra. Finally, it is shown that compressibility significantly enhances the diffusion coefficient in the case of acoustic turbulence and reduces it in the second case. *© 1998 American Institute of Physics.* *© 1998 American Institute of Physics.* [S1063-7761(98)01209-8]

1. INTRODUCTION

The problem of transport of impurity particles and a magnetic field in a turbulent medium is an important part of the theory of turbulence.[1] The case of incompressible turbulence is the most thoroughly investigated. Here the general formulas for the turbulent diffusion coefficients D_T were first derived in the Lagrangian representation of velocities $\mathbf{v}(\mathbf{r},t)$ (see Refs. 2–4), and only much later was the Euler representation in terms of the velocities $\mathbf{u}(\mathbf{r},t)$ developed.[5] In calculating D_T in the Euler representation (which, practically speaking, is the most important), the first step was the use of the solutions of the simplest nonlinear equation for the average Green's function $\langle G(1,2)\rangle \equiv \langle G(\mathbf{r}_1,t_1;\mathbf{r}_2,t_2)\rangle$, first proposed by Kraichnan[6] and then thoroughly developed by Roberts[7] and Kraichnan.[8,9] This equation is now known as the DIA (direct interaction approximation) equation. Calculations that use the numerical solutions of this equation[10–13] have revealed that DIA solutions allow the values of the turbulent diffusion coefficient to be found for the range of all admissible values of the Strouhal number $\xi_0 = u_0 p_0 \tau_0$, where u_0, p_0, and τ_0 are the characteristic values of the turbulent velocity, the wave number (the characteristic length $R_0 \simeq 1/p_0$), and the lifetime of turbulent pulsations. The corrections to this theory due to the contribution of irreducible fourth-order correlators, which are ignored in the DIA equation, amount to approximately 7% for broad Kolmogorov spectra and to approximately 11% for peak-shaped spectra in the ''frozen'' turbulence limit ($\xi_0 \to \infty$). Being negative, these corrections monotonically decrease to zero as $\xi_0 \to 0$. Note that DIA solutions allow D_T to be calculated even in the case of ''multihumped'' spectra (see Ref. 11), where a simple estimate of D_T is extremely difficult.

The derivation of the DIA equation, the first in a hierarchy of nonlinear equations for the average Green's function,[10] does not depend on whether the turbulent medium is compressible or incompressible. It is then natural to also use the DIA equation to calculate the turbulent diffusion coefficients in the general case of a compressible medium. Moreover, it must be noted that the very problem of calculating D_T for the case of compressible turbulence is still underdeveloped. Of the earlier work mention should be made of Eschrich's paper,[14] in which the case $\xi_0 \ll 1$ is examined, and of others[5,15,16] in which approximate calculations were used for the first time to establish the fact that D_T can be negative in turbulence consisting of a chaotic mixture of shock waves with fairly large values of ξ_0, namely, $\geqslant 2-3$.

An interesting treatment of the problem of turbulent diffusion, not limited to an incompressible medium, can be found in Refs. 17 and 18. The result of this work is surprising: the formulas for D_T are the same, irrespective of whether the medium is incompressible or compressible, and do not explicitly contain correlators with div $\mathbf{u}$, the compression index. The diffusion coefficient is found to depend only on the turbulent energy spectrum. It is obvious that the formula is incomplete: the same energy spectrum in incompressible and compressible media corresponds to entirely different motions and hence to different diffusion coefficients. The term with div$\mathbf{u}$ is absent from the formulas of Ref. 17 because of a direct error in using the Klyatskin–Tatarskiĭ method (note that this method is limited to the $\xi_0 \leqslant 1$ range).

In Ref. 18, devoted to acoustic turbulence, the term with div$\mathbf{u}$ is absent because the molecular diffusion coefficient was ignored from the start, although the value of D_T obtained in that paper is equal, in order of magnitude, to D_m. Below it will be demonstrated that allowance for the term with div$\mathbf{u}$ in the event of low absorption may change the result threefold (of course the conclusion drawn in Ref. 18 that D_T is small remains valid). We also note that in the case of acoustic turbulence turbulent diffusion not only due to wave absorption by the medium but also due to the presence of chaotic irregularities in the phases, a fact not mentioned by Kazantsev *et al.*[18] Generally speaking, these processes are different, but they lead to the same form of the velocity correlator. Probably, the chaotic irregularities in the phases are substantial even if wave absorption is low, which enhances the turbulent diffusion in such a medium.

Formally exact equations, expressed in terms of the stochastic Green's function, for the turbulent diffusion coefficients valid for both incompressible and compressible media were derived in Ref. 5. They explicitly contain correlators of the Green's function with div**u** and form a reliable basis for calculations of turbulent diffusion coefficients for a compressible medium. Approximations are introduced only when a specific method of finding or approximating the exact Green's function is used. In the present work formulas for calculating D_T that allow for all powers of the second- and fourth-order correlators are derived, which is a direct consequence of the nonlinearity of the DIA equation. For the sake of definiteness, diffusion of impurity particles in a turbulent medium is examined.

2. BASIC EQUATIONS

The law of conservation of the number of impurity particles leads to an equation for the impurity particle concentration $n(\mathbf{r},t)$:

$$\left(\frac{\partial}{\partial t}-D_m\nabla^2\right)n(\mathbf{r},t)=-\nabla\cdot(\mathbf{u}(\mathbf{r},t)n(\mathbf{r},t)), \tag{1}$$

where D_m is the molecular diffusion coefficient, and $\mathbf{u}(\mathbf{r},t)$ is the velocity field in the turbulent medium, whose statistical ensemble is assumed homogeneous, isotropic, and stationary. The term on the right-hand side has the meaning of a source of particles, which appear in a volume element because of the transport (convection) of particles from regions with a higher particle concentration and because of compression of the volume element itself (recall that $\mathrm{div}\mathbf{u}(\mathbf{r},t)=\Delta V/Vdt$ has the meaning of the relative variation of the volume element of the medium per unit time).

The solution of Eq. (1) with a given Lagrangian velocity field $\mathbf{v}(\mathbf{a},t)$ for $D_m=0$ has the form[19,20]

$$n(\mathbf{r},t)=\frac{n_0(\mathbf{r}-\mathbf{X}(\mathbf{a},t))}{D(\mathbf{a},t)}\equiv\frac{n(\mathbf{a},t)}{D(\mathbf{a},t)}, \tag{2}$$

where $\mathbf{r}=(x_1,x_2,x_3)$, and $\mathbf{a}$ is the radius vector of a liquid particle at initial time $t=0$. The quantity $n_0(\mathbf{a})$ stands for the initial concentration of the impurity particles. The following relationships hold:

$$\mathbf{r}=\mathbf{a}+\mathbf{X}(\mathbf{a},t)\equiv\mathbf{a}+\int_0^t d\tau\,\mathbf{v}(\mathbf{a},t),\quad \frac{\partial\mathbf{r}}{\partial t}\equiv\mathbf{u}(\mathbf{r},t)\equiv\mathbf{v}(\mathbf{a},t), \tag{3}$$

$$D_{ij}(\mathbf{a},t)=\frac{\partial x_i}{\partial a_j},\quad D(\mathbf{a},t)=\det D_{ij},$$

$$\dot{D}=\frac{\partial D}{\partial t},\quad \frac{\dot{D}}{D}=\mathrm{div}\,\mathbf{u}(\mathbf{r},t), \tag{4}$$

$$D_{sp}\dot{D}_{pq}^{-1}=-\frac{\partial v_s}{\partial a_q},\quad d\mathbf{r}=D(\mathbf{a},t)\,d\mathbf{a}. \tag{5}$$

As usual, summation is implied over repeated indices.

Defining a random field of Lagrangian velocities $\mathbf{v}(a,t)$ determines an ensemble of realizations of turbulent flows over which the solution (2) must be averaged so that equations for the average concentration $\langle n(\mathbf{r},t)\rangle$ can be obtained. For fixed $\mathbf{r}$ and t the vector $\mathbf{a}=\mathbf{r}-\mathbf{X}(\mathbf{a},t)$ is a random quantity, and the problem reduces to averaging of the known function $n_0(\mathbf{a})$ of the random argument $\mathbf{a}$ with the random weighting factor $1/D(\mathbf{a},t)$. The transition to the diffusion approximation presupposes that $n_0(\mathbf{a})$ is smooth over the characteristic length $R_0\simeq\sqrt{\langle X^2\rangle}$ and involves using a Taylor expansion as a series in $\mathbf{X}(\mathbf{a},t)$ (see Refs. 3 and 9). As a result, we arrive at an exact expression for the turbulent diffusion coefficient in the Lagrangian representation:

$$D_T=\frac{1}{3}\int_0^t d\tau\left\langle\frac{\mathbf{v}(\mathbf{a},t)\cdot\mathbf{v}(\mathbf{a},\tau)}{D(\mathbf{a},t)}\right\rangle, \tag{6}$$

where $\langle\cdots\rangle$ denotes averaging over the ensemble of velocity realizations. For an incompressible medium $(D(\mathbf{a},t)\equiv1)$, Eq. (6) becomes the well-known expression derived by Taylor in 1921 (see Ref. 2).

In the Euler representation, the expression for $D_T(t)$ becomes[5]

$$D_T(t_1)=\frac{1}{3}\int d^3R\int_0^{t_1}dt_2\,[\langle u_i(1)G(1,2)u_i(2)\rangle$$
$$-\langle\mathbf{R}\cdot\mathbf{u}(1)G(1,2)\mathrm{div}\,\mathbf{u}(2)\rangle]. \tag{7}$$

Here and below we the following convenient notation: $f(1)=f(\mathbf{r}_1,t_1)$, $f(1-2)=f(\mathbf{r}_1-\mathbf{r}_2,t_1-t_2)$, $dn=d\mathbf{r}_n\,dt_n$, $\mathbf{R}=\mathbf{r}_1-\mathbf{r}_2$, $\tau=t_1-t_2$, etc., and $G(1,2)$ is the Green's function of Eq. (1), whose formal expression in the Lagrangian notation has the form

$$G(1,2)=\frac{D(\mathbf{a},t_2)}{D(\mathbf{a},t_1)}\delta\left(\mathbf{r}_1-\mathbf{r}_2-\int_{t_2}^{t_1}d\tau\,\mathbf{v}(\mathbf{a},\tau)\right). \tag{8}$$

Substituting this into Eq. (7) and using (4) and (5), we again obtain Eq. (6). We see that the exact formulas for D_T in the Lagrangian and Euler representations contain the correlators div**u** (or div**v**), which agrees with the qualitative treatment in the Introduction. Below we use only the Euler formula (7), which, unlike the Lagrangian representation, allows the effect of molecular diffusion to be treated, which is mandatory if we wish to examine acoustic turbulence.

The exact linear integral equation for the stochastic Green's function $G(1,2)$ of Eq. (1) has the form

$$G(1,2)=G_m(1-2)-\int d3\,G_m(1-3)$$
$$\times\nabla^{(3)}\mathbf{u}(3)G(3-2). \tag{9}$$

The molecular Green's function $G_m(1-2)\equiv G_m(\mathbf{R},\tau)$ is given by the simple expression

$$G_m(\mathbf{R},\tau)=\frac{H(\tau)}{(4\pi D_m\tau)^{3/2}}\exp\left[-\frac{R^2}{4D_m\tau}\right], \tag{10}$$

where $H(\tau)=1$ for $\tau>0$ and $H(\tau)=0$ for $\tau<0$ (the well-known Heaviside unit function). Inserting the iterates of this equation into (7) yields a power series in the parameter $\xi_0 = u_0 p_0 \tau_0$, an asymptotically divergent series, in which keeping only the first term is meaningful since the other terms are either very large (acoustic turbulence) or simply diverge (turbulence with a broad Kolmogorov spectrum). For more details see Ref. 12.

The stochastic equation (1) shows that the average particle concentration $\langle n(\mathbf{r},t)\rangle$ is related to the fluctuating particle concentration, and vice versa. Hence an attempt to write a single closed equation solely for the average Green's function $\langle G(1,2)\rangle \equiv G(1-2)$ leads to a hierarchy of nonlinear equations (see Refs. 10 and 12) . Equation (9) can be written in a renormalized form by using, in particular, the average Green's function $G(1-2)$ (or, more precisely, the solution of some truncated equation in the hierarchy) in both the constant term and the kernel. Substitution of the iterations of this renormalized equation into (7) results in an asymptotically convergent series, which can be used to calculate D_T for all values of the Strouhal number $(0<\xi_0<\infty)$ (for details see Refs. 10 and 12). The simplest equation in the hierarchy of nonlinear equations is the DIA equation:

$$G(1-2)=G_m(1-2)+\int d3\int d4\, G_m(1-3)\nabla_i^{(3)} \times G(3-4)\nabla_j^{(4)}B_{ij}(3-4)G(4-2), \tag{11}$$

where $B_{ij}(1-2)\equiv\langle u_i(1)u_j(2)\rangle$ is the two-point velocity correlator. Below we find it convenient to use the Fourier transforms of this correlator and of the Green's function $F(\mathbf{R},\tau)\equiv H(\tau)g(R,\tau)$, i.e., we write

$$B_{nk}(\mathbf{R},\tau)=\frac{1}{(2\pi)^3}\int d^3p\,\exp(i\mathbf{p}\cdot\mathbf{R})\,\tilde{B}_{nk}(\mathbf{p},\tau),$$

$$\tilde{B}_{nk}(\mathbf{p},\tau)=(\delta_{nk}p^2-p_np_k)f(p,\tau)+p_np_kW(p,\tau) +ie_{nkq}p_qD(p,\tau). \tag{12}$$

The above formula for $B_{nk}(\mathbf{R},\tau)$ is the most general expression for the two-point velocity correlator in the case of homogeneous, isotropic, and stationary turbulence.[21] The function D describes the helicity of the turbulent motion $(h=\langle\mathbf{u}(1)\cdot\text{curl}\,\mathbf{u}(2)\rangle\neq 0)$, so that in a reflection-symmetric medium D vanishes. In an incompressible medium $(\text{div}\,\mathbf{u}=0)$, $W=0$. Note that helicity does not enter into the DIA equation (11). Hence, in the absence of helicity in the medium, this equation also describes the diffusion of the impurity's magnetic field. The generalized turbulent spectrum,

$$\langle\mathbf{u}(\mathbf{r},t)\cdot\mathbf{u}(\mathbf{r},t+\tau)\rangle=\frac{1}{\pi^2}\int_0^\infty dp\left[p^4f(p,\tau)+\frac{p^4W(p,\tau)}{2}\right], \tag{13}$$

is the sum of the spectrum of incompressible motions $(\text{div}\,\mathbf{u}=0)$ and the spectrum of compressible, irrotational $(\text{curl}\,\mathbf{u}=0)$ motions. We denote these spectra as follows:

$$E_{\text{inc}}(p,\tau)=\frac{p^4f(p,\tau)}{\pi^2},\quad E_{\text{comp}}(p,\tau)=\frac{p^4W(p,\tau)}{2\pi^2}. \tag{14}$$

Numerically it is convenient to solve the equation for the function $g(\mathbf{R},\tau)$ that has been Fourier-transformed with respect to $\mathbf{R}$ and Laplace-transformed with respect to τ, which we denote by $\tilde{g}(p,s)$. The DIA equation transformed in this manner becomes

$$\tilde{g}(p,s)=\Bigg\{s+D_mp^2+\int_0^\infty dq\int_{-1}^1 d\mu\int_0^\infty d\tau \times\left[\frac{(1-\mu^2)p^2E_{\text{inc}}(q,\tau)}{4}+\frac{\mu p(\mu p-q)E_{\text{comp}}(q,\tau)}{2}\right] \times\tilde{g}(|\mathbf{p}-\mathbf{q}|,\tau)\exp(-s\tau)\Bigg\}^{-1}, \tag{15}$$

where $\mu=\mathbf{p}\cdot\mathbf{q}/pq$ is the cosine of the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$, and $\tilde{g}(p,\tau)$ is the Fourier transform of the function $g(\mathbf{R},\tau)$ with respect to $\mathbf{R}$. The sequence of iterations of this equation constitutes a rapidly converging continued fraction.

3. DIA EXPRESSIONS FOR THE DIFFUSION COEFFICIENTS

The DIA expressions for the turbulent diffusion coefficients can be obtained directly from Eqs. (11) and (15) if we introduce the diffusion approximation, or from the general formula (7) with the exact Green's function $G(1,2)$ replaced by its DIA approximation $G(1-2)$. As a result we get

$$D_T^{(0)}=\frac{1}{3}\int_0^\infty dp\int_0^\infty d\tau\Bigg\{[E_{\text{inc}}(p,\tau)+E_{\text{comp}}(p,\tau)]\tilde{g}(p,\tau) +E_{\text{comp}}(p,\tau)p\frac{\partial}{\partial p}\tilde{g}(p,\tau)\Bigg\}. \tag{16}$$

Here and in what follows we limit ourselves to calculating the stationary values of D_T, assuming all along that the upper limit in the integrals with respect to time τ is equal to infinity. The term with the derivative of the Green's function describes the effect of the compressibility of the medium. This term is absent from the formulas in Refs. 17 and 18.

To allow for the corrections to D_T determined by the contribution of the irreducible fourth-order correlators,[13] which are not present in (16), we write the corresponding formulas. We have

$$D_T=D_T^{(0)}+D_{\text{inc}}^{(1)}+D_{\text{comp}}^{(1)}+D_{\text{mix}}^{(1)}+D_h^{(1)}+\cdots.$$

The additional terms with the corresponding subscripts denote the contributions from the spectra of incompressible, compressible, mixed, and helical motions. In calculating the correction, we allowed for the explicit form of the second term of the hierarchy of nonlinear equations for the average Green's function, a term similar in structure to the one used in Ref. 13 (the paper is devoted to the study of magnetic-field diffusion). We have assumed that the velocity ensemble is Gaussian, i.e., the odd-order correlators are assumed to be equal to zero, while the even-order correlators are assumed to be expressed in terms of all possible pair correlators:

$$D_{\rm inc}^{(1)}=\frac{1}{24}\int_0^\infty dp\int_0^\infty dq\int_{-1}^1 d\mu\int_0^\infty d\tau_1\int_0^\infty d\tau_2$$
$$\times\int_0^\infty d\tau_3\, pq\mu(1-\mu^2)E_{\rm inc}(q,\tau_1+\tau_2)$$
$$\times E_{\rm inc}(p,\tau_2+\tau_3)\tilde{g}(q,\tau_1)\tilde{g}(p,\tau_3)\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2), \qquad (17)$$

$$D_h^{(1)}=\frac{1}{24}\int_0^\infty dp\int_0^\infty dq\int_{-1}^1 d\mu\int_0^\infty d\tau_1\int_0^\infty d\tau_2$$
$$\times\int_0^\infty d\tau_3\,(1-\mu^2)E_h(q,\tau_1+\tau_2)$$
$$\times E_h(p,\tau_2+\tau_3)\tilde{g}(q,\tau_1)\tilde{g}(p,\tau_3)\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2), \qquad (18)$$

$$D_{\rm comp}^{(1)}=-\frac{1}{6}\int_0^\infty dp\int_0^\infty dq\int_{-1}^1 d\mu\int_0^\infty d\tau_1\int_0^\infty d\tau_2$$
$$\times\int_0^\infty d\tau_3\,\mu E_{\rm comp}(q,\tau_1+\tau_2)E_{\rm comp}(p,\tau_2+\tau_3)$$
$$\times\{[2pq+(p^2+q^2)\mu+pq\mu^2]\tilde{g}(q,\tau_1)\tilde{g}(p,\tau_3)$$
$$\times\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2)+pq(q+p\mu)[\tilde{g}'(q,\tau_1)\tilde{g}(p,\tau_3)$$
$$\times\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2)+\mu\tilde{g}(q,\tau_1)\tilde{g}'(p,\tau_3)\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2)$$
$$+(q+p\mu)|\mathbf{p}+\mathbf{q}|^{-1}\tilde{g}(q,\tau_1)\tilde{g}(p,\tau_3)\tilde{g}'$$
$$\times(|\mathbf{p}+\mathbf{q}|,\tau_2)]\}, \qquad (19)$$

where $\tilde{g}'(p,\tau)=\partial\tilde{g}(p,\tau)/\partial p$, and

$$D_{\rm mix}^{(1)}=-\frac{1}{12}\int_0^\infty dp\int_0^\infty dq\int_{-1}^1 d\mu\int_0^\infty d\tau_1\int_0^\infty d\tau_2$$
$$\times\int_0^\infty d\tau_3\, pE_{\rm inc}(q,\tau_1+\tau_2)E_{\rm comp}(p,\tau_2+\tau_3)$$
$$\times\{2(q\mu+p)(1-\mu^2)\tilde{g}(q,\tau_1)\tilde{g}(p,\tau_3)\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2)$$
$$+pq\mu(1-\mu^2)\tilde{g}(q,\tau_1)[\tilde{g}'(p,\tau_3)\tilde{g}(|\mathbf{p}+\mathbf{q}|,\tau_2)$$
$$+p|\mathbf{p}+\mathbf{q}|^{-1}\tilde{g}(p,\tau_3)\tilde{g}'(|\mathbf{p}+\mathbf{q}|,\tau_2)]\}. \qquad (20)$$

In Eqs. (17)–(20)we can use the DIA solution $\tilde{g}(p,\tau)$ rather than the solution of the second equation in the hierarchy of nonlinear equations, since the differences between the two solutions (as related to the calculation of D_T) appear only in the terms that allow for irreducible sixth-order correlators.

The above formulas show that $D_h^{(1)}>0$ and $D_{\rm inc}^{(1)}<0$. Here we assumed that $\tilde{g}(p,\tau)$ is a monotonically decreasing function of both arguments (this is actually the case in the diffusion approximation) and that the generalized spectra assume their maximum values at $\tau=0$. It seems natural that helicity enhances the turbulent diffusion coefficient: its presence means that there is additional regularity in turbulent motions, which ensures that impurities are transported over greater distances. We recall that by the helicity spectrum we mean

$$\langle\mathbf{u}(\mathbf{r},t)\cdot{\rm curl}\,\mathbf{u}(\mathbf{r},t+\tau)\rangle=\int_0^\infty dp\,E_h(p,\tau), \qquad (21)$$

$$E_h(p,\tau)=-\frac{p^4D(p,\tau)}{\pi^2}.$$

The DIA Green's function $\tilde{g}(p,\tau)$ for p small (large distance) and τ large becomes the Green's function of the diffusion equation with a diffusion coefficient $D=D_T^{(0)}+D_m$. On the other hand, large-scale turbulent motions provide the leading contribution to the transport of impurity particles. Hence a good way to estimate D_T is to replace the exact value of $\tilde{g}(p,\tau)$ in (16) by its diffusion expression $\tilde{g}_{\rm diff}=\exp(-D_sp^2\tau)$ with an unknown diffusion coefficient D_s. Equation (16) then becomes a nonlinear equation for finding the self-consistent diffusion coefficient D_s. Phythian and Curtis[22] were probably the first to introduce this method.

According to the DIA equation (15), the average value $\langle R^2(t)\rangle$ of the spread of the initial point impurity-particle source ($\langle R^2(t)\rangle=-\nabla_p^2\tilde{g}(p,t)_{p=0}$) has the form

$$\langle R^2(t)\rangle=6D_mt+2\int_0^\infty dp\int_0^t d\tau\,(t-\tau)\Bigg\{E_{\rm inc}(p,\tau)\tilde{g}(p,\tau)$$
$$+E_{\rm comp}(p,\tau)\left[\tilde{g}(p,\tau)+p\,\frac{\partial\tilde{g}(p,\tau)}{\partial p}\right]\Bigg\}. \qquad (22)$$

For $t\gg\tau_0$ this formula becomes the well-known diffusion expression $\langle R^2(t)\rangle=6(D_m+D_T^{(0)})t$. Equation (22) can serve as a tool for checking whether the approximation for the Green's function $\tilde{g}(p,\tau)$ has been chosen correctly.

4. ACOUSTIC TURBULENCE

By acoustic turbulence one usually means a medium with randomly propagating acoustic waves. For example, such turbulence is present in the solar corona. If the turbulence is stationary, and it is this case that we will be studying here, the impurity particles primarily perform periodic oscillatory movements together the main liquid or gas. No diffusion is present here, provided that we ignore the weak molecular diffusion. Three processes, in addition to molecular diffusion, serve as sources of diffusion. First, molecular diffusion is stronger if the particles perform oscillatory movements instead of being at rest. Second, the decrease in the amplitude of the acoustic waves (absorption) and the occurrence of the random irregularities in the phases caused, for instance, by collisions with randomly moving obstacles, lead to shifts in the oscillation centers, which violates the periodicity of motion of an impurity particle, whose behavior is similar to that of a drunkard from the well-known model of random walk motion examined by Taylor in his famous paper.[2] However, the behavior of the particle is different: it still oscillates most of the time, and particle hopping is a rare event. Wave absorption due to viscosity[23] is small and is described by the absorption coefficient $\gamma=\kappa p^2$ (s^{-1}), where p is the wave number, and κ coincides, in order of magnitude, with the molecular diffusion coefficient D_m. Probably, irregularities in the phases occur more easily than true absorption of the wave's energy, but they strongly depend on the way in which acoustic turbulence is realized. It is important to note that functionally the phase irregularities provide the same contribution to the turbulent spectrum as true absorption (see the Appendix).

Stationary homogeneous, isotropic acoustic turbulence is described by the correlator[18]

$$\widetilde{B}_{nk}(\mathbf{p},\tau)=p_n p_k\Phi(p)\cos[\omega(p)\tau]\exp[-\gamma(p)\tau], \tag{23}$$

where $\omega(p)=cp$ is the wave frequency, with c the velocity of the wave. Here and in what follows by τ we mean the absolute value of the time difference: $\tau=|t_1-t_2|$. The spectrum $E\equiv E_{\mathrm{comp}}(p,\tau)$ corresponding to (23) has the form

$$E_{\mathrm{comp}}(p,\tau)=\frac{1}{2\pi^2}p^4\Phi(p)\cos[\omega(p)\tau]\exp[-\gamma(p)\tau]. \tag{24}$$

Since D_T is sure to be small, we assume, as in Ref. 18, that $\widetilde{g}(p,\tau)\simeq\exp(-D_m p^2\tau)$, but unlike Kazantsev *et al.*,[18] we will not set D_T to zero, since the values of D_T are of the same order as the molecular diffusion coefficient D_m. Substituting this value of $\widetilde{g}(p,\tau)$ into the general formula (16), we get

$$D_T^{(0)}=\frac{1}{3}\int_0^\infty dp\, E_{\mathrm{comp}}(p,0)\left\{\frac{\gamma(p)+D_m p^2}{(\gamma+D_m p^2)^2+c^2p^2}\right. \\ \left. -\frac{2D_m p^2[(\gamma+D_m p^2)^2-c^2p^2]}{[(\gamma+D_m p^2)^2+c^2p^2]^2}\right\}. \tag{25}$$

We see that generally the second term in the braces, which allows for the compressibility of the medium, is of the same order as the first term, which is the only term present in Refs. 17 and 18. To make a more specific estimate we assume, as in Ref. 18, that the decay of the correlator (23) with the passage of time is determined by the true absorption of the wave energy, $\gamma(p)=\kappa p^2$, with $\kappa\approx D_m$, i.e., we ignore the contribution of phase irregularities to absorption. We also allow for the fact that $(\kappa+D_m)p/c\approx l/r\ll 1$, where l is the mean free path of the molecules. Equation (25) then yields

$$D_T^{(0)}=\tfrac{1}{3}M^2(\kappa+3D_m), \tag{26}$$

where $M=u_0/c$ is the Mach number. We see that ignoring molecular diffusion reduces the turbulent diffusion coefficient threefold and that the contribution of compensability (more precisely, the second term in (16)), equal to the term $2D_m$ in the parentheses, is substantial. In acoustic turbulence, compressibility enhances the turbulent diffusion coefficient, in contrast to the case of the purely random succession of compressions and dilations of the medium. Probably, compressibility in this regular wave motion of the medium locally (within one wavelength) enhances the impurity concentration gradient, and molecular diffusion proceeds more effectively.

Here we have not mentioned the complicated problem of the accuracy of Eqs. (25) and (26), which can be fully solved only by obtaining the exact solution of the DIA equation (15) and by calculating the correction (19). Note that the solution of the DIA equation with a rapidly oscillating function in the kernel is a complicated mathematical problem. The fact that these equations are fairly accurate is confirmed (true, indirectly) by two facts. First, the substitution of $\widetilde{g}_m(p,\tau)=\exp(-D_m p^2\tau)$ into the DIA equation (15) leads to a diffusion solution, close in magnitude to the initial one, which suggests that selecting $\widetilde{g}_m(p,\tau)$ as the starting approximation was a correct step. Second, the use of a self-consistent procedure for calculating D_T leads to the formula

$$D_s=\left(D_m+\frac{1}{3}M^2\kappa\right)\frac{1}{1-M^2}, \tag{27}$$

which at small Mach numbers yields the same results as (26) (it will be recalled that D_s is the total diffusion coefficient, i.e., the sum of the molecular and turbulent diffusion coefficients). Note also that (26) can easily be derived from (15) and (16) as the first terms in the asymptotic series in powers of $1/(cp)^2$ by using only the normalization property $\widetilde{g}(p,0)=1$.

5. NONACOUSTIC TURBULENCE

By nonacoustic turbulence one usually means a turbulent medium with chaotic incompressible and/or compressible motions of the liquid or gas without any indications of periodicity. Such turbulence can be interpreted as the limit opposite to acoustic turbulence. It is observed, for instance, in regions where stars are formed. Of course, there are also various intermediate cases. The spectra $E_{\mathrm{comp}}(p,\tau)$ and $E_{\mathrm{inc}}(p,\tau)$ in such turbulence may vary substantially. For the sake of definiteness, we limit ourselves to spectra that decay exponentially with the passage of time τ. What we want to do is to see how different the turbulent diffusion coefficients are in the limits of purely incompressible and purely compressible (irrotational) turbulent media formally described by the same generalized spectra:

$$E_{\mathrm{inc,comp}}(p,\tau)=E(p)\exp\left(-\frac{\tau}{\tau_0}\right). \tag{28}$$

For our calculations we have chosen several types of $E(p)$:

$$E(p)=u_0^2\delta(p-p_0), \tag{29}$$

$$E(p)=\left(\frac{u_0^2}{p_0}\right)\frac{0.651\,59x^4}{1+x^{17/3}}. \tag{30}$$

The spectra (29) and (30) represent two limits: a peaked spectrum and a broad spectrum (of Kolmogorov form in the inertial region). In the spectrum (30) and below, $x=p/p_0$. We also considered the following intermediate cases:

$$E(p)=\frac{8}{3\sqrt{\pi}}\frac{u_0^2}{p_0}x^4\exp(-x^2), \tag{31}$$

$$E(p)=\frac{128}{3}\frac{u_0^2}{p_0}x^4\exp(-4x), \tag{32}$$

$$E(p)=5\frac{u_0^2}{p_0}\frac{x^4}{(1+x)^6}. \tag{33}$$

The Green's functions $\widetilde{g}(p,\tau)$ were calculated by the DIA equation (15) and the diffusion coefficients $D_T^{(0)}$, $D_{\mathrm{inc}}^{(1)}$, and $D_{\mathrm{comp}}^{(1)}$, by Eqs. (16), (17), and (19), respectively. The results of calculations of the dimensionless turbulent diffusion coefficient $\overline{D}_T$ $(D_T=(u_0/p_0)\overline{D}_T)$ are depicted in Figs. 1–5.

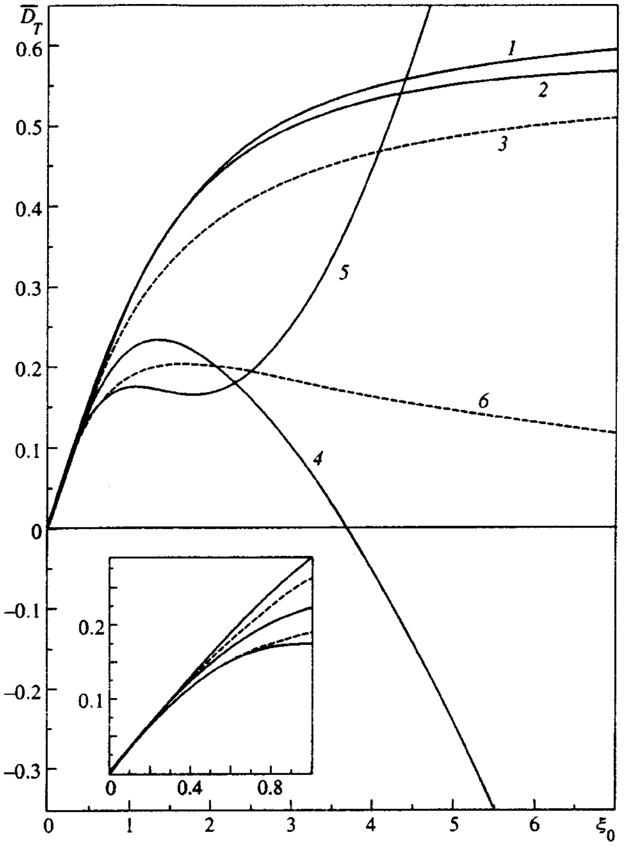


FIG. 1. The dimensionless turbulent diffusion coefficient $\overline{D}_T=(p_0/u_0)D_T$ for the case of turbulence with the delta-function energy spectrum (29). Curves *1*, *2*, and *3* represent, respectively, $\overline{D}_T^{(0)}$, $\overline{D}_T^{(0)}+\overline{D}_{\text{inc}}^{(1)}$, and $\overline{D}_s$ for the case of an incompressible medium. Curves *4*, *5*, and *6* represent the similar quantities for the case of compressible turbulence with the same energy spectrum. The coefficient $\overline{D}_T^{(0)}$ (curve *4*) becomes negative at $\xi_0=u_0p_0\tau_0>3$. The inset depicts the initial sections of the curves drawn to a larger scale.

Curves *1*, *2*, and *3* represent, respectively, the coefficients $\overline{D}_T^{(0)}$, $\overline{D}_T^{(0)}+\overline{D}_{\text{inc}}^{(1)}$, and $\overline{D}_s$ for the case of an incompressible medium. Here $\overline{D}_T^{(0)}$ yields a somewhat overestimated value of the turbulent diffusion coefficient, while the negative correction $\overline{D}_{\text{inc}}^{(1)}$ reduces it slightly. In the limit $\xi_0\to\infty$, all the curves tend to finite limits[12] corresponding to what is known as frozen turbulence, determined by long-lived vortices. The monotonic rise of the turbulent diffusion coefficients corresponds to this increase in the lifetime of turbulent vortices.

Curves *4*, *5*, and *6* representing the case of compressible (irrotational) turbulence have quite a different shape. The striking feature here is that $\overline{D}_T^{(0)}$ assumes negative values for $\xi_0\geqslant 3$ for sharp spectra and for $\xi_0\geqslant 5$ for broad spectra. Does this mean that there is negative diffusion, i.e., a flow of impurity particles into regions with a higher concentration? The answer is no, for two reasons. The first is a formal: the next approximation $\overline{D}_{\text{comp}}^{(1)}$ is positive in the region where $\overline{D}_T^{(0)}$ is negative, and the net result is positive. We also see now that for $\xi_0\geqslant 1$ the value of $\overline{D}_{\text{comp}}^{(1)}$ cannot be considered a correction, i.e., for $\xi_0\geqslant 1$ the theory becomes invalid. The second reason has a physical explanation and at the same time explains why the DIA equation does not ''work'' in the

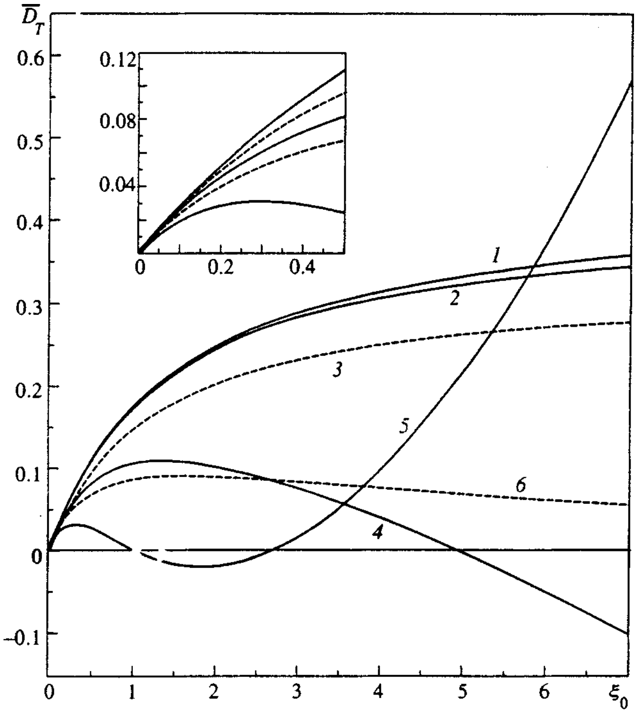


FIG. 2. The dimensionless turbulent diffusion coefficient $\overline{D}_T=(p_0/u_0)D_T$ for the case of turbulence with the energy spectrum (30). The notation is the same as in Fig. 1.

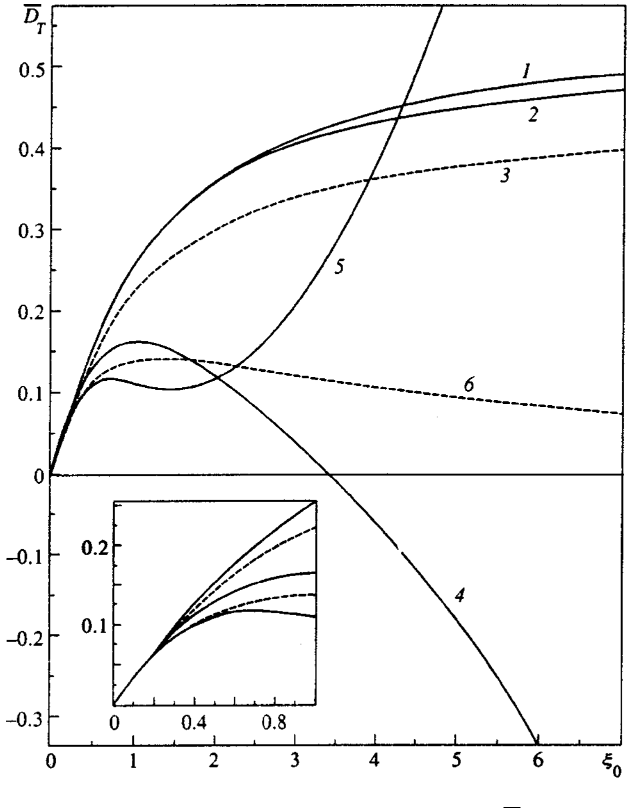


FIG. 3. The dimensionless turbulent diffusion coefficient $\overline{D}_T=(p_0/u_0)D_T$ for the case of turbulence with the energy spectrum (31). The notation is the same as in Fig. 1.

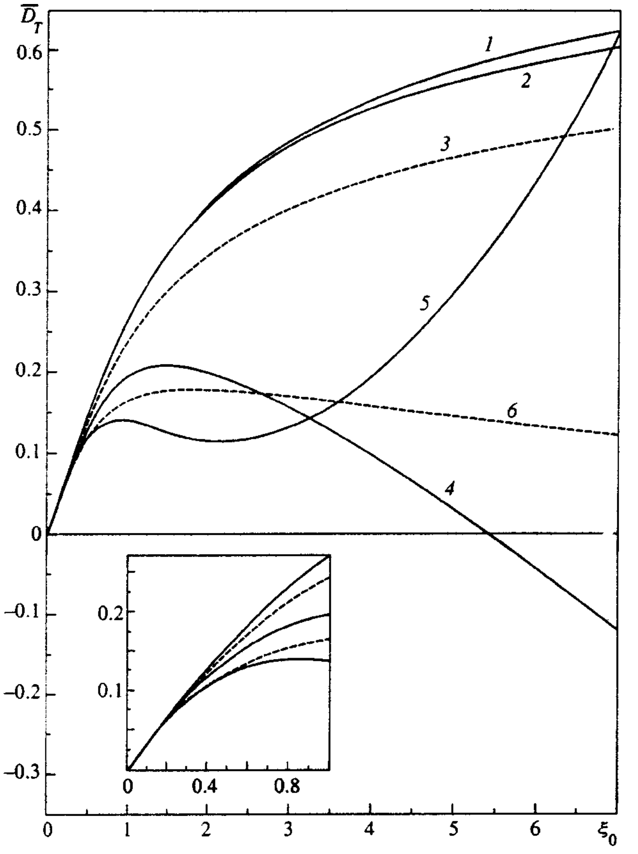


FIG. 4. The dimensionless turbulent diffusion coefficient $\overline{D}_T=(p_0/u_0)D_T$ for the case of turbulence with the energy spectrum (32). The notation is the same as in Fig. 1.

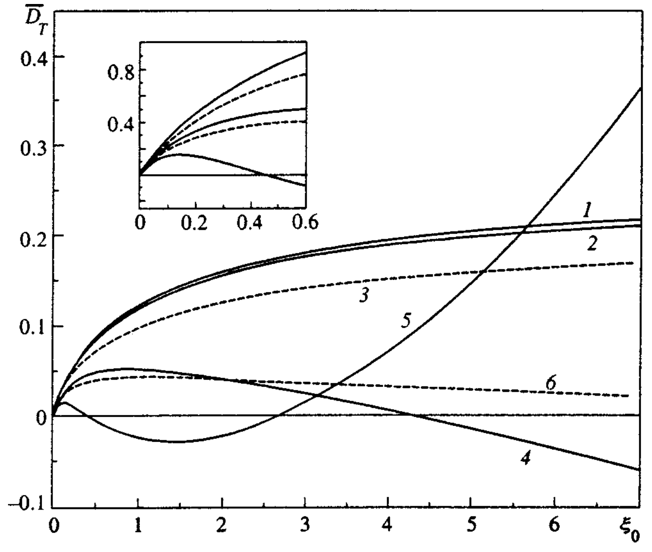


FIG. 5. The dimensionless turbulent diffusion coefficient $\overline{D}_T=(p_0/u_0)D_T$ for the case of turbulence with the energy spectrum (33). The notation is the same as in Fig. 1.

region $\xi_0 \geqslant 1$. The point is that if we consider turbulence as being a chaotic mixture of nonoscillating compressions and dilations (our correlators describe just such motion), it is meaningless to examine cases with $\xi_0 \geqslant 1$, since they are not realized in nature. If long-lived vortices can exist in an incompressible medium, i.e., if cases with $\xi_0 \gg 1$ exist in nature, bulk nonoscillating compressions and dilations have a lifetime τ_0 directly related to the compression (or dilation) ratio $|\Delta V|/V$. To estimate τ_0 we use the well-known relationship

$$\left(\frac{\Delta V}{V\tau_0}\right)^2 \cong \langle \mathrm{div}^2\, \mathbf{u}(\mathbf{r},t)\rangle \equiv \int_0^\infty dp\, p^2 E_{\mathrm{comp}}(p) \approx u_0^2 p_0^2. \tag{34}$$

This leads to the desired expression $\tau_0 \approx \langle |\Delta V/V|\rangle / u_0 p_0$ and hence $\xi_0 = u_0 p_0 \tau_0 \approx \langle |\Delta V/V|\rangle$. The value of the average relative variation of volume determining the parameter ξ_0 depends on the specific physical conditions of compression and dilation of the gas. For instance, for adiabatic processes[24] we have $|\Delta V/V| < 2\gamma/(1+\gamma)$, where γ is the adiabatic exponent ($p=\mathrm{const}\times\rho^\gamma$) @. For a monatomic gas, $\gamma=5/3$ and $|\Delta V/V|<5/4$.

Figures 1–5 shows that the compressibility in nonacoustic turbulence significantly reduces the value of the turbulent diffusion coefficient D_T compared to the case of incompressible turbulence, which has the same generalized spectrum in the $\xi_0 \geqslant 0.5$ range. This explains the ineffectiveness of compressions and dilations in the transport of impurities from one point to another in comparison with convective motion in an incompressible medium. Note that the very occurrence of negative values of D_T in DIA solutions for a compressible medium is natural, since it is in such a medium that large-scale turbulent structures become prominent.[25]

In reality the compressible and incompressible turbulent movements coexist. The above general formulas allow the diffusion coefficients to be calculated in such cases, too. More than that, in a mixed case the convergence of the method of successive approximations in solving the DIA equation is better than in purely irrotational turbulence.

6. CONCLUSION

We have derived general formulas for calculating the turbulent diffusion coefficient D_T in a compressible medium, formulas that allow for the contributions of all powers of the second- and fourth-order velocity correlators. We have also found that the additional term, ignored by other researchers, that describes the correlation of div**u** with the stochastic Green's function contributes substantially to D_T. For the first time we have numerically solved the nonlinear DIA equation for a number of models of compressible turbulence and have shown that compressibility substantially reduces D_T in the case of nonacoustic turbulence in comparison to the value of D_T for incompressible turbulence with the same energy spectrum. On the other hand, for acoustic turbulence, due to compressibility, the value of D_T is larger than that of the molecular diffusion coefficient D_m, the two remaining comparable in order of magnitude, as established with a lower accuracy by other studies.[18] We have found that the DIA equation is quite suitable for finding D_T in compressible media if one keeps to the physically admissible values of the parameter $\xi_0 = u_0 p_0 \tau_0$. All the results of this paper are also fully applicable to the turbulent diffusion coefficients in the case of magnetic-field diffusion in a medium without helicity.

APPENDIX

We perform a simple derivation of an expression for the velocity correlator in the case of acoustic turbulence and show that the decay of the correlator with the passage of time is due to the random irregularities in the phases. Bearing in mind that $u_i(\mathbf{r},t)=\nabla_i\varphi(\mathbf{r},t)$, where $\varphi(\mathbf{r},t)$ is the stochastic potential, we can write the chaotic wave ensemble in the form[26]

$$\varphi(\mathbf{r},t)=\sum_n \varphi_n(\mathbf{p}_n,t)\cos[\mathbf{p}_n\cdot\mathbf{r}-\omega_n t+\alpha_{0n}+\alpha_n(t)], \tag{A1}$$

where $\mathbf{p}_n$ is the wave vector of the nth wave, and $\omega_n=cp_n$ is the wave frequency. Since the waves are absorbed by the medium, the amplitude φ_n depends on time t, with α_{0n} the random initial phase of the wave, and $\alpha_n(t)$ the current phase as the wave passes through the medium (this quantity is assumed to change only slightly in the course of a time interval equal to the wave period). The velocity correlator is $B_{ij}(\mathbf{R},\tau)=-\nabla_i\nabla_j W(\mathbf{R},\tau)$, where $W(\mathbf{R},\tau)=\langle\varphi(\mathbf{r},t)\varphi(\mathbf{r}',t')\rangle$ is the potential correlator, $\mathbf{R}=\mathbf{r}-\mathbf{r}'$, $\tau=t-t'$, and

$$\begin{aligned}\langle\varphi(\mathbf{r},t)\varphi(\mathbf{r}',t')\rangle=\frac{1}{2}\sum_{n,m}\langle\varphi_n\varphi_m\rangle\{&\cos[\mathbf{p}_n\cdot\mathbf{r}+\mathbf{p}_m\cdot\mathbf{r}'-\omega_n t\\ &-\omega_m t'+\alpha_{0n}+\alpha_{0m}+\alpha_n(t)+\alpha_m(t')]\\ &+\cos[\mathbf{p}_n\cdot\mathbf{r}-\mathbf{p}_m\cdot\mathbf{r}'-\omega_n t+\omega_m t'+\alpha_{0n}\\ &-\alpha_{0m}+\alpha_n(t)-\alpha_m(t')]\}.\end{aligned} \tag{A2}$$

Averaging with the weight $1/2\pi$ over the independent initial phases α_{0n} and α_{0m} yields $n=m$, and averaging the amplitudes yields $\langle\varphi_n(t)\varphi_n(t')\rangle=2\Phi(p_n,\tau)$ in the stationary case. The expression for $W(R,\tau)$ becomes

$$W(\mathbf{R},\tau)=\sum_n \Phi(p_n,\tau)\langle\cos[\mathbf{p}_n\cdot\mathbf{R}-\omega_n\tau+\alpha_n(t)-\alpha_n(t')]\rangle. \tag{A3}$$

If we assume that $\alpha_n(t)$ is a stationary Gaussian process, i.e., $\langle\alpha_n(t)\alpha_n(t+\tau)\rangle=\alpha_0^2\alpha(|\tau|)$, where $\alpha(\tau)$ is equal to unity at $\tau=0$ and decays with the passage of time, e.g., as $\exp(-\tau/\tau_0)=1-\tau/\tau_0$. It is known[26] that

$$\begin{aligned}&\langle\exp\{\pm i[\alpha_n(t)-\alpha_n(t')]\}\rangle\\ &\quad=\exp\{-\alpha_0^2[1-\alpha(\tau)]\}\simeq\exp\left(-\frac{\alpha_0^2\tau}{\tau_0}\right).\end{aligned} \tag{A4}$$

As a result, (A3) acquires the final form

$$W(\mathbf{R},\tau)=\sum_n \Phi(p_n,\tau)\cos(\mathbf{p}_n\cdot\mathbf{R}-\omega_n\tau)\exp\left(-\frac{\alpha_0^2\tau}{\tau_0}\right). \tag{A5}$$

If we pass from the sum to an integral with respect to the wave vectors, we arrive at an expression equivalent to (23).

*)E-mail: silant@inaoep.mx

---

[1] A. S. Monin and A. M. Yaglom, *Statistical Hydrodynamics*, Vol. 2, M.I.T. Press, Cambridge, Mass. (1975).

[2] G. I. Taylor, Proc. London Math. Soc. A **20**, 196 (1921).

[3] H. K. Moffatt, J. Fluid Mech. **65**, 1 (1974).

[4] R. H. Kraichnan, J. Fluid Mech. **75**, 657 (1976).

[5] A. Z. Dolginov and N. A. Silant'ev, Geophys. Astrophys. Fluid Dyn. **63**, 139 (1992).

[6] R. H. Kraichnan, J. Fluid Mech. **5**, 497 (1959).

[7] P. H. Roberts, Fluid Mech. Res. **11**, 257 (1961).

[8] R. H. Kraichnan, Phys. Fluids **13**, 22 (1970).

[9] R. H. Kraichnan, J. Fluid Mech. **77**, 753 (1976).

[10] N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **101**, 1216 (1992) [Sov. Phys. JETP **74**, 650 (1992)].

[11] N. A. Silant'ev, Geophys. Astrophys. Fluid Dyn. **75**, 183 (1994).

[12] N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **111**, 871 (1997) [JETP **84**, 479 (1997)].

[13] N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **112**, 1312 (1997) [JETP **85**, 712 (1997)].

[14] K. O. Eschrich, Astron. Nachr. **299**, 137 (1978).

[15] N. A. Silant'ev, Pis'ma Astron. Zh. **16**, 860 (1990) [Sov. Astron. Lett. **16**, 367 (1990)].

[16] N. A. Silant'ev and M. D. Korolkov, Astron. Nachr. **311**, 107 (1990).

[17] S. I. Vaĭnshteĭn, Ya. B. Zel'dovich, and A. A. Ruzmaĭkin, *Turbulent Dynamo in Astrophysics* [in Russian], Nauka, Moscow (1980).

[18] A. P. Kazantsev, A. A. Ruzmaĭkin, and D. D. Sokolov, Zh. Éksp. Teor. Fiz. **88**, 487 (1985) [Sov. Phys. JETP **61**, 285 (1985)].

[19] P. H. Roberts, *An Introduction to Magnetohydrodynamics*, Elsevier, New York (1967).

[20] E. N. Parker, *Cosmical Magnetic Fields*, Clarendon Press, Oxford (1979).

[21] G. K. Batchelor, *The Theory of Homogeneous Turbulence*, Cambridge Univ. Press, London (1953).

[22] R. Phythian and W. D. Curtis, J. Fluid Mech. **89**, 241 (1978).

[23] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon, Oxford (1987); Russ. original Nauka, Moscow (1988).

[24] J. I. Katz, *High Energy Astrophysics*, Addison-Wesley, Reading, MA (1987).

[25] S. S. Moiseev, R. Z. Sagdeev, A. V. Tur, G. A. Khomenko, and V. V. Yanovskiĭ, Zh. Éksp. Teor. Fiz. **85**, 1979 (1983) [Sov. Phys. JETP **58**, 1147 (1983)].

[26] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics* [in Russian], Nauka, Moscow (1981).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Helicity generation in turbulent MHD flows

O. G. Chkhetiani*) and S. S. Moiseev

Space Research Institute, Russian Academy of Sciences, 117810 Moscow, Russia

E. I. Golbraikh

Center for MHD Studies, Ben-Gurion University Beersheba, 84105, Israel

(Submitted 31 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 946–955 (September 1998)

We study turbulent flow of a conducting liquid in a uniform external magnetic field. It is shown that intense helicity generation is possible in the presence of a mean shear flow. It is noted that even though the mean helicity of the initial flow can be zero, the presence of internal topological structure of the flow, for example the presence of helicity of different signs at different scales, is nevertheless necessary for helicity generation. © *1998 American Institute of Physics.* [S1063-7761(98)01309-2]

1. INTRODUCTION

Helicity ($H=\langle\mathbf{v}\cdot\nabla\times\mathbf{v}\rangle$) has a substantial effect on the stability and evolution of both laminar and turbulent flows.[1] Helical cascades, first introduced in Ref. 2 and analyzed in Ref. 3 for various special cases, are closely related to helical turbulence. Two basic limiting cases[2] have been identified—parallel energy and helicity fluxes along the spectrum

$$E(k)\propto\overline{\varepsilon}^{2/3}k^{-5/3},\quad H(k)\propto\overline{\eta\varepsilon}^{-1/3}k^{-5/3},$$

which correspond to a Kolmogorov cascade, and a helicity flux without an energy flux

$$E(k)\propto\overline{\eta}^{2/3}k^{-7/3},\quad H(k)\propto\overline{\eta}^{2/3}k^{-4/3},$$

which is a purely helical cascade ($\overline{\varepsilon}$ and $\overline{\eta}$ are the average dissipation of energy and helicity).

For a long time, helical cascades were treated as purely theoretical curiosities irrelevant to existing experimental data. However, a substantial amount of baggage consisting of experimental laboratory and natural observations of the $-7/3$ spectrum over a wide range of spatial scales has now been accumulated. Such spectra are observed in laboratory MHD flows,[4–6] rotating liquids,[7] stratified turbulence behind a grid,[8] boundary layers,[9] direct measurements of atmospheric turbulence at various altitudes,[10–12] and the tropical pre-typhoon atmosphere.[13] We note that two scaling regions often exist in both atmospheric and laboratory MHD turbulence spectra: $(-5/3)$ and $(-7/3)$.

The existence of a ''helical'' spectrum raises the problem of the origin of helicity. It is well known that helicity is an invariant of the ''pure'' Euler equation (which neglects the influence of stratification, background rotation, and other factors). Helicity generation in stellar and planetary atmospheres is probably basically due to the rotation of the planet and stratification,[14–16,3] whereas the nature of the onset of helicity in MDH flows is as yet unclear. The origin of the helicity observed in laboratory flows is also not obvious. In Ref. 17, where a qualitative analysis of the observed helical spectra was performed, it is argued in this regard that reflection symmetry is spontaneously broken in three-dimensional turbulence.

Investigations of MHD turbulence have a long history. It is widely believed that turbulence becomes two-dimensional under the action of a magnetic field. Many experimental results[6,18] do not fit into the two-dimensional picture and can be explained only if three-dimensionality is taken into account. Such three-dimensional effects also include the observed helical spectra.[4–6] We note that attempts to give a two-dimensional interpretation of the results of laboratory measurements and numerical simulation of MHD turbulence are not the only possibilities. Spectra that can be explained as a manifestation of a two-dimensional -3 spectrum also correspond, to the same degree of accuracy, to a $-7/3$ spectrum.

The problem of helicity generation in MHD flows was first examined in the context of turbulent dynamo theory.[19] It was shown at that time that helicity opposite in sign to the original helicity is generated, intensifying perturbations of the large-scale magnetic field. Somewhat later this effect was explained from the standpoint of the conservation of the magnetic helicity invariant $H_m=\langle\nabla^{-1}\mathbf{h}\cdot\mathbf{h}\rangle$. Helicity was generated only in fields with $(\mathbf{B}\cdot\nabla\times\mathbf{B})\neq 0$ (here $\mathbf{B}$ is the large-scale magnetic field). In laboratory MHD flows, magnetic helicity vanishes, as a rule. Nonetheless, a helical energy cascade $E(k)\propto k^{-7/3}$ is observed over a wide range of external magnetic field values. The present paper is devoted to determining a possible mechanism of helicity generation in turbulent MHD flows.

2. STATEMENT OF THE PROBLEM

We examine turbulent MHD flow of an incompressible fluid in a uniform magnetic field $\mathbf{B}^0$ ($\mu=1$) and with a time-independent mean component of the mean velocity $\mathbf{U}^0$:

$$\partial_t\mathbf{v}-\mathbf{v}\times\operatorname{curl}\,\mathbf{v}+\nabla\frac{\mathbf{v}^2}{2}=-\nabla P/\rho$$
$$+\operatorname{curl}\mathbf{H}\times\mathbf{H}/4\pi\rho+\nu\Delta\mathbf{v},\qquad(1)$$

$$\mathrm{div}\,\mathbf{v}=0, \tag{2}$$

$$\partial_t\mathbf{H}+(\mathbf{v}\cdot\nabla)\mathbf{H}-(\mathbf{H}\cdot\nabla)\mathbf{v}=\nu_m\Delta\mathbf{H}, \tag{3}$$

$$\mathrm{div}\,\mathbf{H}=0. \tag{4}$$

In what follows, we work in the Alfvén variables

$$\mathbf{H}'=\mathbf{H}/\sqrt{4\pi\rho},\quad \mathbf{H}=\mathbf{H}'.$$

In the absence of magnetic fields and neglecting kinematic viscosity, hydrodynamic helicity is a conserved quantity:

$$\partial_t\langle\mathbf{v}\cdot\mathrm{curl}\,\mathbf{v}\rangle=0. \tag{5}$$

The magnetic field leads to the emergence of ''sources'' in the helicity balance equation:

$$\begin{aligned}&\partial_t\langle\mathbf{v}\cdot\mathrm{curl}\,\mathbf{v}\rangle=-2\nu\langle\mathbf{w}\cdot\,\mathrm{curl}\,\mathbf{w}\rangle+M,\\ &M=2\langle\mathbf{w}\times\,\mathrm{curl}\,\mathbf{h}\rangle\cdot\mathbf{B}^0+2\langle\mathbf{h}\times\mathbf{w}\rangle\mathrm{curl}\,\mathbf{B}^0+2\boldsymbol{\Omega}^0\\ &\qquad\cdot\langle\mathrm{curl}\,\mathbf{h}\times\mathbf{h}\rangle+2\langle(\mathbf{w}\,\,\mathrm{curl}\,\mathbf{h})\times\mathbf{h}\rangle.\end{aligned} \tag{6}$$

Here

$$\mathbf{v}=\mathbf{U}^0+\mathbf{u},\quad \langle\mathbf{v}\rangle=\mathbf{U}^0,$$

$$\mathrm{curl}\,\mathbf{v}=\mathbf{U}^0+\mathbf{w},\quad \langle\mathrm{curl}\,\mathbf{v}\rangle=\boldsymbol{\Omega}^0,$$

$$\mathbf{H}=\mathbf{B}^0+\mathbf{h},\quad \langle\mathbf{h}\rangle=0.$$

In the present paper we study the limit of large Alfvén velocities ($|\mathbf{B}^0|\gg|\mathbf{U}^0|$) and neglect in Eq. (6) the corresponding contribution of the mean velocity. We note that $\langle\mathbf{v}\cdot\nabla\times\mathbf{v}\rangle$ is the mean helicity of a flow only in homogeneous and isotropic turbulence. In the general inhomogenous case, $\langle\mathbf{v}\cdot\nabla\times\mathbf{v}\rangle$ corresponds to an ensemble-averaged helicity density. We consider the mean flow with zero helicity $(\mathbf{U}^0\cdot\boldsymbol{\Omega}^0)=0$. Then

$$\langle\mathbf{v}\cdot\,\mathrm{curl}\,\mathbf{v}\rangle=\langle\mathbf{u}\cdot\,\mathrm{curl}\,\mathbf{u}\rangle.$$

The first term in Eq. (6) was first studied by Vaĭnshteĭn in Ref. 19. Only the anisotropy introduced into the correlations of the turbulent velocity field by a quasistationary magnetic field was taken into account. In a uniform magnetic field $\nabla\times\mathbf{B}^0=0$. In the theory of a nonlinear dynamo with $\nabla\times\mathbf{B}^0\neq0$, this term is responsible for nonlinear saturation of large-scale instability.

The cubic nonlinearity in Eq. (6) can be approximated in the spirit of the Orszag approximation as a τ-relaxing term. Its influence reduces to redistribution of energy and helicity over both scales and directions (to renormalization of turbulent visocity and magnetic diffusion; see, for example, Ref. 20). In the present paper we examine in detail the first three terms on the right-hand side of Eq. (6).

3. INDUCTION EQUATION IN A SHEAR FLOW

The fluctuation magnetic field $\mathbf{h}$ excited by a turbulent velocity field is described by the induction equation

$$\partial_t\mathbf{h}-\mathrm{curl}\,[(\mathbf{U}^0+\mathbf{u})\times\mathbf{h}]=(\mathbf{B}^0\cdot\nabla)\mathbf{u}+\nu_m\Delta\mathbf{h}. \tag{7}$$

Neglecting magnetic viscosity, the solution of this equation has the form

$$\mathbf{h}(t)=\mathbf{e}^{-(\hat{A}+\hat{F})t}\,\mathbf{h}(0)+\int_0^t e^{-(\hat{A}+\hat{F})(t-s)}(\mathbf{B}^0\cdot\nabla)\mathbf{u}(s)\,ds, \tag{8}$$

where

$$\hat{A}\mathbf{h}=\mathrm{curl}\,[\mathbf{U}^0\times\mathbf{h}],\quad \hat{F}\mathbf{h}=\mathrm{curl}\,[\mathbf{u}\times\mathbf{h}].$$

Examining the series expansion of the exponential operator in the integrand in Eq. (8), it is easy to see that the expansion of the fluctuation part of the exponential (the $\hat{F}$ operator), starting with the second term (in powers of the velocity), is equivalent to taking the higher-order moments of the turbulent velocity field into account. The expansion parameter is the turbulent Strouhal number $S=u_{\mathrm{tur}}\tau_{cor}/\lambda$, which is based on the main characteristics of the turbulent flow.

For small Strouhal numbers ($S\ll1$), the term $\nabla\times(\mathbf{u}\times\mathbf{h})$ in the magnetic induction equation (7) can be dropped. This corresponds to so-called first-order scaling theory (first-order smoothing approximation—FOSA), employed in the quasilinear theory of plasma and in dynamo theory.[21] In this case there are no restrictions on the magnetic Reynolds number Re_m. Under ordinary conditions the Strouhal number is of order 1. But the basic physical effects can be ''fished out'' even for small Strouhal numbers. More complicated approximations, which as a rule also include elements of numerical simulation, merely renormalize the specific numerical values of the turbulent transport coefficients (diffusion, viscosity, helicity generation coeffiicents, and so on). Neglecting $\nabla\times(\mathbf{u}\times\mathbf{h})$ is also correct in another limiting case—low magnetic Reynolds numbers, $\mathrm{Re}_m\ll1$, and arbitrary Strouhal numbers.[21]

Henceforth we work in the two-scale approximation

$$L\gg\lambda,\quad T\gg\tau,$$

where L, T and λ,τ are characteristic space-time scales of mean and fluctuating quantities. In this case we can take the shear of the mean velocity to be constant (uniform) in the small-scale equations. The magnetic field $\mathbf{h}$ in a flow with constant shear can be determined in terms of the Green's function, first obtained by K.-H. Rädler:[21]

$$G_{ij}(\mathbf{x},\xi,t,\tau)=\gamma_{ij}(t-\tau)\,G(\mathbf{x}-\xi,t-\tau|\mathbf{U}^0).$$

Here

$$\gamma_{ij}(t)=\delta_{ij}+\frac{\partial U_i}{\partial x_j}\,t,$$

$$G(\mathbf{x},t|\mathbf{U}^0)=\begin{cases}\left(\dfrac{1}{4\pi\nu_m t}\right)^{3/2}\exp\left\{-\dfrac{d_{pq}x_px_q}{4\nu_m t}\right\}, & t\geq0,\\ 0, & t<0.\end{cases}$$

For short correlation times the term $d_{pq}(t)$ has the form (see Ref. 21)

$$d_{pq}(t)=\delta_{pq}-\frac{1}{2}\left(\frac{\partial U_p}{\partial x_q}+\frac{\partial U_q}{\partial x_p}\right)t,$$

and $\gamma_{ij}\approx\delta_{ij}$. In this case the solution of Eq. (7) is

$$\mathbf{h}(t)=\int G(\mathbf{x}-\xi,t)\,\mathbf{h}(\xi,0)\,d\xi+\int\!\!\int G(\mathbf{x}-\xi,t-\tau)\times(\mathbf{B}^0\cdot\nabla)\mathbf{u}(\xi,\tau)\,d\xi\,d\tau. \tag{9}$$

For times $T \gg \tau$ the initial values of the fluctuation magnetic field $\mathbf{h}$ can be neglected.

4. CALCULATION OF TWO-POINT CORRELATIONS

Taking account of Eq. (9), we obtain

$$\mathbf{\Omega}^0\langle \operatorname{curl}\mathbf{h}\times\mathbf{h}\rangle=\mathbf{\Omega}^0\hat{\mathcal{G}}_1\hat{\mathcal{G}}_2(\mathbf{B}^0\cdot\nabla_1)(\mathbf{B}^0\cdot\nabla_2)\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle \tag{10}$$

and

$$\langle\mathbf{w}\times\operatorname{curl}\mathbf{h}\rangle\cdot\mathbf{B}^0=\hat{\mathcal{G}}_1\langle\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1\rangle\cdot\mathbf{B}^0. \tag{11}$$

Here $|_1$and $|_2$ correspond to $(\mathbf{x}_1,t_1)$ and $(\mathbf{x}_2,t_2)$ respectively. The following two-point correlations must be determined:

$$(\mathbf{B}^0\cdot\nabla_1)(\mathbf{B}^0\cdot\nabla_2)\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle,\quad \langle\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1\rangle. \tag{12}$$

In homogeneous isotropic turbulence only the term $\langle\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1\rangle$ is nonvanishing (see Ref. 19). The effects of weak anisotropy due to a uniform magnetic field were also studied in Ref. 22. The presence of a mean shear flow leads to inhomogeneities and additional anisotropy of the correlation properties of a turbulent velocity field. The principal linear terms of the Reynolds stresses have a similar gradient form in the various models of turbulent closure. Small differences are present in the definition of the transport coefficients (K–l, K–ε models and others). Nonlinear generalizations of Reynolds stresses make sense only for flows with solid boundaries (rectangular and circular channels, etc.). In the FOSA approximation the nonlinearity in the equation, which governs the anisotropic properties of the turbulence, can be neglected.

Comparing the terms in the equation for the velocity, we see that the main parameter characterizing the degree of influence of the magnetic field is the dimensionless combination $(B^{02}/U^0 v_{\text{tur}})S$, where B^0 (in Alfvén variables) equals the Alfvén velocity in magnitude, U^0 is the characteristic mean velocity, v_{tur} is the characteristic turbulent velocity, and S is the Strouhal number. Henceforth we assume that

$$\frac{B^{02}}{U^0 v_{\text{tur}}}S\gg 1. \tag{12a}$$

In particular, in the laboratory setup at the Center for MHD Studies at Ben-Gurion University, the typical magnetic fields were $(3-4)\times 10^3$ G, the velocity of mercury flow along a cylindrical pipe 10 cm in diameter was 20 cm/s, and the typical velocity of the turbulent pulsations was 1–2 cm/s. Strouhal numbers greater than 5×10^{-3} are sufficient to satisfy the criterion (12a). The prerequisite $S\ll 1$ for first-order smoothing theory is satisfied for these values.

We note that this criterion is also satisfied in the solar plasmasphere.

Accordingly, the effect on two-point correlations of shear in the mean velocity can be treated as a linear correction to the underlying state determined by the magnetic field.

The effect of a uniform magnetic field on the correlation characteristics of turbulence was analyzed in Refs. 18, 19, 21, and 23. The main effect is that a uniform magnetic field draws energy from turbulent pulsations and reradiates it in the form of Alfvén waves propagating along the magnetic field lines. The structure of the two-point correlation tensor remains unchanged in Fourier space; the only change is that the coefficients become dependent on the angle between the wave vector and the magnetic field. Note that the form of the correlation tensor obtained in Refs. 19 and 23 is valid in weak magnetic fields and can be obtained by an appropriate expansion of the more general expression obtained in Ref. 21.

We assume that turbulent pulsations of the velocity field in the absence of an external magnetic field are homogeneous and isotropic. Then the velocity correlation tensor in Fourier space has the form

$$\hat{\mathcal{Q}}^0_{ij}(\mathbf{k},\omega)=\frac{E(k,\omega)}{4\pi k^2}\left(\delta_{ij}-\frac{k_ik_j}{k^2}\right)+i\frac{H(k,\omega)}{8\pi k^4}\varepsilon_{ijk}k_k. \tag{13}$$

The mean energy and helicity of the turbulent flow are

$$\langle\mathbf{u}\cdot\mathbf{u}\rangle=\int E(k,\omega)\,dk\,d\omega,$$

$$\langle\mathbf{u}\cdot\operatorname{curl}\mathbf{u}\rangle=\int H(k,\omega)\,dk\,d\omega. \tag{14}$$

Under conditions in which first-order smoothing theory (FOSA) is applicable, $\min(S,\mathrm{Re}_m)\ll 1$, the correlation tensor, taking account of the uniform external magnetic field, has the form (see Ref. 21)

$$\hat{\mathcal{Q}}^h_{ij}(\mathbf{k},\omega)=\frac{\hat{\mathcal{Q}}^0_{ij}(\mathbf{k},\omega)}{1+(\mathbf{B}^0\cdot\mathbf{k})^2\dfrac{2\nu\nu_mk^4-2\omega^2+(\mathbf{B}^0\cdot\mathbf{k})^2}{(\nu_m^2k^4+\omega^2)(\nu^2k^4+\omega^2)}}. \tag{15}$$

The field of turbulent pulsations can provisionally be separated into homogeneous (neglecting shear) $\mathbf{u}^h$ and inhomogeneous (taking shear into account) $\mathbf{u}^{nh}$ components. The inhomogeneous component satisfies the linearized equation

$$(\partial_t-\nu\Delta)\mathbf{u}^{nh}=-\nabla p^{nh}+\operatorname{curl}\mathbf{h}\times\mathbf{B}^0-(\mathbf{U}^0\cdot\nabla)\mathbf{u}^h-(\mathbf{u}^h\cdot\nabla)\mathbf{U}^0, \tag{16}$$

where $\mathbf{h}$ is given by (9). In the present case, to obtain the form of the inhomogeneous Reynolds stresses, one can neglect the effect of the mean velocity on the magnetic field induced by turbulent pulsations. Then the equation for $\mathbf{u}^{nh}$ has the form

$$(\partial_t-\nu\Delta-(\mathbf{B}^0\cdot\nabla)^2\hat{G})\mathbf{u}^{nh}=-\nabla p^{nh}-(\mathbf{U}^0\cdot\nabla)\mathbf{u}^h-(\mathbf{u}\cdot{}^h\nabla)\mathbf{U}^0. \tag{17}$$

1. Calculation of $(\mathbf{B}^0\cdot\nabla_1)(\mathbf{B}^0\cdot\nabla_2)\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle$. Multiplying the equation for $\mathbf{w}(\mathbf{x}_1,t_1)$ by $\mathbf{u}(\mathbf{x}_2,t_2)$ and carrying out the spatial averaging, we obtain

$$(\partial_{t_1}-\nu\Delta_{\xi}-(\mathbf{B}^0\cdot\nabla_{\xi})^2\hat{G})\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle^{nh}$$
$$=\langle[\mathrm{curl}\,[\mathbf{U}^0\times\mathbf{w}]+[\mathbf{u}\times\mathbf{\Omega}^0]|_1\times\mathbf{u}|_2]\rangle,$$
$$\xi=\mathbf{x}_2-\mathbf{x}_1\,.$$

After a number of calculations we obtain

$$(\mathbf{B}^0\cdot\nabla_1)(\mathbf{B}^0\cdot\nabla_2)\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle|_i^{nh}$$
$$=\hat{\mathcal{G}}^{B^0}B_u^0B_v^0\varepsilon_{ijk}\varepsilon_{jlm}\varepsilon_{tpr}(\delta_{tm}\delta_{ks}+\delta_{tk}\delta_{ms})$$
$$\times\frac{\partial}{\partial x_l^1}\left\{\frac{\partial U_p^0}{\partial x_u^1}\left\langle w_r\frac{\partial u_s}{\partial\xi_v}\right\rangle+U_p^0\Big|_1\left\langle\frac{\partial w_r}{\partial\xi_u}\frac{\partial u_s}{\partial\xi_v}\right\rangle\right.$$
$$\left.+\Omega_r^0\Big|_1\left\langle\frac{\partial u_p}{\partial\xi_u}\frac{\partial u_s}{\partial\xi_v}\right\rangle+\frac{\partial\Omega_r^0}{\partial x_u^1}\left\langle u_p\frac{\partial u_s}{\partial\xi_v}\right\rangle\right\}. \tag{18}$$

Here

$$(\hat{\mathcal{G}}^{B^0}\mathbf{v})_{\mathbf{k},w}=\left[-i\omega+\nu k^2+\frac{(\mathbf{B}^0\cdot\mathbf{k})^2}{-i\omega+\nu_m k^2}\right]^{-1}\hat{\mathbf{v}}(\mathbf{k},\omega).$$

For the homogeneous component of the turbulent velocity field we obtain

$$(\mathbf{B}^0\cdot\nabla_1)(\mathbf{B}^0\cdot\nabla_2)\langle\mathbf{w}|_1\times\mathbf{u}|_2\rangle|_i^{h}$$
$$=B_{0m}B_{0n}\left\{\left\langle\frac{\partial^2u_i}{\partial\xi_m\partial\xi_k}\frac{\partial u_k}{\partial\xi_n}\right\rangle-\left\langle\frac{\partial^2u_k}{\partial\xi_m\partial\xi_i}\frac{\partial u_k}{\partial\xi_n}\right\rangle\right\}.$$

After integrating over angles the homogeneous part vanishes.

For low Strouhal numbers (short correlation times), the mean value theorem can be used to calculate the integrals—the spatial derivatives of the mean and turbulent velocities can be removed from the integrand and set equal to their average (local) values. Taking the scalar product of $\mathbf{\Omega}^0$ with Eq. (19) we obtain

$$M_A=\chi_1(\mathbf{\Omega}^{0^2}\mathbf{B}^{0^2}-(\mathbf{\Omega}^0\cdot\mathbf{B}^0)^2+\mathbf{\Omega}^0(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\times\mathbf{U}^0))$$
$$+\chi_0(\mathbf{\Omega}^0\cdot\nabla)(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\cdot\mathbf{\Omega}^0)$$
$$+\varepsilon_1(\mathbf{\Omega}^0\cdot\nabla)(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\cdot\mathbf{U}^0)+\varepsilon_2(2\mathbf{B}^{0^2}(\mathbf{\Omega}^0\cdot\Delta\mathbf{U}^0)$$
$$-(\mathbf{\Omega}^0\cdot(\mathbf{B}^0\cdot\nabla)^2\mathbf{U}^0)-(\mathbf{\Omega}^0\cdot\nabla)(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\cdot\mathbf{U}^0)). \tag{19}$$

Here

$$\chi_0\simeq\frac{2}{3}\tau_A^3\langle\mathbf{u}\cdot\mathbf{w}\rangle,\quad \chi_1\simeq\frac{2}{15}\tau_A^3\langle\mathbf{w}\cdot\mathrm{curl}\,\mathbf{w}\rangle \tag{20}$$

$$\varepsilon_1\simeq\frac{1}{3}\tau_A^3\langle\mathbf{w}\cdot\mathbf{w}\rangle,\quad \varepsilon_2\simeq\frac{2}{15}\tau_A^3\langle\mathrm{curl}\,\mathbf{w}\cdot\mathrm{curl}\,\mathbf{w}\rangle \tag{21}$$

are proportional to the turbulent helicity, superhelicity, enstrophy, and superenstrophy, respectively, and

$$\tau_A^3\simeq\int G(\mathbf{x}-\mathbf{x}_1,t-t_1)G(\mathbf{x}-\mathbf{x}_2,t-t_2)\hat{\mathcal{G}}^{B^0}$$
$$\times(\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}^*,t_1-t^*)g(\mathbf{x}^*,t^*)$$
$$\times d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{x}^*\,dt_1\,dt_2\,dt^*. \tag{22}$$

Here $g(\mathbf{x}^*,t^*)$ is the space-time dependence for the corresponding correlation function.

2. Calculation of $\langle\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1\rangle$. Multiplying the equation for $\mathbf{w}(\mathbf{x},t)$ by $\mathbf{w}(\mathbf{x}_1,t_1)$ and carrying out the spatial averaging, we obtain

$$(\partial_t-\nu\Delta_{\xi}-(\mathbf{B}^0\cdot\nabla_{\xi})^2\,\hat{G})\langle\mathbf{w}\times\mathbf{w}|_1\rangle^{nh}$$
$$=\langle\mathrm{curl}\,(\mathbf{U}^0\times\mathbf{w}+\mathbf{u}\times\mathbf{\Omega}^0)\times\mathbf{w}|_1\rangle,\quad \xi=\mathbf{x}-\mathbf{x}_1\,. \tag{23}$$

After a number of calculations we obtain

$$\langle\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1\rangle\,|_i^{nh}$$
$$=\hat{\mathcal{G}}^{B^0}B_t^0(\delta_{kl}\delta_{im}-\delta_{km}\delta_{il})(\varepsilon_{kfn}\delta_{mpr}$$
$$+(\delta_{pk}\delta_{rf}-\delta_{pf}\delta_{rk})\delta_{mn})$$
$$\times\left\{\frac{\partial U_p^0}{\partial x_l}\left\langle w_r\frac{\partial^2u_n}{\partial\xi_f\partial\xi_t}\right\rangle+\frac{\partial\Omega_r^0}{\partial x_l}\left\langle u_p\frac{\partial^2u_n}{\partial\xi_f\partial\xi_t}\right\rangle\right\}. \tag{24}$$

For the homogeneous component of the turbulent velocity field we have

$$M_B^0=\hat{\mathcal{G}}^1\langle(\mathbf{w}\times(\mathbf{B}^0\cdot\nabla_1)\mathbf{w}|_1)\,\rangle|_i^h$$
$$\simeq-\frac{1}{6}(\mathbf{B}^{0^2})\,\tau_0\langle\mathbf{w}\cdot\mathrm{curl}\,\mathbf{w}\rangle, \tag{25}$$

where

$$\tau_0\simeq\int G(\mathbf{x},t)g(\mathbf{x},t)\,d\mathbf{x}\,dt.$$

This term is proportional to the turbulent superhelicity $\langle\mathbf{w}\cdot\nabla\times\mathbf{w}\rangle$. In Ref. 19, it was treated only as a dissipative term. However, since helicity is not a sign-definite quantity, it can be both dissipative and productive. Indeed, a mean helicity of zero does not mean that the superhelicity is zero. Positive and negative helicity density on different scales can be a source of both subsequent redistribution of helicity and generation of helicity. For large Reynolds numbers the magnetic generation of helicity is much greater than the dissipative helicity.

Multiplying Eq. (24) by $\mathbf{B}^0$ and using the mean value theorem, we obtain

$$M_B\simeq-\varepsilon_1'(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\cdot\mathbf{\Omega}^0)+\chi_1'(\mathbf{B}^0\cdot\nabla)(\mathbf{B}^0\cdot\mathbf{U}^0). \tag{26}$$

Here

$$\varepsilon_1'\simeq\frac{1}{3}\tau_B^2\langle\mathbf{w}\cdot\mathbf{w})\rangle,\quad \chi_1'\simeq\frac{1}{3}\tau_B^2\langle\mathbf{w}\cdot\mathrm{curl}\,\mathbf{w})\rangle$$

are proportional to the turbulent enstrophy and superhelicity, respectively, and

$$\tau_B^2\simeq\int G(\mathbf{x}-\mathbf{x}_1,t-t_1)\,\hat{\mathcal{G}}^{\,B^0}(\mathbf{x}-\mathbf{x}_1-\mathbf{x}^*,t-t^*)$$
$$\times g(\mathbf{x}^*,t^*)\,d\mathbf{x}_1\,d\mathbf{x}^*\,dt_1\,dt^*.$$

Finally, the helicity balance equation takes the form

$$\partial_t\langle\mathbf{v}\cdot\mathrm{curl}\,\mathbf{v}\rangle=M_A+M_B^0+M_B-2\nu\langle\mathbf{w}\cdot\mathrm{curl}\,\mathbf{w}\rangle.$$

The new terms in Eqs. (20) and (26) in the helicity balance equation are functions of the external magnetic field, the shear (vorticity) of the mean velocity, and the correlation functions of the turbulent velocity field. Helicity production is highly nonlinear. The dependence on the external magnetic field is more complicated because of its influence on the coefficients ε and χ.

Hence, turbulent flow with nonuniform shear of the mean velocity and a uniform external magnetic field can generate hydrodynamic helicity.

5. CONCLUSIONS

We have derived a balance equation for hydrodynamic helicity in turbulent MHD flow. A direct analysis reveals that the shear of the mean velocity and the resulting anisotropy of the correlation properties of the random velocity field are sufficient conditions for helicity generation. We emphasize that helical spectra have been observed in laboratory MHD flows only with a grid (honeycomb) placed in the flow, resulting in the emergence of abrupt spatial gradients of the mean velocity. We note especially that under these circumstances, either a regular mean or turbulent velocity component in the main flow should have internal (hidden) topological structure—specifically, nonvanishing superhelicity $\langle \mathbf{w}\cdot\nabla\times\mathbf{w}\rangle$.

Superhelicity is responsible for the well-known dissipative mechanism of helicity generation, first demonstrated in Ref. 24. For large Reynolds numbers this effect can be the main mechanism for the emergence of helicity. Nonzero superhelicity is a common property of systems with helicity fluctuations. The helicity balance equation contains terms that are related to other similar topological characteristics. We stress that we are talking not only about the mean turbulent superhelicity (or some other similar quadratic combination), but also about superhelicity of the mean flow. In other words, the presence of hidden topological properties of large- and small-scale helical motions under the action of an external magnetic field is the principal source leading to the generation of mean hydrodynamic helicity. An external magnetic field serves as the trigger for helicity generation.

In our view, helicity generation is based on the following. We assume that in the absence of an external magnetic field a balance of left- and right-handed regular (random) motions exists. In an external magnetic field, random Alfvén modes propagate, their spatial spectrum reflecting the spatial structure of the unperturbed regular and random motions. However, the frequency of the Alfvén waves depends on their scale, i.e., these modes now enter with different ''weight,'' which ultimately breaks the symmetry between the left- and right-handed helical motion. This effect is similar to the mechanism of dissipative helicity generation[24], where the difference in dissipation at different scales likewise leads to breaking of the initial reflection symmetry.

As we have already emphasized, the helicity growth rate depends on the correlation characteristics of the main state, which in turn decrease with increasing magnetic field as $\propto(1+aB^{0^2}+bB^{0^4})^{-1}$. Accordingly, helicity generation goes as $\propto(B^{0^2}/(1+aB^{0^2}+bB^{0^4})$ and clearly has critical values of the magnetic field with maximum rate of growth. The magnitude of the critical magnetic field depends on the principal average characteristics of the turbulent flow. Such behavior has been observed in the experiments performed at the Center for MHD Studies (Beersheba, Israel): As the external magnetic field increases, the stably observed $-7/3$ spectrum is replaced by a two-dimensional -4 spectrum.

We thank A. Eidelman (Beersheba, Israel) for providing the laboratory data and for fruitful discussions. This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 96-02-19560).

*)E-mail: ochkheti@mx.iki.rssi.ru

---

[1] H. K. Moffat and A. Tsinober, Annu. Rev. Fluid Mech. **24**, 281 (1992).
[2] A. Brissaud *et al.*, Phys. Fluids **16**, 1363 (1973).
[3] S. S. Moiseev and O. G. Chkhetiani, Zh. Éksp. Teor. Fiz. **110**, 357 (1996) [JETP **83**, 192 (1996)].
[4] I. Platnieks and S. F. Seluto, in *Liquid Metal Magnetohydrodynamics*, J. Lielpeters and R. Moreay (Eds.), Kluwer, Dordrecht (1989).
[5] C. Henoch, M. Hoffert, H. Branover, and S. Sukoriansky, Prog. Astronaut. Aeronaut. **149**, 190 (1993).
[6] H. Branover, A. Eidelman, M. Nagorny, and M. Kireev, Prog. Astronaut. Aeronaut. **162**, 64 (1994).
[7] M. Mory and E. J. Hopfinger, Phys. Fluids **29**, 2140 (1986).
[8] E. C. Itsweire and K. N. Hellavei, in *Proc. Second Symp. on Turbulence and Diffusion*, Colorado (1985), p. 172.
[9] T. Wei and W. W. Willmarth, J. Fluid Mech. **204**, 57 (1989).
[10] S. J. Caughey and S. G. Palmer, Q. J. R. Meteorol. Soc. **805**, 811 (1979).
[11] G. I. Boer and T. G. Sheperd, J. Atmos. Sci. **40**, 164 (1983).
[12] K. S. Gage and G. D. Nastrom, J. Atmos. Sci. **43**, 729 (1986).
[13] I. N. Klepikov, I. V. Pokrovskaya, and E. A. Sharkov, Issled. Zemli iz Kosmosa, No. 3, 13 (1995).
[14] D. K. Lilly, J. Atmos. Sci. **43**, 126 (1986).
[15] R. Hide, Geophys. Astrophys. Fluid Dyn. **48**, 69 (1989).
[16] M. V. Kurganskiĭ, Fiz. Atm. Okeana **29**, 464 (1993).
[17] A. Bershadskii, E. Kit, and A. Tsinober, Proc. R. Soc. London, Ser. A **441**, 147 (1993).
[18] R. Moreau and A. Alemany, in *MHD Flows and Turbulence, Proc. of the Beer-Sheeva Int. Seminar* (1975), p. 51.
[19] S. I. Vaĭnshteĭn, Zh. Éksp. Teor. Fiz. **61**, 612 (1971) [Sov. Phys. JETP **34**, 327 (1972)].
[20] E. Golbraikh, O. G. Chkhetiani, and S. S. Moiseev, Zh. Éksp. Teor. Fiz. **114**, 171 (1998) [JETP **87**, 95 (1998)].
[21] F. Krause and K.-H. Rädler, *Magnetic Hydrodynamics of Mean Fields and Dynamo Theory* [Russian translation], Mir, Moscow (1984).
[22] L. L. Vaĭnshteĭn and S. I. Vaĭnshteĭn, Geomagn. Aeron. **13**, 149 (1973).
[23] W. H. Matthaeus, Phys. Rev. A **24**, 2135 (1981).
[24] J. C. Andre and M. Lesieur, J. Fluid Mech. **81**, 207 (1977).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

The temperature jump and slow evaporation in molecular gases

A. V. Latyshev and A. A. Yushkanov

Moscow Pedagogical University, 107005 Moscow, Russia

(Submitted 7 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 956–971 (September 1998)

We set up model transport equations that describe the behavior of molecular (diatomic and polyatomic) gases with a molecule collision rate proportional to the molecular velocity. In deriving these equations we allow for the internal (rotational) degrees of freedom, while the vibrational degrees of freedom are assumed ''frozen.'' We also set up an exact equation for the problem of the temperature jump with allowance for slow evaporation from the liquid surface into the saturated vapor atmosphere. Finally, we derive explicit formulas for calculating the coefficients of the temperature jump and gas-density jump above a flat surface and do the necessary numerical calculations. © *1998 American Institute of Physics.*
[S1063-7761(98)01409-7]

1. INTRODUCTION. DERIVATION OF THE BASIC EQUATIONS AND STATEMENT OF THE PROBLEM

The Smoluchowski problem (the problem of the temperature jump) has been attracting much attention for a long time (the history of this problem can be found in Refs. 1 and 2). For a simple gas this problem has been solved by analytic methods involving the use of model equations[3] and by approximate and numerical methods for the Boltzmann transport equation.[4–7] In addition to the Smoluchowski problem, the behavior of a gas when there is slow evaporation from a surface is also of interest.[8–11]

The work we have just mentioned deals with a monatomic gas. At the same time, it would be interesting to study the behavior of a molecular gas in such processes near a surface. As is known, transport processes in a molecular gas are much more complicated than in a simple gas.[12] This places more emphasis on the role of model collision integrals in describing transport processes, since the elastic and inelastic cross sections that enter into the Boltzmann collision integral have yet to be well enough studied to be used for a detailed quantitative description of the processes.

The structure of model collision integrals can in principle depend on the nature of the phenomena that play the most important role in the given problem. In what follows we consider a temperature range in which the vibrational degrees of freedom are effectively ''frozen,'' while the rotational degrees of freedom can be described by a classical approach. The temperature range in which these conditions are met extends from tens to thousands of kelvins.[13]

There are many approaches to setting up model transport equations for molecular gases (see, e.g., Refs. 14–17). Some of these incorporate the discrete structure of the levels of the internal energy of molecules. Allowance for such a structure is important at low temperatures, while at high temperatures the discrete level structure does not manifest itself and is superfluous, not to mention the fact that allowing for this structure is inconvenient for practical applications. In this connection we will use an approach, proposed in Ref. 17, that is based on the possibility of using the semiclassical approximation to describe the rotational degrees of freedom at high temperatures (a similar but somewhat different approach is developed in Refs. 14 and 15).

In Ref. 17 we discussed a model of a molecular diatomic gas with a constant rate of molecular collisions (this rate is assumed independent of the velocity of the molecules). Another interesting model is that with a constant mean free path of the molecules, which is close to the model of a gas of hard spheres.[2,11] The fact that the mean free path is constant means that the collision rate is proportional to the velocity of the molecules.

In the present paper we use a model Boltzmann transport equation of the Bhatnagar–Gross–Krook (BGK) type to describe molecular gases (diatomic and polyatomic) in which the molecular collision rate is proportional to the molecular velocity. On the basis of this model we set up analytic solutions for the Smoluchowski and slow evaporation problems.

The collision integral for a simple gas with a collision rate proportional to the speed of molecules has the form[2,11]

$$J[f]=\frac{w}{\lambda_0}(f_{\mathrm{eq}}-f). \tag{1}$$

Here λ_0 is the characteristic mean free path of the molecules, $w=|\mathbf{v}-\mathbf{u}_0(\mathbf{r})|$ is the speed of a gas molecule in a reference frame in which the gas at the given point $\mathbf{r}$ is at rest, i.e., has the mass-weighted mean velocity, which is zero; $\mathbf{v}$ is the molecular velocity of the gas in the laboratory reference frame, and $\mathbf{u}_0(\mathbf{r})$ is the mass-weighted mean velocity of the gas at point $\mathbf{r}$ in the laboratory reference frame. The function f_{eq} can be written

$$f_{\mathrm{eq}}=n_{\mathrm{eq}}\left(\frac{m}{2\pi kT_{\mathrm{eq}}}\right)^{3/2}\exp\left[-\frac{m}{2kT_{\mathrm{eq}}}(\mathbf{v}-\mathbf{u}_{\mathrm{eq}})^2\right],$$

where m is the mass of a molecule, and k is Boltzmann's constant.

The quantities n_{eq}, T_{eq}, and $\mathbf{u}_{\mathrm{eq}}$ can be found from the condition that the collision integral (1) conserve the number of molecules, the momentum, and the energy. These requirements can be written in the form of an equation:

$$\int wM_1 f_{\text{eq}}\, d^3v = \int wM_1 f\, d^3v, \tag{2}$$

where $M_1 = 1$, $m\mathbf{v}$, or $mv^2/2$.

The collision integral for a diatomic gas can also be written in a form similar to (1). Here f_{eq} is given by

$$f^* = n^* \left(\frac{m}{2\pi kT^*} \right)^{3/2} \frac{J}{2kT^*} \times \exp\left[-\frac{J\omega^2}{2kT^*} - \frac{m}{2kT^*}(\mathbf{v}-\mathbf{u}^*)^2 \right],$$

where J is the total moment of inertia of a molecule, and ω is the frequency of the molecule's rotation.

The quantities n^*, T^*, and $\mathbf{u}^*$ can also be found from conservation laws that are generalizations of (2):

$$\int wM_2 f^* \omega\, d^3v\, d\omega = \int wM_2 \omega\, d^3v\, d\omega, \tag{3}$$

where $M_2 = 1$, $m\mathbf{v}$, or $mv^2/2 + J\omega^2/2$.

For molecular gases whose molecules contain more than two atoms (below we call such gases polyatomic), the function f_{eq} can be written

$$f^* = n^* \left(\frac{m}{2\pi kT^*} \right)^{3/2} \frac{(J_1 J_2 J_3)^{1/2}}{(2\pi kT^*)^{3/2}} \times \exp\left[-\frac{m}{2kT^*}(\mathbf{v}-\mathbf{u}^*)^2 - \frac{J_1\omega_1^2 + J_2\omega_2^2 + J_3\omega_3^2}{2kT^*} \right],$$

where J_1, J_2, and J_3 are the principal moments of inertia of a molecule, and $(\omega_1, \omega_2, \omega_3)$ is the vector of the angular velocity of rotation of a molecule.[13,18]

The quantities n^*, T^*, and $\mathbf{u}^*$ can be found from conservation laws, which is this case can be written

$$\int wM_3 f^*\, d^3v\, d^3\omega = \int wM_3 f\, d^3v\, d^3\omega, \tag{4}$$

where $M_3 = 1$, $m\mathbf{v}$, or $mv^2/2 + (J_1\omega_1^2 + J_2\omega_2^2 + J_3\omega_3^2)/2$.

The concentration n, velocity $\mathbf{u}$, and temperature T of the gas are given by the relationships[1,13]

$$n = \int f\omega\, d^3v\, d\omega, \quad \mathbf{u} = \frac{1}{n} \int f\mathbf{v}\, d^3v\, d\omega,$$

$$T = \frac{2}{5n} \int f\left[\frac{1}{2} m(\mathbf{v}-\mathbf{u})^2 + \frac{1}{2} J\omega^2 \right] \omega\, d^3v\, d\omega$$

for a diatomic gas, and

$$n = \int f\, d^3v\, d^3\omega, \quad \mathbf{u} = \frac{1}{n} \int f\mathbf{v}\, d^3v\, d^3\omega,$$

$$T = \frac{1}{3n} \int f\left[\frac{1}{2} m(\mathbf{v}-\mathbf{u})^2 + \frac{1}{2}(J_1\omega_1^2 + J_2\omega_2^2 + J_3\omega_3^2) \right] d^3v\, d^3\omega$$

for a polyatomic gas.

Now let us suppose that in the region under consideration the gas temperature T changes slowly, i.e., the relative variation of the temperature over the mean free path l of molecules of the gas is small:

$$l\,|\nabla \ln T| \ll 1. \tag{5}$$

Below we study the interaction of the gas and the surface of the condensed phase. We assume that in the reference frame in which the condensed phase is at rest, the speed of the gas u is much less than the speed of sound, i.e.,

$$\sqrt{\frac{m}{2kT}}\, u \ll 1. \tag{6}$$

It can be shown[1,2] that the conditions (5) and (6) yield

$$\sqrt{\frac{m}{2kT}}\, u^* \ll 1. \tag{7}$$

The condition (5) makes it possible to isolate a region of the volume occupied by the gas whose size is much larger than the mean free path and in which the relative temperature variations are small, i.e.,

$$\frac{|T - T_0|}{T_0} \ll 1. \tag{8}$$

Here T_0 is the temperature of the gas at a certain point of that region.

If conditions (7) and (8) are met, the transport equation can be linearized. Here the absolute value of a molecule's velocity, w, on the right-hand side of Eq. (1) can be replaced in the linear approximation by v. The distribution function f can be written $f = f_0(1+\varphi)$, where

$$f_0 = n_0 \left(\frac{m}{2\pi kT_0} \right)^{3/2} \frac{J}{2kT_0} \exp\left[-\frac{mv^2}{2kT_0} - \frac{J\omega^2}{2kT_0} \right]$$

for a diatomic gas, and

$$f_0 = n_0 \left(\frac{m}{2\pi kT_0} \right)^{3/2} \frac{(J_1 J_2 J_3)^{1/2}}{(2\pi kT_0)^{3/2}} \times \exp\left[-\frac{mv^2}{2kT_0} - \frac{J_1\omega_1^2 + J_2\omega_2^2 + J_3\omega_3^2}{2kT_0} \right]$$

for a polyatomic gas.

If the conditions (5), (7), and (8) are met, the function φ is small, i.e., $|\varphi| \ll 1$. Note that according to the definition of T_0, there is a certain ambiguity in selecting this temperature. In point of fact, instead of T_0 and n_0 we can take other parameters T_0' and n_0' such that $|T_0 - T_0'| \ll \max(T_0, T_0')$ and $|n_0 - n_0'| \ll (n_0, n_0')$. Here the entire linearization procedure remains valid. Only φ changes, but the linearization condition ($|\varphi| \ll 1$) remains valid.

The steady-state linearized Boltzmann transport equation with a model collision integral in the BGK form for a molecular gas can be written

$$(\mathbf{v}\cdot\nabla\varphi)+\frac{v\varphi}{\lambda_0}=\frac{v}{\lambda_0}\left[\frac{\delta n}{n_0}+\frac{\delta T}{T_0}\left(\frac{mv^2}{2kT_0}+\frac{[J\omega^2]}{2kT_0}-l\right)\right.$$
$$\left.+\frac{m}{kT_0}\mathbf{u}^*\cdot\mathbf{v}\right]. \qquad (9)$$

Here $\delta n=n^*-n_0$, $\delta T=T^*-T_0$, and $\varphi=\varphi(\mathbf{r},\mathbf{v},\omega)$, with $l=3$ and $[J\omega^2]=J\omega^2$ for a diatomic gas, and $l=7/2$ and $[J\omega^2]=J_1\omega_1^2+J_2\omega_2^2+J_3\omega_3^2$ for a polyatomic gas.

In the Smoluchowski and slow evaporation problems there is a flat boundary (interface) separating the gas from the condensed phase. At the surface there may be evaporation and condensation. Let us examine a region occupied by the gas that is adjacent to the surface. The size L of this region must be much larger than the mean free path of the gas molecules, but at the same it must small enough for condition (8) to be met. The condition (5) guarantees that such a region exists. We assume that there is a heat flux normal to the surface. Then far from the surface (outside the Knudsen layer whose thickness is of order of the mean free path) in this region there exists a linear temperature gradient perpendicular to the surface.[1,19] We introduce a Cartesian coordinate system with its origin at the surface and with its x axis perpendicular to the surface, so that the region (half-space) filled by the gas corresponds to the positive part of the x axis. Then $T=T_e+Ax$, $l\ll x\ll L$, where $A=(dT/dx)_{\mathrm{as}}$ is the asymptotic value of the temperature gradient. We denote the surface temperature by T_s and the concentration of the saturated vapor at the surface temperature by n_s. Then T_e-T_s and n_e-n_s are called the jumps in temperature and concentration, where in the second case it is assumed that evaporation or condensation takes place at a vanishing rate.

In the linear approximation these two jumps are proportional to the temperature gradient. The temperature jump is equal to the difference between the gas temperature linearly extrapolated to the surface, and the surface temperature itself. In other words, the temperature jump is the difference between the ‘‘hydrodynamic’’ (without allowance for the Knudsen layer) gas temperature and the surface temperature. Our goal is to calculate the relative values of the temperature and concentration jumps, $\varepsilon_T=T_e/T_s-1$ and $\varepsilon_n=n_e/n_s-1$, which in the linear approximation are proportional to the relative temperature gradient, i.e., $\varepsilon_T=c_tK$ and $\varepsilon_n=c_nK$, with $K=A/T_s$.

Far from the surface the gas can move toward or away from the surface, which corresponds to condensation or evaporation. We denote the evaporation or condensation rate by U (note that it is perpendicular to the surface). In this case, in the linear approximation the relative jumps in the temperature and concentration of the gas are proportional to U: $\varepsilon_T=s_t(2U)$ and $\varepsilon_n=s_n(2U)$.

The quantities c_t, c_n, s_t, and s_n must be found by solving the transport equation.

The statement of the problem implies that the distribution function depends on only one spatial variable, x. It is convenient to introduce the dimensionless variables

$$x^*=\frac{x}{\lambda_0},\quad \xi=\sqrt{\frac{m}{2kT_0}}v,\quad \omega^*=\sqrt{\frac{[J\omega^2]}{2kT_0}}\omega,$$

$$\mu=\frac{v_x}{v},\quad U^*=\sqrt{\frac{m}{2kT_0}}U.$$

Below we write the variables x^* and ω^* and the quantity U^* without the asterisk. Then, if we combine (3), (4) and (9), we obtain

$$\mu\frac{\partial\varphi}{\partial x}+\varphi(x,\mu,\xi,\omega)=\int_{-1}^{1}\int_0^\infty\int_0^\infty\exp(-\xi'^2-\omega'^2)k$$
$$\times(\mu,\xi,\omega;\mu',\xi',\omega')\varphi(x,\mu',\xi',\omega')$$
$$\times d\mu'\,d\xi'\,d\omega', \qquad (10)$$

where

$$k(\mu,\xi;\mu',\xi')=\frac{2r}{\pi^{(r-1)/2}}\left[1+\frac{3}{2}\mu\xi\mu'\xi'+\frac{r}{4r-1}\right.$$
$$\times\left(\xi^2+\omega^2-\frac{4r-1}{r}\right)$$
$$\left.\times\left(\xi'^2+\mu'^2-\frac{4r-1}{r}\right)\right],$$

with $r=1$ for a diatomic gas $r=2$ for a polyatomic gas, and ω is the absolute value of the angular velocity of rotation of a molecule in both cases.

Far from the wall the function φ has the form

$$\varphi_{\mathrm{as}}(x,\mu,\xi,\omega)=\varepsilon_n+\varepsilon_T\left(\xi^2+\omega^2-\frac{7r-2}{2r}\right)$$
$$+\left(2U-\frac{2K}{3\sqrt{\pi}}\right)\mu\xi+K(x-\mu)$$
$$\times\left(\xi^2+\omega^2-\frac{9r-2}{2r}\right).$$

Generally, the boundary condition problem for a molecular gas at the surface is extremely complicated.[20] Below we limit ourselves to the case in which the accommodation of the molecules by the surface is total.[1,2] Then the boundary condition at the surface takes the simple form

$$\varphi(0,\mu,\xi,\omega)=0,\quad 0<\mu<1, \qquad (11)$$

while far from the wall $(x\to+\infty)$ we have

$$\varphi(x,\mu,\xi,\omega)=\varphi_{\mathrm{as}}(x,\mu,\xi,\omega)+o(1),\quad -1<\mu<0. \qquad (12)$$

2. DIATOMIC GAS

We examine the boundary value problem (10)–(12) for a diatomic gas, i.e., we put $r=1$ in these equations. We expand the function φ in three orthogonal directions:

$$\varphi=\varphi_{\mathrm{as}}(x,\mu,\xi,\omega)+h_1(x,\mu)+\xi h_2(x,\mu)$$
$$+(\xi^2+\omega^2-3)h_3(x,\mu). \qquad (13)$$

Orthogonality is understood in the sense that the scalar product

$$(f,g)=2\int_{-1}^{1}\int_{0}^{\infty}\int_{0}^{\infty}\exp(-\xi^2-\omega^2)\xi^3\omega f(\mu,\xi,\omega) \times g(\mu,\xi,\omega)\,d\mu\,d\xi\,d\omega \tag{14}$$

is zero. Plugging (13) into (10), we obtain a set of transport equations:

$$\mu\frac{\partial h_1}{\partial x}+h_1=(1,h_1)+4\alpha(1,h_2), \tag{15a}$$

$$\mu\frac{\partial h_2}{\partial x}+h_2=3\mu[2\alpha(\mu',h_1)+(\mu',h_2)+\alpha(\mu',h_3)], \tag{15b}$$

$$\mu\frac{\partial h_3}{\partial x}+h_3=\frac{2}{3}\alpha(1,h_2)+(1,h_3). \tag{15c}$$

The scalar product (14) becomes simpler if the inner integrals can be calculated, and in relation to the set of equations (15) it means the following:

$$(\mu^k,h_l)=\frac{1}{2}\int_{-1}^{1}\mu'^k h_l(x,\mu')\,d\mu',\quad k=0,1,\quad l=1,2,3.$$

Now we use the conservation laws (3). Conservation of the number of particles and energy makes it possible to simplify the set of equations (15). These laws yield $(\mu,\varphi)=0$ and $(\mu(\xi^2+\omega^2),\varphi)=0$. Substituting (13) into (15), we get

$$(\mu',h_1)=-4\alpha(\mu',h_2),\quad (\mu',h_3)=-\frac{2}{3}\alpha(\mu',h_2).$$

Then Eq. (15b) simplifies:

$$\mu\frac{\partial h_2}{\partial x}+h_2=3c\mu(\mu',h_2), \tag{16}$$

where $c=1-39\pi/128$.

Equations (15a) and (15c) imply that instead of the function h_3 it is more convenient to study the difference $\tilde{h}_3=h_1-6h_3$, for which we have

$$\mu\frac{\partial \tilde{h}_3}{\partial x}+\tilde{h}_3=(1,\tilde{h}_3) \tag{17}$$

with the boundary conditions

$$\tilde{h}_3(0,\mu)=0,\quad 0<\mu<1,$$

and, as $x\to\infty$,

$$\tilde{h}_3(x,\mu)=\varepsilon_n-\frac{11}{2}\varepsilon_T-\frac{13}{2}K(x-\mu),\quad -1<\mu<0.$$

We write the set of equations (15a), (16), and (17) in vector form:

$$\mu\frac{\partial h}{\partial x}+h(x,\mu)=\frac{1}{2}\int_{-1}^{1}K(\mu,\mu')\,h(x,\mu')\,d\mu'. \tag{18}$$

Here h is a vector with elements (components) h_1, h_2, and $\tilde{h}_3$, while $K(\mu,\mu')=K_0+3c\mu\mu' K_1$ is the kernel of Eq. (18), with

$$K_0=\begin{bmatrix}1 & 4\alpha & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},\quad K_1=\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

The boundary conditions (11) and (12) for Eq. (18) can be written

$$h(0,\mu)=-h_{\mathrm{as}}(0,\mu),\quad 0<\mu<1, \tag{19}$$

$$h(\infty,\mu)=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix},\quad -1<\mu<0, \tag{20}$$

where

$$h_{\mathrm{as}}(x,\mu)=\begin{bmatrix}\varepsilon_n+\varepsilon_T/2-K(x-\mu)/2\\ (2U-2K/3\sqrt{\pi})\mu\\ \varepsilon_n-11\varepsilon_T/2-13K(x-\mu)/2\end{bmatrix}.$$

We now turn to the solution of the boundary value problem (18)–(20) Separation of variables in Eq. (18) immediately yields the particular solutions

$$h_\eta(x,\mu)=\exp\left(-\frac{x}{\eta}\right)\Phi(\eta,\mu),$$

with $\Phi(\eta,\mu)$ the eigenvector function of the characteristic equation

$$(\eta-\mu)\Phi(\eta,\mu)=\tfrac{1}{2}\eta[K_0n_0(\eta)+3c\mu K_1n_1(\eta)], \tag{21}$$

where

$$n_k(\eta)=\int_{-1}^{1}\mu^k\Phi(\eta,\mu)\,d\mu,\quad k=0,1. \tag{22}$$

We seek a solution of Eqs. (21) and (22) for $\eta\in(-1,1)$ in the space of the generalized functions $\Phi(\eta,\mu)=F(\eta,\mu)n_0(\eta)$ (see Ref. 21), where

$$F(\eta,\mu)=\frac{1}{2}\eta K(\mu\eta)P\frac{1}{\eta-\mu}+\Lambda(\eta)\delta(\eta-\mu). \tag{23}$$

Here Px^{-1} denotes the principal value of the integral of x^{-1}, $\delta(x)$ is the Dirac delta function, and $\Lambda(z)$ is the dispersion matrix function

$$\Lambda(z)=E+\frac{1}{2}z\int_{-1}^{1}K(\mu z)\frac{d\mu}{\mu-z},$$

where E is the identity matrix, or $\Lambda(z)=\lambda_c(z)K(z^2)+K_2$, with

$$K_2=\begin{bmatrix}0 & -4\alpha & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix},$$

$$\lambda_c(z)=1+zt(z),\quad t(z)=\frac{1}{2}\int_{-1}^{1}\frac{d\tau}{\tau-z};$$

$\lambda_c(z)$ is Case's dispersion function (see Ref. 22).

According to Ref. 22, the dispersion function $\lambda(z)$ is the determinant of the dispersion matrix function:

$$\lambda(z)\equiv\det\Lambda(z)=\lambda_c^2(z)\,\omega(z), \tag{24}$$

where $\omega(z)=1+3cz^2\lambda_c(z)$.

Using the argument principle from the theory of analytic functions,[23] we see that $\omega(z)$ has two real zeros $\pm\eta_0$, where according to Ref. 2, pp. 358–359, $\eta_0=1+\varepsilon$, with $\varepsilon\approx 10^{-8}$. Expanding $\lambda(z)$ in a series in the neighborhood of the point at infinity $z=\infty$, we see that this point is a zero of order 4. Hence the discrete spectrum of the characteristic equation, being the set of zeros of the dispersion function, consists of a point of order 4 and the two points $\pm\eta_0$. There are six discrete solutions of Eq. (18) corresponding to these zeros:

$$h^{(1)}(x,\mu)=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad h^{(2)}(x,\mu)=\begin{bmatrix}0\\0\\1\end{bmatrix},$$

$$h^{(i)}(x,\mu)=(x-\mu)h^{(i-2)}(x,\mu),\quad i=3,4,$$

$$h_{\pm\eta_0}(x,\mu)=\frac{1}{2}\,\frac{\pm\eta_0K(\pm\mu\eta_0)}{\pm\eta_0-\mu}\exp\left(-\frac{x}{\pm\eta_0}\right)n(\pm\eta_0).$$

Plugging $h_{\eta_0}(x,\mu)$ into Eq. (18), we find that the vector $n(\eta_0)$ satisfies the homogeneous equation

$$\Lambda(\eta_0)n(\eta_0)=\begin{bmatrix}0\\0\\0\end{bmatrix}, \tag{25}$$

with $\Lambda(\eta_0)=0$. The vector equation (25) is equivalent to three scalar equation, from which we find that

$$n(\eta_0)=\begin{bmatrix}4\alpha\eta_0t(\eta_0)\\-\lambda_c(\eta_0)\\0\end{bmatrix}.$$

Let us show that the solution of the boundary value problem (18)–(20) can be expressed as an expansion in the eigenvectors (solutions),

$$h(x,\mu)=A_0h_{\eta_0}(x,\mu)+\int_0^1\exp\left(-\frac{x}{\eta}\right)F(\eta,\mu)\,A(\eta)\,d\eta, \tag{26}$$

where A_0 is an unknown constant (called the discrete spectrum coefficient), and $A(\eta)$ is an unknown vector function (called the continuous spectrum coefficient). The expansion (26) satisfies the condition (20) automatically. Using condition (19), we obtain a singular integral equation with a Cauchy kernel:

$$\frac{1}{2}\int_0^1\eta K(\mu\eta)A(\eta)\frac{d\eta}{\eta-\mu}+\Lambda(\mu)A(\mu)+A_0h_{\eta_0}(0,\mu)+h_{\text{as}}(0,\mu)=\begin{bmatrix}0\\0\\0\end{bmatrix},\quad 0<\mu<1. \tag{27}$$

We introduce the auxiliary vector function

$$N(z)=\frac{1}{2}\int_0^1\eta\,K(z\eta)\,A(\eta)\frac{d\eta}{\eta-z}, \tag{28}$$

whose boundary values above and below the cut from (0 to 1) are related by the Sochozki–Plemelj formulas:

$$N^+(\mu)-N^-(\mu)=\pi i\mu\,K(\mu^2)\,A(\mu),$$

$$N^+(\mu)+N^-(\mu)=\int_0^1\eta\,K(\mu\eta)\,A(\eta)\frac{d\eta}{\eta-\mu},\quad 0<\mu<1.$$

We can also write similar formulas for the dispersion matrix function:

$$\Lambda^+(\mu)-\Lambda^-(\mu)=\pi i\mu\,K(\mu^2),$$

$$\Lambda^+(\mu)+\Lambda^-(\mu)=2\Lambda(\mu).$$

Using these formulas, we reduce Eq. (27) to the vector Riemann–Hilbert boundary value problem

$$L^+(\mu)[N^+(\mu)+A_0h_{\eta_0}(0,\mu)+h_{\text{as}}(0,\mu)]=L^-(\mu)[N^-(\mu)+A_0h_{\eta_0}(0,\mu)+h_{\text{as}}(0,\mu)],\quad 0<\mu<1, \tag{29}$$

with the matrix coefficient

$$G(\mu)=[L^+(\mu)]^{-1}L^-(\mu)=K(\mu^2)(\Lambda^+(\mu))^{-1}\Lambda^-(\mu)K^{-1}(\mu^2),$$

where $L(z)=K(z^2)\,\Lambda(z)\,K^{-1}(z^2)$. We now have the problem of factoring the coefficients $G(\mu)$:

$$G(\mu)=X^+(\mu)[X^-(\mu)]^{-1},\quad 0<\mu<1, \tag{30}$$

where $X(z)$ is an unknown matrix function analytic in the complex z plane with a cut along the segment $[0,1]$. Equation (30) is the homogeneous problem corresponding to the inhomogeneous problem (29). The method of Eq. (30) was developed by the present authors in Ref. 17; here we simply write the solution:

$$X(z)=\begin{bmatrix}U(z)&4\alpha(U(z)-V(z))&0\\0&V(z)&0\\0&0&U(z)\end{bmatrix}.$$

In this matrix,

$$U(z)=z\exp(-u(z)),$$

$$u(z)=\frac{1}{\pi}\int_0^1[\theta(\tau)-\pi]\frac{d\tau}{\tau-z},\quad \theta(\tau)=\arg\lambda_c^+(\tau),$$

and

$$V(z)=z\exp(-v(z)),$$

$$v(z)=\frac{1}{\pi}\int_0^1[\varepsilon(\tau)-\pi]\frac{d\tau}{\tau-z},\quad \varepsilon(\tau)=\arg\omega^+(\tau),$$

where $\theta(\tau)$ and $\varepsilon(\tau)$ stand for the continuous branches of the arguments of the functions $\lambda_c^+(\tau)$ and $\omega^+(\tau)$, specified by the conditions $\theta(0)=0$ and $\varepsilon(0)=0$. We establish the asymptotic behavior of the functions $U(z)$ and $V(z)$ in the neighborhood of the point at infinity:

$$U(z)=z-U_1+o(1),\quad |z|\to\infty;$$

$$V(z)=z-V_1+o(1),\quad |z|\to\infty.$$

Here

$$U_1=-\frac{1}{\pi}\int_0^1[\theta(\tau)-\pi]\,d\tau,\quad V_1=-\frac{1}{\pi}\int_0^1[\varepsilon(\tau)-\pi]\,d\tau$$

are the Laurent coefficients of z^{-1} in the expansions of $U(z)$ and $V(z)$ in the neighborhood of the point at infinity. Integration by parts transforms the formula for U_1 into the well-known formula (4.21) of Ref. 2, p. 334:

$$U_1=\frac{1}{2}\int_0^1\frac{d\tau}{(1-\tau^2)[\lambda_c^2(\tau)+(\pi\tau/2)^2]}.$$

Hence the asymptotic behavior of the matrix $X(z)$ in the neighborhood of the point at infinity is given by

$$X(z)=zE-\begin{bmatrix}U_1 & 4\alpha(U_1-V_1) & 0\\ 0 & V_1 & 0\\ 0 & 0 & U_1\end{bmatrix}+o(1),\quad |z|\to\infty.$$

We now return to the solution of the inhomogeneous problem (29). Using (30), we transform (29) into the problem of determining the analytic vector function $N(z)$ from a zero jump:

$$\begin{aligned}&[X^+(\mu)]^{-1}[N^+(\mu)+A_0h_{\eta_0}(0,\mu)+h_{\mathrm{as}}(0,\mu)]\\&\quad=[X^-(\mu)]^{-1}[N^-(\mu)+A_0h_{\eta_0}(0,\mu)+h_{\mathrm{as}}(0,\mu)],\end{aligned}\tag{31}$$

where $0<\mu<1$. Allowing for the behavior of the vectors and matrices in the problem (31) and their asymptotic behavior at the point at infinity, we find the general solution of this problem:

$$N(z)=-h_{\mathrm{as}}(0,z)-A_0h_{\eta_0}(0,z)+X(z)\left[\frac{B}{z-\eta_0}+C\right],\tag{32}$$

where B and C are arbitrary vectors with elements b_i and c_i $(i=1,2,3)$.

The solution (32) has simple poles at the points η_0 and $z=\infty$. On the other hand, the vector function $N(z)$ introduced earlier in (28) is analytic at the point η_0, and at the point $z=\infty$ its first (upper) and third (lower) elements vanish and its second element has a finite limit. To remove these singularities we pass to the solvability conditions and employ the fact that the solution (32) has constant parameters. To this end we write the solution (32) in the explicit form

$$\begin{aligned}\begin{bmatrix}N_1(z)\\ N_2(z)\\ N_3(z)\end{bmatrix}&=-\begin{bmatrix}\varepsilon_n+\varepsilon_T/2+Kz/2\\ (2U-2K/3\sqrt{\pi})z\\ \varepsilon_n-11\varepsilon_T/2+13Kz/2\end{bmatrix}\\&\quad-\frac{1}{2}\,\frac{A_0}{\eta_0-z}\begin{bmatrix}-4\alpha\eta_0\\ z\\ 0\end{bmatrix}\\&\quad+\begin{bmatrix}U(z) & 4\alpha(U(z)-V(z)) & 0\\ 0 & V(z) & 0\\ 0 & 0 & U(z)\end{bmatrix}\\&\quad\times\left\{\frac{1}{z-\eta_0}\begin{bmatrix}b_1\\ b_2\\ b_3\end{bmatrix}+\begin{bmatrix}c_1\\ c_2\\ c_3\end{bmatrix}\right\}.\end{aligned}\tag{33}$$

The expressions (32) and (33) suggest that the pole at point η_0 can be removed by a single vector condition

$$\frac{1}{2}A_0\,\eta_0K(\eta_0^2)\,n(\eta_0)+X(\eta_0)B=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix},$$

which is equivalent to three scalar conditions:

$$U(\eta_0)b_1+4\alpha(U(\eta_0)-V(\eta_0))b_3=2\alpha\eta_0A_0,$$

$$V(\eta_0)b_2=-\frac{1}{2}A_0\eta_0,\quad b_3=0.$$

We use these equations to set up the vector B:

$$B=\frac{1}{2}\,\frac{\eta_0A_0}{V(\eta_0)}\begin{bmatrix}4\alpha\\ -1\\ 0\end{bmatrix}.$$

The pole at $z=\infty$ can be removed by the conditions

$$c_1=\frac{1}{2}K,\quad c_2=2U-\frac{2K}{3\sqrt{\pi}},\quad c_3=\frac{15}{2}K.$$

Next we equate the constant terms in the expansions of the general solution and the auxiliary function $N(z)$ introduced by (28). This yields a set of three equations:

$$\varepsilon_n+\frac{1}{2}\varepsilon_T=b_1-c_1U_1-4\alpha(U_1-V_1)c_2,\tag{34}$$

$$\frac{3c}{2}\int_0^1\eta^2A_2(\eta)\,d\eta=-\frac{1}{2}A_0-b_2+c_2V_1,\tag{35}$$

$$\varepsilon_n-\frac{13}{2}\varepsilon_T=-c_3U_1.\tag{36}$$

To calculate the integral in (35), we take the function $N_2(z)$ from Eq. (28),

$$N_2(z)=\frac{3c}{2}z\int_0^1\eta^2A_2(\eta)\frac{d\eta}{\eta-z},$$

so that the general solution (33) yields

$$N_2(z)=-\left(2U-\frac{2K}{3\sqrt{\pi}}\right)z-\frac{1}{2}\,\frac{A_0z}{\eta_0-z}+V(z)\left(\frac{b_2}{z-\eta_0}+c_2\right).$$

From the Sochozki–Plemelj formula for $N_2(z)$ and the above equalities we find that

$$\frac{3c}{2}\int_0^1\eta^2A_2(\eta)\,d\eta=\frac{1}{2\pi i}\int_0^1[V^+(\eta)-V^-(\eta)]\times\left(\frac{b_2}{\eta-\eta_0}+c_2\right)\frac{d\eta}{\eta}.$$

To calculate this integral we use contour integration, with a complicated contour that encloses the cut $[0,1]$ and the point at infinity. Omitting the proof, we simply write

$$\begin{aligned}\frac{3c}{2}\int_0^1 \eta^2 A(\eta)\,d\eta &= \frac{b_2}{\eta_0}\frac{1}{2\pi i}\int_0^1 [V^+(\eta)-V^-(\eta)]\frac{d\eta}{\eta-\eta_0} \\ &\quad + \left(c_2-\frac{b_2}{\eta_0}\right)\frac{1}{2\pi i}\int_0^1 [V^+(\eta)-V^-(\eta)]\frac{d\eta}{\eta} \\ &= \frac{b_2}{\eta_0}(V(\eta_0)-\eta_0+V_1)+\left(c_2-\frac{b_2}{\eta_0}\right)(V(0)+V_1) \\ &= \frac{b_2}{\eta_0}(V(\eta_0)-\eta_0-V(0))+c\,(V(0)+V_1).\end{aligned}$$

Plugging this into (35) yields

$$\frac{b_2}{\eta_0}(V(\eta_0)-V(0))+c_2 V(0)=-\frac{1}{2}A_0,$$

from which we get

$$A_0=-2c_2V(\eta_0)=-2\left(2U-\frac{2K}{3\sqrt{\pi}}\right)V(\eta_0),$$

so that

$$b_1=-4\alpha\eta_0c_2=-4\alpha\eta_0\left(2U-\frac{2K}{3\sqrt{\pi}}\right).$$

Combining Eqs. (34) and (36), we find that

$$\varepsilon_T=KU_1-\frac{2\alpha}{3}\left(2U-\frac{2K}{3\sqrt{\pi}}\right)(U_1-V_1+\eta_0),$$

$$\varepsilon_n=\frac{11}{2}\varepsilon_T-\frac{13}{2}KU_1,$$

or, with the notation $\beta=-2\alpha(U_1-V_1+\eta_0)/3$, we obtain $\varepsilon_T=2U\beta+K(U_1-2\beta/3\sqrt{\pi})$ and $\varepsilon_n=2U(11\beta/2)+K(-U_1-11\beta/3\sqrt{\pi})$. Making use of the results of numerical calculations, ($U_1=0.71045$ and $V_1=0.98340$), we have the final formulas for the jumps in temperature and concentration:

$$\varepsilon_T=2U(-0.16330)+K(0.77187),$$

$$\varepsilon_n=2U(-0.89815)+K(-0.37263).$$

3. POLYATOMIC GAS

Now we examine the boundary value problem (10)–(12) for a polyatomic gas, i.e., we put $r=2$ in these equations. We obtain a boundary value problem that entails solving the transport equation

$$\begin{aligned}\mu\frac{\partial\varphi}{\partial x}+\varphi(x,\mu,\xi,\omega)=\frac{4}{\sqrt{\pi}}\int_{-1}^1\int_0^\infty\int_0^\infty &\exp(-\xi'^2-\omega'^2) \\ &\times k(\mu,\xi,\omega;\mu',\xi',\omega') \\ &\times\varphi(x,\mu',\xi',\omega') \\ &\times d\mu'\,d\xi'\,d\omega' \end{aligned}\tag{37}$$

with boundary conditions

$$\varphi(0,\mu,\xi,\omega)=0,\quad 0<\mu<1,\tag{38}$$

and, as $x\to\infty$,

$$\varphi(x,\mu,\xi,\omega)=\varphi_{\rm as}(x,\mu,\xi,\omega)+o(1),\quad -1<\mu<0.\tag{39}$$

Here

$$k(\mu,\xi,\omega;\mu',\xi',\omega')=1+\frac{3}{2}\mu\xi\mu'\xi'+\frac{2}{7}\left(\xi^2+\omega^2-\frac{7}{2}\right)\times\left(\xi'^2+\omega'^2-\frac{7}{2}\right)$$

is the kernel of Eq. (37), and

$$\varphi_{\rm as}(x,\mu,\xi,\omega)=\varepsilon_n+\varepsilon_T(\xi^2+\omega^2-3)+\left(2U-\frac{2K}{3\sqrt{\pi}}\right)\mu\xi+K(x-\mu)(\xi^2+\omega^2-4)$$

is the asymptotic part of the distribution function.

Next we expand the function φ in three orthogonal directions:

$$\varphi=h_1(x,\mu)+\xi h_2(x,\mu)+\left(\xi^2+\omega^2-\frac{7}{2}\right)h_3(x,\mu)+\varphi_{\rm as}(x,\mu,\xi,\omega).$$

Here orthogonality is understood in the sense that the scalar product

$$(f,g)=\frac{4}{\sqrt{\pi}}\int_{-1}^1\int_0^\infty\int_0^\infty\exp(-\xi^2-\omega^2)\,\xi^3\omega^2f(\mu,\xi,\omega)\times g(\mu,\xi,\omega)\,d\mu\,d\xi\,d\omega$$

is zero. Plugging the decomposition of φ into Eq. (37), we obtain the transport equations

$$\mu\frac{\partial h_1}{\partial x}+h_1=(1,h_1)+4\alpha(1,h_2),\tag{40}$$

$$\mu\frac{\partial h_2}{\partial x}+h_2=3\mu[2\alpha(\mu',h_1)+(\mu',h_2)+\alpha(\mu',h_3)],\tag{41}$$

$$\mu\frac{\partial h_3}{\partial x}+h_3=\frac{4}{7}\alpha(1,h_2)+(1,h_3),\tag{42}$$

where, as before, $\alpha=3\sqrt{\pi}/16$.

The set of equations (40)–(42) can be simplified if we take advantage of two conservation laws, for the number of molecules and the energy:

TABLE I.

Problem	Coefficient	Monatomic gas	Diatomic gas	Polyatomic gas
Smoluchowski problem	c_t	0.79954	0.77187	0.76269
	c_n	-0.39863	-0.37263	-0.37092
Slow evaporation problem	s_t	-0.23687	-0.16330	-0.13888
	s_n	-0.82905	-0.89815	-0.90272

$$\int \exp(-\xi^2-\omega^2)\,\xi^3\omega^2\mu\,\varphi(x,\mu,\xi,\omega)\,d\mu\,d\xi\,d\omega=0$$

and

$$\int \exp(-\xi^2-\omega^2)\,\xi^3\omega^2(\xi^2+\omega^2)$$
$$\times\mu\,\varphi(x,\mu,\xi,\omega)\,d\mu\,d\xi\,d\omega=0.$$

Plugging the decomposition of the function φ into these relationships, we obtain two equations, from which we find that $(\mu',h_1)=-4\alpha(\mu',h_2)$ and $(\mu',h_3)=-(4/7)\alpha(\mu',h_2)$. Hence Eq. (41) simplifies:

$$\mu\frac{\partial h_2}{\partial x}+h_2=3\,c\mu(\mu',h_2),$$

where $c=1-135\,\pi/448$. We make a linear substitution similar to the one in the diatomic case: $h_1-7h_3\rightarrow h_3$. As a result, instead of (42) we have the equation

$$\mu\frac{\partial h_3}{\partial x}+h_3=(1,h_3).$$

The resulting set of transport equations can be written in vector form:

$$\mu\frac{\partial h}{\partial x}+h(x,\mu)=\frac{1}{2}\int_{-1}^{1}K(\mu,\mu')h(x,\mu')\,d\mu', \tag{43}$$

where h is a column vector with elements h_1, h_2, and h_3, while $K(\mu,\mu')$ is the same matrix as in Sec. 2. The boundary conditions now assume the form

$$h(0,\mu)=-h_{\mathrm{as}}(0,\mu),\quad 0<\mu<1, \tag{44}$$

$$h(\infty,\mu)=\begin{bmatrix}0\\0\\0\end{bmatrix},\quad -1<\mu<0, \tag{45}$$

where

$$h_{\mathrm{as}}(x,\mu)=\begin{bmatrix}\varepsilon_n+\varepsilon_T/2-K(x-\mu)/2\\(2U-2K/3\sqrt{\pi})\mu\\\varepsilon_n-13\varepsilon_T/2-15K(x-\mu)/2\end{bmatrix}.$$

Formally, the problem (43)–(45) is the same as problem (18)–(20). Hence we can simply write the expressions for the jumps in temperature and concentration:

$$\varepsilon_T=2U\beta+K(U_1-2\beta/3\sqrt{\pi}),$$

$$\varepsilon_n=2U(13\beta/2)+K(-U_1-13\beta/3\,\sqrt{\pi}).$$

Here $\beta=-4\alpha(U_1-V_1+\eta_0)/7$, and η_0 is the zero of the function $\omega(\eta)=1+3cz^2\lambda_c(z)$. Again taking advantage of numerical calculations ($V_1=0.97915$ and $U_1=0.71045$), we write the final formulas for calculating the jumps in temperature and concentration:

$$\varepsilon_T=2U(-0.13888)+K(0.76269),$$

$$\varepsilon_n=2U(-0.90272)+K(-0.37092).$$

The results for diatomic and polyatomic gases are listed in Table I. In the same table we also list the results for a monatomic gas taken from our earlier paper.[11]

Note that in the Smoluchowski problem of the temperature jump, which is the most important problem from the standpoint of applications, the boundary condition is usually written[1,2] in the form $T_e-T_s=C_T lA$. Here l has the sense of a mean free path, and various authors give different definitions of this quantity. Here we use the definition that in a monatomic gas coincides with Cercignani's definition:[2] $l=\Pr\chi\sqrt{\pi m/2kT}$ where χ is the thermal diffusivity, and Pr is the Prandtl number. Then at $\Pr=2/3$ we have $C_T=1.99885$ for a monatomic gas, $C_T=1.86763$ for a diatomic gas, and $C_T=1.82963$ for a polyatomic gas. Clearly, when a monatomic gas is replaced by a diatomic gas, the jump in temperature decreases by 6.6 %, and when a diatomic gas is replaced by a polyatomic gas, it decreases by 2 %. This decrease in the rate of variation of the jump in temperature is related to the fact that when we replace a monatomic gas by a diatomic gas, the number of degrees of freedom of the molecule changes by 67 % (from 3 to 5), and when a diatomic gas is replaced by a polyatomic gas, the number of degrees of freedom of the molecules changes by only 20 % (from 5 to 6).

In conclusion, we note that in the present paper we have for the first used a consistent analytic approach to solve the fundamental Smoluchowski problem for molecular gases with differing numbers of atoms in a molecule, and have also solved the slow evaporation problem. The two problems have been examined in the same setting.

ACKNOWLEDGMENTS

This work was made possible by a grant from the Russian Fund for Fundamental Research (Projects 96-15-96904 and 97-01-00333).

[1] M. N. Kogan, *Dynamics of Rarefied Gas* [in Russian], Nauka, Moscow (1967).
[2] C. Cercignani, *Theory and Application of the Boltzmann Equation*, Plenum, New York (1975).
[3] A. V. Latyshev, Prikl. Mat. Mekh. **54**, 581 (1990).
[4] E. P. Gross, E. A. Jackson, and S. Ziering, Ann. Phys. (N.Y.) **1**, 141 (1957).
[5] Yu. Yu. Abramov, Teplofiz. Vys. Temp. **8**, 1013 (1970).
[6] E. G. Mayasov, A. A. Yushkanov, and Yu. I. Yalamov, Pis'ma Zh. Tekh. Fiz. **14**, 498 (1988) [Sov. Tech. Phys. Lett. **14**, 220 (1988)].
[7] S. K. Loyalka, Physica A **163**, 213 (1990).
[8] Y. Sone and Y. Onishi, J. Phys. Soc. Jpn. **35**, 1773 (1973).
[9] Y. Onishi and Y. Sone, J. Phys. Soc. Jpn. **44**, 1981 (1978).
[10] E. B. Dolgosheina, A. V. Latyshev, and A. A. Yushkanov, Izv. Ross. Akad. Nauk, Ser. Mekh. Zhidk. Gaza No. 1, 163 (1992).
[11] A. V. Latyshev and A. A. Yushkanov, Izv. Ross. Akad. Nauk, Ser. Mekh. Zhidk. Gaza No. 3, 140 (1996).
[12] V. M. Zhdanov and M. Ya. Alievskiĭ, *Transport and Relaxation Processes in Molecular Gases* [in Russian], Nauka, Moscow (1989).
[13] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, Oxford (1969).
[14] V. A. Rykov, Izv. Akad. Nauk SSSR, Ser. Mekh. Zhidk. Gaza No. 6, 105 (1975).
[15] I. N. Larina and V. A. Rykov, in *Numerical Methods in Rarefied Gas Dynamics* [in Russian], Issue 4, Computer Center of the USSR Academy of Sciences, Moscow (1979), p. 52.
[16] K. Barwinkel and U. Thelker, in *20th Int. Symp. of Rarefied Gas Dynamics*, 1996 Book of Abstracts, H. 3, Beijing (1996).
[17] A. V. Latyshev and A. A. Yushkanov, Teor. Mat. Fiz. **95**, 530 (1993).
[18] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).
[19] L. D. Landau and E. M. Lifshitz, *Hydrodynamics* [in Russian], Nauka, Moscow (1991).
[20] I. N. Larina and V. A. Rykov, Izv. Akad. Nauk SSSR, Ser. Mekh. Zhidk. Gaza No. 5, 141 (1986).
[21] V. S. Vladimirov, *Generalized Functions in Mathematical Physics*, Mir Publishers, Moscow (1979).
[22] K. Case and P. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, MA (1967).
[23] F. D. Gakhov, *Boundary Value Problems*, Addison-Wesley, Reading, MA (1966).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Critical dynamics of slightly disordered spin systems

V. V. Prudnikov,*) S. V. Belim, A. V. Ivanov, E. V. Osintsev, and A. A. Fedorenko

Omsk State University, 644077 Omsk, Russia

(Submitted 4 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 972–984 (September 1998)

This paper is a field theoretic description of the critical dynamics of spin systems with frozen nonmagnetic impurities. For three-dimensional systems the dynamical critical exponent is found directly by employing the three-loop approximation with the Padé–Borel summation technique. The results are compared with those obtained by calculating the dynamical exponent for homogeneous systems in the four-loop approximation, and with the values obtained by computer simulation of the critical dynamics by Monte Carlo methods. Calculations of the dynamical exponent for the two-dimensional Ising model in the four-loop approximation are also presented. © *1998 American Institute of Physics.* [S1063-7761(98)01509-1]

As is known, only for Ising magnetic materials do phase transitions in homogeneous spin systems change when randomly distributed frozen nonmagnetic impurities are introduced into such systems.[1] The ε-expansion method makes it possible to calculate the values of the critical exponents for dilute magnetic materials.[2] However, the asymptotic convergence of the ε-expansionseries in this case is even slower than for homogeneous systems.[3] The renormalization group approach to the description of slightly disordered spin systems, applied directly to three-dimensional systems by Mayer *et al.*,[4,5] has made it possible to obtain static critical exponents in the four-loop approximation. However, no calculations of equal accuracy exist for the description of the dynamics of disordered systems, the reason being that the computational load is extremely large even in the lowest perturbation-theory orders.

The present paper is a field theoretic description of the critical dynamics of slightly disordered three-dimensional spin systems in the three-loop approximation. The adopted model is a classical spin system with the nonmagnetic impurity atoms frozen at the lattice sites. The system Hamiltonian is

$$H=\frac{1}{2}\sum_{ij} J_{ij}p_ip_j\mathbf{S}_i\cdot\mathbf{S}_j,$$

where $\mathbf{S}_i$ is an n-component spin variable, J_{ij} are the coupling constants of the translation-invariant short-range ferromagnetic interaction, and p_i is a random variable described by the distribution function

$$P(p_i)=p\,\delta(p_i-1)+(1-p)\,\delta(p_i)$$

with $p=1-c$ (here c is the concentration of the nonmagnetic impurity atoms). Thermodynamically this model is equivalent to the $O(n)$-symmetric Ginzburg–Landau–Wilson model, which has the Hamiltonian

$$H[\varphi,V]=\int d^dx\left\{\frac{1}{2}[|\nabla\boldsymbol{\varphi}|^2+r_0\boldsymbol{\varphi}^2+V(x)\boldsymbol{\varphi}^2]+\frac{g_0}{4!}\boldsymbol{\varphi}^4\right\}, \quad (1)$$

where $\boldsymbol{\varphi}(x,t)$ is the n-componentorder parameter, $V(x)$ is the potential of the random impurity field, $r_0\sim T-T_{0c}(p)$, with T_{0c} the system's critical temperature determined in the mean-field theory, g_0 is a positive constant, and d is the number of dimensions of the system. We assume that the impurity potential is specified by a Gaussian distribution:

$$P_V=A_V\exp\left[-(8\delta_0)^{-1}\int d^dx\, V^2(x)\right],$$

where A_V is a normalization constant, and δ_0 is a positive constant proportional to the impurity concentration and the square of the impurity potential.

The dynamical behavior of the system in the relaxation regime near the critical temperature can be described by the Langevin transport equation for the order parameter:[6]

$$\frac{\partial\boldsymbol{\varphi}}{\partial t}=-\lambda_0\frac{\delta H}{\delta\boldsymbol{\varphi}}+\boldsymbol{\eta}+\lambda_0\mathbf{h}, \quad (2)$$

where λ_0 is the transport coefficient, $\boldsymbol{\eta}(x,t)$ is the Gaussian random force, which is a measure of the effect of the heat reservoir and is specified by the distribution function

$$P_\eta=A_\eta\exp\left[-(4\lambda_0)^{-1}\int d^dx\,dt\,\boldsymbol{\eta}^2(x,t)\right]$$

with normalization constant A_η, and $\mathbf{h}(t)$ is an external field thermodynamically conjugate to the order parameter. The temporal correlation function $G(x,t)$ of the order parameter can be found by solving Eq. (2) for $\boldsymbol{\varphi}[\boldsymbol{\eta},\mathbf{h},V]$, with $H[\boldsymbol{\varphi},V]$ given by (1), averaging the result over the Gaussian random force $\boldsymbol{\eta}$ via P_η and over the random potential $V(x)$ of the impurity field via P_V, and isolating the part of the solution that is linear in $\mathbf{h}(0)$, i.e.,

$$G(x,t)=\frac{\delta}{\delta\mathbf{h}(0)}\cdot\langle\boldsymbol{\varphi}(x,t)\rangle_{\mathrm{imp}}|_{h=0},$$

where

$$\langle\boldsymbol{\varphi}(x,t)\rangle_{\mathrm{imp}}=B^{-1}\int D\{\boldsymbol{\eta}\}D\{V\}\boldsymbol{\varphi}(x,t)P_\eta P_V,$$

$$B=\int D\{\boldsymbol{\eta}\}D\{V\}P_{\eta}P_{V}.$$

Significant difficulties are encountered when the standard renormalization group method is applied to this dynamical model. However, as shown by De Dominicis[7] for inhomogeneous systems in the absence of disorder introduced by impurities, in describing the critical dynamics the model based on the Langevin equation is equivalent to the standard Lagrange system[8] with the Lagrangian

$$L=\int d^{d}x\,dt\left\{\lambda_0^{-1}\boldsymbol{\varphi}^{2}+i\boldsymbol{\varphi}^{*}\cdot\left(\lambda_0^{-1}\frac{\partial\boldsymbol{\varphi}}{\partial t}+\frac{\delta H}{\delta\boldsymbol{\varphi}}\right)\right\},$$

where we have introduced the auxiliary field $\boldsymbol{\varphi}^{*}$. Here the correlation function $G(x,t)$ of the order parameter for a homogeneous system is given by

$$G(x,t)=\langle\boldsymbol{\varphi}(0,0)\cdot\boldsymbol{\varphi}(x,t)\rangle$$

$$=\Omega^{-1}\int D\{\boldsymbol{\varphi}\}D\{\boldsymbol{\varphi}^{*}\}\boldsymbol{\varphi}(0,0)\cdot\boldsymbol{\varphi}(x,t)\exp(-L[\boldsymbol{\varphi},\boldsymbol{\varphi}^{*}]),$$

where

$$\Omega=\int D\{\boldsymbol{\varphi}\}D\{\boldsymbol{\varphi}^{*}\}\exp(-L[\boldsymbol{\varphi},\boldsymbol{\varphi}^{*}]).$$

A generalization of the given field theoretic approach and the details of applying such an approach to the critical dynamics of disordered spin systems with frozen point impurities and extended defects are discussed in the context of the ε-expansion in a paper by one of the present authors.[9]

Instead of examining the correlation function proper, it is convenient to study its vertex part, which in the context of the Feynman diagram approach can be written in the three-loop approximation as follows:

$$\Gamma^{(2)}(k,\omega;r_0,g_0,\delta_0,\lambda_0)=r_0+k^{2}-\frac{i\omega}{\lambda_0}-4\delta_0D_1$$

$$-\frac{n+2}{18}g_0^{2}D_2+\frac{4(n+2)}{3}g_0\delta_0D_3-16\delta_0^{2}(D_4+D_5)$$

$$+\frac{(n+2)(n+8)}{108}g_0^{3}\left(\sum_{i=6}^{8}D_i\right)-\frac{2(n+2)^{2}}{9}$$

$$\times g_0^{2}\delta_0\left(\sum_{i=9}^{18}D_i\right)+\frac{16(n+2)}{3}$$

$$\times g_0\delta_0^{2}\left(\sum_{i=19}^{31}D_i\right)-64\delta_0^{3}\left(\sum_{i=32}^{39}D_i\right). \qquad (3)$$

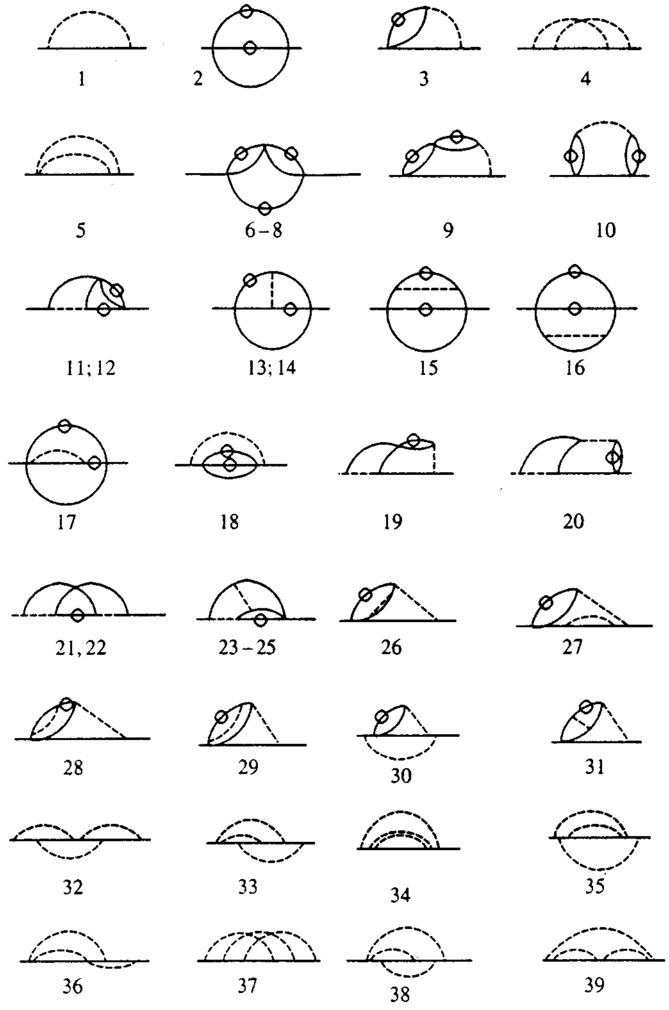


FIG. 1. Diagrammatic representation of the contributions to the vertex function $\Gamma^{(2)}(k,\omega)=G^{-1}(k,\omega)$ in the three-loop approximation. The solid lines correspond to $G_0(k,\omega)=(r_0+k^{2}-i\omega/\lambda_0)^{-1}$, the solid lines with a circle correspond to $C_0(k,\omega)=2\lambda^{-1}((r_0+k^{2})^{2}+(\omega/\lambda_0)^{2})^{-1}$, a four-leg vertex $\times$ corresponds to g_0, and an impurity four-leg vertex $\rangle\!-\!\langle$ corresponds to $\delta_0\delta(\omega)$.

The diagrams corresponding to the D_i are depicted in Fig. 1. The Feynman diagrams contain d-dimensional integrals with respect to momenta and are characterized near the critical point by an ultraviolet divergence at high momenta $\mathbf{k}$ with pole singularities. To remove these poles one usually employs a dimensional regularization scheme, which involves introducing renormalized quantities.[10] We define the renormalized order parameter as $\boldsymbol{\varphi}=Z^{-1/2}\boldsymbol{\varphi}_0$. Then the renormalized vertex functions have the generalized form

$$\Gamma_R^{(m)}(k,\omega;r,g,\delta,\lambda,\mu)=Z^{m/2}\Gamma^{(m)}(k,\omega;r_0,g_0,\delta_0,\lambda_0) \qquad (4)$$

with the renormalized coupling constants g and δ, temperature r, and transport coefficient λ:

$$g_0=\mu^{4-d}Z_g g,\qquad \delta_0=\mu^{4-d}Z_\delta\delta,$$

$$r_0=\mu^{2}Z_r r,\qquad \lambda_0{}^{-1}=\mu^{2}Z_\lambda\lambda^{-1}, \qquad (5)$$

where the scaling parameter μ is introduced so that the quantities reduce to dimensionless form. In Eq. (4), $\Gamma^{(2)}$ corresponds to the reciprocal correlation function of the order parameter $G(k,\omega)$, and $\Gamma^{(4)}$ corresponds to the four-leg vertex functions $\Gamma_g^{(4)}$ and $\Gamma_\delta^{(4)}$ for the coupling constants g and δ, respectively; the Z-factors can be found from the requirement that the renormalized vertex functions be regular, which is expressed in the normalization conditions

$$Z\left.\frac{\partial\Gamma^{(2)}(k)}{\partial k^{2}}\right|_{k^{2}=0}=1,\quad Z^{2}\Gamma_g^{(4)}|_{k_i=0}=\mu^{4-d}g,$$

TABLE I. Values of the derivatives of the diagrams depicted in Fig. 1, $D_i'=\frac{\partial D_i}{\partial(-i\omega/\lambda)}|_{k=0,\omega=0}$.

D_1'/J	−1.000 000	D_{14}'/J^3	−0.032 279	D_{27}'/J^3	−0.666 667
D_2'/J^2	−0.130 768	D_{15}'/J^3	0.061 515	D_{28}'/J^3	0.584 625
D_3'/J^2	−0.666 667	D_{16}'/J^3	0.004 666	D_{29}'/J^3	−0.092 766
D_4'/J^2	−2.000 000	D_{17}'/J^3	−0.333 557	D_{30}'/J^3	−0.074 202
D_5'/J^2	−1.000 000	D_{18}'/J^3	0.042 034	D_{31}'/J^3	−0.194 407
D_6'/J^3	−0.104 778	D_{19}'/J^3	−2.053 736	D_{32}'/J^3	−2.053 736
D_7'/J^3	−0.032 835	D_{20}'/J^3	−2.053 736	D_{33}'/J^3	−2.053 736
D_8'/J^3	−0.032 835	D_{21}'/J^3	−1.142 275	D_{34}'/J^3	−1.000 000
D_9'/J^3	−0.519 431	D_{22}'/J^3	−0.396 553	D_{35}'/J^3	0.666 667
D_{10}'/J^3	−0.519 431	D_{23}'/J^3	−1.142 275	D_{36}'/J^3	0.666 667
D_{11}'/J^3	−0.276 601	D_{24}'/J^3	−0.396 553	D_{37}'/J^3	−2.053 736
D_{12}'/J^3	−0.468 697	D_{25}'/J^3	−0.396 553	D_{38}'/J^3	−0.074 202
D_{13}'/J^3	−0.032 279	D_{26}'/J^3	0.226 932	D_{39}'/J^3	0.000 000

$$Z^2\Gamma_\delta^{(4)}|_{k_i=0}=\mu^{4-d}g,\qquad Z\frac{\partial\Gamma^{(2)}(k,\omega)}{\partial(-i\omega)}\bigg|_{k^2,\omega=0}=\lambda^{-1}. \tag{6}$$

This regularization of the vertex functions can be carried out in the three-loop approximation. To this end we write the values of the vertex functions in the normalization conditions as follows:

$$\Gamma_g^{(4)}|_{k_i=0}=g_0\sum_{i,j=0}^{3}A_{ij}g_0{}^i\delta_0{}^j,$$

$$\Gamma_\delta^{(4)}|_{k_i=0}=\delta_0\sum_{i,j=0}^{3}B_{ij}g_0{}^i\delta_0{}^j,$$

$$\frac{\partial\Gamma^{(2)}}{\partial k^2}\bigg|_{k^2=0}=\sum_{i,j=0}^{3}C_{ij}g_0{}^i\delta_0{}^j,$$

$$\frac{\partial\Gamma^{(2)}}{\partial(-i\omega/\lambda)}\bigg|_{k=0,\omega=0}=\sum_{i,j=0}^{3}D_{ij}g_0{}^i\delta_0{}^j, \tag{7}$$

where the coefficients are sums of the corresponding diagrams or their derivatives at zero external momenta and frequencies. For instance, the numerical values of the derivatives of the diagrams (Fig. 1),

$$D_i'=\frac{\partial D_i}{\partial(-i\omega/\lambda)}\bigg|_{k=0,\omega=0},$$

which comprise the coefficients D_{ij} and are obtained as a result of applying the calculation method adopted in Ref. 11, are listed in Table I, where

$$J=\int\frac{d^dq}{(q^2+1)^2}=\frac{S_d}{2}\Gamma\left(\frac{d}{2}\right)\Gamma\left(2-\frac{d}{2}\right)$$

is the one-loop integral with $S_d=2\pi^{d/2}/(2\pi)^d\Gamma(d/2)$; $\Gamma(x)$ is the gamma function. We write the expansion of g_0, δ_0, Z, and Z_λ in powers of the renormalization coupling constants g and δ,

$$g_0=g\sum_{i,j=0}^{3}a_{ij}g^i\delta^j,\qquad \delta_0=\delta\sum_{i,j=0}^{3}b_{ij}g^i\delta^j,$$

$$Z=\sum_{i,j=0}^{3}c_{ij}g^i\delta^j,\qquad Z_\lambda=\sum_{i,j=0}^{3}d_{ij}g^i\delta^j, \tag{8}$$

where the unknown expansion coefficients a_{ij}, b_{ij}, c_{ij}, and d_{ij} can be expressed in terms of A_{ij}, B_{ij}, C_{ij}, and D_{ij} via the normalization conditions.

The next step in the field theoretic approach amounts to determining the scaling functions $\beta_g(g,\delta)$, $\beta_\delta(g,\delta)$, $\gamma_r(g,\delta)$, $\gamma_\varphi(g,\delta)$, and $\gamma_\lambda(g,\delta)$ that specify the renormalization group differential equation for the vertex functions:

$$\left[\mu\frac{\partial}{\partial\mu}+\beta_g\frac{\partial}{\partial g}+\beta_\delta\frac{\partial}{\partial\delta}-\gamma_r r\frac{\partial}{\partial r}+\gamma_\lambda\lambda\frac{\partial}{\partial\lambda}-\frac{m}{2}\gamma_\varphi\right]\times\Gamma^{(m)}(k,\omega;r,g,\delta,\lambda,\mu)=0.$$

For the discussion of the dynamical behavior that follows we will need only the functions β_g and β_δ and the dynamical scaling function γ_λ determined by the following relationships:

$$4-d+\beta_g\frac{\partial\ln gZ_g}{\partial g}+\beta_\delta\frac{\partial\ln gZ_g}{\partial\delta}=0,$$

$$4-d+\beta_g\frac{\partial\ln\delta Z_\delta}{\partial g}+\beta_\delta\frac{\partial\ln\delta Z_\delta}{\partial\delta}=0,$$

$$\gamma_\lambda=\beta_g\frac{\partial\ln Z_\lambda}{\partial g}+\beta_\delta\frac{\partial\ln Z_\lambda}{\partial\delta}. \tag{9}$$

The explicit form of the functions β_g and β_δ in the four-loop representation was obtained by Mayer,[5] who introduced the coupling constants v and u, related to g and δ by $v=(n+8)Jg/6$ and $u=-16J\delta$. Next we specify the functions β and γ_λ:

$$\beta_v=v\sum_{i,j=0}^{3}\beta_{ij}^{(v)}v^iu^j,\qquad \beta_u=u\sum_{i,j=0}^{3}\beta_{ij}^{(u)}v^iu^j,$$

$$\gamma_\lambda=\sum_{i,j=0}^{3}\gamma_{ij}v^iu^j; \tag{10}$$

the values of the expansion coefficients for a three-dimensional Ising model ($n=1$) are listed in Table II. The nature of the critical point for each value of n and d is fully

TABLE II. Values of the coefficients in the expressions for the scaling functions.

(i,j)	$\beta_{i,j}^{(u)}$	$\beta_{i,j}^{(v)}$	$\gamma_{i,j}$
(0,0)	-1	1	0
(1,0)	1	3/2	-0.25
(0,1)	2/3	1	0
(2,0)	$-95/216$	$-185/216$	0.053 240
(1,1)	$-50/81$	$-104/81$	0.030 862
(0,2)	$-92/729$	$-308/729$	0.008 400
(3,0)	0.389 922	0.916 667	$-0.049\,995$
(2,1)	0.857 363	2.132 996	$-0.152\,964$
(1,2)	0.467 388	1.478 058	$-0.044\,167$
(0,3)	0.090 448	0.351 069	$-0.012\,642$

specified by the stable fixed point (v^*,u^*) for the coupling constants, which is fixed by the requirement that the functions β vanish, i.e.,

$$\beta_v(v^*,u^*)=0, \quad \beta_u(v^*,u^*)=0.$$

The order of the quantities v^* and u^* is $4-d$, so that the expansion series in v and u for the scaling functions are asymptotically convergent if $d=3$.

These series are normally summed using the the Padé–Borel method.[12] Numerical analysis of the equations for determining the fixed points and of the stability conditions shows that in contrast to the ε-expansion of Khmel'nitskiĭ[2] and Jayaprakash and Katz,[3] for $d=3$ there is no accidental degeneracy of the fixed points at $n=1$. Only two of the four fixed points are of interest here: the fixed point for homogeneous systems, $(v^*\neq 0, u^*=0)$, and the impurity fixed point $(v^*\neq 0, u^*\neq 0)$, which specifies new critical properties of disordered systems. The impurity fixed point is stable only if $n=1$, while for $n\geqslant 2$ the presence of disorder related to the presence of frozen-in impurities is unimportant for critical behavior. The impurity fixed point for the three-dimensional Ising model in the three-loop approximation is given by $v^*=2.256938$ and $u^*=-0.728168$.

By plugging the values of the coupling constants at the fixed point into the scaling function $\gamma_\lambda(v,u)$ we can determine the dynamical critical exponent z, which is the measure of the critical retardation of relaxation processes,

$$z=2+\gamma_\lambda(v^*,u^*). \tag{11}$$

However, the expansion of $\gamma_\lambda(v^*,u^*)$ in powers of v^* and u^* at $d=3$ is asymptotically convergent at best, and summing the series directly does not yield reasonable values. To sum the series one can employ the generalized Padé–Borel method, which amounts to applying the Borel transformation to the series

$$\gamma_\lambda(v,u)=\sum_{i,j}\gamma_{ij}v^iu^j=\int_0^\infty e^{-t}\Gamma_\lambda(vt,ut)\,dt,$$

$$\Gamma_\lambda(x,y)=\sum_{i,j}\frac{\gamma_{ij}}{(i+j)!}x^iy^j, \tag{12}$$

and using the Padé–Chisholm approximants

$$[M,N/K,L]=\sum_{i=0}^{M}\sum_{j=0}^{N}a_{ij}v^iu^j\left(\sum_{p=0}^{K}\sum_{q=0}^{L}b_{pq}v^pu^q\right)^{-1}.$$

The resulting expansion for $\gamma_\lambda(v,u)$ in powers of v and u in the three-loop approximation allows using approximants of the form $[1,1/1,1]$ and $[2,2/1,1]$. Application of the approximants $[1,1/1,1]$ corresponds to the description of the critical dynamics of disordered magnetic materials in the two-loop approximation,[13] and yields a dynamical exponent $z_{\text{imp}}^{(2)}=2.169849$. Using the approximants $[2,2/1,1]$ makes it possible to obtain the exponent z in the form

$$\begin{aligned}z=2&+\frac{\alpha_1u}{\beta}+\frac{\beta-1}{\beta^2}(\alpha_2u^2+\alpha_3uv+\alpha_4v^2)\\&+\frac{2\beta^2-\beta+1}{\beta^3}(\alpha_5u^2v+\alpha_6uv^2)-\frac{1}{\beta}\left[\alpha_1u+\frac{1}{\beta}(\alpha_2u^2\right.\\&\left.+\alpha_3uv+\alpha_4v^2)+\frac{1}{\beta^2}(\alpha_5u^2v+\alpha_6uv^2)\right]{}_2F_0(1,1,\beta),\end{aligned} \tag{13}$$

where ${}_2F_0(1,1,\beta)$ is the confluent hypergeometric function, and

$$\alpha_1=\gamma_{1,0}, \quad \alpha_2=\frac{\gamma_{2,0}}{2}-\frac{\gamma_{1,0}\gamma_{3,0}}{3\gamma_{2,0}},$$

$$\alpha_3=\frac{\gamma_{1,1}}{2}-\frac{\gamma_{0,1}\gamma_{0,3}}{3\gamma_{0,2}}, \quad \alpha_4=\frac{\gamma_{0,2}}{2},$$

$$\alpha_5=\frac{\gamma_{2,1}}{6}-\frac{\gamma_{1,1}\gamma_{3,0}}{6\gamma_{2,0}}-\frac{\gamma_{2,0}\gamma_{0,3}}{6\gamma_{0,2}},$$

$$\alpha_6=\frac{\gamma_{1,2}}{6}-\frac{\gamma_{1,1}\gamma_{0,3}}{6\gamma_{0,2}}-\frac{\gamma_{0,2}\gamma_{3,0}}{6\gamma_{2,0}},$$

$$\beta=\beta_1u+\beta_2v, \quad \beta_1=-\frac{\gamma_{3,0}}{3\gamma_{2,0}}, \quad \beta_2=-\frac{\gamma_{0,3}}{3\gamma_{0,2}}.$$

Using the values of the coupling constants at the impurity fixed point, $v^*=2.256938$ and $u^*=-0.728168$, we obtain a dynamical exponent

$$z_{\text{imp}}^{(3)}=2.165319. \tag{14}$$

The fact that the difference in the values of z_{imp} calculated in the three- and two-loop approximations is small suggests that allowing for higher-order corrections can lead only to negligible changes in this value. At the same time, the calculations in Ref. 13 using an ϵ-expansion in the two-loop approximation yielded $z_{\text{imp}}^{(2)}=2.336$, which justifies the use of the renormalization group procedure in describing the critical behavior of dilute magnetic material for the case where $d=3$.

To establish the effect of impurities on dynamical critical behavior we must compare the values of z for disordered and homogeneous systems. As is known,[6] fluctuation corrections to the mean-field value of the dynamical exponent $z^{(0)}=2$ emerge in homogeneous systems only in the two-loop approximation, while in disordered systems the dynami-

TABLE III. Values of the coefficients in the expressions for the vertex functions.

Coefficient	$d=3$	$d=2$
A_1	−1.0	−1.0
A_2	1.222 2222	1.375 0699
A_3	−1.705 3479	−2.305 4548
B_1	0.005 4869	0.008 4916
B_2	−0.007 0112	−0.011 6591
B_3	0.010 1430	0.017 9966
C_1	0.009 6865	0.015 2547
C_2	−0.012 6257	−0.021 3740
C_3	0.016 9420	0.035 2450

cal effects of scattering of magnetization fluctuations by impurities show up by the first-order approximation.

The effectiveness of summation methods for asymptotically convergent series is largely determined by the number of known terms in the series. Therefore, when the summation methods are applied to the dynamical scaling function, the accuracy of the three-loop approximation for the exponent z of a disordered system can only correspond to the four-loop approximation for z in a homogeneous spin system. With this mind, we calculated the dynamical critical exponent for homogeneous three-dimensional spin systems in the four-loop approximation. The expressions for the vertex functions in (7) for homogeneous systems become much simpler, and in the four-loop approximation assume the form

$$\Gamma_v^{(4)}|_{k_i=0}=v_0+A_1v_0^2+A_2v_0^3+A_3v_0^4,$$

$$\left.\frac{\partial\Gamma^{(2)}}{\partial k^2}\right|_{k^2=0}=1+B_1v_0^2+B_2v_0^3+B_3v_0^4,$$

$$\left.\frac{\partial\Gamma^{(2)}}{\partial(-i\omega)}\right|_{k=0,\omega=0}=1+C_1v_0^2+C_2v_0^3+C_3v_0^4, \tag{15}$$

where $v_0=(n+8)Jg_0/6$. The values of the coefficients at $n=1$ are listed in Table III. The four-loop diagrams that producezx the coefficient C_3 are shown in Fig. 2. Carrying out the calculations, these diagrams split into 48 $4d$-fold integrals, whose numerical values are listed in Table IV. To describe the dynamical behavior we require only the functions $\beta(v)$ and $\gamma_\lambda(v)$:

$$\beta(v)=-(4-d)\left[\frac{\partial\ln Z_v v}{\partial v}\right]^{-1},\quad \gamma_\lambda(v)=\beta(v)\,\frac{\partial\ln Z_\lambda}{\partial v}. \tag{16}$$

The explicit form of the first function in the six-loop approximation was obtained by Baker *et al.*[12] By consistently applying the above field theoretic approach, we were able to derive an expression for the dynamical scaling function $\gamma_\lambda(v)$ in the four-loop approximation:

$$\begin{aligned}\gamma_\lambda(v)=&-(4-d)v[2(B_1-C_1)+(3B_2-3C_2-4A_1B_1\\&+4A_1C_1)v+(4B_3-4C_3-9A_1B_2+9A_1C_2\\&+10A_1^2B_1-10A_1^2C_1-4A_2B_1+4A_2C_1\\&-8B_1D_1+6B_1^2-2C_1^2)v^2].\end{aligned} \tag{17}$$

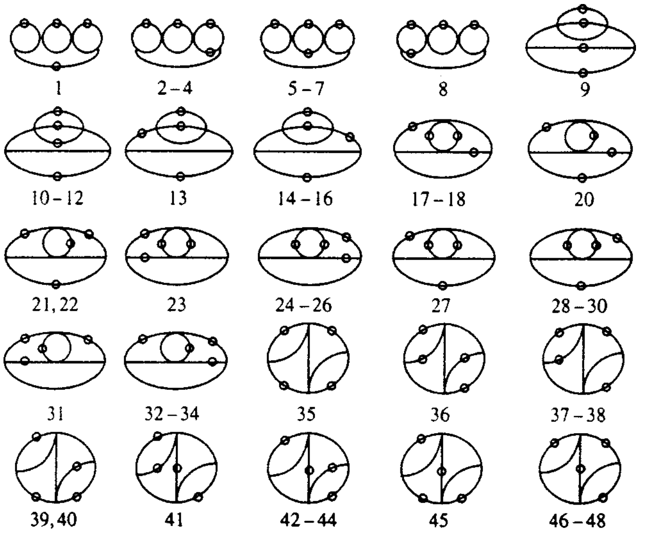


FIG. 2. Four-loop diagrams contributing to the vertex function $\Gamma^{(2)}(k,\omega)$.

Inserting the values of the coefficients listed in Table III for $d=3$ and $n=1$, we obtain

$$\gamma_\lambda(v)=0.008399v^2-0.000045v^3+0.020423v^4. \tag{18}$$

In accordance with Ref. 12,

$$\beta(v)=-v+v^2-0.422497v^3+0.351069v^4-0.376527v^5. \tag{19}$$

To calculate the values of the coupling constant v^* at the fixed point and the dynamical critical exponent z, we used the Padé–Borel summation method with the approximants [4/1] and [3/1], respectively. As a result, for $d=3$ and $n=1$ we obtained

$$v^*=1.4299,\quad z_{\text{pure}}^{(4)}=2.017.$$

Comparison of the results revealed a significant difference between the values of the dynamical exponent z for the homogeneous and disordered Ising models. This makes it possible to study the effect of impurities on the dynamical critical behavior in a real physical experiment and via Monte Carlo simulation.

Let us compare our value of the dynamical exponent $z_{\text{imp}}^{(3)}$ with computer simulations of the dynamical critical behavior of a disordered Ising model.[14–16] In Refs. 14 and 15, critical magnetization relaxation was numerically simulated for a system with dimensions of 48^3 and impurity concentrations $0.4\leqslant p\leqslant 1$. Jan *et al.*[17] combined the Monte Carlo method with the dynamical renormalization group method to determine the dynamical critical exponent z. The following values of the critical exponent were obtained for the homogeneous system and two slightly disordered systems with $p=0.95$ and 0.8:

$$z(1.0)=1.97\pm0.08,\quad z(0.95)=2.19\pm0.07,$$

$$z(0.8)=2.20\pm0.08,$$

which are in good agreement with the numerical results. Heuer[16] obtained the values of z by analyzing the asymptotic properties of the dynamical autocorrelation function for a

TABLE IV. Values of four-loop diagrams.

No.	$d=3$	$d=2$	No.	$d=3$	$d=2$	No.	$d=3$	$d=2$
1	0.104 869	0.165 307	17	0.001 108	0.004 131	33	0.002 527	0.007 463
2	0.004 166	0.009 670	18	0.000 923	0.003 307	34	0.014 580	0.029 449
3	0.008 180	0.022 921	19	0.000 932	0.003 343	35	0.039 776	0.070 254
4	0.029 674	0.059 714	20	0.019 410	0.034 609	36	0.002 378	0.006 421
5	0.003 264	0.003 943	21	0.019 189	0.034 135	37	0.004 691	0.012 723
6	0.015 354	0.010 076	22	0.004 177	0.011 294	38	0.003 820	0.007 370
7	0.014 330	0.028 777	23	0.001 928	0.004 644	39	0.011 650	0.027 311
8	0.011 627	0.016 314	24	0.000 706	0.005 891	40	0.005 377	0.013 297
9	−0.002 506	−0.006 853	25	0.003 421	0.010 167	41	0.003 981	0.007 464
10	0.000 823	0.002 744	26	0.000 862	0.003 535	42	0.003 314	0.010 303
11	0.003 444	0.009 238	27	0.000 551	0.002 471	43	0.009 470	0.023 519
12	0.003 745	0.010 685	28	0.003 898	0.011 209	44	0.003 866	0.010 905
13	−0.004 883	−0.012 280	29	0.001 077	0.003 405	45	0.023 730	0.038 420
14	−0.004 883	−0.012 280	30	0.003 815	0.011 007	46	0.033 485	0.062 921
15	0.007 527	0.017 1805	31	0.007 379	0.012 666	47	0.007 121	0.021 633
16	−0.005 471	−0.014 199	32	0.004 177	0.009 667	48	0.004 760	0.011 691

system that is in a state of equilibrium and exhibits strong magnetization fluctuations. For instance, it was found that

$$z(1.0)=2.095\pm 0.008$$

for the homogeneous system

$$z(0.95)=2.16\pm 0.01,\quad z(0.9)=2.232\pm 0.004,$$

$$z(0.8)=2.38\pm 0.01,$$

for slightly disordered systems, and

$$z(0.6)=2.93\pm 0.03$$

for a system with $p=0.6$. Believing that the fixed point of the critical behavior of a slightly disordered system, which is independent of the impurity concentration, is also such a point for any impurity concentration, Heuer[16] estimated the asymptotic value of the dynamical exponent z to be 2.4 ± 0.1. The value of z for a homogeneous system obtained by Heuer[16] differs drastically from the results of the field theoretic approach, while for a system with $p=0.95$ the agreement is unexpectedly good. Our view on the universality of the critical behavior of disordered systems has been explained in Refs. 14 and 15, where we proposed separating the universal critical behavior of slightly disordered systems from that of highly disordered systems and hypothesized that the critical exponents of three-dimensional disordered systems exhibit stepped universality.

The predictions of the theory concerning the effect of impurities on the dynamical critical behavior of magnetic materials (a higher value of $z_{\mathrm{imp}}(d=3)$ compared to the value of $z_{\mathrm{pure}}(d=3)$) can be corroborated by several experimental methods: inelastic neutron scattering (the linewidth $\omega_\varphi \propto |T-T_c|^{z\nu}$ at $q=0$ and $\omega_\varphi \propto q^z$ at $T=T_c$), EPR and NMR (the resonance linewidth $\Delta\omega \propto |T-T_c|^{(d-2+\eta-z)\nu}$, where η is the Fisher exponent), measurements of the dynamic susceptibility in an external high-frequency magnetic field ($\chi(\omega)\propto \omega^{-\gamma/z\nu}$ at $T=T_c$, where γ is the susceptibility exponent), and ultrasound measurements (the sound absorption coefficient $\alpha(\omega)\propto |T-T_c|^{-(\alpha+z\nu)}\omega^2 g(\omega/|T-T_c|^{z\nu})$ and the acoustic dispersion $C^2(\omega - C^2(0)\propto |T-T_c|^{-\alpha} f(\omega/|T-T_c|^{z\nu}))$. Unfortunately, we know of no experimental work in which the dynamical critical behavior of slightly diluted Ising-like magnetic materials is studied.

The critical dynamics of a slightly disordered two-dimensional Ising model in the relaxation regime does not differ from the dynamics of the homogeneous model.[13] An analysis of the critical dynamics of the two-dimensional Ising models shows that the values z span a broad range: $2.08\leqslant z\leqslant 2.24$. For instance, $z=2.14\pm 0.02$ (Ref. 18), 2.13 ± 0.03 (Ref. 19), 2.076 ± 0.005 (Ref. 20), 2.24 ± 0.04 (Ref. 21), 2.24 ± 0.07 (Ref. 22), and 2.16 ± 0.04 (Ref. 23) in computer simulation; $z=2.126$ (Ref. 24) in the field theoretic approach in the two-loop approximation with the interpolation of the results of the $1+\varepsilon$- and $4-\varepsilon$-expansions; and $z=2.183\pm 0.005$ (Ref. 25) in the same approach with interpolation of the results of the high-temperature expansion.

We calculated the dynamical exponent z for a homogeneous two-dimensional Ising model in the four-loop approximation in the context of the field theoretic approach. The corresponding values of the coefficients in the expressions (15) for the vertex functions and the numerical values of the four-loop diagrams for the two-dimensional model are listed in Tables III and IV. As a result we arrived at the following expressions for the scaling functions:

$$\gamma_\lambda(v)=0.027053v^2-0.004184v^3+0.022130v^4,$$

$$\beta(v)=-v+v^2-0.716174v^3+0.930766v^4 -1.582388v^5. \tag{20}$$

Summing by the Padé–Borel method, we found the values of the coupling constant v^* at the fixed point and of z:

$$v^*=1.8836,\quad z^{(4)}(d=2)=2.093.$$

We see that the exponent z is at the lower edge of the range mentioned earlier. The adopted procedure of calculating the exponents is assumed to be the most accurate, so that we expect the calculated values to be the benchmarks for computer simulations of homogeneous systems and to be used in developing simulation methods for disordered systems.

ACKNOWLEDGMENTS

This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant 97-02-16124).

*) E-mail: prudnikv@univer.omsk.su

---

[1] A. B. Harris, J. Phys. C **7**, 1671 (1974).
[2] D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **68**, 1960 (1975) [Sov. Phys. JETP **41**, 981 (1975)].
[3] C. Jayaprakash and H. J. Katz, Phys. Rev. B **16**, 3987 (1977).
[4] I. O. Mayer, A. I. Sokolov, and B. N. Shalaev, Ferroelectrics **95**, 93 (1989).
[5] I. O. Mayer, J. Phys. A **22**, 2815 (1989).
[6] P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).
[7] C. De Dominicis, Nuovo Cimento Lett. **12**, 567 (1975).
[8] E. Brezin, J. C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **8**, 434, 2418 (1973).
[9] V. V. Prudnikov and I. D. Lawrie, J. Phys. C **17**, 1655 (1984).
[10] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, Wiley, New York (1979).
[11] S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **68**, 273 (1975) [Sov. Phys. JETP **41**, 133 (1975)].
[12] G. A. Baker Jr., B. G. Nickel, and D. I. Meiron, Phys. Rev. B **17**, 1365 (1978).
[13] V. V. Prudnikov and A. N. Vakilov, Zh. Éksp. Teor. Fiz. **101**, 1853 (1992) [Sov. Phys. JETP **74**, 990 (1992)].
[14] A. N. Vakilov and V. V. Prudnikov, JETP Lett. **55**, 741 (1992).
[15] V. V. Prudnikov and A. N. Vakilov, Zh. Éksp. Teor. Fiz. **103**, 962 (1993) [JETP **76**, 469 (1993)].
[16] H.-O. Heuer, J. Phys. A **26**, L-341 (1993).
[17] N. Jan, L. L. Moseley, and D. Stauffer, J. Stat. Phys. **33**, 1 (1983).
[18] C. Kalle, J. Phys. A **17**, L-801 (1984).
[19] J. K. Williams, J. Phys. A **18**, 49 (1985).
[20] M. Mori and Y. Tsuda, Phys. Rev. B **37**, 5444 (1988).
[21] P. H. Poole, N. Jan, and D. I. Meiron, J. Phys. A **23**, L-453 (1990).
[22] V. V. Prudnikov and O. N. Markov, J. Phys. A **28**, 1549 (1995).
[23] F. Wang, N. Hatane, and M. Suzuki, J. Phys. A **28**, 4543 (1995).
[24] R. Bausch, V. Dohm, H. K. Janssen, and R. K. P. Zia, Phys. Rev. Lett. **47**, 1837 (1981).
[25] B. Daman and I. D. Reger, Z. Phys. B **98**, 97 (1995).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

d-symmetry superconductivity due to valence bond correlations

A. A. Ovchinnikov,*) M. Ya. Ovchinnikova,†) and E. A. Plekhanov

Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia

(Submitted 24 December, 1997)
Zh. Éksp. Teor. Fiz. **114**, 985–1005 (September 1998)

It is shown that *d*-symmetry superconductivity due to valence bond correlations is possible. Valence bond correlations are compatible with antiferromagnetic spin order. In order to explictly construct a homogeneous state with the valence bond structure in the two-dimensional Hubbard model for an arbitrary doping, we have used the variational method based on unitary local transformation. Attraction between holes in the *d*-channel is due to modulation of hopping by the site population in course of the valence bond formation, and corresponding parameters have been calculated variationally. An important factor for the gap width is the increase in the density of states on the Fermi level due to antiferromagnetic splitting of the band. The gap width and its ratio to the T_c are $2\Delta \simeq 0.1t$ and $2\Delta/kT_c \simeq 4.5-4$ for $U/t \simeq 8$. The correspondence between the theoretical phase diagram and experimental data is discussed. The dependence of T_c on the doping $\delta = |n-1|$ and the Fermi surface shape are highly sensitive to the weak interaction t' leading to diagonal hoppings. In the case of $t'>0$ and *p*-doping, the peak on the curve of $T_c(\delta)$ occurs at the doping δ_{opt}, when the energy of the flattest part of the lower Hubbard subband crosses the Fermi level at $k \sim (\pi,0)$. In underdoped samples with $\delta < \delta_{\mathrm{opt}}$, the anisotropic pseudogap in the normal state corresponds to the energy difference $|E(\pi,0)-\mu|$ between this part of the spectrum and the Fermi level.
[S1063-7761(98)01609-6]

1. INTRODUCTION

In recent years, a lot of useful information about electron spectra of high-temperature superconductors (HTSC) has been obtained through high-resolution angle-resolved photoemission spectroscopy (ARPES),[1,2] neutron scattering experiments,[3] and phase-sensitive measurements.[4,5] The latest results are concerned with the discovery of the ''small'' Fermi surface[6] and anisotropic pseudogap in excitation spectra of ''underdoped'' HTSC in the normal state.[7,8] The approach based on the energy band model, which is a natural language for describing these experiments, should necessarily take into account all types of correlations whose role is expected to be significant, as indicated by analysis in the localized limit $U \to \infty$ or numerical calculations in the model of small clusters.

The aim of all theoretical investigations is to answer the most important questions, such as whether the attraction among charge carriers and pairing mechanism are controlled by correlations, and what is the role of antiferromagnetic correlations and valence bond correlations.[9–12] The idea that the superconducting pairing is controlled by correlation was put forth and developed by Zaitsev.[13]

The aim of the present work is to investigate these questions using a variational approach and a representation of a correlated state in an explicit form. We hope to elucidate the role of specific types of correlations, in particular, those of the valence bond type. Our analysis is based on the classic model of strongly correlated systems, namely, the Hubbard two-dimensional model:

$$H = -t \sum_{\langle nm \rangle,\sigma} (c^{\dagger}_{n\sigma} c_{m\sigma} + \mathrm{H.c.}) + \sum_{n} U n_{n\uparrow} n_{n\downarrow}. \tag{1}$$

Its adequacy to the electronic structure of the CuO_2 plane and its basic parameters have been established with a fair degree of certainty.[14,15]

The states of resonant valence bonds (RVB), which were introduced by Anderson,[16,17] implied that the system configuration was composed of singlet components of states with two particles localized at sites connected by a valence bond. Later[18–20] variational functions were constructed for the Hubbard model, which are the band analogues of states with a structure of periodic dimers and homogeneous states of valence bonds.[20] In these solutions, unlike RVB, formation of singlet states leads to changes in the charge states of sites connected by the bonds. Some features of such state have also been investigated, such as their compatibility with antiferromagnetism, the spectrum of hole-like excitations and its impact on photoemission spectra, etc. The problem of the superconducting ordering has not been addressed. Meanwhile, the method of unitary local transformations,[19,20] unlike nonunitary transformations, such as the Gutzwiller ansatz,[21] allows one not only to construct a correlated function, but also to derive an explicit expression for the effective Hamiltonian, and hence to check whether a superconducting ordering of the $d_{x^2-y^2}$ symmetry is feasible. A lot of experiments using various techniques[10] have indicated that this symmetry really takes place in some cuprates ($YBa_2Cu_3O_{7-\delta}$, $Bi_2Sr_2CaCu_2O_{8+\delta}$). For this reason and because the one-site repulsion suppresses the *s*-symmetry ordering but does not affect d-symmetry correlations, we will consider only the latter type of symmetry.

The attraction between holes is often associated with correlated (or modulated) hopping interactions, first described by Hirsch.[22] In contrast to this model[22] and similar interactions in the $t-J$ model,[12] in this paper the form and amplitude of such interactions are determined using a variational method by solving an appropriate self-consistent problem.

2. CONSTRUCTION OF A HOMOGENEOUS VALENCE BOND STATE AND SELF-CONSISTENT SOLUTION OF THE PROBLEM

A wave function Ψ with valence bond correlations is derived from the function Φ of the uncorrelated state using a unitary transformation:

$$\Psi=\hat{W}(\alpha)\Phi, \quad \hat{W}(\alpha)=\exp(\alpha Z), \quad Z=\sum_{\langle nm\rangle} Z_{nm}. \qquad (2)$$

The local anti-Hermitian operator Z_{nm} acting on the bond $\langle nm\rangle$ of neighboring sites is defined by the formula

$$Z_{nm}=[g^{\dagger}_{nm,\sigma}g^{\dagger}_{nm,-\sigma}u_{nm,-\sigma}u_{nm,\sigma}-\mathrm{H.c.}]$$

$$\equiv -\frac{1}{2}\sum_{\sigma} j_{nm\sigma}\Delta_{nm,-\sigma}, \qquad (3)$$

$$j_{nm\sigma}=c^{\dagger}_{n\sigma}c_{m\sigma}-c^{\dagger}_{m\sigma}c_{n\sigma}, \quad \Delta_{nm,-\sigma}=n_{n-\sigma}-n_{m-\sigma},$$

$$g_{nm,\sigma}(u_{nm,\sigma})=\frac{c_{n\sigma}\pm c_{m\sigma}}{\sqrt{2}}. \qquad (4)$$

Here g and u denote the even and odd combinations of orbitals of neighboring centers. The operator Z_{nm} acts only on singlet components with two particles at neighboring sites in the full wave function Φ. It includes various configurations of singlet pairs of neighboring sites, similarly to the RVB state.[16] But in our model, unlike the RVB, the charges of sites connected by a bond are changed (optimized). The transformation (2) may be expected to be effective because, in the case of two sites with two holes localized on them, the wave function $\Psi(nm)=\exp(\alpha Z_{nm})\Phi$, which is derived from the uncorrelated function $\Psi=|g^{\dagger}_{\uparrow}g^{\dagger}_{\downarrow}\rangle$ (for $t<0$), is an exact singlet function of the system for optimal $\alpha=-0.5\tan^{-1}(U/4t)$.

Since the transformation in Eq. (2) is unitary, the initial Hubbard Hamiltonian in the basis of correlated states Ψ is exactly equivalent to the transformed Hamiltonian

$$\tilde{H}(\alpha)=W^{\dagger}(\alpha)HW(\alpha) \qquad (5)$$

in the basis of functions Φ. The variational parameter α in $W(\alpha)$ is, in essence, an order parameter for the structure of valence bonds. By analyzing the new problem (5) in the mean-field approximation, one can investigate states with correlations of this type for an arbitrary doping.

Unlike the periodic structures of valence bonds with isolated dimers discussed above,[19] in the case of a homogeneous state (2) the local operators Z_{nm} do not commute each other. Therefore we cannot determine the effective Hamiltonian in all orders in α. We can, however, find an exact expression for $\tilde{H}(\alpha)$ in terms of Fermi operators to within terms $\sim\alpha^2$:

$$\tilde{H}(\alpha)\approx H+\alpha[H,Z]+\frac{\alpha^2}{2}[[H,Z],Z]$$

$$=H^{(0)}+\alpha H^{(1)}+\frac{\alpha^2}{2}H^{(2)}, \qquad (6)$$

and derive a self-consistent solution of the new problem in the mean-field approximation. Unlike the case of an isolated dimer in a 2D lattice, each site is involved in four bonds. Therefore, as will be shown below, the optimal α in the transformation remains small even for large U/t ($\alpha\leqslant 0.22$ for $U/t\leqslant 8$), which enables us to use the expansion (6). In addition to the smallness of α, there is another circumstance which justifies the use of the Hamiltonian (6) which is second order in α. Specifically, the optimal α depends largely on U/t and changes little with doping. This means that for fixed α Hamiltonian (6) can be treated as a rough approximation, or, if you will, as an empirical Hamiltonian for a new model, which may provide better understanding of the real situation. The full expression for $\tilde{H}(\alpha)$ is rather lengthy.

Let us analyze in greater detail the contribution $H^{(1)}$ to the effective Hamiltonian (6) which is of first order in α. It is expressed in terms of fermion operators as

$$H^{(1)}=[H,Z]=H^{(1)}_U+T^{(1)}, \qquad (7)$$

$$H^{(1)}_U=-\frac{U}{2}\sum_{\langle nm\rangle,\sigma} t_{nm\sigma}\Delta^2_{nm,-\sigma}, \qquad (8)$$

$$T^{(1)}=t\sum_{\langle nm\rangle,\sigma}\Bigg\{[\Delta_{nm\sigma}\Delta_{nm,-\sigma}+j_{nm\sigma}j_{nm,-\sigma}]$$

$$+\sum_{m_i\in\langle mm_i\rangle}A(n,m,m_i,\sigma)$$

$$+\sum_{n_i\in\langle nn_i\rangle}A(m,n,n_i,\sigma)\Bigg\}. \qquad (9)$$

In the last two three-site terms we have $m_i\neq n$ and $n_i\neq m$, respectively. The operators $j_{nm\sigma}$ and $\Delta_{nm\sigma}$ are defined in Eq. (4), and the operators A and $t_{nm\sigma}$ are defined by the expressions

$$A(n,m,m_i,\sigma)=-\frac{1}{2}[t_{nm_i\sigma}\Delta_{mm_i-\sigma}+j_{nm\sigma}j_{nm_i-\sigma}], \qquad (10)$$

$$t_{nm\sigma}=(c^{\dagger}_{n\sigma}c_{m\sigma}+\mathrm{H.c.}). \qquad (11)$$

The term in Eq. (6) which is second order in α contains contributions from two, three, and four sites. Its exact expression is given in the Appendix [Eqs. (32)–(40)].

Let us investigate the most general class of uncorrelated BCS states with anomalous averages of the d-symmetry and a double magnetic unit cell in order to test the possibility of the antiferromagnetic spin ordering and $d_{x^2-y^2}$-type superconducting ordering. For a function Φ of this general form and an effective Hamiltonian (5), the mean energy

$\overline{H}(y_i)=\langle\Psi H\Psi\rangle=\langle\Phi\tilde{H}\Phi\rangle$ is calculated exactly. This is a function of the following one-electron normal and anomalous averages with respect to Φ:

$$\{y_i\}=\{r_0,r_1,r_{\sqrt{2}},r_2,r_{\sqrt{5}},r_3,d_0,d_{\sqrt{2}},d_2,w_1,w_2,w_{\sqrt{5}},w_3\}_i, \tag{12}$$

where

$$r_l=\frac{1}{2}\sum_{\sigma}\langle c^{\dagger}_{n\sigma}c_{n+l,\sigma}\rangle, \tag{13}$$

$$d_l=\frac{1}{2}\sum_{\sigma}\frac{\sigma}{|\sigma|}(-1)^n\langle c^{\dagger}_{n\sigma}c_{n+l,\sigma}\rangle, \tag{14}$$

$$w_l=\frac{1}{2}\sum_{\sigma}\frac{\sigma}{|\sigma|}\operatorname{sign}(l_x^2-l_y^2)\langle c^{\dagger}_{n\sigma}c^{\dagger}_{n+l,\sigma}\rangle=w_l^{*}. \tag{15}$$

Let us take into account all symmetry types of the studied state, namely the translational symmetry, the invariance with respect to reflection $x\rightarrow -x$ or $y\rightarrow -y$, the equivalence between the odd and even sublattices when $\sigma\rightarrow-\sigma$. Thus, the parameters defined by Eqs. (13)–(15) are real, in addition r_l, d_l, and $|w_l|$ are functions of $|l|$ only; $w_l=0$ at $l_x=\pm l_y$; $d_l=0$ for odd l_x+l_y. According to Eq. (15), only the sign in front of w_l depends on the direction of $\mathbf{l}$. Any $2n$-fermion operator averaged over Φ is exactly expressed in terms of the one-electron averages r_l, d_l, and w_l.

In calculating the mean energy, we took only terms of to the second order in the anomalous averages w_l by virtue of the smallness of both T_c and superconducting gap in comparison with the total band width $\sim t$ and the antiferromagnetic gap $\sim Ud_0$. As a result, the energy averaged over the most general uncorrelated state taking into acount the double magnetic cell and anomalous averages of the d-symmetry is expressed in terms of one-electron averages (12). The mean energy per site is

$$\overline{H}(y_i)=\mathscr{H}(r_i,d_i)+\mathscr{H}^{\mathrm{SC}}(w_i,r_i,d_i), \tag{16}$$

$$\begin{aligned}\mathscr{H}(r_i,d_i)=&\{U(r_0^2-d_0^2)-8tr_1\}+\alpha\{-8r_1U[r_0(1-r_0)\\&+d_0^2+r_1^2]+16t[d_0^2+2d_0d_{\sqrt{2}}+d_0d_2)]\}\\&+\frac{1}{2}\alpha^2\{\mathscr{H}_U^{(2)}+\mathscr{T}^{(2)}\},\end{aligned} \tag{17}$$

$$\mathscr{H}^{\mathrm{SC}}=\sum_{ij}k_{ij}w_iw_j=-8\alpha Ur_1w_1^2+\alpha^2\sum_{ij}k_{ij}^{(2)}w_iw_j. \tag{18}$$

Here we have shown explicitly only the terms of zeroth and first orders in α. The expressions for $\mathscr{H}_U^{(2)}$ and $\mathscr{T}^{(2)}$, which are of second order in α, and a full expression for $\mathscr{H}^{\mathrm{SC}}$ are given in Eqs. (42), (44), and (45)–(47) in Appendix. All arguments in $\overline{H}$ are given in (12). One can easily check that $\overline{H}$ is invariant with respect to each of replacements ($r_l\rightarrow\delta_{l,0}-r_l, t\rightarrow -t$), or $d_l\rightarrow -d_l$, or $w_l\rightarrow -w_l$ taken separately. The first of these corresponds to the electron–hole symmetry of the Hubbard model.

It is worthwhile to discuss in detail the origin of the terms with anomalous averages in Eq. (18). Consider the contribution of the term $\alpha H^{(1)}$ of first order in α to the effective Hamiltonian [Eqs. (7)–(9)]. The calculation indicates that terms with anomalous averages are absent in $\langle\alpha T^{(1)}\rangle$. This is quite natural because neither the transformation $W(\alpha)$ nor the superconducting ordering can lower the energy of noninteracting particles. But the average $\langle\alpha H_U^{(1)}\rangle$ is a function of anomalous order parameters. Specifically, expression (8) for the operator $H_U^{(1)}$ contains terms like

$$c^{\dagger}_{n\sigma}c_{m\sigma}(n_{n,-\sigma}+n_{m,-\sigma})-2c^{\dagger}_{n\sigma}c_{m\sigma}n_{n,-\sigma}n_{m,-\sigma}. \tag{19}$$

These terms describe versions of the correlated hopping interaction modulated by the populations of the lattice sites. The first term in Eq. (19) has the form of the interaction suggested by Hirsch.[22] The term with anomalous averages of this operator is zero,

$$\langle c_{n\sigma}c_{m\sigma}(n_{n,-\sigma}+n_{m,-\sigma})\rangle=-2w_0w^{(s)}_{|n-m|},$$

since the s-type superconductivity is forbidden ($w_0=\langle c^{\dagger}_{n\uparrow}c^{\dagger}_{n\downarrow}\rangle=0$). The occurrence of an s-symmetry superconducting ordering ($w_0\neq 0$) in the desired state Φ would increase the energy by $\Delta\overline{H}=Uw_0^2$ per lattice site. The contribution of d-symmetry anomalous parameters is due to operators like the second term in Eq. (19). This decisive contribution is defined explicitly by Eq. (18). In reality, we have used in our calculations the exact expression (45)–(47) for the contribution of anomalous averages to the mean energy.

The presence of the term linear in α in $\overline{H}$ indicates that a minimum of $\overline{H}(\alpha)$ corresponds to a nonzero value of α. We can determine signs of some parameters for $t>0$ in Eq. (1) using Eqs. (16)–(18):

$$tr_1>0,\quad r_1>0,\quad \alpha r_1>0,\quad \alpha>0,\quad k_{11}\simeq-8\alpha r_1U<0. \tag{20}$$

Thus, the requirement that the energy should be lowered as a result of formation of valence bonds determines the sign of the transformation parameter, which gives rise to the minus sign in front of the main constant k_{11} in the d-symmetry superconducting ordering in Eq. (18), hence implies the possibility of d-symmetry superconductivity.

The one-determinant uncorrelated function Φ which minimizes the energy $\overline{H}-\mu\overline{N}$ is a product of one-particle eigenfunctions $\chi^{\dagger}_{k\lambda\sigma}$ of the linearized Hamiltonian

$$\tilde{H}_L=\sum_{\nu}\frac{\partial\overline{H}}{\partial y_{\nu}}(\hat{y}_{\nu}-y_{\nu})+\overline{H}(y_i)=\sum_{k}{}^{F}\hat{h}_k+\text{const.} \tag{21}$$

Operators $\hat{y}_i$, which correspond to the averages y_i [Eqs. (12)–(15)] are given by

$$\begin{aligned}\hat{r}_l&=\frac{1}{2n_l}\sum_{l,\sigma}c^{\dagger}_{n,\sigma}c_{n+l,\sigma},\\ \hat{d}_l&=\frac{1}{2n_l}\sum_{l,\sigma}(-1)^n\frac{\sigma}{|\sigma|}c^{\dagger}_{n,\sigma}c_{n+l,\sigma},\\ \hat{w}_l&=\frac{1}{4n_l}\sum_{l,\sigma}\operatorname{sign}(l_x^2-l_y^2)\frac{\sigma}{|\sigma|}[c^{\dagger}_{n,\sigma}c^{\dagger}_{n+l,-\sigma}+\text{H.c.}].\end{aligned} \tag{22}$$

Here n_l is the number of all vectors $\mathbf{l}$ of length $l=|\mathbf{l}|$, over which the sum is performed.

As a result, the linearized Hamiltonian (21) in the momentum operator basis

$$b^{\dagger}_{k,j}=\{c^{\dagger}_{k\uparrow},c^{\dagger}_{\tilde{k}\uparrow},c_{-k\downarrow},c_{-\tilde{k}\downarrow}\}_j, \quad j=1,\ldots 4, \quad \tilde{k}=(\pi,\pi)+k, \tag{23}$$

is expressed as

$$H_L=\sum_k{}^F \hat{h}_k+\text{const}, \quad \hat{h}_k=\sum_{i,j=1}^{4} L_{ij}b^{\dagger}_{ki}b_{kj}. \tag{24}$$

Here we have set $\tilde{k}=k+(\pi,\pi)$, and the superscript F on the summation sign indicates that the sum over k is performed inside the magnetic Brillouin zone $|k_x\pm k_y|<\pi$. The matrix L_{ij} is defined by the formula

$$L_{ij}=\begin{pmatrix} \epsilon_k-\mu & \Delta_k & W_k & 0 \\ \Delta_k & \epsilon_{\tilde{k}}-\mu & 0 & W_{\tilde{k}} \\ W_k & 0 & -(\epsilon_k-\mu) & \Delta_k \\ 0 & W_{\tilde{k}} & \Delta_k & -(\epsilon_{\tilde{k}}-\mu) \end{pmatrix}, \tag{25}$$

$$\epsilon_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial r_l}g_l(k), \quad \Delta_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial d_l}g_l(k),$$

$$W_k=\frac{1}{2}\sum_l \frac{\partial\overline{H}}{\partial w_l}q_l(k). \tag{26}$$

In the sums over l for ϵ_k, Δ_k, and W_k, the subscript l runs through all r_l, or d_l, or w_l of the full set (12) of one-electron averages, which determine $\overline{H}$. Equations (25) and (26) include the following notations:

$$g_l(k)=\frac{1}{n_l}\sum_l \cos k_x l_x \cos k_y l_y,$$

$$q_l(k)=\frac{1}{n_l}\sum_l \operatorname{sign}(l_x^2-l_y^2)\cos k_x l_x \cos k_y l_y. \tag{27}$$

Here n_l is the number of all vectors $\mathbf{l}$ of length $l=|\mathbf{l}|$, over which the sum is performed in Eq. (27), and the functions defined by Eq. (27) have the following symmetry properties:

$$g_l(\tilde{k})=(-1)^{l_x+l_y}g_l(k), \quad q_l(\tilde{k})=(-1)^{l_x+l_y}q_l(k).$$

The one-electron eigenfunctions $\chi_{k\lambda}$ and spectrum $\mathscr{E}_\lambda(k)$ of the linearized Hamiltonian are calculated by diagonalizing the matrix (25):

$$\chi^{\dagger}_{k\lambda}=\sum_j b^{\dagger}_{kj}S_{j,\lambda}, \quad \sum_j L_{ij}S_{j,\lambda}=S_{i,\lambda}\mathscr{E}_\lambda, \tag{28}$$

for each k within the magnetic Brillouin zone. Given the eigenfunctions and spectrum, we can complete the self-consistent procedure, i.e., calculate the desired averages $y_i=\langle\hat{y}_i\rangle_\Phi$ by the formulas

$$r_l=\frac{1}{2N}\sum_k{}^F [g_l(k)(U_{11}+1-U_{33})+g_l(\tilde{k})(U_{22}+1-U_{44})],$$

$$d_l=\frac{1}{2N}\sum_k{}^F g_l(k)[U_{12}+U_{21}+U_{34}+U_{43}],$$

$$w_l=\frac{1}{2N}\sum_k{}^F [q_l(k)(U_{13}+U_{31})+q_l(\tilde{k})(U_{24}+U_{42})], \tag{29}$$

where

$$U_{ij}=\sum_\lambda S^{*}_{i\lambda}S_{j\lambda}f_F(\mathscr{E}_\lambda/kT),$$

and $f_F(x)$ are Fermi distribution functions; g and q are defined in Eq. (27).

The self-consistent procedure described above minimizes the energy with respect to Φ. Then minimization with respect to α yields the desired variational correlated state (2) and the optimal effective Hamiltonian (6) in the basis on uncorrelated states. Self-consistent calculations of two types, which were matched to one another, were performed: the complete calculation involving all normal and anomalous averages, which yielded the superconducting gap $2\Delta_0(U,\delta,T)$ for states of valence bonds with both the antiferromagnetic and superconducting ordering (AF + VB + SC), and the calculation of the critical temperature $T_c(U,\delta)$ based on equations linear in w_l using similar states (AF + VB) without anomalous averages. In the latter case, the critical temperature of the superconducting transition was obtained using the standard equations of the perturbation theory in terms of incipient anomalous order parameters:

$$\operatorname{Det}|D_{ij}-\delta_{ij}|=0, \quad D_{ij}=\partial w_i/\partial w_j|_{w_l=0}. \tag{30}$$

The matrix D_{ij} was calculated using Eqs. (48) and (49) given in Appendix.

An independent test of the procedure (or the basis for an alternative iteration procedure) is the relationship $\alpha=-\overline{H}^{(1)}(y_i)/\overline{H}^{(2)}(y_i)$ between the optimal parameter α and the averages of the contributions to the effective Hamiltonian (6) with self-consistent values of y_i. The coincidence of all physical parameters for equal levels of p- and n-doping was also checked. In the presence of antiferromagnetic ordering $(d_0\neq 0)$, the results of numerical diagonalization of matrix (28) and the approximate solution for S_{ij} [Eq. (52) in Appendix] coincide. This means that states of only the lower (upper) Hubbard subband participate to any great extent in superconducting pairing in the case of $p(n)$-doping.

3. DISCUSSION OF RESULTS

The region $U/t\leqslant 9$ studied in the present work is limited by the condition that the optimal parameter α should be small. The description of the CuO_2 plane based on the single-band Hubbard model suggests that $U/t\sim 8$.[14,15] In this region, valence bond correlations actually lead to a lower system energy and the homogeneous state of valence bonds is lower than the similar state with a periodic dimer structure, and both these states are compatible with antiferromagnetism.

The characteristic energy effect is shown by Fig. 1 for a system with $U/t=8$. The points in this graph are results of

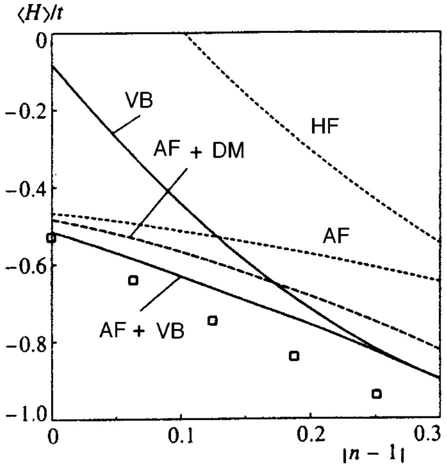


FIG. 1. Average energies per lattice site for antiferromagnetic and paramagnetic states with valence bond correlations (the solid curves AF + VB and VB), and for similar uncorrelated Hartree–Fock states (curves AF and HF, respectively). The dashed curve AF + DM corresponds to a periodic dimer structure of valence bonds. The squares plot the results of exact Hamiltonian diagonalization for a 4×4 cluster.[22]

an exact diagonalization for 4×4 clusters.[23] The main negative contribution to the energy difference between the simple Hartree–Fock paramagnetic state and the Hubbard state in the limit $n\rightarrow1$ is due to the alternation of spin orientations. The region of antiferromagnetic ordering on the diagram plotted in coordinates of interaction constant and doping is shown in Fig. 2a. The critical doping $\delta_c=|n_c-1|$ at which the antiferromagnetic ordering disappears is $\delta_c\sim0.26$–0.3 for $U/t=6$–8. These values are lower than the result of the generalized Hartree–Fock method without valence bond correlations, $\delta_c\sim0.4$–0.45, but they are higher than the critical doping $\delta_c\sim0.05$ at which antiferromagnetic ordering in available HTSC crystals is destroyed. The value $\delta_c\sim0.05$ in the diagram of Fig. 2a would corresponds to a very small parameter $U/t\leqslant2$.

Several explanations of this discrepancy can be proposed.

The simplest hypothesis suggests that the region of antiferromagnetic ordering is too large because the Hubbard model does not take into account those interactions that break perfect nesting. Meanwhile, the presence of the hopping interaction t' between non-nearest neighbors in cuprates is supported by various investigations, in particular, by details of the one-band mapping[14] of the Emery three-band model, by empirical tight-binding models,[24] which reproduce the shapes of the Fermi surface and energy band obtained from the ARPES data, and by calculations of levels in finite clusters based on the $t-t'-J$ model and LDA calculations.[25] The empirical value[24,25] of the diagonal hopping parameter $t'/t\sim0.2$ is larger than theoretical estimates[14] and is different for different HTSC ceramics.[26] Note, however, that the empirical values were obtained by fitting the ARPES data to a single band without a splitting between the upper and lower Hubbard subbands. Under the conditions of antiferromagnetic splitting and flat bands, the sensitivity of the Fermi surface shape and other characteristics to t' is greater, so the actual values of t'/t may be lower. Our calculations, however, indicate that the interaction t' has little effect on the doping corresponding to the boundary between the antiferromagnetic and paramagnetic states.

One might think that the mean-field approximation [in this case, applied to the effective Hamiltonian $\widetilde{H}(\alpha)$] cannot describe antiferromagnetic correlations with a large but finite correlation length. Meanwhile, the slave-boson technique yields two phases of spin ordering near the transition to the paramagnetic state, one with a short-range and another with a long-range antiferromagnetic ordering,[27,28] and the boundary of the true paramagnetic state is very close to that shown in Fig. 1. There are other independent arguments in favor of the hypothesis of a large area of antiferromagnetic spin ordering in separate CuO_2 planes, in contrast to the narrow area of bulk antiferromagnetism observed in experiments.[10,29] The most spectacular facts are the transformation of the Fermi surface to a small Fermi surface in the underdoped region of $Bi_2Sr_2CaCu_2O_{8+\delta}$ detected by the ARPES technique[6] and observation of ''shadow'' pieces of the Fermi surface centered around the point $\Gamma(0,0)$ using the scanning version of the ARPES technique.[30] There are other arguments in favor of a wide area of antiferromagnetic ordering in CuO_2 planes,[31] and they have been used as a basis for the most recent superconductivity theory.[32]

In the mean-field approximation, we can calculate neither the radius R_{AF} of the area in which the antiferromagnetic quantization axis has a constant direction, nor the dynamics of fluctuations of this direction. But the energy parameters of the process (see Fig. 1) lead us to a conclusion that this

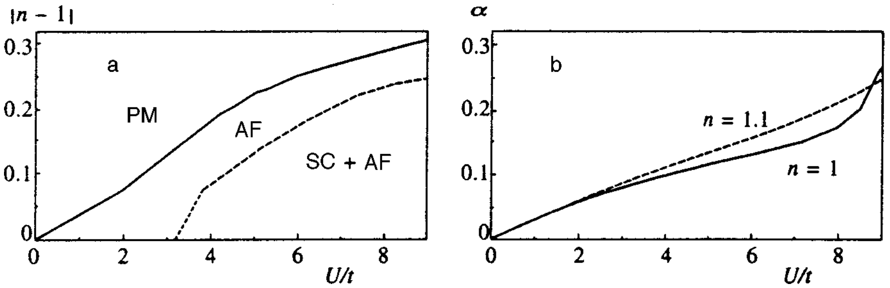


FIG. 2. (a) Phase diagram for solutions with valence bond correlations in the''doping versus interaction U/t'' plane. The solid curve corresponds to the critical doping, at which antiferromagnetic ordering disappears. The dashed line is the boundary of the existence of an anisotropic superconducting gap calculated at $kT=0.002t$. The lower part of this line is shown schematically (short dashes) because of the small gap width and poor convergence near T_c. (b) Optimal transformation parameter α versus U/t for an undoped system ($n=1$) and at a doping $|n-1|=0.1$ (solid and dashed lines, respectively).

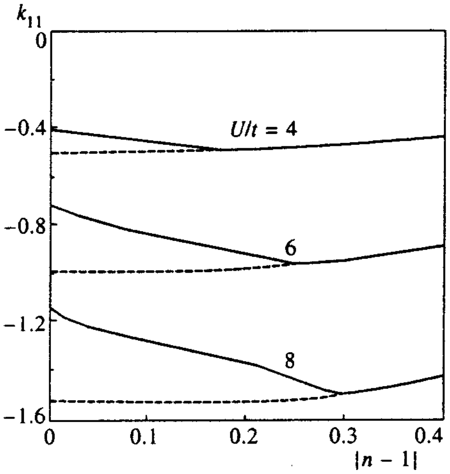


FIG. 3. The main constant k_{11} of the d-symmetry superconducting ordering in the average of the effective Hamiltonian (10), (11) versus doping for antiferromagnetic and paramagnetic states of valence bonds at $U/t=4$, 6, and 8. The dashed lines correspond to paramagnetic solutions in regions where the antiferromagnetic states has a lower energy.

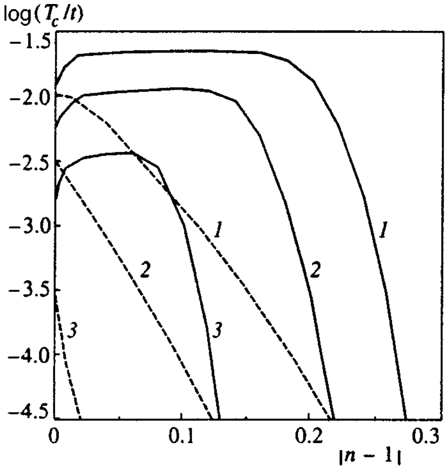


FIG. 4. Logarithm of the critical temperature T_c versus doping. The solid and dashed lines correspond to the antiferromagnetic and paramagnetic states of valence bonds, respectively. Curves 1, 2, and 3 correspond to $U/t=8$, 6, and 4.

radius is much larger than the lattice constant ($R_{\mathrm{AF}} \gg a$). Therefore, from the standpoint of short-range valence bond-correlations and attraction between holes due to formation of valence bonds between neighboring sites, we cannot endorse the hypothesis of the reality of ''two-dimensional antiferromagnetism'' in CuO_2 planes over a wide range of doping, contrary to the assumption about the narrow region of bulk antiferromagnetism in cuprates.

Figure 2b shows characteristic values of the optimal transformation parameter α. It depends largely on U/t and changes little with n. The curve of $\alpha(U)$ indicates that the expansion (6) can be used down to $U/t \leqslant 8$.

Now let us proceed to the main task of this work, namely, the investigation of the feasibility of d-symmetry superconducting ordering. Given the fast drop of anomalous averages w_l with increasing l, even the sign and absolute value of factor k_{11} in the main contribution $[k_{11}w_1^2]$ to the superconducting component (18) of the average energy (16) are quite significant. Figure 3 shows the factor k_{11} as a function of the doping for several values of U/t for antiferromagnetic or paramagnetic states of valence bonds. The data for paramagnetic states, whose energy is considerably higher, are given only to demonstrate that the cause of attraction between holes ($k_{11}<0$) is not long-range antiferromagnetic correlations, but valence bond correlations. In particular, $k_{11}=0$ for $\alpha=0$. In paramagnetic states, the factor k_{11} is also negative, and its absolute value is even larger. Nonetheless, owing to the difference in the densities of states on the Fermi surface, only in the antiferromagnetic state is there a sufficiently wide range of the doping in which the system has a wide superconducting gap and high critical temperature T_c.

The critical temperature $T_c(U,\delta)$ of the superconducting transition is determined using Eq. (30) as a point at which anomalous averages w_l occur superposed on the normal state AF + VB. The T_c value coincides with temperature at which the superconducting gap $2\Delta(T)$ and anomalous averages in the self-consistent calculations turn to zero. Figure 4 shows the logarithm $\log T_c(n)$ as a function of the doping for $U/t=8$ and 6. The entire domain of the superconducting state is within the antiferromagnetic ordering region, i.e., below the boundary $\delta_c(U/t)$ between the states AF + VB and PM + VB in Fig. 2a.

At the same time, similar functions $\log T_c(n)$ for paramagnetic states drop rapidly as $T_c \to 0$ even at low doping. The cause is that the density of one-electron states of the corresponding effective linearized version of the problem is high only in a narrow region about the Van-Hove singularity at $k=(0,\pm\pi)$, $(\pm\pi,0)$. On the contrary, in antiferromagnetic states the splitting of the initial band into two subbands leads to a considerable widening of the region with a high density of states near the point $(0,\pi)$. As a consequence, the region of high critical temperatures also broadens (see the plateau on curves of Fig. 4).

There is a problem that cannot be solved in the mean-field approximation, namely, whether there is, in addition to the antiferromagnetic doubling of the unit cell, an alternative mechanism of splitting the initial band into upper and lower Hubbard subbands, which could widen flat parts of the spectrum near the Fermi level. Essentially, the same problem had to be solved in discussing the difference between boundaries of the ''two-dimensional'' antiferromagnetic ordering in our calculations and the bulk antiferromagnetism in real cuprates.

Let us reconsider properties of the superconducting state with the $d_{x^2-y^2}$ symmetry due to valence bond correlations. Figure 5 shows the superconducting gap $2\Delta(T)$ as a function of temperature for several model parameters. The ratio $\xi = 2\Delta(0)/kT_c$ ranges between 3.9 and 4.5 for $U/t=8$ instead of the BCS value 3.5. Our calculations of ξ are smaller than $\xi=2\Delta(0)/kT_c \sim 10-12$ obtained using approximations like self-consistent techniques of spectral functions[33–35] or Green's functions.[36] The ratio ξ measured in various experiments is highly anisotropic and depends on the z-component of quasimomentum, whereas the value for the ab-plane of cuprates varies in the range $\xi \sim 5-7$.[29,37]

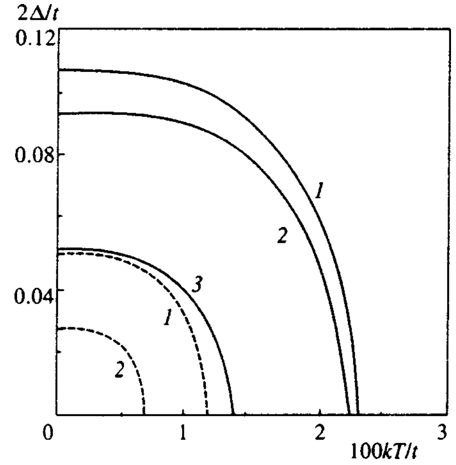


FIG. 5. Superconducting gap $2\Delta(T)$ versus temperature at $U/t=8$ (solid curves) and $U/t=6$ (dashed curves). Curves 1, 2, and 3 correspond to doping values $|n-1|=0.1$, 0.15, and 0.2.

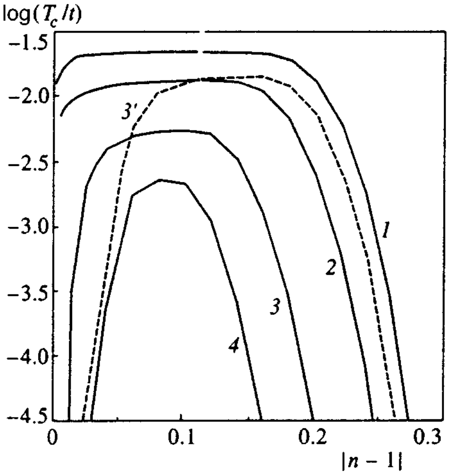


FIG. 6. Logarithm of the critical temperature T_c versus doping for the extended Hubbard model (31) at $U/t=8$, $t'=0$, and various values of parameter V characterizing interaction between particles at neighboring sites. Solid curves 1, 2, 3, and 4 correspond to $V/t=0$, 0.1, 0.2, and 0.25. The dashed line shows calculations for $U/t=10$, $V/t=0.2$, and $t'=0$.

Thus, according to mean-field calculations of $\tilde{H}(\alpha)$, the region of the d-symmetry superconductivity is within the region of the ''two-dimensional'' (or hidden) antiferromagnetic spin ordering determined in the same approximation. As was noted above, in our approach, this region projected to real objects can become a region of long-range spin correlations, but not necessarily of a real long-range ordering. The dashed line in Fig. 2a is the boundary of the superconductivity region in terms of the doping at temperature $kT=0.002t$. Note that at $U/t=8$ the maximum critical temperature is $kT_c=0.023t$. The corresponding gap width is $2\Delta(0)=0.107t$, which makes $2\Delta(0)=53$ meV and $kT_c=133$ K for $t\sim0.5$ eV (Ref. 14).

An unexpected result of our calculations is the large width of the anisotropic d-symmetry superconducting gap extending down to very small values $\delta=|n-1|\sim0.03$–0.04 for $U/t=6$–8. In the initial Hubbard model, the transition to the dielectric state with $T_c=0$ as $n\rightarrow1$ occurs only for very small doping, when the chemical potential is slightly higher than the edge of the lower Hubbard subband and $\delta=|n-1|$ is determined by the tail of the distribution function. In the region $\delta>0.05$ the values T_c and $2\Delta(0)$ are constant or even increase with decreasing δ. This contradicts the phase diagram of real cuprates,[10] where $T_c(\delta)$ drops rapidly in the range of doping smaller than the optimal parameter δ_{opt} corresponding to the maximum of T_c.

This contradiction can be interpreted in various ways.

One can appeal to the arguments of Refs. 38 and 39. They suggested the presence of states with nonzero anomalous averages and anisotropic gaps, but without superconducting properties owing to quantum fluctuations. This hypothesis was put forth in attempts to reconcile the absence of superconductivity and the presence of an anisotropic pseudogap in the normal state of underdoped samples. It seems, however, more constructive to search for real interactions that suppress superconducting ordering at low doping.

Suppose that we have come to this discrepancy as a result of using an ideal model (1), more specifically, the initial Hubbard model did not take into account interactions disrupting the perfect nesting. For example, we did not take account of hopping interaction t' between non-nearest neighbors (diagonal neighbors). We also neglected the similar Coulomb interaction between particles at different sites. Our attention is focused on the t' interaction for several reasons. First, it does not disrupt the perfect nesting in underdoped systems, but has a considerable effect on the Fermi surface shape owing to the very flat bands. Second, the value and even the sign of t' depend on the composition of the material because of the competition between two channels of such diagonal hoppings in the CuO_2 plane, namely, through direct interaction t_{pp} with oxygen orbits and through the process of second order in the $p-d$ hybridization. Therefore comparative analysis of the effect of t' in different cuprates can be of great importance.[26]

In connection with the arguments given above, we have performed self-consistent calculations of the phase diagram and several system characteristics with an extended effective Hamiltonian:

$$H_{\text{eff}}(\alpha)=\tilde{H}(\alpha,U,t)+V\sum_{\langle nm\rangle}n_n n_m+t'\sum_{\langle\langle nm\rangle\rangle}\sum_{\sigma}(c_{n\sigma}^{\dagger}c_{m\sigma}+\text{H.c.}),\tag{31}$$

i.e., we added to the initial effective Hamiltonian (6) the Coulomb-like interaction V between the nearest neighboring cells and hopping interaction t' between the nearest diagonal sites $\langle\langle nm\rangle\rangle$ with $|n-m|=\sqrt{2}$. For simplicity, we have included only terms of zeroth order in α for these interactions in our approximate estimates.

Figure 6 plots T_c as a function of the doping for $V=(0$–$0.25)t$. The Coulomb interaction between neighboring centers really destroys the superconducting ordering at low doping and at the same time decreases the maximum critical temperature T_c. In particular, at $V=0.2t$, $U/t=8$, and $|n-1|=0.1$, the critical temperature drops to $T_c=0.0057t$. Assuming $t\sim0.5$ eV,[14] we have $T_c\sim30$ K, which

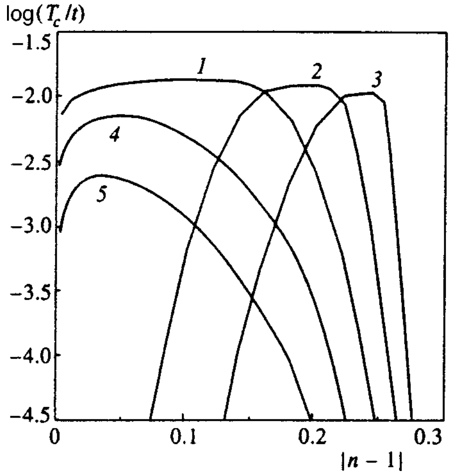


FIG. 7. Effect of parameter t' on the phase diagram $T_c(\delta)$ for the extended model (31) at $U/t=8$ and $V/t=0.1$. Curves 1, 2, and 3 correspond to $t'/t=0$, 0.05, and 0.1, curves 4 and 5 to negative values of $t'/t=-0.05$ and -0.1.

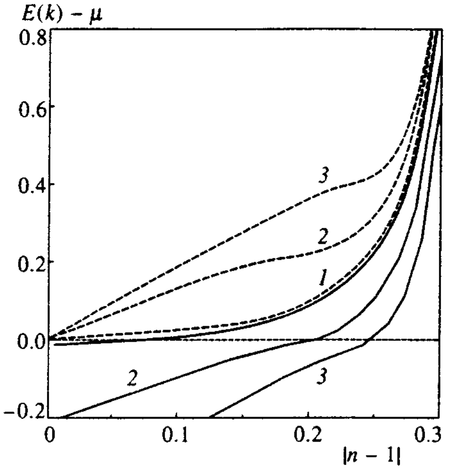


FIG. 8. Band energy $E(k)-\mu$ measured with respect to the chemical potential at two point of the magnetic Brillouin zone boundary, namely $k=(\pi,0)$ (solid curves) and $k=(\pi/2,\pi/2)$ (dashed lines) for the model with $U/t=8$, $V/t=0.1$, and various t'. Curves 1, 2, and 3 correspond to $t'/t=0$, 0.05, 0.1, and phase curves 1, 2, and 3 in Fig. 7.

is lower than the experimental data $T_c^{\max}\sim 100\,\mathrm{K}$. The latter value could be obtained for $U/t\sim 10$ and $V/t=0.2$ (see the dashed curve in Fig. 6), but with these parameters the transformation parameter $\alpha\sim 0.27$ is not small. The decrease in T_c with V is understandable. A positive constant ($\Delta k_{11}=4V$) is added to the main superconducting parameter, i.e., factor k_{11} in front of w_1^2 in the expression for the average energy. The destruction of the superconducting ordering at low doping can be attributed to the weakening of the correlation interaction constant $k_{11}\approx -8\alpha U r_1+4V$ with decreasing r_1 when antiferromagnetic localization of holes takes place as $n\rightarrow 1$.

Figure 7 demonstrates a strong effect of diagonal hopping interaction t' on the shape of $T_c(\delta)$. This effect is clearly seen when the sign of t' is varied.

For $t'>0$ the curve of $T_c(\delta)$ is like the curve of experimental data, namely, T_c drops rapidly on both sides of the optimal doping δ_{opt}. As t' increases, the peak position δ_{opt} shifts toward higher values without changing the maximum value $T_c(\delta_{\mathrm{opt}})$. This behavior is understandable if we take into account that for $t'\neq 0$ the energy $E_1(k)$ of the lower Hubbard subband is not constant along the boundary $|k_x\pm k_y|=\pi$ of the magnetic Brillouin zone:

$$E_1(k)-E_1(\pi,0)=4t'(\cos k_x \cos k_y+1).$$

Note that, although the boundary of the Brillouin zone in dielectric undoped materials is just a line of the minimum dielectric gap, it is seen in ARPES spectra as a blurred Fermi boundary since the intensities of photoemission in the first and extended Brillouin zones are different, notwithstanding the equal energies of bands after the antiferromagnetic doubling of the unit crystal cell.

It is remarkable that, in the case of $t'>0$ and *p*-doping, the flattest part of the lower Hubbard subband crosses the Fermi surface at the optimal doping (at the maximum T_c):

$$E_1(\pi,0)-\mu=0 \quad \text{at} \quad \delta=\delta_{\mathrm{opt}}, \qquad n<1.$$

Figure 8 shows the band energy $E_1(k)-\mu$ with respect to the chemical potential at two points $k=(\pi,0)$ and $k=(\pi/2,\pi/2)$ on the boundary of the magnetic Brillouin zone. At $t'=0$ these energies are almost equal, so the ''large'' Fermi surface centered at $\Gamma(0,0)$ and the corresponding ''shadow'' piece about $Y(\pi,\pi)$ emerge as a result of doping. For $t'\neq 0$ the evolution of the Fermi surface with doping is radically different. For $t'>0$ we have

$$E_1(\pi/2,\pi/2)-E_1(\pi,0)\sim 4t'>0.$$

This means that for low doping $\delta<\delta_{\mathrm{opt}}$ the ''small'' Fermi surface emerges in the form of pockets about the point $(\pi/2,\pi/2)$, whereas the part of the spectrum for $k\sim(\pi,0)$, where the density of states is the highest, are below the Fermi level. Only at the optimal doping, $\delta=\delta_{\mathrm{opt}}$, the spectrum crosses the Fermi level at $(\pi,0)$, after which the ''large'' Fermi surface emerges. Given that parts of the magnetic Brillouin zone boundary, i.e., the dielectric parts, are seen in the ARPES experiments as blurred Fermi boundaries, the ARPES data can be interpreted at low doping in terms of a generalized ''large'' Fermi boundary composed of the dielectric parts of the boundary and real (non-shadow in terms of intensity) metallic parts of the ''small'' Fermi surface.

In this interpretation, the pseudogap detected in recent years[7,8] in the normal state of underdoped samples of $\mathrm{Bi_2Sr_2CaCu_2O_{8+\delta}}$ is nothing but the energy $|E_1(\pi,0)-\mu|$ needed to eject electrons from parts $k\sim(\pi,0)$, where the density of states is the highest. In underdoped samples, these parts of the spectrum are below the Fermi level, hence they are populated. In overdoped samples, $\delta>\delta_{\mathrm{opt}}$, these parts of the spectrum are above the Fermi level and unpopulated. More accurate calculations and interpretation of photoemission spectra should become goals of a dedicated investigation.

Systems with the opposite sign of interaction, $t'<0$, demonstrate a radically different behavior of the phase diagram $T_c(\delta)$ (Fig. 7) and Fermi boundary. The region of the *d*-symmetry superconductivity remain sufficiently wide with

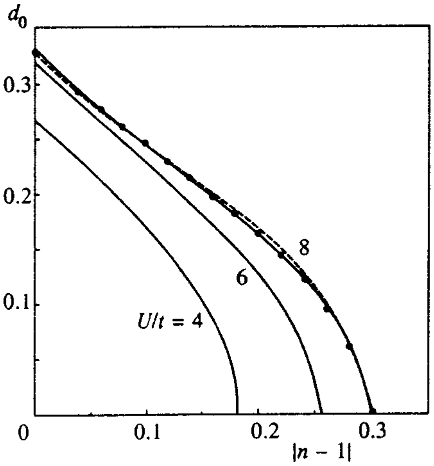


FIG. 9. Alternating spin density d_0 as a function of doping for $U/t=4$, 6, 8, $V=0$, and $t'=0$ (solid curves). The dashed curve and points correspond to $U/t=8$, $V/t=0.1$, $t'/t=0.1$ and -0.1.

the increase in $|t'|$ and shifts to the region of low doping. At the same time, the maximum critical temperature decreases with increasing $|t'|$. This shift of the superconducting region and decrease in the maximal T_c for $t'<0$ demands an interpretation. The simplest hypothesis is that the region of the ''two-dimensional'' antiferromagnetism becomes more narrow, consequently, the superconducting region shifts. This assumption, however, is not true: actually, the region of antiferromagnetic ordering changes little as t' varies between -0.1 and 0.1. This can be seen in Fig. 9, which shows results for the alternating spin density d_0 in the case of self-consistent AF + VB solutions. Therefore, let us turn again to features of Hubbard subbands and Fermi boundaries.

For $t'<0$, the part of the spectrum in the lower Hubbard subband with the smallest slope, $k\sim(\pi,0)$, does not cross the Fermi level, but only approaches it (Fig. 10). The increase in the slope $\partial E_1(k,\delta)/\partial\delta|_{k=(\pi,0)}$ with $|t'|$ leads to a decrease in density of states and T_c. Figure 10 also demonstrates that for $t'<0$ and a certain doping $\delta^*=\delta^*(t')$ the spectrum crosses the Fermi level at the point $k=(\pi/2,\pi/2)$. But the region near this point does not contribute to the $d_{x^2-y^2}$-symmetry superconducting ordering since the gap is zero on lines $k_x=\pm k_y$. Therefore the parameter δ^* at which $E_1(k)-\mu=0$ at the point $k=(\pi/2,\pi/2)$ is in no way related to the maximum of T_c for $t'<0$. The Fermi surface also has a radically different shape fornegative t'. At low doping $\delta<\delta^*$, the generalized Fermi boundary consists of the dielectric parts of the magnetic Brillouin zone boundary about $k\sim(\pi/2,\pi/2)$ and metallic parts of the ''small'' Fermi boundary around points $(\pi,0)$. The results discussed in this paper are concerned with the state with a definite structure of valence bonds, in particular, with valence bonds between neighboring sites, which leads to superconductivity with the $d_{x^2-y^2}$ symmetry. For $t'>0$ this model provides a reasonable interpretation of the phase diagram $T_c(\delta)$ and anisotropic pseudogap. For $t'<0$, however, the calculated and measured phase diagrams are different, although there is a good reason to suppose that this case is realized in $La_{2-x}Sr_xCuO_4$. This is supported by the interpretation of inelastic neutron scattering for $\omega\to0$.[40,41] This discrepancy raises the question of whether alternative structures of valence bonds and/or other symmetry types of superconducting pairing are possible.

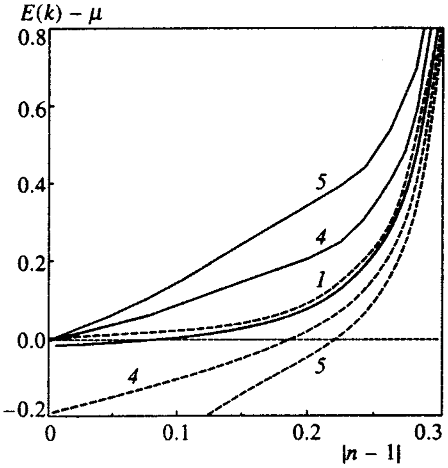


FIG. 10. Same as in Fig. 8, but with $t'\leqslant0$: curves 1, 4, and 5 correspond to $t'/t=0$, 0.05, -0.1, and phase curves 1, 4, and 5 in Fig. 7.

4. CONCLUSIONS

In conclusion, note the following.

1. The unitary transformation taking into account valence bond correlations yields a model Hamiltonian in which hoppings are modulated by site population, similarly to the Hirsch model.[22] It is this modulation of hoppings which leads to attractive interaction between holes in the *d*-channel, but, unlike the Hirsch model,[22] the interaction parameters are calculated by variational techniques. In contrast to the $t-J$ model, which excludes double filling of sites, the resulting effective Hamiltonian can be processed in the mean-field approximation.

2. The gain in the energy at low doping leads to doubling of the unit cell owing to alternation of spin orientations, and also to formation of valance bonds between neighboring sites. The problems of the small difference between the calculated energy and the results of simulations of small systems, and of the efficiency of other possible correlation types not included in our model have remained unsolved.

3. The mechanism of attraction between holes due to modulation of hoppings by site population cannot be reduced to exchange of antiferromagnetic fluctuations, which has been discussed intensely[42,43] as a cause of this attraction. The effective attraction constants have been calculated variationally and found to be close for both antiferromagnetic and paramagnetic states of valence bonds. At the same time, it turned out that the density of states near the Fermi level is high enough to produce substantial values of the superconducting gap and critical temperature $[2\Delta(0)\approx4.5$ and $kT_c\leqslant0.023t$ at $U/t=8]$ only in the case of antiferromagnetic splitting of the initial energy band.

4. In the classical Hubbard model, $H(U,t)$, the $d_{x^2-y^2}$-symmetry superconducting ordering occurs over a wide range of doping extending to $n\to1$ within the range of ''two-dimensional'' or hidden antiferromagnetism. This fact

is in contradiction with the phase diagram of real cuprates.

5. The calculations based on the generalized Hubbard model have shown that the phase diagram $T_c(\delta)$ and Fermi surface shape are very sensitive to the magnitude and sign of the hopping interaction t' between the non-nearest neighbors. The optimal doping δ_{opt} for $t'>0$ is the doping at which the flattest part of the lower Hubbard subband crosses the Fermi level for $k\sim(\pi,0)$. In this case, the anisotropic pseudogap of the normal state in underdoped samples should be associated with the energy difference $|E(\pi,0)-\mu|$ between the lower subband at $k\sim(\pi,0)$ and the Fermi level.

The reported investigation has been possible thanks to support from the Russian Fund for Fundamental Research (projects 7-03-33727A and 96-15-97492). The authors are indebted to V. Ya. Krivnov for helpful discussions and R. O. Zaĭtsev for constructive and useful criticism.

APPENDIX A:

With the effective Hamiltonian (6) the operator $H^{(2)}$ at $\alpha^2/2$ is

$$H^{(2)}=H_U^{(2)}+T^{(2)}, \tag{32}$$

$$H_U^{(2)}=-U\sum_{\langle nm\rangle}\left\{R(n,m)+\sum_{m_i}Q(n,m,m_i)+\sum_{n_i}Q(m,n,n_i)\right\}. \tag{33}$$

In the latter three-site terms, the sums are performed over $m_i\in\langle mm_i\rangle$, $m_i\neq n$, and $n_i\in\langle nn_i\rangle$, $n_i\neq m$, and the operators R and Q are defined by

$$R(n,m)=\Delta_{nm\uparrow}\Delta_{nm\downarrow}+j_{nm\uparrow}j_{nm\downarrow}, \tag{34}$$

$$\begin{aligned}Q(n,m,m_i)=-\frac{1}{4}\sum_{\sigma}\{&t_{nm_i\sigma}\Delta^2_{nm,-\sigma}\Delta_{mm_i,-\sigma}\\&+(c^\dagger_{n\sigma}c_{m\sigma}c^\dagger_{m,-\sigma}c_{m_i,-\sigma}+\text{H.c.})\\&-t_{nm\sigma}t_{mm_i,-\sigma}(n_{n,-\sigma}+n_{m_i\sigma}-2n_{n,-\sigma}n_{m_i\sigma})\},\end{aligned} \tag{35}$$

$$T^{(2)}=-t\sum_{\sigma,\langle nn'\rangle}\{[2c^{\dagger(1)}_{n\sigma}c^{(1)}_{n'\sigma}+c^{\dagger(2)}_{n\sigma}c_{n'\sigma}+c^\dagger_{n\sigma}c^{(2)}_{n'\sigma}]+\text{H.c.}\}, \tag{36}$$

$$c^{(1)}_{n\sigma}=-\frac{1}{2}\sum_{m\in\langle nm\rangle}(c_{n\sigma}j_{nm,-\sigma}+c_{m\sigma}\Delta_{nm,-\sigma}), \tag{37}$$

$$c^{(2)}_{n\sigma}=\sum_{m\in\langle nm\rangle}\left\{L(n,m)+\sum_{m_i}M(n,m,m_i)+\sum_{n_i}N(n,m,n_i)\right\}. \tag{38}$$

In the three-site terms the sums are performed over $m_i\in\langle mm_i\rangle$, $m_i\neq n$, and $n_i\in\langle nn_i\rangle$, $n_i\neq m$, and the operators in Eq. (38) are given by

$$L(n,m)=-\frac{1}{2}[c_{m\sigma}(1-2n_{n\sigma})t_{nm-\sigma}+c_{n\sigma}\Delta^2_{nm-\sigma}],$$

$$\begin{aligned}M(n,m,m_i)=\frac{1}{4}[&c_{n\sigma}j_{mm_i\sigma}t_{nm-\sigma}+c_{m_i\sigma}\Delta_{nm-\sigma}\Delta_{nm_i-\sigma}\\&+c_{n\sigma}\Delta_{mm_i\sigma}j_{nm_i-\sigma}+c_{m\sigma}\Delta_{nm-\sigma}j_{mm_i-\sigma}\\&+c_{m\sigma}n_{m_i\sigma}t_{mm_i-\sigma}],\end{aligned} \tag{39}$$

$$\begin{aligned}N(n,m,n_i)=\frac{1}{4}[&c_{n_i\sigma}j_{nm-\sigma}\Delta_{nn_i-\sigma}+c_{n_i\sigma}n_{n\sigma}t_{nm-\sigma}\\&+c_{n\sigma}j_{nm-\sigma}j_{nn_i-\sigma}+c_{n\sigma}n_{n_i\sigma}j_{mn_i-\sigma}\\&+c_{m\sigma}\Delta_{nn_i\sigma}t_{nn_i-\sigma}].\end{aligned} \tag{40}$$

The operators $j_{nm\sigma}$, $\Delta_{nm\sigma}$, and $t_{nm\sigma}$ are defined by Eqs. (4) and (11).

The averaging of expressions (33) and (36) per lattice site yields

$$\langle H^{(2)}\rangle=[\mathscr{H}_U^{(2)}+\mathscr{T}^{(2)}]|_{w=0}+\sum_{ij}k^{(2)}_{ij}w_iw_j, \tag{41}$$

$$\begin{aligned}\mathscr{H}_U^{(2)}=4U\{&2d_0^2+3r_1^2[(1-2r_0)^2-4d_0^2]+(\overline{r_\nu}-2\overline{\phi_{\nu0}})\\&\times[r_0(1-r_0)+d_0^2+6r_1^2]+[\overline{d_\nu}d_0+8r_1^2\overline{\phi_\nu}\\&+\overline{r_\nu\phi_\nu}-2\overline{\phi_\nu\phi_{\nu0}}]\},\end{aligned} \tag{42}$$

$$\begin{aligned}&\phi_0=r_0^2-d_0^2-r_1^2,\quad \phi_l=r_l^2-d_l^2,\qquad \phi_{0l}=r_0r_l-d_0d_l,\\&f_0=r_0^2+d_0^2,\quad f_l=r_l^2+d_l^2,\quad \phi_{l,l'}=r_lr_{l'}-d_ld_{l'}.\end{aligned} \tag{43}$$

Any of the symbols $\overline{A_\nu}=\{\overline{r_\nu},\overline{d_\nu},\overline{\phi_\nu},\overline{\phi_{0\nu}},\overline{f_\nu},\overline{r_\nu\phi_\nu},\overline{\phi_\nu\phi_{0\nu}}\}$ denotes

$$\overline{A_\nu}=2A_{\sqrt{2}}+A_2.$$

In this notation, we have

$$\begin{aligned}\mathscr{T}^{(2)}=-4t\{&r_1[-51r_0(1-r_0)-93(d_0^2+2r_1^2)-18(\overline{r_\nu}\\&-2\overline{\phi_{0\nu}})+5\overline{f\nu}+6f_{\sqrt{2}}+2(4\phi_{\sqrt{2}}+4\phi_{2,\sqrt{2}}+\phi_2)\\&+6(\overline{r_\nu})^2+2(\overline{d_\nu})^2-8d_0\overline{d_\nu}]+(r_3+6r_{\sqrt{5}})[-r_0\\&\times(1-r_0)-7d_0^2-10r_1^2]+6r_{\sqrt{5}}\overline{f_\nu}+3r_3f_2\}.\end{aligned} \tag{44}$$

The contribution of anomalous order parameters to the average energy (16) per lattice site calculated to second order in w_l is

$$\overline{H}^{SC}=[-8\alpha Ur_1]w_1^2+\alpha^2[S_U+S_T], \tag{45}$$

$$\begin{aligned}S_U=U\{&4w_1^2[2r_0(1-r_0)+\overline{r_\nu}(1-2r_0)+2d_0\overline{d_\nu}+4d^2_{\sqrt{2}}\\&-2d_2^2]+8w_1w_2[r_1(1-2r_0)]+2w_2^2[r_2(1-2r_0)\\&+8r_1^2-2d_0d_2]\},\end{aligned} \tag{46}$$

$$\begin{aligned}S_T=-2t\{&w_1^2[-34r_1-18r_{\sqrt{5}}-3r_3]+w_2^2[r_1+8r_{\sqrt{5}}\\&+3r_3]+w^2_{\sqrt{5}}[-6r_{\sqrt{5}}]+w_3^2[-r_3]+w_1w_5[4r_1]\\&+w_1w_3[2r_1]\}.\end{aligned} \tag{47}$$

The matrix $D_{i,j}=\partial w_i/\partial w_j$ at $w_l=0$ in Eq. (30) is calculated using solutions AF + VB:

$$D_{ij}=B_{i\nu}k_{\nu j},$$

where $k_{\nu j}$ are coefficients of the quadratic form $H^{SC}=k_{\nu j}W-\nu w_j$ with anomalous averages (18), and elements of matrix B_{ij} are expressed in terms of energies $E_{1(2)}(k)$ of the upper and lower Hubbard bands with an antiferromagnetic gap $2\Delta_k$ between them:

$$B_{\nu j}=-N^{-1}\sum_k^F R(l_\nu,l_j,k), \quad l_j=\{1,2,\sqrt{5},3\}_j, \tag{48}$$

$$R(l,l',k)=q_lq_{l'}\left(\frac{1-2f_1}{2E_1}+\frac{1-2f_2}{2E_2}\right), \quad l,l'\neq 2, \tag{49}$$

$$R(2,2,k)=q_2^2\left[\cos^2\gamma_k\left(\frac{1-2f_1}{2E_1}+\frac{1-2f_2}{2E_2}\right)+\sin^2\gamma_k 2\frac{1-f_1-f_2}{E_1+E_2}\right],$$

$$R(2,l,k)=R(l,2,k)=q_lq_2\cos\gamma_k\left(\frac{1-2f_1}{2E_1}-\frac{1-2f_2}{2E_2}\right),$$

$$l\neq 2.$$

Here $f_{1(2)}$ are the Fermi distribution functions for the band energies (measured with respect to the chemical potential):

$$E_{1(2)}(k)=(\epsilon_k+\epsilon_{\tilde{k}})/2\mp\sqrt{(\epsilon_k-\epsilon_{\tilde{k}})^2/4+\Delta_k^2}-\mu, \tag{50}$$

the parameters γ_k, ϵ_k, and Δ_k are determined by the equation $\tan\gamma_k=(\epsilon_k-\epsilon_{\tilde{k}})/\Delta_k$ and Eq. (26), respectively. In the absence of antiferromagnetism and with only one anomalous order parameter, the equation for T_c goes over to the standard BCS equation, the only difference being the weight factor $q_1^2(k)$ in all sums over k to reflect the d-symmetry of the superconducting ordering. In this case, the two contributions due to E_1 and E_2 with summation over k in the halved Brillouin zone are equivalent to summation over the entire Brillouin zone with the unsplit energy band. In a p- or n-doped system with antiferromagnetic splitting, only one of the E_1 and E_2 subbands determines T_c and determinant (30).

Since the superconducting gap is small in comparison with both the total band width and antiferromagnetic splitting, in solving the full problem AF + VB + SC, without loss of accuracy one can use approximate expressions for the eigenvalues $\mathscr{E}_\lambda$ and eigenfunctions (28) of the linearized Hamiltonian (24). They are determined by the approximate solutions of Eq. (28):

$$\mathscr{E}_\lambda=\{\mathscr{E}_1,\mathscr{E}_2,-\mathscr{E}_1,-\mathscr{E}_2\}_\lambda, \tag{51}$$

$$\mathscr{E}_i=-\sqrt{(E_i-\mu)^2+W_i^2}, \quad i=1,2,$$

$$S_{i\lambda}(k)=\begin{pmatrix} c_\varphi & s_\varphi & 0 & 0\\ -s_\varphi & c_\varphi & 0 & 0\\ 0 & 0 & c_\varphi & -s_\varphi\\ 0 & 0 & s_\varphi & c_\varphi \end{pmatrix}_{ij} \times\begin{pmatrix} c_1 & 0 & -s_1 & 0\\ 0 & c_2 & 0 & -s_2\\ s_1 & 0 & c_1 & 0\\ 0 & s_2 & 0 & c_2 \end{pmatrix}_{j\lambda}. \tag{52}$$

Here

$$s_\varphi=\sin\varphi, \quad c_\varphi=\cos\varphi, \quad c_i=\cos\theta_i, \quad s_i=\sin\theta_i$$

$$\tan 2\varphi=\frac{2\Delta_k}{\epsilon_k-\epsilon_{\tilde{k}}}, \quad \tan 2\theta_i=\frac{2W_i}{E_i-\mu}, \quad i=1,2,$$

$$W_{1(2)}=\frac{1}{2}[W_k+W_{\tilde{k}}+\cos 2\varphi(W_k-W_{\tilde{k}})].$$

The parameters ϵ_k, Δ_k, W_k, and $E_{1(2)}$ are determined by Eqs. (26) and (50).

*)E-mail: ovchin@glas.apc.org

†)E-mail: movchin@center.chph.ras.ru

---

[1] Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
[2] J. W. Allen, R. Claessen, R. O. Anderson *et al.*, in *The Physics of the Hubbard Model*, ed. by D. K. Campbel, J.M.P. Carmelo and F. Guinea, Plenum Press, New York (1994).
[3] T. E. Mason, G. Aeppli, S. M. Hayden *et al.*, Phys. Rev. Lett. **71**, 919 (1993).
[4] J. R. Kirtley, C. C. Tsuei, J. Z. Sun *et al.*, Nature (London) **373**, 225 (1995).
[5] D. A. Brawner, C. Mancer, and H. R. Ott, Phys. Rev. **55**, 2788 (1997).
[6] D. S. Marshall, D. S. Dessau, A. G. Loeser *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
[7] A. G. Loeser, Z.-X. Shen, D. S. Dessau *et al.*, Science **273**, 325 (1996).
[8] H. Ding, T. Yokoya, J. C. Campuzano *et al.*, Nature (London) **382**, 51 (1996).
[9] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[10] D. J. Scalapino, Phys. Rep. **250**, 329 (1995).
[11] E. Dagotto and T. M. Rice, Science **271**, 618 (1996).
[12] Yu. A. Izyumov, Usp. Fiz. Nauk **167**, 465 (1997).
[13] R. O. Zaĭtsev, JETP Lett. **55**, 135 (1992); **56**, 339 (1992).
[14] J. H. Jefferson, H. Eskes, and L. F. Feiner, Phys. Rev. B **45**, 7959 (1992).
[15] H. B. Schuttler and A. J. Fedro, Phys. Rev. B **45**, 7588 (1992).
[16] P. W. Anderson, Science **235**, 1196 (1987).
[17] E. J. Mele, Phys. Rev. B **38**, 8940 (1988).
[18] I. I. Ukrainskii, Int. J. Quantum Chem. **52**, 413 (1994).
[19] A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **110**, 342 (1996) [JETP **83**, 141 (1996)].
[20] A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **112**, 1409 (1997) [JETP **85**, 767 (1997)].
[21] M. C. Gutzwiller, Phys. Rev. **137**, 1726 (1965).
[22] J. E. Hirsch, Phys. Rev. Lett. **54**, 1317 (1985).
[23] E. Dagotto, A. Moreo, F. Ortolani *et al.*, Phys. Rev. B **67**, 10741 (1992).
[24] R. J. Radke and M. R. Norman, Phys. Rev. B **50**, 9554 (1994).
[25] M. S. Hybertsen, E. B. Stechel, M. Schluter, and D. R. Jennison, Phys. Rev. B **41**, 11068 (1990).
[26] D. Duffy and A. Moreo, Phys. Rev. B **52**, 15607 (1995).
[27] U. Trapper, D. Ihle, and H. Fenke, Phys. Rev. B **52**, R11553 (1995).
[28] G. Baumgartel, J. Schmalian, and K. H. Benemann, Europhys. Lett. **24**, 601 (1993).

[29] G. M. Eliashberg, in *Physical Properties of High-Temperature Superconductors*, ed. by D. M. Ginzberg, World Scientific, Singapore (1966); Russ. translation, Mir, Moscow (1990).
[30] P. Aebi, J. Osterwalder, P. Schaller *et al.*, J. Phys. Chem. Solids **56**, 1845 (1995).
[31] N. Nagaosa, Science **275**, 1078 (1997).
[32] S. C. Zhang, Science **275**, 1089 (1997).
[33] N. E. Brickers and S. R. White, Phys. Rev. B **43**, 8044 (1991).
[34] C.-H. Pao and N. E. Brickers, Phys. Rev. Lett. **72**, 1870 (1997).
[35] P. Monthoux and D. J. Scalapino, Phys. Rev. Lett. **72**, 1874 (1997).
[36] St. Lenck, J. P. Carbotte, and R. C. Dynes, Phys. Rev. B **50**, 10149 (1994).
[37] M. C. Schabel, C. M. Park, A. Matsuura *et al.*, Phys. Rev. **55**, 2796 (1997).
[38] V. J. Emery and S. A. Kivelson, Nature (London) **373**, 434 (1995).
[39] S. Doniach and M. Inui, Phys. Rev. B **41**, 6668 (1990).
[40] P. Benard, L. Chen, and A.-M. S. Tremblay, Phys. Rev. B **47**, 589 (1993).
[41] Q. Si, Y. Zha, K. Levin, and J. P. Lu, Phys. Rev. B **47**, 9055 (1993).
[42] D. Pines, J. Phys. Chem. Solids **56**, 1651 (1995).
[43] P. Montoux and D. Pines, Phys. Rev. B **50**, 16015 (1994).

Translation provided by the Russian Editorial office.

Dipolar Zeeman mixing in weak magnetic fields: An experiment with ice protons

N. V. Zavaritskiĭ†) and I. S. Solodovnikov*)

P. L. Kapitza Institute for Physical Problems, Russian Academy of Sciences, 117973 Moscow, Russia

(Submitted 16 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1006–1017 (September 1998)

The evolution of the longitudinal magnetization of nuclear spins in a cw high-frequency magnetic field has been measured using a SQUID magnetometer at liquid-helium temperatures in magnetic fields H_0 of up to 57 Oe. The time T_m for thermal mixing of the Zeeman and dipolar systems has been found to range between 0.05 and 4×10^2 s. For $T_m>1$ s the function $T_m(H_0)$ is exponential. The proton NMR spectra near the fundamental and twice the Larmor frequency have been obtained. The shift in the resonance with respect to the Larmor frequency is close to the theoretical prediction. © *1998 American Institute of Physics.* [S1063-7761(98)01709-0]

1. INTRODUCTION

The state of a spin system in a solid can be characterized by a single temperature in a weak magnetic field comparable to the local field.[1] In magnetic fields substantially greater than the local field, the spin system is usually characterized by the temperatures of the Zeeman reservoir and the dipolar reservoir, the latter being related to the secular part of the dipole–dipole interaction. The time required for the system to relax to the ''quasi-equilibrium'' state with two temperatures is of order of the spin–spin relaxation time T_2 (Ref. 1, Ch. 4). In magnetic fields comparable to the local field, the temperatures of these two subsystems equilibrate owing to the nonsecular part of the dipole–dipole interaction. This process is called thermal mixing or dipolar Zeeman cross relaxation [Ref. 1 (Ch. 6) and Ref. 2]. When the mixing time is relatively short, $T_m\lesssim10^5T_2$, it is described by a Gaussian function of the magnetic field: $T_m(H_0)\sim T_2\exp(H_0^2/M_2^*)$, where M_2^* is near the second moment M_2 of the resonance line (Ref. 1, Ch. 6). Measurements of longer times T_m are usually hampered by spin–lattice relaxation.[3] In the case of cross relaxation between different spin systems in a rotating reference frame, T_m is usually an exponential function of the effective magnetic field (Ref. 4, Ch. 1). Zobov and Lundin[5] concluded that the exponential shape of the strong-field asymptotic form of the cross relaxation rate is a universal feature of spin dynamics. The related issue of the exponential shape of NMR line edges was studied both theoretically and experimentally.[6] Therefore it seemed interesting to test the conclusions of Ref. 5 in the case of the dipolar Zeeman cross relaxation in a system of spins of one type, namely, to measure the function $T_m(H_0)$ over the widest possible time interval.

A suitable object of such an investigation is ice: the concentration of nuclei with nonzero spin other than protons is negligible, namely 1×10^{-2}% of deuterium and 4×10^{-2}% of ^{17}O. The long spin–lattice relaxation time[7] $T_1\gtrsim10$ h measured in a strong magnetic field ($H_0=7$ kOe) at $T\approx150$ K holds out the possibility of measuring long cross relaxation times. In the present work T_1 was measured at liquid-helium temperatures in weak magnetic fields $H_0<55$ Oe. The impact of residual oxygen (gas dissolved in water from which ice crystals were grown) on the spin–lattice relaxation has been investigated.

The shape of the proton NMR line in ice has been experimentally studied by many researchers.[7–11] These experiments used both cw NMR measurements in which external magnetic field was scanned[7,8,10] and Fourier transforms of decaying nuclear induction signals.[9,11] The experiments were performed with both polycrystalline samples and single crystals. A technique for calculating M_2 in the conventional hexagonal lattice with a configuration of disordered proton locations was suggested.[9,10] In these experiments we have obtained the proton resonance line shape in ice by measuring the rates of spin heating by hf magnetic field and compared our results with published data. It is also of interest tthat the ''forbidden'' line at twice the NMR frequency, which was observed in some materials[12–14] in weak magnetic fields, was detected.

2. EXPERIMENTAL

A diagram of the experimental chamber is given in Fig. 1. The inside volume defined by an outside shell in the form of a 7-mm tube fabricated from a Mylar film and epoxy cement is insulated from the copper cold conductor by a Mylar ring. Two coils of niobium–titanium wire are wound on top of the Mylar tube and fastened with epoxy cement. One of them is the SQUID reception coil, the other is a heater designed to generate a temperature gradient during crystal growth. Four copper wires of diameter 0.2 mm coated with lead–tin soldering alloy are cemented on top of these coils for heat-sinking from the inside volume of the device. Similar wires are used for cooling a coil for excitation of NMR, which is wound with the Nb–Ti wire on a Mylar core. Two small pieces of Nb foil cemented to the lower part of the outer surface of the device serve as a capacitive detector (driven at 0.2 MHz) of the crystallization onset. The temperature in the cell is measured by a copper resistive thermometer mounted on the cold-conductor. The cell is shielded

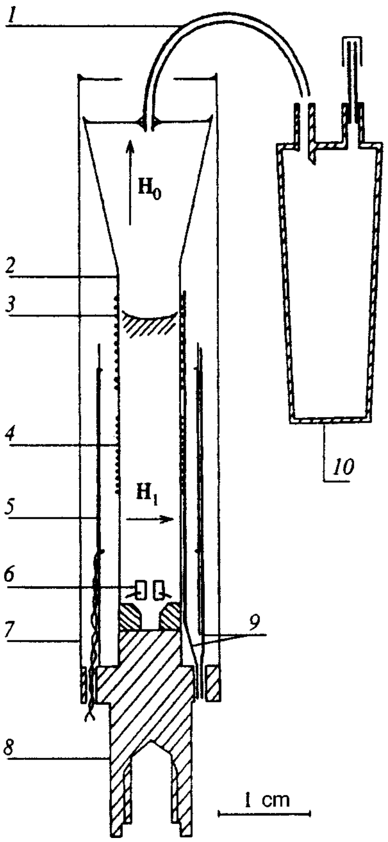


FIG. 1. Diagram of the experimental chamber: (*1*) filling pipe; (*2*) outer shell of the cell; (*3*) heater; (*4*) reception coil of the SQUID magnetometer; (*5*) coil generating hf field; (*6*) detector of crystal growth onset; (*7*) electromagnetic shield from copper and niobium foil and Mylar film; (*8*) cold conductor; (*9*) heat-sinking wires; (*10*) polyethylene trap.

by tubes welded from copper and niobium foils. A flexible plastic tube is cemented to the upper lid of the cell, and about 1.2 cm^3 of distilled water is fed through this tube via a Teflon capillary. The filling tube is connected to the vacuum container (jacket) via a trap designed to collect water drops and thus limit the quantity of vapor escaping from the cell.

The magnetometer incorporating a high-frequency SQUID used to measure the longitudinal magnetic moment under conditions of NMR excitation was described elsewhere.[15] The dc magnetic field H_0 is generated by a short-circuited superconducting solenoid. The field H_0 is determined accurately using measurements of NMR in ^{3}He[16] performed with this solenoid and a similar experimental cell. The output signal from the SQUID is recorded by a computer, which also switches on and off the hf magnetic field. In these experiments we have used linearly polarized hf field perpendicular to the dc field. The value H_1 of the rotating component of the hf magnetic field was calibrated using the area of the obtained NMR line.

Before the device was cooled, the vacuum container had been filled with the selected gas. When a small amount of liquid nitrogen was poured into the nitrogen Dewar of the device, the cell cooled down at a rate of 0.25–0.4 °C/min. At $t=3\,^\circ$C a power of 0.2 W was fed to the heater (the power was derived from the 10°C rise in the temperature of the heat conductor). The crystallization onset was detected using the signal from the capacitor and briefly heating the cold conductor, whose temperature just prior to this time ranged between $-7\,^\circ$C and $-12\,^\circ$C owing to the overcooling of the water. At $t\approx-30\,^\circ$C the heater was turned off, the vacuum jacket of the device was pumped off and filled with helium at a pressure of about 10 Torr, which was used as the heat-exchange gas in further cooling. Measurements were usually performed at a fixed value of H_0 in one experimental run. The vacuum container was usually filled with the selected gas at the end of the run at $T\approx100$ K.

The cooling rate while the ice crystal was growing in our experiments was close to that chosen by Evtushenko *et al.*,[17] who reported that their samples were composed of randomly oriented single-crystal grains of hexagonal ice with a typical size of about one millimeter. Pictures of the polycrystalline ice structure obtained at a growth rate of $\sim10^{-3}$ cm/s are given in Hobbs's book.[18]

3. SPIN–LATTICE RELAXATION

In measurements of the spin–lattice relaxation time T_1, we used the conventional technique of NMR saturation by a resonant hf field with subsequent monitoring of the magnetization recovery after the hf field was turned off. An example of the SQUID output versus time with two pulses of the resonant hf field is shown in Fig. 2a. The amplitude of the step generated by the hf pulse is proportional to the sample magnetization M_Z before the pulse. Figure 2b shows the step amplitude as a function of the time interval Δt that has passed after the previous pulse. By approximating these measurements using the formula

$$M_Z=M_0[1-\exp(-\Delta t/T_1)], \tag{1}$$

one can determine the equilibrium magnetization M_0 (in relative units of the SQUID output) and the longitudinal relaxation time. Under the conditions of our experiments, the spin–lattice relaxation time is controlled by oxygen which

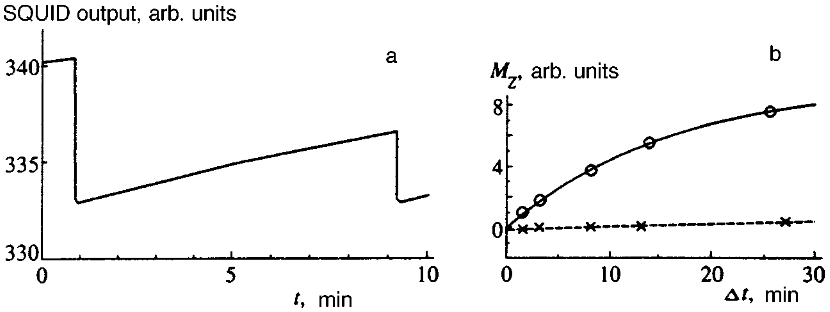


FIG. 2. (a) Trace of SQUID output with two pulses of hf magnetic field at the resonant frequency; $H_0=44.7$ Oe, $T=4.2$ K, $H_1=26$ mOe; the duration of each pulse is 4 s. (b) Amplitude of steps caused by hf pulses as a function of time interval between two pulses for two types of residual gas in the cell: (○) air, $H_0=44.7$ Oe; (×) nitrogen, $H_0=50.9$ Oe. The solid curve shows the fit of Eq. (1) to the experimental points.

probably is trapped in the crystal when water is frozen with air dissolved in it. When the vacuum jacket was filled with air at a pressure of 1 atm, the time T_1 at $T=4.2$ K ranged between 10 and 20 min in different experiments, but was independent of the dc magnetic field in the range of 10 to 55 Oe when T_1 was measured in one experiment. When the jacket was filled with helium or nitrogen instead of air, T_1 was considerably longer: estimates of T_1 using Eq. (1) and measurements at $\Delta t<30$ min yield in this case $T_1\simeq 10$ h. In the temperature range of 1.3–4.2 K the time T_1 increases with decreasing temperature, approximately following the law $T_1\propto T^{-1}$.

4. HEATING OF THE SPIN SYSTEM BY AN HF FIELD

The evolution of the longitudinal magnetization in an hf magnetic field with a frequency close to that of NMR can be described by the Provotorov equations (Ref. 1, Ch. 4, and Ref. 19). In a magnetic field H_0 comparable to the local field, the rate of temperature equilibration between the dipolar and Zeeman reservoirs in the laboratory reference frame should also be taken into account [Ref. 1 (Ch. 6) and Ref. 2]. If the spin–lattice relaxation is neglected,

$$\dot{\alpha}=-(W_1+W_2)\alpha+\left(\frac{\Delta}{\omega_0}W_1+W_2\right)\beta,$$

$$\dot{\beta}=D^{-2}(\omega_0\Delta W_1+\omega_0^2W_2)\alpha-D^{-2}(\Delta^2W_1+\omega_0^2W_2)\beta. \qquad (2)$$

Here α and β are the inverse temperatures of the Zeeman and dipolar subsystems in the laboratory frame, and the longitudinal magnetization satisfies $M_Z\propto\alpha$. The parameter $W_1=\pi\gamma^2H_1^2g(\omega)$ determines the rate at which $\alpha\omega_0/\Delta$ and β are equalized by the hf field [this process is usually treated (Ref. 1, Ch. 4) as equilibration (mixing) of the Zeeman and dipolar reservoirs in the reference frame rotating with the hf field]; $g(\omega)$ is the shape function of the resonance line, and $\Delta=\omega_0-\omega$ is the hf field detuning from the resonant frequency ω_0. In a weak magnetic field H_0 the resonant frequency ω_0 is shifted from the Larmor frequency γH_0 because of the nonsecular part of the dipole–dipole interaction.[2] In Eq. (2) $D=\gamma H_L'$ is the local frequency related to the local field in the rotating frame, which is determined by the secular part of the dipolar interaction. The terms with W_1 on the right-hand side of Eq. (2) yield the conventional Provotorov equations, which apply to the case of $H_1\ll H_L'$ (Ref. 1, Ch. 4). The parameter $W_2=T_m^{-1}D^2/(\omega_0^2+D^2)$ describes the rate at which α and β equilibrate in weak magnetic fields owing to the nonsecular part of the dipolar interaction.

When an intense hf field, albeit smaller than H_L', is turned on, the temperatures of the Zeeman and dipolar subsystems in the rotating frame rapidly equilibrate in the time $t_F^0\approx D^2[W_1(D^2+\Delta^2)]^{-1}$, then the inverse temperature (which is unique in this reference frame) tends to zero at the rate determined by W_2. In this section, we consider the case of an hf field which is strong in this sense. Equation (2) with the initial condition $\alpha(0)=\beta(0)$ has a solution describing the longitudinal magnetization as a sum of two decreasing exponential functions when the hf field is on:

$$\frac{M_Z(t)}{M_Z(0)}=C_F^0\exp\left(-\frac{t}{t_F^0}\right)+(1-C_F^0)\exp\left(-\frac{t}{t_S^0}\right). \qquad (3)$$

In the case of a strong magnetic field, when the condition $W_1\gg(1/T_m)(1+\Delta^2/D^2)^{-1}$ holds,

$$C_F^0\approx C_F-\frac{t_F^0}{t_S^0}\frac{2\omega_0}{\omega}\frac{\Delta^2}{D^2+\Delta^2},$$

$$t_S^0\approx t_S+t_F^0\left(\frac{\omega_0\Delta+D^2}{\omega D}\right)^2. \qquad (4)$$

The second terms in the right-hand sides of Eqs. (4) are corrections to the values

$$C_F=\frac{D^2}{D^2+\Delta^2}\frac{\omega}{\omega_0},\qquad t_S=T_m\frac{D^2+\Delta^2}{D^2}\frac{\omega_0^2+D^2}{\omega^2}, \qquad (5)$$

which determine the ''weight'' of the initial fast relaxation stage and the time constant in the second, slow stage of M_Z decay in the limit of a strong hf field.

Traces of the SQUID output as a function of time, which show how the longitudinal magnetization changes in an hf field, are given in Fig. 3a. The signal due to the cell walls and coil holder was about $0.04M_0$ and was probably caused by protons in the Mylar and epoxy cement. The spin–lattice relaxation time under the conditions of our experiments is about 1–10 s. In order to get rid of this signal, we performed two cycles of measurements at each frequency of the hf field. At the end of each cycle, we turned on the hf field at the resonant frequency ω_0 to achieve full saturation of the nuclear spin system. The time before the second cycle of about one minute was sufficient for the spins in the walls to relax. The recorded SQUID signals of the first and second cycle were differenced numerically (Fig. 3b). The resulting curves were approximated using Eq. (3) on the time interval with the hf field on. One can see in Fig. 3b that the weight of the initial stage, C_F^0, and the time constant of the slower stage, t_S^0, are independent of H_1 at higher hf field intensities. With due account of the finite value of the hf field, C_F and t_S^0 were calculated by Eq. (4) using measurements of C_F^0 and t_S^0. These parameters as functions of the frequency are plotted in Fig. 4a, alongside the curves obtained by fitting Eq. (5) to experimental data. The time T_m was obtained by extrapolating t_S to the resonant frequency.

If the hf field is tuned to the resonance, the contribution of the second (slow) stage is zero, and the evolution of the longitudinal magnetization is described by the single time constant $t_F^0=1/W_1$, because the hf field pumps energy only to the Zeeman subsystem and does not affect the dipolar subsystem. When an hf field at frequency $2\omega_0$ is applied, we also measured $M_Z(t)$ described by one decreasing exponential function. When the frequency is detuned from $2\omega_0$, the measured weight C_F of the initial stage and time constant t_S of the second stage under a strong hf field (Fig. 4b) are described by equations similar to Eq. (5), but with Δ replaced by $\omega_0-\omega/2$:

$$C_F\approx D^2[D^2+(\omega_0-\omega/2)^2]^{-1},$$

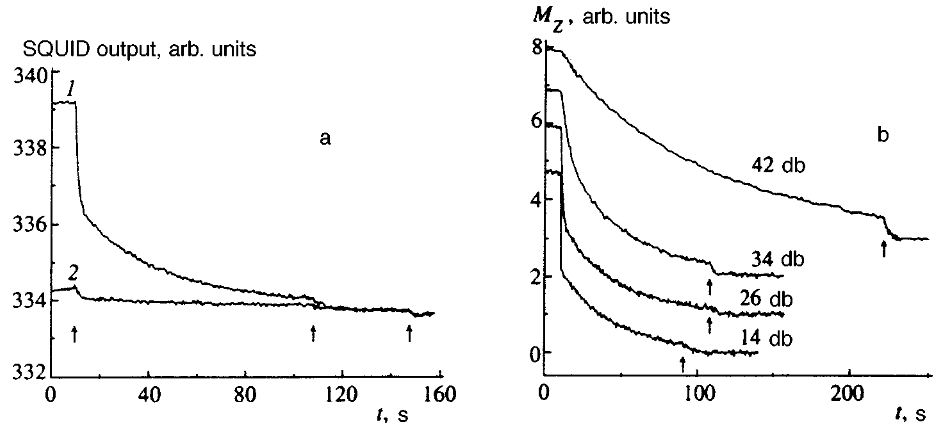


FIG. 3. (a) Records of SQUID output in which the hf field is turned on in two successive cycles. The arrows mark moments when the hf field $H_1 = 11$ mOe (the output attenuator was set to 26 dB) at frequency $\omega/2\pi = 180$ kHz, hf resonant field $H_1 = 8.6$ mOe at 193 kHz were turned on, and when the hf fields were turned off. The record of curve 2 was started in 62 s after the end of the record of curve 1. The signal due to the walls can be seen after the hf field is switched off. (b) Difference between curves 1 and 2 at various hf fields (attenuator settings are given at the curves). The arrows mark times when the resonant hf field was turned on. The shift of each curve with respect to the previous one along the ordinate axis is one unit.

$$t_S \approx T_m[D^2 + (\omega_0 - \omega/2)^2]D^{-2}. \tag{6}$$

The values of D, which were fitting parameters for curves of C_F and t_S versus frequency, around the fundamental NMR line were found to be independent of H_0 within the experimental uncertainty, and we obtained $D/\gamma = 3.7 \pm 0.2$ Oe. This value corresponds[1] to $(M_2/3)^{1/2}$, where M_2 is the second Van Vleck moment of the NMR line. Measurements of ω_0, which is also a fitting parameter, are plotted in Fig. 5. The difference between ω_0 and the Larmor frequency γH_0 is in agreement with the formula[2] $\omega_0 - \gamma H_0 \propto 1/H_0^2$.

The time T_m derived from measurements around the fundamental NMR line is plotted in Fig. 6. In all experiments T_m was within $0.2T_1$. The measurements of T_m were independent of the temperature and composition of the residual gas. This indicates that the spin–lattice relaxation is not important for the mixing of the Zeeman and dipolar subsystems under these conditions. In order to check whether deuterium, whose nucleus has spin 1 and a quadrupole splitting in the ice lattice of 0.2 MHz,[11] has an impact on this process, we experimented with samples containing ten times as much deuterium as natural ice. We have not detected an effect of deuterium concentration on the dipolar Zeeman cross relaxation rate. It seems that ^{17}O nuclei, whose quadrupole splitting is about 1.7 MHz in ice[20] and considerably higher than the proton NMR frequencies in our experiments, are also unimportant for the dipolar Zeeman mixing in the proton spin system. The curve of $T_m(H_0)$ (Fig. 6) for $T_m > 1$ s ($H_0 > 35$ Oe) can be approximated by the formula

$$T_m = T^* \exp(H_0/H^*), \quad T^* = 6 \times 10^{-5}\,\mathrm{s}, \quad H^* = 3.6\ \mathrm{Oe}. \tag{7}$$

These measured value of T_m was obtained in polycrystalline samples and, apparently, is a value averaged over different orientations of grains. Let us estimate the width of their distribution assuming that $H^* \propto M_2^{1/2}$ and taking the data concerning the second-moment anisotropy[9]: without the maximum of M_2 for $\mathbf{H}_0$ aligned with the hexagonal axis, $M_2^{\max}/M_2^{\min} = 1.2$. Then $T_m = 400$ s is a value averaged over a distribution with $T_m^{\max}/T_m^{\min} \simeq 5$. This estimate clearly shows that single crystals should be used in measurements of longer times of thermal mixing between Zeeman and dipolar subsystems.

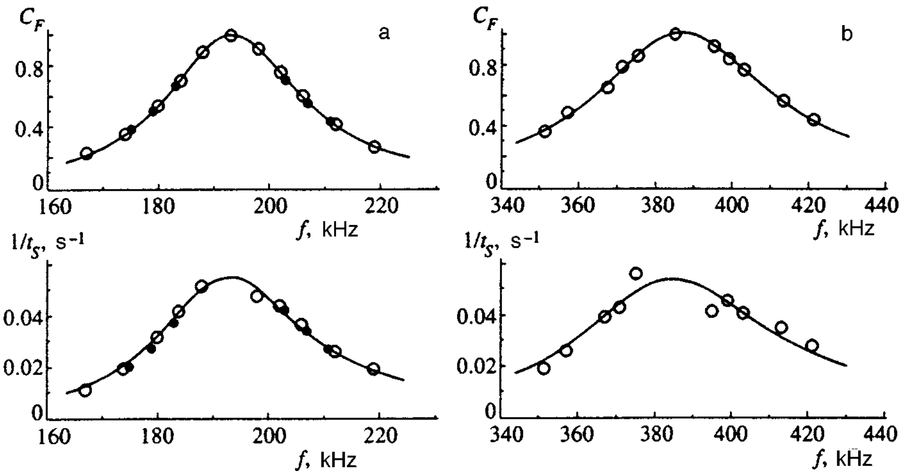


FIG. 4. (a) Parameters C_F and t_S versus frequency obtained using Eq. (4) and curves calculated by Eq. (5) fitted to these data: (○) $T = 4.2$ K; (●) $T = 1.3$ K. Measurements were performed at $H_0 = 44.7$ Oe, $H_1 = 17$ mOe. (b) Measurements at ω about $2\omega_0$ under the same conditions, but at $H_1 = 120$ mOe. The curves are calculations by Eq. (6) with D and T_m derived from curves of C_F and t_S versus frequency plotted in the left panel.

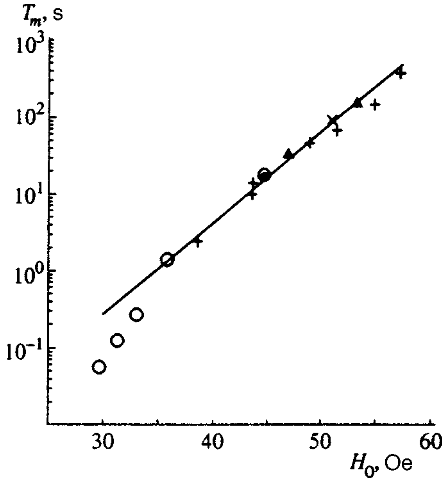


FIG. 6. Time for mixing of the Zeeman and dipolar subsystems at various temperatures and residual gas compositions: (+) 4.2 K, helium; (○) 4.2 K, air; (●) 1.3 K, air; (×) 4.2 K, nitrogen; (black triangles) 0.15% D_2O doping, 1.3 K, air. The straight line is a calculation by Eq. (7).

5. NMR LINE SHAPE

In magnetic fields $H_0 < 30$ Oe the time for the temperatures of the Zeeman and dipolar subsystems to equilibrate is $T_m < 0.1$ s, and the temperatures of these two reservoirs in the laboratory frame remain equal in the presence of a weak hf field. The heating of the spin system by a weak hf field is characterized by a single time constant.[13,14] For $H_0 < 30$ Oe this time constant can be much shorter than $T_1 \gtrsim 10$ min. This case is considered in this section. The time of the spin heating by a weak hf field, in accordance with Eq. (2) is

$$\tau = \frac{1}{\pi \gamma^2 H_1^2 g(\omega)} \frac{\omega_0^2 + D^2}{\omega^2}, \tag{8}$$

provided that $\tau \gg T_m(1 + \Delta^2/D^2)$. Measurements of the spin heating by an hf field were performed in two cycles in order to get rid of the signal due to the walls, as was described in the previous section. The hf field values were chosen to have $\tau \approx 10$ s near the resonant frequency and twice the resonant frequency and $\tau \simeq 100$ s on the edges of these lines. In this case, one can neglect the spin–lattice relaxation. Plots of $1/\tau \propto \omega^2 g(\omega)$ for $H_0 = 29.6$ Oe are given in Fig. 7a. Let us separate the lines around the fundamental and twice the NMR frequency: $g(\omega) = g_1 + g_2$. Given the shape of the fundamental line, g_1, we calculated the resonant frequency

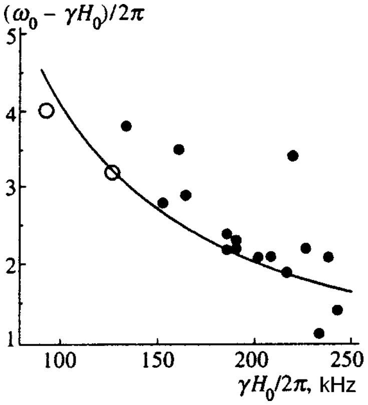


FIG. 5. Resonance shift (●) with respect to the Larmor frequency measured in strong hf fields. (○) Values of $(\gamma/2\pi)M_1$ derived from NMR line shapes. The solid curve is calculated by Eq. (9).

$$\omega_0 = \int \omega g_1 d\omega \Bigg/ \int g_1 d\omega,$$

its shift with respect to the Larmor frequency, $\gamma M_1 = \omega_0 - \gamma H_0$, and the second moment (in units of s^{-2})

$$\gamma^2 M_2 = \int \omega^2 g_1 d\omega \Bigg/ \int g_1 d\omega - \omega_0^2.$$

Integration was performed only over frequencies $\omega > 0$.

The shift of the NMR line with respect to the Larmor frequency is caused by the nonsecular part of the dipole–dipole interaction. This shift can be ascribed[2] to an increase in the effective field acting on the spin owing to local fields perpendicular to $\mathbf{H}_0$. For a polycrystalline sample with random orientation of grains (''powder'')[21]

$$M_1 = \frac{2}{3} \frac{M_2}{H_0}. \tag{9}$$

The measured line shift is close to the calculation by Eq. (9) (Fig. 5). Equation (9) also accounts for the resonant frequency shift $\omega_0 - \gamma H_0$ derived from measurements in strong hf fields.

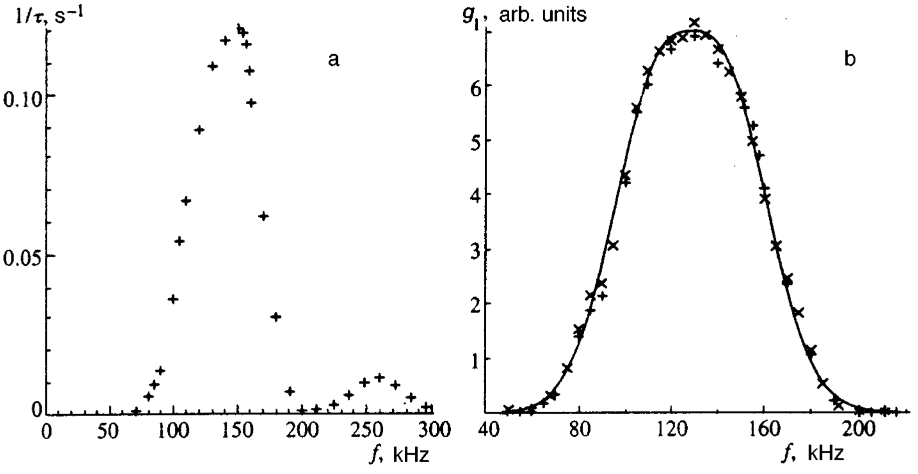


FIG. 7. (a) Rate of spin system heating by the hf field transformed to a single hf field value $H_1 = 18$ mOe; $H_0 = 29.6$ Oe, $T = 4.2$ K. (b) Shape of the fundamental NMR line, $H_0 = 29.6$ Oe, (+, ×) measurements obtained in two different runs. The solid line shows the convolution of the Pake doublet with the Gaussian.

TABLE I.

H_0, Oe	$\frac{\gamma H_0}{2\pi}$, kHz	M_2, Oe^2	$\int g_2 d\omega / \int g_1 d\omega$		$\frac{g_2(2\omega_0)}{g_1(\omega_0)}$
			Calculations by Eq. (10)	Measurements	
21.8	93	33.0	0.048	0.039	0.057
29.6	126	34.6	0.026	0.021	0.027
44.7	190	–	0.011	–	0.012

The measurements of M_2 and relative line intensity near twice the Larmor frequency are summarized in Table 1. Our measurements of M_2 are close to the value $M_2 = 33.8$ Oe^2 obtained in samples of polycrystalline hexagonal ice.[11] The relative total intensity of the line about twice the frequency calculated for a powder by Cheng[22] is

$$\frac{\int g_2 d\omega}{\int g_1 d\omega} = \frac{2}{3}\frac{M_2}{H_0^2}. \tag{10}$$

The calculations of this intensity by Eq. (10) are given in Table 1. Measurements of this intensity are in agreement with calculations by Eq. (10). Note that, within the experimental uncertainty, the relative total intensity of this line near twice the Larmor frequency equals the ratio between the peaks of g_2 and g_1, which was determined experimentally also in a stronger magnetic field H_0.

The hf field H_1 was calibrated using the area under the curve of $(\tau\omega^2)^{-1}$ versus frequency for the fundamental NMR line using Eq. (8) and the normalization condition $\int g_1 d\omega = 1$, which is valid to within $\simeq M_2/H_0^2$ (see Ref. 14).

The shape of the fundamental NMR line (Fig. 7b) can be approximated (Ref. 23, Ch. VII) by a convolution of the Pake doublet, which describes a spectrum of a powder containing pairs of spins with a fixed internuclear distance, and a Gaussian $\exp[-(\omega-\omega_0)^2/(2b^2)]$. This approach is used in describing shapes of proton resonance lines in crystal hydrates, where the effective (with due account of molecular vibrations) interproton distances in the water molecule is taken equal to $R_{PP} = 1.58$ Å.[24] The second moment of the Pake doublet for this distance is

$$\frac{9}{20}\frac{\gamma^2\hbar^2}{R_{PP}^6} = 23 \text{ Oe}^2.$$

The parameter b can be derived from the measurement of M_2: $b/\gamma \approx 3.3$ Oe. Thus, the intramolecular contribution to the second moment is about $0.7M_2$. All in all, the results concerning the fundamental NMR line obtained in measurements of the heating rate of the spin system by an hf magnetic field are in fair agreement with the data by Ripmeester *et al.*[11]

Barnaal and Lowe[9] considered the effect of water molecules undergoing tunneling rotations through an angle of 180° around the bisector of the H–O–H angle on the shape of the NMR line in ice and concluded that the frequency of such tunneling transitions is smaller than the dipolar frequency $\sim 10^4$ Hz. The frequency of the tunneling rotations can be estimated, assuming a sine-wave shape of the curve of potential energy versus angle, using the following relationship:

$$\nu_t \simeq \nu_l \exp(-2\sqrt{2IU}/\hbar), \tag{11}$$

where $\nu_l \simeq 10^{13}$ Hz is the frequency of librational oscillations,[24] and I is the moment of inertia of the molecule. The barrier height U can be equated to the activation energy in the temperature dependence of protons' T_1 in undoped ice measured by Kume[7] in the region of thermally induced rotations for $t > -40$°C: $U = 0.44$ eV. The calculation by Eq. (11) yields $\nu_t \sim 10^{-1}$ Hz. It seems that such low frequencies of tunneling rotations are also unimportant for dipolar Zeeman cross relaxation.

6. CONCLUSIONS

We have measured the times for cross relaxation between the Zeeman and dipolar subsystems for protons in ice in magnetic fields ranging between 30 and 57 Oe. The cross relaxation time T_m as a function of a dc magnetic field in the interval between 1 s and 4×10^2 s is adequately described by Eq. (7). This result supports the theoretical prediction[5] that the function $T_m(H_0)$ should be exponential in the limit of large T_m for the dipolar Zeeman cross relaxation in a system of spins of one type.

Our measurements of the second moment are close to the data by Ripmeester *et al.*.[11] The measured shift of the fundamental NMR line position and the intensity of the line near twice the resonant frequency are in agreement with theoretical estimates.[21,22]

One of the authors (I.S.S.) would like to thank V. A. Atsarkin for discussions of this work. The work is part of the *Magnetometer* project supported by the Fundamental Metrology R&D Program sponsored by the government of Russia.

[1)] E-mail: solodov@kapitza.ras.ru
[†)] Deceased.

---

[1] M. Goldman, *Spin Temperature and Nuclear Magnetic Resonance in Solids*, Clarendon Press, Oxford (1970).
[2] V. A. Atsarkin, Fiz. Tverd. Tela **27**, 656 (1985) [Sov. Phys. Solid State **27**, 406 (1985)].
[3] V. A. Safin, V. A. Skrebnev, and V. M. Vinokurov, Zh. Éksp. Teor. Fiz. **87**, 1889 (1984) [Sov. Phys. JETP **60**, 1087 (1984)].
[4] A. Abragam and M. Goldman *Nuclear Magnetism: Order and Disorder*, Clarendon Press, Oxford (1982).
[5] V. E. Zobov and A. A. Lundin, Zh. Éksp. Teor. Fiz. **106**, 1097 (1994) [JETP **79**, 595 (1994)].
[6] M. I. Bulgakov, A. D. Gul'ko, F. S. Dzheparov *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 614 (1993) [JETP Lett. **58**, 592 (1993)].
[7] K. Kume, J. Phys. Soc. Jpn. **15**, 1493 (1960).
[8] K. Kume and R. Hoshino, J. Phys. Soc. Jpn. **16**, 290 (1961).
[9] D. E. Barnaal and I. J. Lowe, J. Chem. Phys. **46**, 4800 (1967).
[10] S. W. Rabideau, E. D. Finch, and A. B. Denison, J. Chem. Phys. **49**, 4660 (1968).
[11] J. A. Ripmeester, C. I. Ratcliffe, and D. D. Klug, J. Chem. Phys. **96**, 8503 (1992).
[12] M. Kohl, M. Odehnal, V. Petricek *et al.*, J. Low Temp. Phys. **72**, 319 (1988).
[13] L. S. Vlasenko, N. V. Zavaritskiĭ, S. V. Sorokin, and V. G. Fleĭsher, Zh. Éksp. Teor. Fiz. **91**, 1496 (1986) [Sov. Phys. JETP **64,** 881 (1986)].
[14] A. G. Anderson, Phys. Rev. **125**, 1517 (1962).
[15] O. Lounasmaa, *Experimental Principles and Methods below 1* K, Academic Press, London (1974).

[16] I. S. Solodovnikov and N. V. Zavaritskiĭ, Zh. Éksp. Teor. Fiz. **110**, 2047 (1996) [JETP **83**, 1127 (1996)].
[17] A. A. Evtushenko, M. B. Martirosyan, and V. F. Petrenko, Fiz. Tverd. Tela **30**, 2133 (1988) [Sov. Phys. Solid State **30**, 1229 (1988)].
[18] P. V. Hobbs, *Ice Physics*, Clarendon Press, Oxford (1974), p. 592–595.
[19] B. N. Provotorov, Zh. Éksp. Teor. Fiz. **41**, 1582 (1961) [Sov. Phys. JETP **14**, 1126 (1962)].
[20] Y. Margalit and M. Shporer, J. Magn. Reson. **43**, 112 (1981).
[21] L. L. Buishvili, N. P. Giorgadze, and M. D. Zviadadze, Zh. Éksp. Teor. Fiz. **72**, 750 (1977) [Sov. Phys. JETP **45**, 392 (1977)].
[22] H. Cheng, Phys. Rev. **124**, 1359 (1961).
[23] A. Abragam *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford (1961).
[24] S. P. Gabuda and A. F. Rzhavin, *Nuclear Magnetic Resonance in Crystal Hydrates and Hydrated Proteins* [in Russian], Nauka, Novosibirsk (1978).

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Magnetically induced spatial dispersion in the cubic magnetic semiconductors $Cd_{1-x}Mn_xTe$

B. B. Krichevtsov,*) R. V. Pisarev,†) A. A. Rzhevskiĭ, and V. N. Gridnev‡)

A. F. Ioffe Physicotechical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

H.-J. Weber

Physics Department, Dortmund University, 44221 Dortmund, Germany

(Submitted 22 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1018–1033 (September 1998)

In the transverse geometry we have detected birefringence that is linear in the magnetic field **B** and the light wave vector **k** in the cubic magnetic semiconductors $Cd_{1-x}Mn_xTe$ $(0\leqslant x\leqslant 0.52)$. The effect was found to be large, ~ 1 (deg cm^{-1} T^{-1}), and highly anisotropic, in contrast to the Faraday and Voigt effects. The phenomenon is represented by terms of type $\gamma_{ijkl}B_k k_l$ in the permittivity tensor ε_{ij} and can be described by two parameters, A and g. Spectral studies have shown that the normalized parameters A/x and g/x are independent of x, i.e., the effect can be related to the Mn^{2+} ions. Below the edge E_g of the forbidden band, the dispersion of A is described by a $(E_g-E)^{-1.4}$-dependence, while the dispersion of g is nil. Theoretical analysis has shown that the spectral curves for A and g can be explained by the special features of the dispersion laws for electrons and holes (features related to the fact that there is no inversion center) and by the dependence of the parameters of the exchange interaction on the electron wave vector. © *1998 American Institute of Physics.* [S1063-7761(98)01809-5]

1. INTRODUCTION

In the past decade the magnetic, optical, and magnetooptic properties of magnetic semiconductors have been under intensive study. The interest in these crystals is due, in particular, to the huge values of the magnetooptic Faraday, Voigt, and Kerr effects, circular dichroism, etc.[1–3] These phenomena have been studied by many researchers, but their microscopic nature in many cases remains open to discussion. For instance, the interpretation of the contribution of band-to-band transitions to the Faraday effect meets with serious difficulties,[4,5] since the observed spectral curves differ substantially from those predicted by the theory.[6,7]

The Faraday, Kerr, and other effects, which are linear in the magnetic field **B**, can be described phenomenologically by an axial tensor of rank 3, which is allowed in crystals of all classes and in disordered media. However, many magnetic semiconductors, such as $Cd_{1-x}Mn_xTe$, crystallize into a noncentrally symmetric cubic structure of the zinc-blende type $(\bar{4}3m)$, and when a magnetic field **B** is applied, the phenomenon of magnetically induced spatial dispersion related to the bilinear terms of the $\Delta\varepsilon_{ij}=\gamma_{ijkl}k_kB_l$ type in the permittivity tensor ε_{ij} (**k** is the wave vector of the light) is permitted. One example is nonreciprocal birefringence (the kB-effect). One characteristic of any noncentrally symmetric crystal is that the axial tensor γ_{ijkl} is symmetric in the indices i and j (Refs. 8–10). Since the kB-effect is a spatial dispersion effect linear in **k**, far from excitonic resonances it has an additional (in relation to the Faraday effect) small parameter a/λ, where a is the interatomic distance, and λ is the wavelength in the medium.

We know of only a few publications that report detecting optical phenomena associated with magnetically induced spatial dispersion in nonmagnetic semiconductors at low temperatures. Among these is the effect of magnetic field inversion[11,12] and the bleaching, by a transverse magnetic field, of a crystal placed between crossed polarizers.[13,14] These phenomena were observed in CdS, CdSe, and GaAs in the exciton absorption region. Up to now theoretical studies of the microscopic mechanisms have been limited to allowing for terms in the exciton or polariton dispersion law that are linear in **k** and **B** (Refs. 13–15). It is known, however, that the action of an external magnetic field in magnetic semiconductors is substantially enhanced by the $sp-d$ exchange interaction, which leads to anomalously high values of the Faraday effect (linear in the magnetic field)[1,2] and of the Voigt effect (quadratic in the magnetic field).[3] One can assume that the size of the kB-effect in magnetic semiconductors is somewhere in the middle between the size of the effect in dia- or paramagnetic materials, where it is small, and that in magnetically ordered crystals, where it is determined by strong internal exchange fields and is therefore sufficiently large to study in experiments.[16–18] Below we also show that the kB-effect is more sensitive to the electron structure of magnetic semiconductors than the Faraday effect, which suggests that by studying the dispersion of the kB-effect we can extract detailed information concerning the band structure, especially the asymmetry of the bands.

These and other ideas served as a basis for our present research, in which we study, both theoretically and experimentally, optical phenomena associated with magnetically induced spatial dispersion in the magnetic semiconductors $Cd_{1-x}Mn_xTe$. Since the experiments of Ivchenko *et al.*[13] and

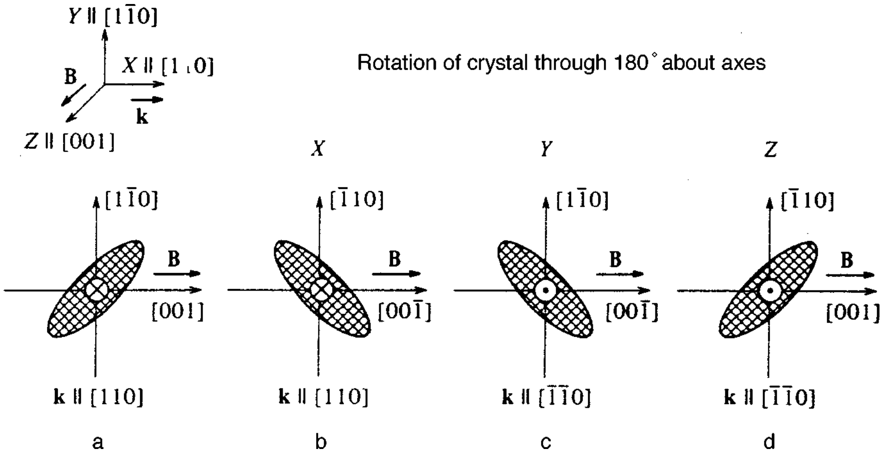


FIG. 1. Cross section of the indicatrix in the laboratory system of coordinates XYZ for $\mathbf{k}\|[110]$ and $\mathbf{B}\|[001]$ (a), and its variation under 180° rotations of the crystal about the X axis (b), the Y axis (c), and the Z axis (d).

Gogolin *et al.*,[14] based on measurements of the intensity of the light that has traveled through crossed polarizers, cannot, strictly speaking, conclusively prove that the observed phenomenon is odd in **k** and **B**, in this work we pay special attention to substantiating a method that makes it possible to obtain direct proof that this is indeed the case. In this method the measured quantity is determined directly by the product k_iB_j, and a change in sign of either of the vectors changes the sign of the effect. We present the results of measuring the field, angular, spectral, and concentration curves for birefringence linear in the magnetic field and for the Voigt and Faraday effects. The experimental data on the kB-effect are interpreted by a theory that takes into account the special features of the dispersion laws for the conduction and valence-band electrons, and the dependence of the exchange parameters on the electron wave vector.

2. PHENOMENOLOGICAL APPROACH

The permittivity tensor for a dia- or paramagnetic crystal in an external magnetic field **B**, with allowance for terms up to the second order in **B** and **k**, can be written[8–10]

$$\Delta\varepsilon_{ij}=\varepsilon_{ij}^{0}+\alpha_{ijk}B_k+\beta_{ijk}k_k+\gamma_{ijkl}B_kk_l+\nu_{ijkl}k_kk_l+\mu_{ijkl}B_kB_l, \tag{1}$$

where ε_{ij}^{0} is the permittivity tensor in zero magnetic field and without spatial dispersion. Here β_{ijk}, ν_{ijkl}, and μ_{ijkl} are polar tensors and α_{ijk} and γ_{ijkl} are axial tensors. In the transparent region the tensor α_{ijk} describes the Faraday effect, β_{ijk} the optical activity, γ_{ijkl} the effect of magnetically induced spatial dispersion (the kB-effect), ν_{ijkl} the Lorentz birefringence, and μ_{ijkl} the quadratic Voigt effect. The components of the tensors α_{ijk}, ν_{ijkl}, and μ_{ijkl} can be nonvanishing in crystals with any symmetry, while the components of β_{ijk} can only be nonvanishing in non-centrosymmetric crystals with optical activity. The tensor γ_{ijkl} is nonzero in any non-centrosymmetric crystal.

Let us examine the change in the optical properties of a crystal of the $\bar{4}3m$ class induced by a magnetic field. Optical activity is forbidden, so that changes in the optical properties are determined by the tensors α_{ijk}, μ_{ijkl}, and γ_{ijkl} The tensor α has only one component, μ has three components,[19] and γ has two: $A=\gamma_{xxyy}$ and $g=\gamma_{xyxy}$ (Ref. 15). In longitudinal geometry ($\mathbf{k}\|\mathbf{B}$), the Faraday effect is present, and in transverse geometry ($\mathbf{k}\perp\mathbf{B}$), the Voigt effect. The contribution of magnetically induced spatial dispersion (or the kB-effect) to the permittivity tensor has the form

$$\Delta\varepsilon_{ij}^{\gamma}=\gamma_{ijkl}B_kk_l=\begin{pmatrix} A(B_yk_y-B_zk_z) & g(B_xk_y-B_yk_x) & g(B_zk_x-B_xk_z)\\ g(B_xk_y-B_yk_x) & A(B_zk_z-B_xk_x) & g(B_yk_z-B_zk_y)\\ g(B_zk_x-B_xk_z) & g(B_yk_z-B_zk_y) & A(B_xk_x-B_yk_y)\end{pmatrix}, \tag{2}$$

where the x, y, and z axes correspond to the directions [100], [010], and [001] in the crystal. This contribution leads to birefringence linear in the magnetic field, which occurs both in the Faraday geometry and in the Voigt geometry. Obviously, in the Faraday geometry the linearly induced birefringence, being a higher-order effect, leads to a relatively small ellipticity compared to the Faraday rotation of the polarization plane. In the Voigt geometry there is no Faraday effect, and the kB-effect and the quadratic Voigt effect are of second order and lead to birefringence that can be accurately separated (as shown below) into individual contributions by using their differing symmetry properties and differing magnetic field dependence.

Below we analyze the most important cases of variation of the optical indicatrix when $\varepsilon_{ij}^{\gamma}$ is taken into account.

2.1. k∥[110] and B∥[001]

The tensor $\varepsilon_{ij}^{\gamma}$ can be reduced to principal axes by a 45° rotation about the z axis ($x'\|\mathbf{k}$) and a 45° rotation about the x' axis. The principal directions of the indicatrix ellipse are oriented at 45° to the direction of the magnetic field **B** (Fig. 1a). The amount of birefringence $\Delta n=gBk/n$, i.e., it depends only on the parameter g. Let us assume that in the laboratory system of coordinates XYZ, **k** points in the X direction and **B** in the Z direction. Using the transformation properties of $\varepsilon_{ij}^{\gamma}$, we can show that a 180° rotation about the X and Y axes results in a 90° rotation of the principal directions of the indicatrix (Figs. 1b and 1c), while a rotation about the Z axis does not alter these directions (Fig. 1d).

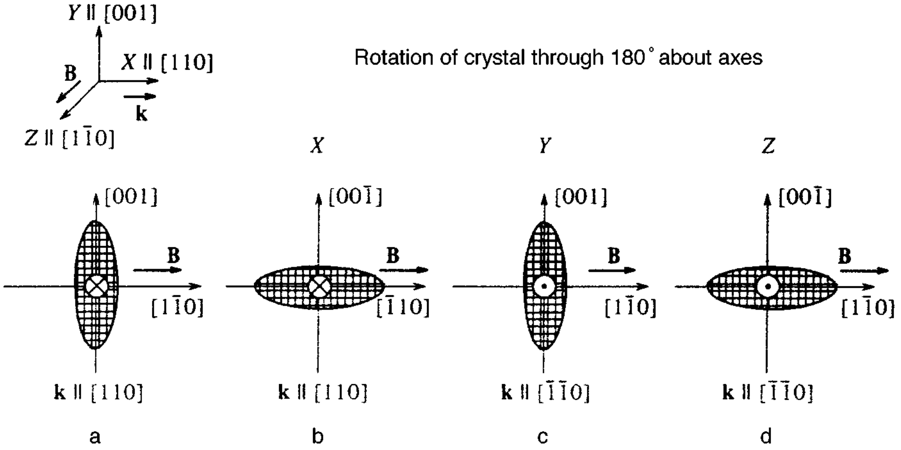


FIG. 2. Cross section of the indicatrix in the laboratory system of coordinates XYZ for $\mathbf{k}\|[110]$ and $\mathbf{B}\|[1\bar{1}0]$ (a), and its variation under 180° rotations of the crystal about the X axis (b), the Y axis (c), and the Z axis (d).

2.2. k∥[110] and B∥[1$\bar{1}$0]

The tensor $\varepsilon_{ij}^{\gamma}$ can be reduced to principal axes by a single 45° rotation about the z axis. One of the principal directions proves to be parallel to the magnetic field **B**, and the other is perpendicular to **B** (Fig. 2a). The amount of birefringence $\Delta n=(3A+2g)Bk/4n$ is determined by the parameters A and g. A rotation of the crystal about the X and Z axes leads to a rotation of the principal directions of the indicatrix (Figs. 2b and 2d), and a rotation about the Y axis does not change these directions (Fig. 2c). Note that the above changes in the orientation of the indicatrix are due to the linearity of the tensor $\Delta\varepsilon_{ij}^{\gamma}$ in **k** and **B**, and can be invoked as experimental proof that the effect is odd in **k**. The reversal of the sign of the effect upon reversal of the sign of **B** proves that the effect is also odd in **B**.

A 90° rotation in the (110) plane from the [001] axis to the [1$\bar{1}$0] axis forces the principal directions of the indicatrix to rotate by 45°. The dependence of the rotation angle φ of the principal directions of the indicatrix on the direction θ of the magnetic field is given by

$$\cot 2\varphi=Q\tan\theta, \tag{3}$$

where $Q=(3A+2q)/4g$; this is depicted in Fig. 3 for various values of Q. At $Q=\pm 1$ the principal directions of the indicatrix rotate through half the deviation of the magnetic field from the direction [001]. The ellipticity of the indicatrix does not change at $Q=1$. For $Q\neq\pm 1$, the rotation of the indicatrix is nonlinear with respect to field rotations, and the ellipticity depends on the direction of **B**.

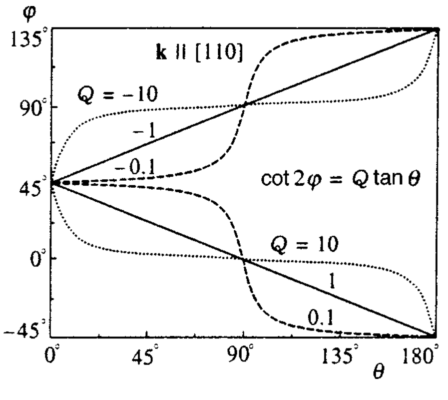


FIG. 3. Dependence of the azimuthal angle φ of the principal indicatrix direction on the direction (angle θ) of the magnetic field **B** for $\mathbf{k}\|[110]$. The values of φ and θ are measured from the [001] direction.

2.3. k∥[111] and B⊥[111]

In this case the principal directions are oriented at an angle of 45° when the magnetic field **B** is parallel to the [112] direction. When **B** is parallel to [110], one of the principal directions is parallel to **B** and the other is perpendicular to **B**. In both cases $\Delta n=(A+2g)kB/\sqrt{6}\,n$.

The presence of symmetry elements can lead to a situation in which an effect is forbidden in certain geometries. For instance, the kB-effect is forbidden if **k** is parallel to a fourfold symmetry axis, say of the [100] type. If **B** lies in a reflection plane perpendicular to the crystal surface, the indicatrix axes will be oriented at an angle of 45° to the field. If **B** is perpendicular to this plane, one of the principal directions of the indicatrix will be parallel to **B**, and the other will be perpendicular to **B**.

In the presence of the quadratic Voigt effect, variations of the optical indicatrix induced by the field are more complicated. Nevertheless, as we show in Sec. 3, the kB-effect and the Voigt effect can be separated.

3. MEASUREMENT METHODOLOGY

The optical layout of the experimental setup is depicted in Fig. 4. Light from the source (a helium–neon laser with $\lambda=0.633\,\mu$m or an Al_2O_3:Ti laser with $\lambda=0.7-0.83\,\mu$m) travels through a polarizer, through the sample in the gap of an electromagnet in which $\mathbf{B}\perp\mathbf{k}$, a quarter-wave plate, a Faraday modulator, and an analyzer, anending up at a photodiode. Such a layout is commonly used to measure birefringence,[20] and the polarization of the light after the polarizer must be parallel to one of the axes of the quarter-wave plate and form an angle of 45° with the principal directions of the cross section of the indicatrix. The angle of rotation of the analyzer corresponding to extinction is determined by birefringence and is half the phase shift between linearly polarized normal waves.

Two geometries were used in our experiments, $\mathbf{E}\|\mathbf{B}$ and $\widehat{\mathbf{EB}}=45°$ (Figs. 4a and 4b). In the $\mathbf{E}\|\mathbf{B}$ geometry (Fig. 4a), the input polarization $\mathbf{E}_1$ is parallel to **B** and the principal

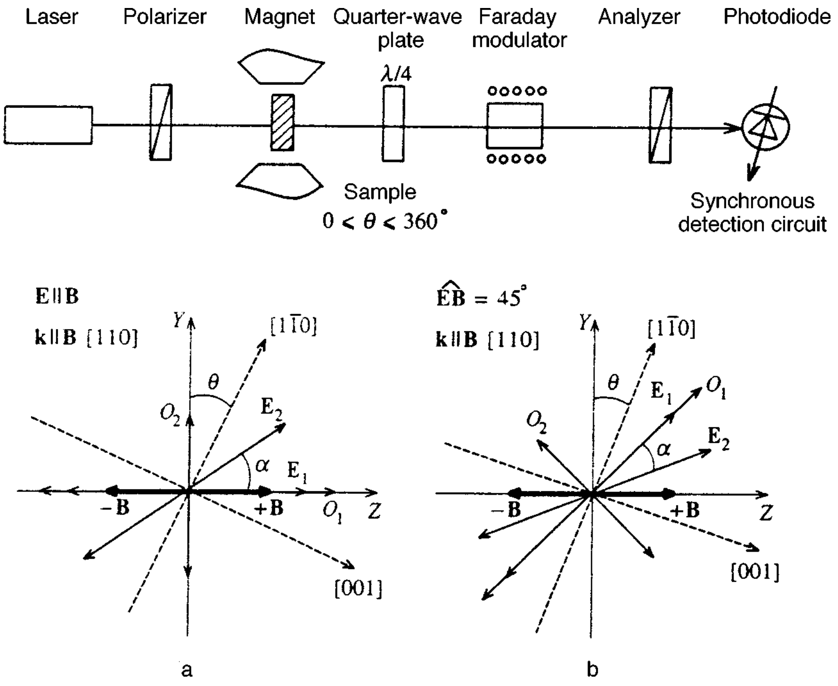


FIG. 4. Experimental setup for measuring the kB-effect. The mutual orientations of the magnetic field **B**, the input polarization $\mathbf{E}_1$ of the light, the principal directions O_1 and O_2 of the quarter-wave plate, and the output polarization $\mathbf{E}_2$ of the light are depicted in the lower half of the figure in the geometries $\mathbf{E}\|\mathbf{B}$ (a) and $\widehat{\mathbf{E}\,\mathbf{B}}=45°$ (b).

directions O_1 of the quarter-wave plate. In this case there is no quadratic Voigt effect, since its principal directions are parallel and perpendicular to **B**, but birefringence induced by the kB-effect is present. In the $\widehat{\mathbf{EB}}=45°$ geometry (Fig. 4b), the input polarization and the axis of the quarter-wave plate is oriented at 45° to **B**. In this case we have both the Voigt effect and the kB-effect, in whose geometry the axes are parallel and perpendicular to **B**. What is measured in both geometries is the rotation α of the output polarization $\mathbf{E}_2$ of light that has traveled through the crystal sample and the quarter-wave plate.

The crystal could be rotated in the gap of the electromagnet about an axis in the direction of **k**, with its azimuthal angle varying in the $0<\theta<360°$ range. It could also be rotated 180° about an axis perpendicular to **B** and **k** and about an axis parallel to **B**. The magnetic field could be varied between $-1.5\,\mathrm{T}$ and $+1.5\,\mathrm{T}$. The magnitude of the linear birefringence due to internal stresses in the crystal and Lorentz birefringence was determined in zero magnetic field. The sensitivity in measuring the rotation of the polarization plane was $10''$. The measurements were carried out at $T=294\,\mathrm{K}$. We found it important to eliminate the Faraday effect and magnetic circular dichroism resulting from slight misalignments of the magnetic field relative to **k**, since these phenomena are also linear in the magnetic field and can substantially exceed the kB-effect. To this end, in both geometries we measured the field dependence of $\alpha(B)$ without the quarter-wave plate. The fact that there was no rotation of the polarization plane indicated that **B** and **k** were strictly perpendicular.

We studied the spectral dependence of the Faraday effect and the absorption spectra for all crystal samples. The dispersion of the refractive index was investigated in the transparency region by measuring the rotation of the polarization plane of light reflected by the surface of the crystal for different angles of incidence. The refractive index n was calculated by the Fresnel formulas.

The parameters of $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ single crystals ($x=0$, 0.25, 0.35, 0.42, and 0.52) used in our experiments are listed in Table I. E_g was calculated using the formulas given in Ref. 1, with the values of the cell parameters taken from Table I. The samples were cut in the (100), (110), and (111) planes and polished down to $2\times3\times0.7\,\mathrm{mm}^3$ plates. The orientation of the samples was determined radiographically by the Bragg reflection method. To check for crystal defects and the presence of twinning and aggregates, we studied the Laue diffraction patterns in reflected light. Note that since the kB-effect is anisotropic, the presence of twins or aggregates can weaken the observed effect and distort the nature of the anisotropy (e.g., the effect can show up in forbidden geometries ($\mathbf{k}\|[100]$)). Spontaneous birefringence Δn of all samples was at most roughly 5×10^{-6}.

4. EXPERIMENTAL RESULTS

For samples of type (110) with $\mathbf{E}\|\mathbf{B}$ geometry, the observed birefringence was linear in the magnetic field. As a result of azimuthal rotation of the crystal about the X axis, which was parallel to **k**, the slope of the linear α *(B)* curves was found to vary substantially. Figure 5a depicts the function $\alpha(B)$ at various azimuthal angles θ for a crystal with

TABLE I. Parameters of the $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ samples.

x	Cell parameter, Å	E_g (300 K), eV
0.25	6.449(1)	1.86
0.35	6.435	1.99
0.42	6.423	2.08
0.52	6.409	2.21

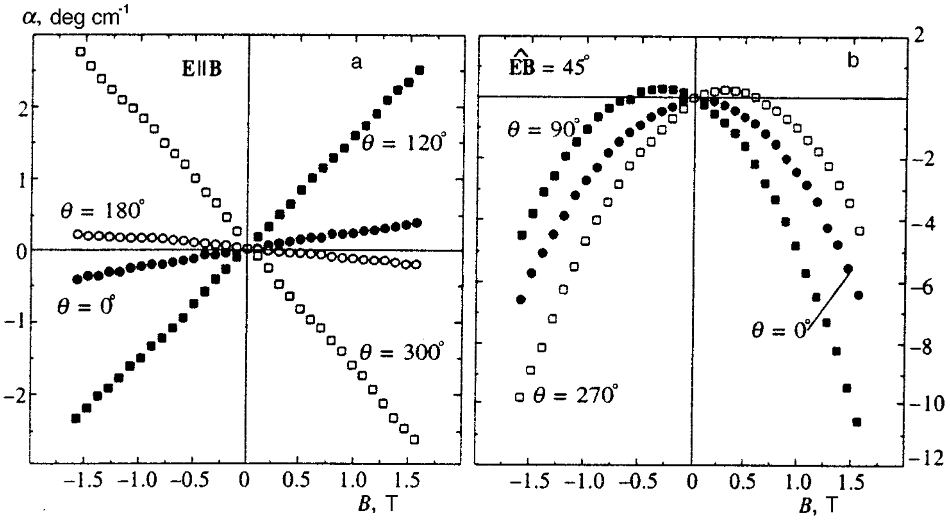


FIG. 5. Magnetic-field dependence of the kB-effect in $Cd_{1-x}Mn_xTe(x=0.42)$ for various directions of the magnetic field in the geometries $\mathbf{E}\|\mathbf{B}$ (a) and $\widehat{\mathbf{EB}}=45°$ (b).

$x=0.42$ with $\lambda=0.633\,\mu$m. The angle dependence of $\partial\alpha/\partial B$ under a rotation of the crystal about the X axis is shown in Fig. 1a. A 180° rotation about the X axis changes the sign of the derivative, and the $\partial\alpha/\partial B\ (\theta)$ dependence can be described by a combination of first- and third-order harmonics $\cos\theta$ and $\cos 3\theta$. The effect disappears at $\mathbf{B}\|[110]$, when the magnetic field is perpendicular to the symmetry plane normal to the sample's surface. The slope of the $\alpha\ (B)$ curve changes sign when the crystal is rotated 180° about the Y axis, and does not change when the rotation is about the Z axis.

In the $\widehat{\mathbf{EB}}=45°$ geometry, the $\alpha(B)$ field dependence for crystals of the (110) type are asymmetric with respect to the value $B=0$ and are described by the sum of the contribution quadratic and linear in the magnetic field. Figure 5b depicts the function $\alpha(B)$ at different values of the azimuthal angle θ for a crystal sample with $x=0.42$. In all crystal samples the contribution quadratic in the magnetic field is, within experimental error, independent of the direction of $\mathbf{B}$, which indicates that the Voigt effect is isotropic. This is certainly not a trivial result since cubic symmetry allows for an anisotropic Voigt effect. The contribution to $\alpha(B)$ that is linear in the magnetic field depends on the orientation of $\mathbf{B}$. Figure 6b depicts the angular dependence of $\partial\alpha/\partial B$. The effect disappears when the magnetic field is perpendicular to the symmetry plane normal to the surface. A 180° rotation of the crystal about the Z axischanges the sign of $\partial\alpha/\partial B$, while a 180° rotation about the Y axisdoes not.

For samples of the (111) type, the angular dependence of the slope of the $\alpha\ (B)$ curve, i.e., the $\partial\alpha/\partial B\ (\theta)$ dependence, whose contribution is linear in B, is described in both geometries, $\mathbf{E}\|\mathbf{B}$ and $\widehat{\mathbf{EB}}=45°$, by the third harmonics.

Note that the observed patterns, which are linear in the magnetic field, cannot be linked to the Faraday effect or magnetic circular dichroism, even in conjunction with spontaneous linear birefringence. There are several facts that suggest this:

1. The Faraday effect and magnetic circular dichroism in a cubic crystal are isotropic. Indeed, it has been established in experiments that in $Cd_{1-x}Mn_xTe$ crystals the size of the Faraday effect is independent of the direction of $\mathbf{k}$ with respect to the crystallographic axes. Hence all possible spurious effects associated with a combination of the Faraday effect and spontaneous birefringence are described by the even harmonics of the angle θ and do not change sign when the crystal is rotated 180° about the X axis.

2. Rotations of the polarization plane due to the Faraday effect, which are related to the fact that $\mathbf{B}$ is not exactly perpendicular to $\mathbf{k}$, cannot lead to a rotation of the polarization plane after the light has traveled through the quarter-

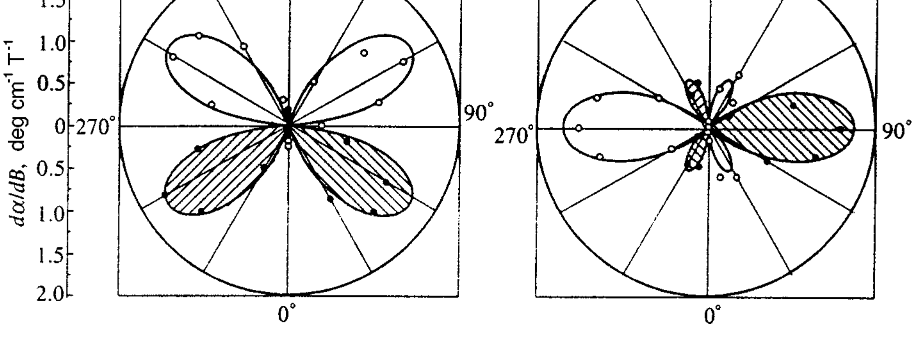


FIG. 6. Rotational anisotropy of the kB-effectin $Cd_{1-x}Mn_xTe$ $(x=0.42)$ in the geometries $\mathbf{E}\|\mathbf{B}$ (a) and $\widehat{\mathbf{EB}}=45°$ (b). Solid curves correspond to calculated dependence.

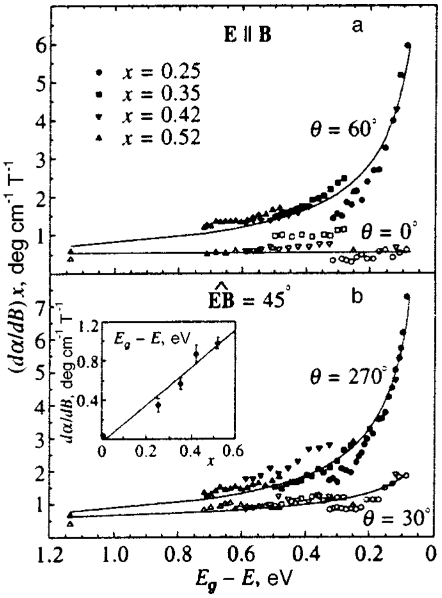


FIG. 7. Normalized kB-effect $(\partial\alpha/\partial B)/x$ as a function of E_g-E at extrema of the angular dependence in the geometries $\mathbf{E}\|\mathbf{B}$ (a) and $\widehat{\mathbf{EB}}=45°$ (b). The inset depicts the concentration dependence of $(\partial\alpha/\partial B)/x$ for $(E_g-E)=0.45$ eV.

wave plate. Such rotations lead to changes in ellipticity, which do not show up in the adopted measuring methodology.

3. It has been established in experiments that rotations of the magnet through angles of about 1°, which lead to a Faraday effect, have an extremely small influence on the α *(B)* dependence in both geometries.

Figures 7a and 7b depict the normalized kB-effect $(\partial\alpha/\partial B)/x$ as a function of E_g-E in both geometries, $\mathbf{E}\|\mathbf{B}$ and $\widehat{\mathbf{EB}}=45°$, where E_g is the band gap and E is the phonon energy. The measurements were made at angles θ corresponding to the extrema of the angular dependence of the kB-effect (see Figs. 6a and 6b). The inset in Fig. 7 depicts the concentration dependence of the kB-effect, $\alpha(x)$, at $E_g-E=0.45$ eV, including data on undoped CdTe $(x=0)$. The refractive index n of $Cd_{1-x}Mn_xTe$ in the transparent region is close to 3, and was found to vary by 10 %in the spectral range being investigated (see Table II).

5. DISCUSSION

Our results clearly show that $Cd_{1-x}Mn_xTe$ crystals exhibit a nonreciprocal kB-effect, which is described by the tensor γ_{ijkl} in (1). This is suggested by: (1) the linear dependence of α on the magnetic field; (2) the specific behavior of the optical indicatrix when the crystal is rotated 180° about an axis perpendicular to both $\mathbf{B}$ and $\mathbf{k}$ and about an axis parallel to $\mathbf{B}$; and (3) the azimuthal dependence of $\alpha(B)$ when the crystal is rotated about the X axis, which is parallel to $\mathbf{k}$ (in particular, the change in the sign of the effect as $\theta\rightarrow\theta+180°$). The behavior of the indicatrix under a rotation of the crystal about the X, Y, and Z axes is in full agreement with the conclusions of the symmetry study in Sec. 2, and proves that the effect is odd in $\mathbf{k}$.

The angular dependence of the kB-effect was demonstrated by two different approaches. The first, based on the Jones matrix method, made it possible to calculate the angular and field curves for $\alpha(B)$ in the geometries $\mathbf{E}\|\mathbf{B}$ and $\widehat{\mathbf{EB}}=45°$ in the (110) and (111) planes with allowance for the Voigt effect and spontaneous birefringence Δn of about 5×10^{-6}. The calculations showed that in this case, over the range of field strengths used in the experiments, this dependence on field and angle has no effect on contributions to $\alpha(B)$ linear in the magnetic field. This provided an easy way to separate the effects linear in the magnetic field from the quadratic effects. The second approach, based on approximate expressions, made it possible to calculate the angular and field curves for $\alpha(B)$ with allowance for spontaneous birefringence and the Voigt effect for an arbitrary direction of $\mathbf{k}$. In the (110) and (111) planes both approaches were found to yield the same result. For an arbitrary direction of $\mathbf{k}$, the $\partial\alpha(\theta)/\partial B$ dependence is described by first and third harmonics in θ, whose amplitudes are determined by A and g and the direction cosines of $\mathbf{k}$.

We can calculate the dispersion of the parameters A and g by using the $\partial\alpha(E_g-E)/\partial B$ dependence at the angles θ corresponding to the extrema of angular curves for A and g in the geometries $\mathbf{E}\|\mathbf{B}$ and $\widehat{\mathbf{EB}}=45°$. The spectral curves for the normalized parameters A/x and g/x calculated from the $\partial\alpha/\partial B$ vs. (E_g-E) curves in Figs. 7a and 7b are depicted in Fig. 8a. The spectral curves for A/x and g/x indicate that the dispersion of the parameter Q describing the anisotropy of the kB-effect is large: Q changes from 2.1 at $E_g-E=1.1$ eV to 11.8 at $E_g-E=0.1$ eV. The anisotropy of the kB-effectincreases as the difference E_g-E decreases, and as we move away from E_g the kB-effect approaches the isotropic case $Q=1$.

Figure 8b depicts the spectral curves for the normalized values of the Faraday effect (FR/x) and the Voigt effect (VB/x^2). The curves in Figs. 8a and 8b are remarkable because they make it possible for the first time to compare the value and dispersion of three different magnetooptic effects

TABLE II. Refractive index of $Cd_{1-x}Mn_xTe$ crystals.

	E, eV					
x	1.96	1.71	1.67	1.63	1.56	1.50
0*					3.05	3.00
0.25		3.22(2)	3.13(2)	3.11(2)	3.08(1)	3.00(1)
0.42	3.26(5)	3.05(2)		3.00(3)	2.99(2)	2.93(2)
0.52	2.97(2)	2.99(2)	2.85(5)		2.86(5)	

**Note.* Values of n for CdTe from Ref. 30.

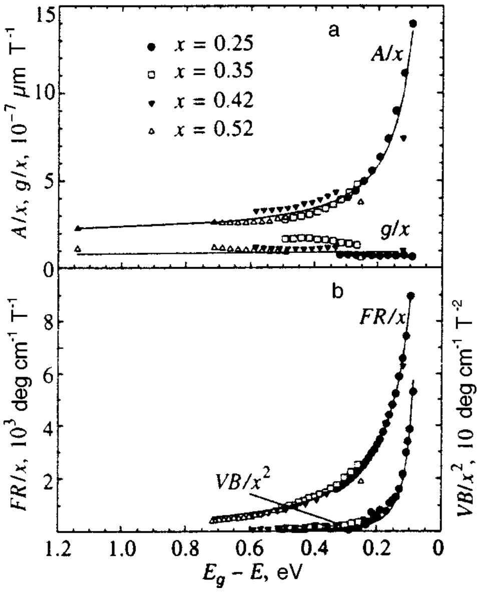


FIG. 8. Normalized parameters A/x and g/x and the normalized values of the Faraday effect (FR/x) and the Voigt effect (VB/x^2) as functions of (E_g-E). Solid curves correspond to calculated dependence.

across a group of crystals that have differing concentrations x. The Voigt effect is approximately a thousand times larger than the kB-effect. Near the edge of the forbidden band in a magnetic field $B=1$ T, the kB-effect and the Voigt effect are approximately of the same size, but as we move away from the edge, the Voigt effect decreases much faster than the kB-effect and at $E_g-E=0.3\,\mathrm{eV}$ can be ignored. Eunsoon *et al.*[3] studied the behavior of the Voigt effect at low temperatures. As the temperature decreases, the size of the Voigt effect increases, and at $T\simeq 10\,\mathrm{K}$ it is approximately a hundred times larger than at $T=294\,\mathrm{K}$.

The dependence of the normalized parameters A/x and g/x on E_g-E for crystals with different manganese content "fit" universal, i.e., x-independent, curves (Fig. 8a). Universal curves are also a characteristic feature of the normalized Faraday effect FR/x and the normalized Voigt effect VB/x^2 (Fig. 8b). The existence of universal curves indicates that for a fixed difference E_g-E the Faraday effect and the kB-effectare linear in x and the Voigt effect is quadratic in x. The fact that the kB-effect is a linear function of x and the fact the kB-effect in undoped CdTe is tens times smaller than in crystals with manganese are indications that the Mn^{2+} ions provide the leading contribution to the kB-effect.

The dispersion of the three magnetooptic effects in the different samples can be described by the function $d+t(E_g-E)^{-\tau}$, where d, t, and τ are parameters. For the normalized component A/x we have $\tau=1.4$, for the Faraday effect $\tau=1.5$, and for the Voigt effect $\tau=3.5$. The off-diagonal component g/x depends very weakly on the frequency ($t=0$). For A/x we have $d\simeq 2.0\times 10^{-7}\,\mu\mathrm{m\,T}^{-1}$ and for g/x we have $d\simeq 1.1\times 10^{-7}\,\mu\mathrm{m\,T}^{-1}$, while for the Faraday and Voigt effects we have $d=0$.

Note that the overall sign of the kB-effect is undefined even if we know the type of plane from which the crystal has been cut, the direction of $\mathbf{k}$, and the direction of the magnetic field $\mathbf{B}$. The sign of the kB-effect can be determined only for an unambiguously fixed orientation of the unit cell in the laboratory system of coordinates. Note that this is a difficult task for both radiographic and neutron-diffraction methods. In principle, the kB-effect can be employed in orienting a crystal if we use a standard sample.

6. THEORY

The early theoretical work on the kB-effectanalyzed the microscopic mechanisms involved in the excitonic[13,14] or intraband[21,22] transitions, so that these results cannot be used to analyze our data. In accordance with the experimental conditions ($E<E_g$), here we theoretically analyze the kB-effect that is the result of transitions from the valence band Γ_8 to the conduction band Γ_6.

Equation (1) shows that the tensor γ_{ijkl} is the derivative of the permittivity tensor $\varepsilon_{ij}(w,\mathbf{k},\mathbf{B})$ (Ref. 9) with respect to k_l and B_k:

$$\gamma_{ijkl}=\frac{4\pi\hbar^2}{E^2V}\frac{\partial}{\partial k_l\,\partial B_k}\sum_{r,s,\mathbf{q}}\left[\frac{J^i_{s\mathbf{q},r\mathbf{q}+\mathbf{k}}(\mathbf{k})J^j_{r\mathbf{q}+\mathbf{k},s\mathbf{q}}(-\mathbf{k})}{E_{r\mathbf{q}+\mathbf{k}}-E_{s\mathbf{q}}-E}+\frac{J^i_{r\mathbf{q}-\mathbf{k},s\mathbf{q}}(\mathbf{k})J^j_{s\mathbf{q},r\mathbf{q}-\mathbf{k}}(-\mathbf{k})}{E_{r\mathbf{q}-\mathbf{k}}-E_{s\mathbf{q}}+E}\right]\Bigg|_{\mathbf{k},\mathbf{B}\to 0}, \tag{4}$$

where V is the crystal volume, $r=\pm 1$ and $s=\pm 1,\pm 3$ label the states of the bands Γ_6 and Γ_8, respectively, and $\mathbf{J}(\mathbf{k})$ is the Fourier transform of the current operator.

To calculate the tensor $\hat{\gamma}$ we must take into account the noncentral nature of the crystal's symmetry and the external magnetic field. However, if we allow only for the second factor, we can calculate the spectrum and wave functions of the electrons only numerically.[23] When we calculate $\hat{\gamma}$, the problem becomes even more complicated because in addition to the magnetic field we must take into account the noncentral symmetry of the crystal. Nevertheless, the most important characteristic of $\hat{\gamma}$, the frequency behavior of $\hat{\gamma}$ near E_g, can be determined knowing only the dependence of the electron energy and the matrix elements of the current operator on $\mathbf{q}$ near the center of the Brillouin band.

We take into account only the first term in (4), since the second term yields only a contribution to $\hat{\gamma}$ that is weakly frequency-dependent and therefore is negligible near the absorption edge. For the same reason we calculate only the most rapidly varying contributions (with frequency) to $\hat{\gamma}$ near E_g, which emerge in the process of differentiation with respect to k_l and B_k. In particular, we ignore the dependence of the matrix elements of the current operator on the magnetic field, i.e., in (4) only the energy denominator is differentiated with respect to B_k. Here we use the dependence of the band energies $E_{r\mathbf{q}+\mathbf{k}}$ and $E_{s\mathbf{k}}$ on $\mathbf{B}$ in the limit where $\mathbf{B}$ is small. Under such conditions the magnetic field leads to anisotropic splitting of the bands of light holes (lh) and heavy holes (hh) and to isotropic splitting of the conduction band (c) (see Refs. 23 and 24):

$$\Delta E_{lh}(\mathbf{q},\mathbf{B})=\pm b\sqrt{4-3\cos^2\theta}, \tag{5}$$

$$\Delta E_{hh}(\mathbf{q},\mathbf{B})=\pm 3b\cos\theta, \tag{6}$$

$$\Delta E_c(\mathbf{q},\mathbf{B}) = \pm 3a, \tag{7}$$

where θ is the angle between the electron wave vector $\mathbf{q}$ and the average spin vector of Mn^{2+} ions, or $\langle \mathbf{S}^{\mathrm{Mn}} \rangle$. The constants a and b are proportional to S_z^{Mn} ($\langle \mathbf{S}^{\mathrm{Mn}} \rangle \| z$) and describe the exchange interaction of Mn^{2+} ions and band electrons:[24]

$$a = -N_0 \alpha \langle S_z^{\mathrm{Mn}} \rangle x, \quad b = -N_0 \beta \langle S_z^{\mathrm{Mn}} \rangle x, \tag{8}$$

where N_0 is the number of unit cells per unit volume, and α and β are the exchange integrals for the conduction and valence bands.

As for differentiation with respect to k_l, in addition to allowing for the dependence of the electron energy in the conduction band, $E_{c\,\mathbf{q}+\mathbf{k}}$, we must also allow in (4) for the $\mathbf{k}$-dependence of the matrix elements of the current operator, i.e., we must differentiate with respect to k_l not only the denominator but also the numerator. Let us consider the problem in detail.

In crystals with a zinc-blende structure there is no $\mathbf{q}$-linear splitting of the conduction band; the effect of the noncentral symmetry of the crystal, which manifests itself only when one allows for spin–orbit coupling, is described by terms starting with the one that is cubic in $\mathbf{q}$.[25] Furthermore, there is no $\mathbf{q}$-linear splitting of the band in a magnetic field either, i.e., the dispersion law for the conduction electrons does not contain a term bilinear in $\mathbf{q}$ and $\mathbf{B}$ (see, e.g., Ref. 26). The derivative $\partial E_c(\mathbf{q})/\partial \mathbf{q}$, which appears in (4) when we differentiate the energy denominator with respect to $\mathbf{k}$, has the form

$$\frac{\partial E_{r\mathbf{q}}}{\partial \mathbf{q}} = \frac{\hbar^2 \mathbf{q}}{m_c} + r\,\delta_0 \mathbf{f}(\mathbf{q}), \tag{9}$$

where $\mathbf{f}$ is a quadratic function of $\mathbf{q}$, and δ_0 is the inversion asymmetry parameter, responsible for the spin splitting of the conduction band. Integrating, we see that the contribution to $\gamma_{xxyy} = A$ proportional to δ_0 varies near E_g as $(E_g - E)^{-1/2}$, while $\gamma_{xyxy} = g = 0$. Note that the valence electron energy in (4) does not depend on the photon momentum $\mathbf{k}$. For this reason the derivative $\partial E_{s\mathbf{q}}(\mathbf{q})/\partial q_l$ does not appear when we differentiate with respect to k_l in (4). Nevertheless, the $\mathbf{q}$-odd splitting of the valence band determines the order of the singularity of the integrand in (4) and has a profound effect on the corresponding contribution to $\hat{\gamma}$. This splitting of the Γ_8 band can be described by terms linear in $\mathbf{q}$ (see, e.g., Ref. 27) thanks to the term

$$\Delta H_v = \frac{4}{\sqrt{3}} C_0 [q_x \{ J_x (J_y^2 - J_z^2) \} + \mathrm{c.p.}] \tag{10}$$

in the effective valence-band Hamiltonian (here J is the matrix of the angular momentum operator in the base Γ_8, c.p. stands for cyclic permutation, and the braces indicate symmetrization). In estimating the contribution of the linear splitting of the valence band, we must allow only for the first term in (9). Calculating $\hat{\gamma}$ in the same approximation as before, i.e., keeping only the most rapidly varying terms (with frequency), we find that the contribution of the $\mathbf{q}$-linear splitting of the valence band exhibits the same frequency dependence as the corresponding contribution from the conduction band.

The lack of $\mathbf{q}$-linear splitting of the Γ_6 band means that we must take into account the $\mathbf{q}$-linear terns in the matrix elements of the current operator in (4). For the band-to-band transitions considered here, the electron velocity operator $\mathbf{v}$, which enters into the current operator $\mathbf{J}(\mathbf{k}) = e(\mathbf{v} e^{-i\mathbf{k}\cdot\mathbf{r}} + e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{v})/2$, has the form

$$v_i = \sqrt{3}/\hbar [P R_i + i B s_{inm} q_n R_m], \tag{11}$$

where $\mathbf{R}$ is the polar vector operator in the $\Psi_{\Gamma_6}, \Psi_{\Gamma_8}$ basis, s_{inm} is a completely symmetric tensor, and P and B are the Kane parameters, with B representing the fact that the $\bar{4}3m$ group has no center of inversion. An estimate of the contribution to $\hat{\gamma}$ of the second term on the right-hand side of Eq. (11) yields a frequency dependence of the components of $\hat{\gamma}$ similar to the one obtained earlier. Note that since there is no linear splitting of the conduction band, we must allow for the $\mathbf{q}$-linear terms in the velocity operator (11).

The fact that $g = 0$ is obviously the result of ignoring the weakly frequency-dependent terms in (4). At the same time, it agrees qualitatively with our experimental data, according to which g is weakly frequency-dependent and is small compared to A, especially near E_g.

Note that thanks to the term (10), the effective Hamiltonian of an exciton coupled to the valence band and the conduction band contains $\mathbf{q}$-linear terms and terms bilinear in $\mathbf{q}$ and $\mathbf{B}$. This leads to the same frequency dependence of A and g ($\sim (E - E_{\mathrm{ex}})^{-2}$) in the excitonic part of the spectrum.[14]

The frequency dependence $A \sim (E_g - E)^{-1/2}$ calculated above proves to slower than that obtained in the experiment, $A \sim (E_g - E)^{-1.4}$. A similar situation emerges in the interpretation of Faraday-rotation experiments in $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ (Ref. 4), where the dependence of the rotation angle observed near E_g, $\phi \sim (E_g - E)^{-3/2}$, proves to be much more rapid than that given by theory, $\phi \sim (E_g - E)^{-1/2}$ (Ref. 6), and observed in cubic nonmagnetic semiconductors (see, e.g., Ref. 7).

Hugonnard-Bruyère *et al.*[4] show that this discrepancy between theory and experiment can be removed if one allows for the dependence of the exchange integrals α and β in (8) on the wave vector.[28] The extent to which this factor influences the magnetooptic effects depends on the size q_0 of the region near the center of the Brillouin zone, where α and β do not change substantially. If in the above estimates of the tensor $\hat{\gamma}$ we allow for the dependence of α and β on q, assuming as in Ref. 4 that

$$\alpha, \beta \sim \frac{q_0^2}{q_0^2 + q^2}, \tag{12}$$

then over the range of photon energies specified by the fact that the parameter $\kappa = m_{ch} E_g / \hbar^2 q_0^2 (1 - E/E_g) \sim 1$ is of order unity, with $m_{ch}^{-1} = m_c^{-1} + m_{hh}^{-1}$, the behavior of the function $A(E)$ changes from $A \sim (E_g - E)^{-1/2}$ for $\kappa \ll 1$ to $A \sim (E_g - E)^{-3/2}$ for $\kappa \gg 1$.

Note that the contribution of light holes to the magnetooptic effects in $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ can be ignored.[4] For $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$, the parameter $m_{ch}E_g/\hbar^2 q_0^2$ is roughly 50 (see Ref. 4), which in our experiment corresponds to $\kappa \gg 1$. If we adopt this assumption, we obtain the dependence

$$A \sim (E_g - E)^{-3/2} + d(E), \tag{13}$$

where $d(E)$ is a slowly varying function of the photon energy, which takes into account the contribution of the terns dropped in (4). This dependence is in good agreement with the result of our experiment. What is important is that the dependence of α and β on the wave vector alters the behavior of the parameter A over a broad range of energies E not too close to E_g. This sets it apart from the influence of the electron–hole Coulomb interaction on dispersion effects,[29] which in the $E \leqslant E_g$ range shows up only near E_g, where the energy deficit $E_g - E$ is comparable to the exciton binding energy $R \simeq 10$ meV.

Note that despite the fact that near E_g the frequency dependence of Faraday rotation coincides with that of the kB-effect, microscopically they are quite different. Our analysis shows that the spectral behavior of the kB-effect is sensitive to the nature of the dispersion of the electron bands, while Faraday rotation is only weakly dependent on the dispersion law.

Since the magnitude of the kB-effect is linear in the inversion asymmetry parameters, it can be used to determine these parameters. However, solving this problem requires calculating dispersion laws and matrix elements, and lies outside the scope of the present paper.

7. CONCLUSION

The principal results of the present investigation are a symmetry analysis of the phenomenon of magnetically induced spatial dispersion in non-centrosymmetric cubic crystals of the $\bar{4}3m$ group, the experimental study of nonreciprocal birefringence of light (the kB-effect) in $\mathrm{Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ crystals $(0 \leqslant x \leqslant 0.52)$, and the development of a microscopic theory of the phenomenon. We have found that in magnetic semiconductors this effect is large even at room temperature, which is related to the strong splitting of electron states due to the sp–d exchange interaction. Although in magnetic semiconductors the kB-effect is several orders of magnitude weaker than the Faraday effect, it is comparable in magnitude to typical values of the Faraday effect in many dia- and paramagnetic materials.

On the other hand, near E_g the Voigt effect and the kB-effect are comparable in magnitude, while far from E_g the kB-effect may substantially exceed the Voigt effect. A characteristic feature of the observed birefringence is its strong anisotropy with respect to the directions of $\mathbf{B}$ and $\mathbf{k}$, which sets it apart from such well-known magnetooptic phenomena as the Faraday effect and the Voigt effect. The spectral curves for the parameters A and g describing the kB-effect exhibit different dispersion as E_g is approached, which suggests that the corresponding microscopic mechanisms are different. The anisotropy parameter Q of the kB-effect has a very large dispersion, with the anisotropy of the effect decreasing as we move away from the band edge.

We have shown that the spectral curves for A obtained in our experiments can be interpreted if we simultaneously allow for direct band-to-band transitions between the heavy-hole band and the conduction band, the dependence of the current operator on the wave vector, and the dependence of the exchange parameters on the electron wave vector. The parameter g is zero in the adopted approximation and requires other mechanisms for its interpretation. Since inherently the kB-effect is due to the non-centrosymmetric nature of the crystal, it can be used to determine the parameter describing the linear dispersion of electrons in semiconductors.

ACKNOWLEDGMENTS

We are grateful to G. K. Averkieva for providing single crystals and to N. F. Kartenko for doing the radiographic research. The present work was supported by the Russian Foundation for Basic Research, the Fundamental Spectroscopy Program, and Deutsche Forschungsgemeinschaft.

*) E-mail: krichev@star.shuv.pti.spb.su
†) E-mail: pisarev@star.shuv.pti.spb.su
‡) E-mail: gridnev@star.shuv.pti.spb.su

---

[1] J. K. Furdyna, J. Appl. Phys. **64**, R29 (1988).

[2] P. I. Nikitin and A. I. Savchuk, Usp. Phys. Nauk **160**, No. 2, 167 (1990) [Sov. Phys. Usp. **33**, 974 (1990)].

[3] Eunsoon Oh, D. U. Bartholomew, A. K. Ramdas, J. K. Furdyna, and U. Debska, Phys. Rev. B **44**, 10 551 (1991).

[4] S. Hugonnard-Bruyère, C. Buss, F. Vouilloz, R. Frey, and C. Flytzanis, Phys. Rev. B **50**, 2200 (1994).

[5] C. Buss, S. Hugonnard-Bruyère, R. Frey, and C. Flytzanis, Solid State Commun. **92**, 929 (1994).

[6] I. M. Boswarva, R. E. Howard, and A. B. Lidiard, Proc. R. Soc. London, Ser. A **269**, 125 (1962).

[7] J. G. Mavroides, in *Optical Properties of Solids*, F. Abelès (Ed.), North-Holland, Amsterdam (1972).

[8] D. L. Portigal and E. Burstein, J. Phys. Chem. Solids **32**, 603 (1971).

[9] V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion and Excitons*, 2nd ed., Springer-Verlag, Berlin (1984).

[10] P. Etchegoin, A. Fainstein, P. Santos, L. C. Lew Yan Voon, and M. Cardona, Solid State Commun. **92**, 505 (1994).

[11] J. J. Hopfield and D. G. Thomas, Phys. Rev. Lett. **4**, 357 (1960).

[12] E. F. Gross, B. P. Zakharchenya, and O. V. Konstantinov, Fiz. Tverd. Tela (Leningrad) **3**, 305 (1961) [Sov. Phys. Solid State **3**, 221 (1961)].

[13] E. L. Ivchenko, V. P. Kochereshko, G. V. Mikhaĭlov, and I. N. Ural'tsev, JETP Lett. **37**, 164 (1983); Phys. Status Solidi B **121**, 221 (1984).

[14] O. V. Gogolin, V. A. Tsvetkov, and E. G. Tsitsishvili, Zh. Éksp. Teor. Fiz. **87**, 1038 (1984) [Sov. Phys. JETP **60**, 593 (1984)].

[15] E. G. Tsitsishvili, Fiz. Tekh. Poluprovodn. **20**, 650 (1986) [Sov. Phys. Semicond. **20**, 412 (1986)].

[16] R. V. Pisarev, V. V. Krichevtsov, and V. V. Pavlov, Phase Transit. **37**, 63 (1991).

[17] V. V. Krichevtsov, V. V. Pavlov, R. V. Pisarev, and V. N. Gridnev, Phys. Rev. Lett. **76**, 4628 (1996).

[18] V. N. Gridnev, V. V. Krichevtsov, V. V. Pavlov, and R. V. Pisarev, JETP Lett. **65**, 68 (1997).

[19] Yu. I. Sirotin and M. P. Shaskol'skaya, *Fundamentals of Crystal Physics*, Mir Publishers, Moscow (1982).

[20] J. Ferré and G. A. Gehring, Rep. Prog. Phys. **47**, 526 (1984).

[21] É. I. Rashba and V. I. Sheka, Fiz. Tverd. Tela (Leningrad) **3**, 1735 (1961) [Sov. Phys. Solid State **3**, 1257 (1961)].

[22] Y.-F. Chen, N. Dobrowolska, J. K. Furdyna, and S. Rodriguez, Phys. Rev. B **32**, 890 (1985).

[23] J. A. Gaj, J. Ginter, and R. R. Galazka, Phys. Status Solidi B **89**, 655 (1978).

[24] J. A. Gaj, in *Diluted Magnetic Semiconductors*, Vol. 25 of *Semiconductors and Semimetals*, J. K. Furdyna and J. Kossut (Eds.), Academic Press, Boston (1988), p. 275.

[25] E. Kane, in *Semiconductors and Semimetals*, Vol. 1, R. K. Willardson and A. C. Beer (Eds.), Academic Press, New York (1966), p. 75.

[26] N. R. Ogg, Proc. Phys. Soc. London B **89**, 431 (1966).

[27] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, Wiley, New York (1974).

[28] A. K. Bhattacharjee, Phys. Rev. B **41**, 5696 (1990).

[29] A. G. Aronov and A. S. Ioselevich, Fiz. Tverd. Tela (Leningrad) **20**, 2615 (1978) [Sov. Phys. Solid State **20**, 1511 (1978)].

[30] D. T. F. Marple, J. Appl. Phys. **35**, 539 (1964).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Rigorous three-dimensional theory of the parametric excitation of space-charge waves in semiconductors

B. I. Sturman,*) A. I. Chernykh, and E. A. Shamonina

International Institute for Nonlinear Studies, Siberian Branch, 630090 Novosibirsk, Russia

(Submitted 27 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1034–1046 (September 1998)

We analyze the stability against parametric excitation of space-charge waves of the space-charge field induced in a semiconducting crystal by a traveling light grating. We show that when the grating velocity is low, an important element of the analysis is the allowance for higher Fourier harmonics of the field. By combining analytic and numerical methods we study the stability against an increase in the amplitude of small three-dimensional perturbations of a general type. We find that instability is possible only within a single range of light-pattern velocities and that it leads to selective excitation of one-dimensional perturbations. Finally, we use the results of our analysis to interpret experiments on the generation of spatial subharmonics in sillenite crystals. © *1998 American Institute of Physics.* [S1063-7761(98)01909-X]

1. INTRODUCTION

The existence of lightly damped low-frequency space-charge waves in semiconductors was predicted about a quarter of a century ago.[1,2] However, at the time the prediction was not directly corroborated by experiments and was soon forgotten. Real interest in space-charge waves emerged unexpectedly in connection with studies of photorefractive phenomena in cubic crystals of the sillenite family $\mathrm{Bi}_{12}\mathrm{RO}_{20}$, where R = Si, Ge, and Te. Experiments revealed[3–6] that when the crystal is exposed to a traveling intensity grating,

$$I=I_0[1+m\cos(\mathbf{K}\cdot\mathbf{r}-\Omega t)],\tag{1}$$

under certain conditions (e.g., in the presence of an external field parallel to the lattice vector **K**) there emerges a space-charge field characterized not only by a period that coincides with that of the external field, $2\pi/K$, but also by fractional spatial frequencies, **K**/2, **K**/3, and **K**/4, i.e., double, triple, and quadruple periods. An alternative way to excite subharmonics[7,8] is to apply an alternating electric field at a zero frequency offset Ω between the light beams forming the interference pattern (1). The subharmonics are detected with the usual photorefractive scheme: due to the Pockels effect, the space-charge field induces a change in the refractive index, and diffraction of the pump waves by the refractive-index grating generates additional light beams (Fig. 1).

An explanation of subharmonic generation was given in Refs. 9–11. There it was shown that the decaying (parametric) nonlinear process determined by the resonance conditions

$$\Omega=\omega_{\mathbf{k}_1}+\omega_{\mathbf{k}_2},\quad \mathbf{K}=\mathbf{k}_1+\mathbf{k}_2\tag{2}$$

can lead to instability in the initial periodic state against an increase in the amplitude of a space-charge wave with wave vectors $\mathbf{k}_{1,2}$ and natural frequencies $\omega_{\mathbf{k}_{1,2}}$. The requirement that the damping of the wave be light, which is needed to overcome the parametric instability threshold, is guaranteed by the large value of the product $\mu\tau$ (μ is the mobility of the photoelectrons and τ is their lifetime) and the high strength E_0 of the applied field. The parametric-resonance conditions (2) is satisfied by proper selection of the detuning Ω. Since in a space-charge wave $\omega_{\mathbf{k}}\propto(\mathbf{k}\cdot\mathbf{E}_0)^{-1}$, the conditions (2) are met only if $\Omega\geqslant 4\omega_{\mathbf{k}}$.

The parametric instability theory linear in the space-charge wave amplitudes and the contrast m of the light pattern has provided an explanation for a broad range of observed laws of subharmonic generation and has made certain predictions.[12–14] Among such predictions is the possibility of longitudinal splitting of the fundamental subharmonic **K**/2 and the broadening of the spatial frequency spectrum (see Refs. 15 and 16 and Figs. 1b and 1c). The foundations of the nonlinear theory of parametrically excited space-charge waves were laid down in a recent paper.[17] In some respects the theory is similar to the well-known theory of excitation of spin waves in ferromagnets.[18]

Parallel to theoretical studies, experiments on the excitation of space-charge waves in sillenite crystals have been conducted. An intriguing experimental result was the detection of transverse splitting of the fundamental subharmonic **K**/2 (see Fig. 1d) in the scheme with a traveling intensity grating.[19,20] Such splitting is inconsistent with the linear theory of parametric excitation of space-charge waves, which predicts a reduction in instability as the transverse (with respect to **K**) component of the wave vectors $\mathbf{k}_{1,2}$ increases.

Recently, Pedersen and Johansen[21,22] have attempted to explain the transverse splitting. They focused on the frequency-offset range $\Omega<4\omega_K$, and in describing the initial periodic state limited themselves, as before,[10,11] to an approximation linear in the contrast m of the light pattern. This approximation yields a one-dimensional harmonic distribution of the space-charge field,

$$E=E_K\exp[i(Kz-\Omega t)]+\text{c.c.},$$

with an amplitude

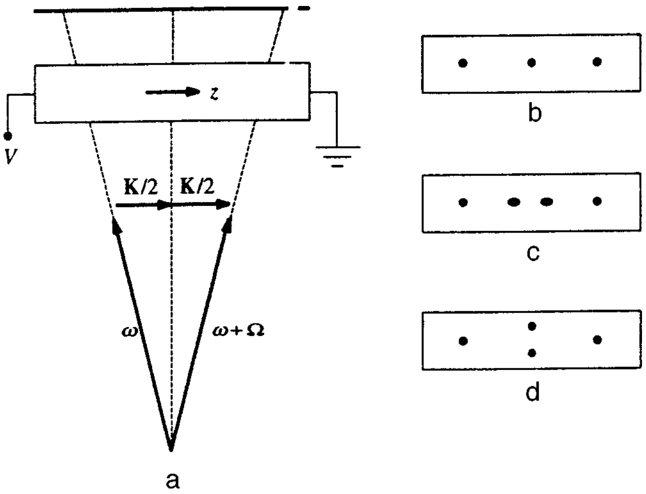


FIG. 1. (a) Experimental setup used to generate subharmonics of a traveling light pattern, and the patterns of light on the screen placed behind the crystal corresponding to the generation of (b) the unsplit subharmonic **K**/2, (c) the longitudinally split subharmonic, and (d) the transversely split subharmonic.

$$E_K=\frac{m}{2}\,\frac{E_0}{\varepsilon^{-1}-1+iQ_K^{-1}},\tag{3}$$

where $\varepsilon=\omega_K/\Omega$ is a dimensionless parameter varied in the experiment, and Q_K is the Q-factor of a space-charge wave with a wave vector K (Ref. 11). Under these assumptions, at $\Omega\simeq\omega_K$ (i.e., at $\varepsilon\equiv\omega_K/\Omega\simeq1$) and $m=1$, a new region of instability against parametric excitation of space-charge waves was found, and it was shown that in this region the most easily excited waves are those with nonzero transverse components of the wave vectors. This made it possible to interpret the results of experimental observations of the transverse splitting of the **K**/2 subharmonic.

Clearly, the approximation employed by Pedersen and Johansen[21,22] (which is linear in the contrast m) breaks down when $\varepsilon\simeq1$ and $m\simeq1$. Indeed, in this case $|E_K|\simeq E_0Q_K/2\gg E_0$, which defies common sense because the total field inside the crystal, $E_0+E(z)$, proves to be alternating and much higher than the applied field. Actually, the contrast-linear approximation is applicable as long as $|E_K|\lesssim E_0$; at $\Omega\simeq\omega_K$ (i.e., in the linear-resonance case) this is equivalent to the constraint $m\lesssim Q_K^{-1}\ll1$. At higher values of m, which always play a significant role in experiments, higher Fourier harmonics of the field E with the spatial frequencies $2K$, $3K$, ... become important, while the amplitude E_K rapidly decreases due to the nonlinear coupling with these harmonics.[11,23] At $\varepsilon\simeq1$ and $m\simeq1$ the number of significant Fourier harmonics in the initial state can be estimated to be Q_K; in sillenite crystals this usually amounts to 6–8(Refs. 10 and 11).

A comment on the results of Pedersen and Johansen[21,22] is in order. What the two researchers examined at $\Omega<4\omega_K$ (i.e., at $\varepsilon>0.25$) was a first-order nonresonant parametric process. The increase in the minimum (in **k**) relative distance to a parametric resonance, $(\omega_{\mathbf{k}}-\omega_{\mathbf{K}-\mathbf{k}}-\Omega)_{\min}/\Omega=4\varepsilon-1$, with increasing ε was balanced by a rapid increase in the amplitude $E_K(\varepsilon)$ due to the approach of linear resonance. It is this increase that gives rise to a new instability region at $\varepsilon\simeq1$.

In actual fact, $E_K(\varepsilon)$ reaches a plateau when $\varepsilon>0.25$ due to excitation of higher spatial harmonics, which reduce the efficiency of the of the parametric process. Moreover, higher-order parametric processes specified by the conditions for synchronism, $s\Omega=\omega_{\mathbf{k}}+\omega_{s\mathbf{K}-\mathbf{k}}$ with $s=2,3,\ldots$, greatly affect the evolution of weak perturbations. In particular, these nonlinear processes may be resonant. Thus, the physical situation in the offset range $\Omega\lesssim4\omega_K$ proves to be much more complicated than the situation for $\Omega\gg4\omega_K$. At the same time, the increase in $E_K(\varepsilon)$ and the effective generation of higher harmonics of the space-charge field at $\varepsilon\approx1$ leaves some hope that the theoretical results of Pedersen and Johansen[21,22] and their interpretation of the transverse splitting of the **K**/2 subharmonic are meaningful, at least qualitatively.

In this paper we give a consistent analysis of the stability of the initial periodic solution for a space-charge field over the full range of the most important frequency offsets. Our analysis takes into account the higher Fourier harmonics in the initial periodic state and allows for three-dimensional small perturbations of a general type. Such an approach is necessary not only for an explanation of transverse splitting—it plays an important role in determining the accuracy of the analytic linear theory of parametric excitation[10,11] built for $\Omega\gtrsim4\omega_K$, which forms the basis of the nonlinear theory of photorefractive subharmonics.[17]

A purely analytic or numerical study of the instability in the general case is extremely difficult. However, a combination of analytic and numerical methods can make the problem much simpler. The main idea of our analysis is as follows. First we numerically solve the one-dimensional problem of finding N Fourier harmonics of the space-charge field, which to high accuracy may be taken as the initial periodic solution. Next we analytically find the characteristic $2N$-by-$2N$ matrix that links the Fourier components of a three-dimensional perturbation. The elements of this matrix contain the instability growth rate Γ, the Fourier harmonics of the initial space-charge field, and the longitudinal and transverse components of the wave vector of the perturbation. The condition that the determinant of this matrix vanish yields a characteristic equation of degree $2N$ for Γ. We solve this equation numerically to find the branch of the solution corresponding to the maximum value of $\Gamma'=\mathrm{Re}\,\Gamma$ and to study the dependence of this maximum value on ε and the wave vector of the perturbation.

Before analyzing instability we examine the properties of the initial dynamical equation for the space-charge field potential. We conclude the paper by discussing the results and the conclusions that follow.

2. INITIAL RELATIONSHIPS

Our study (and, for that matter, the analysis of Refs. 21 and 22) is based on the scalar dynamical equation for the potential φ defined by

$$\mathbf{E}=-\nabla\varphi.$$

That equation, derived in Ref. 11, describes processes involving photoexcitation of electrons from deep traps, the drift of electrons in the field $\mathbf{E}_0+\mathbf{E}$, diffusion, and recombination, and has the form

$$\nabla^2\varphi_{zt}-\frac{\omega_0}{l_s}\nabla^2\varphi-\frac{1}{l_0}\nabla^2\varphi_t+\omega_0\nabla^2\varphi_z+\frac{l_D^2}{l_0}\nabla^4\varphi_t$$
$$=-\frac{4\pi e}{\epsilon_0}\delta g_z+\frac{4\pi e}{\epsilon_0}\frac{1}{E_0}\nabla\cdot(\delta g\,\nabla\varphi)+\frac{1}{E_0}\nabla\cdot(\nabla^2\varphi_t\nabla\varphi),\tag{4}$$

where ∇^2 is the Laplacian, $\omega_0=g_0/N_t$ is the characteristic frequency, N_t is the effective trap concentration, g_0 is the spatially homogeneous part of the electron photoexcitation rate (this quantity can easily be expressed in terms of the light's absorption coefficient and intensity I_0), $\delta g=mg_0\cos(Kz-\Omega t)$ is the spatially oscillating part of the electron photoexcitation rate, ϵ_0 is dc dielectric constant of the crystal, $l_0=\mu\tau E_0$ is the photoelectron drift length, $l_D=\sqrt{k_BT\mu\tau/e}$ is the diffusion length, k_B is Boltzmann's constant, T is the absolute temperature, $l_s=\epsilon_0E_0/4\pi eN_t$ is the screening length, and the subscripts z and t indicate derivatives with respect to the longitudinal coordinate and time.

The first two terms on the left-hand side of Eq. (4) describe nondissipative propagation of space-charge waves, and the third, fourth, and fifth describe the damping of the waves. The terms on the right-hand side characterize the excitation of space-charge waves and nonlinear interaction effects. If we drop these terms and assume that $\varphi\propto\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega_{\mathbf{k}}t-\gamma_{\mathbf{k}}t)$, we can easily find the natural frequency $\omega_{\mathbf{k}}$ and the damping constant $\gamma_{\mathbf{k}}$ of a wave with wave vector $\mathbf{k}$ (Ref. 11):

$$\omega_{\mathbf{k}}=\frac{4\pi eg_0}{\epsilon_0E_0k_z},$$
$$\gamma_{\mathbf{k}}=g_0\left(\frac{1}{N_t}+\frac{4\pi e}{\epsilon_0\mu\tau E_0^2}\frac{1}{k_z^2}+\frac{4\pi k_BT}{\epsilon_0E_0^2}\frac{k^2}{k_z^2}\right).\tag{5}$$

We see that $\omega_{\mathbf{k}}$ and $\gamma_{\mathbf{k}}$ are proportional to I_0, i.e., the Q-factor of the wave, $Q_{\mathbf{k}}=|\omega_{\mathbf{k}}|/\gamma_{\mathbf{k}}$, is independent of the light intensity. The maximum value of $Q_{\mathbf{k}}$ (as a function of $\mathbf{k}$ and E_0) is $Q_{\mathbf{k}}^{\max}=\sqrt{\pi eN_t\mu\tau/\epsilon_0}$. In the numerical estimates and calculations that follow, we assume that $\varepsilon_0=56$ and $N_t=10^{16}\,\mathrm{cm}^{-3}$, while $\mu\tau$ is varied over the range $(4-8)\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$. The adopted values agree with the data on sillenites taken from Refs. 24 and 25 and yield $Q_{\mathbf{k}}^{\max}=6-8$. The values of $Q_{\mathbf{k}}$ close to $Q_{\mathbf{k}}^{\max}$ are reached at $E_0\gtrsim6-7\,\mathrm{kV\,cm^{-1}}$, $k_\perp=0$ ($k_\perp$ is the transverse component of the wave vector), and the wave's period $2\pi/k_z\simeq15-25\,\mu$m. Note that according to (5) an increase in $k_\perp$ always leads to an increase in $\gamma_{\mathbf{k}}$ and a decrease in $Q_{\mathbf{k}}$.

If we ignore the terms that are nonlinear in φ and δg, Eq. (4) becomes the expression (3) for the time-independent amplitude of the space-charge field. Elementary estimates show that nonlinear terms are negligible when $|E_K|\ll E_0$. Equation (3) shows that if $\varepsilon\lesssim0.25$, the above inequality holds up to $m\approx1$.

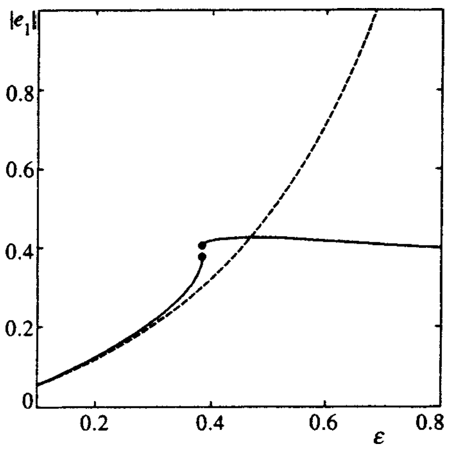


FIG. 2. Comparison of the exact solution for the amplitude $|e_1|$ (solid curve) with the results obtained in an approximation that is linear in the contrast (dashed line).

3. PERIODIC SOLUTION

Equation (4) admits of the one-dimensional steady-state solution $\varphi^{(0)}(Kz-\Omega t)$ with the periodicity of the light pattern (1). The corresponding space-charge field has only a z-component. To satisfactory accuracy this field can be represented by a truncated Fourier series:

$$E=E_0\sum_{\substack{s=-20\\s\neq0}}^{20}e_se^{is\xi},\tag{6}$$

where $\xi=Kz-\Omega t$, and e_s is the sth dimensionless Fourier harmonic of the field. By solving Eq. (4) numerically we found a set $\{e_s(\varepsilon)\}$ and proved that this set corresponds to a unique physical solution. The latter is done by comparing the results obtained via reduction of Eq. (4) to an algebraic system of nonlinear equations for the amplitudes e_s and via temporal evolution of the field $E=-\varphi_z$ to a stationary state. The solid curve in Fig. 2 depicts the $|e_1|$ (ε) dependence calculated for $m=1$, $E_0=7\,\mathrm{kV/cm^{-1}}$, a lattice constant $2\pi/K=20\,\mu$m, and $\mu\tau=6\times10^{-7}/\mathrm{V\,cm^{-1}}$. The dashed curve corresponds to (3) obtained in an approximation that is linear in the contrast m for the same values of the parameters. We see that for $\varepsilon\lesssim0.3$ the curves are essentially identical. At the same time, for $\varepsilon\gtrsim0.6$ the linear approximation yields gross overestimates of $|e_1|$. In actuality, $|e_1(\varepsilon)|$ reaches a plateau when $\varepsilon\gtrsim0.4$. Interestingly, near $\varepsilon=0.4$ the solid curve lies above the dashed curve, i.e., the allowance for higher harmonics leads to an increase in the fundamental amplitude E_K.

We also note the presence of hysteresis at $\varepsilon\simeq0.39$. Such behavior and the fact that the maximum is shifted from $\varepsilon=1$ to $\varepsilon\simeq0.4$ are due to the positive nonlinear frequency shift for a space-charge wave.[9,11,25] As the product $\mu\tau$ increases, the hysteresis and the shift of the maximum become more prominent.

Figure 3 depicts the dependence of $|e_s|$ on the number s for different values of ε and the previous values of the experimental and material parameters. Clearly, the number of significant harmonics rapidly increases as ε varies from 0.35 to 0.4. When $\varepsilon\lesssim0.25$, the higher harmonics are extremely

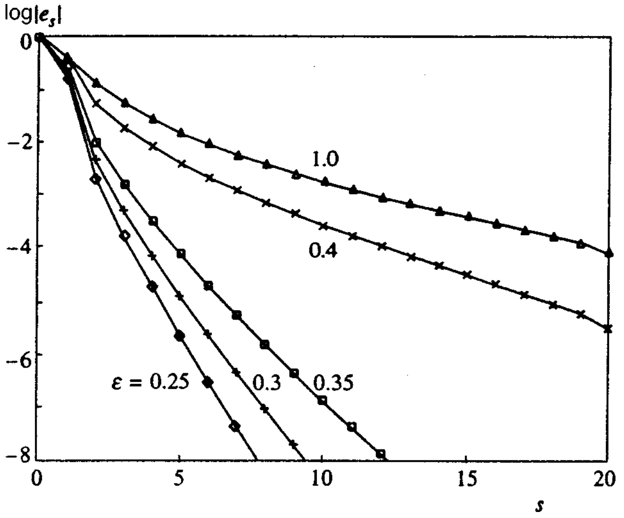


FIG. 3. Dependence of the amplitude $|e_s|$ on the number s for different values of ε. The curves have been drawn for better perception.

small. Figure 4 depicts the dependence of the total field $E_0+E(\xi)$ for $m=1$ and several values of ε (the curves were obtained by summing the Fourier series (6)). At $\varepsilon=0.25$ the distribution of the field is similar to that of the light intensity (1). The increase in the contribution of the higher harmonics with ε is accompanied by an increase in the asymmetry of the field's profile and a shift of the maximum in the positive direction. Note that when ε is very large, the total field approaches zero near the right endpoint of the interval.

4. CHARACTERISTIC EQUATION

To study the stability of the periodic solution found earlier we write the potential as $\varphi=\varphi^{(0)}+\varphi^{(1)}$, where $\varphi^{(1)}$ is a small perturbation that generally depends on the variables ξ, $\mathbf{r}_\perp=(x,y)$, and t. Linearization of the original dynamical equation (4) with respect to $\varphi^{(1)}$ leads to a homogeneous partial differential equation. It is significant is that the coefficients of this equation depend on ξ (through $\varphi^{(0)}(\xi)$) but not on $\mathbf{r}_\perp$ or t. Hence in this equation we can put

$$\varphi^{(1)}=u(\xi)\exp(i\mathbf{k}_\perp\cdot\mathbf{r}_\perp+\Gamma t)+\text{c.c.},$$

where $\mathbf{k}_\perp=(k_x,k_y)$ is the transverse component of the wave vector of the perturbation, and Γ is the instability growth rate. Thus, we arrive at a homogeneous differential equation for $u(\xi)$, with the coefficients of this equation containing $\varphi^{(0)}(\xi)$, Γ, and $\mathbf{k}_\perp$.

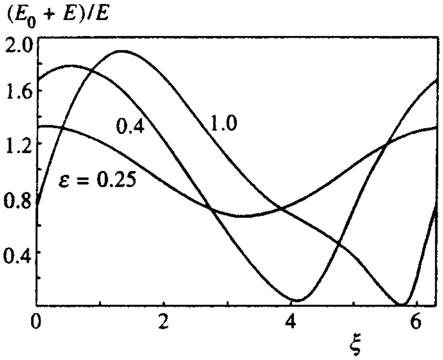


FIG. 4. Spatial profile of the electric field within a single period for several values of ε.

The amplitude $u(\xi)$ can be written

$$u=e^{i\kappa\xi}\sum_{s'=-N}^{N-1}u_{s'}e^{is'\xi}, \tag{7}$$

where κ is the dimensionless longitudinal wave number, which takes values in the range $[0,1]$, and $N\leqslant 20$ is the number of Fourier harmonics in the initial state that we wish to take into account. The expansion (7) is similar to the representation for the Bloch function of an electron in a one-dimensional periodic potential. In fact, it signifies that by taking into account the perturbation of the initial state we allow for spatial frequencies $\mathbf{K}(-N+\kappa)\pm\mathbf{k}_\perp$, ..., $\mathbf{K}(N-1+\kappa)\pm\mathbf{k}_\perp$. The absolute values of the longitudinal components of these vectors are les than NK. When $\kappa=1/2$, the perturbation is the fundamental subharmonic, unsplit or transversely split depending on whether or not the transverse projection $\mathbf{k}_\perp$ of the wave vector is zero.

By using the representation (7) and the Fourier series (6) we finally reduce the differential equation for $u(\xi)$ to a linear matrix equation for u_s:

$$(C_s\delta_{ss'}-A_{ss'})u_{s'}=0. \tag{8}$$

Here $\delta_{ss'}$ is the Kronecker delta, and the coefficients C_s and $A_{ss'}$ are

$$\begin{aligned}C_s=[(s+\kappa)^2+\theta^2]\Bigg\{&-i(s+\kappa)\nu-(s+\kappa)^2+\varepsilon-i\varepsilon(s+\kappa)\\&\times Kl_s+\frac{1}{Kl_0}[\nu-i(s+\kappa)]+\frac{Kl_D^2}{l_0}[(s+\kappa)^2+\theta^2]\\&\times[\nu-i(s+\kappa)]\Bigg\},\end{aligned} \tag{9}$$

$$\begin{aligned}A_{ss'}=-\varepsilon\frac{m}{2}(\delta_{s\,s'+1}+\delta_{s\,s'-1})[(s+\kappa)(s'+\kappa)+\theta^2]\\+e_{s-s'}\{(s-s')^2[(s+\kappa)(s'+\kappa)+\theta^2]+(s+\kappa)\\\times[(s'+\kappa)^2+\theta^2](s'+\kappa+i\nu)\},\end{aligned}$$

where $\nu=\Gamma/\Omega$, and $\theta=|k_\perp|/K$ is the dimensionless transverse wave number. Note that the $2N$-dimensional wave vector C_s comes from the left-hand side of Eq. (4), while the matrix $A_{ss'}$, comprising the numerically determined amplitudes $e_{s-s'}$, comes from the right-hand side.

The system (8) has a solution if

$$\det(C_s\delta_{ss'}-A_{ss'})=0, \tag{10}$$

which is the desired characteristic equation for the instability growth rate. This equation has $2N$ branches of solutions for Γ as a function of ε, κ, and θ. Generally, these branches can be found only numerically. An analytic solution of the characteristic equation is possible only for $N=1$. A simplifying

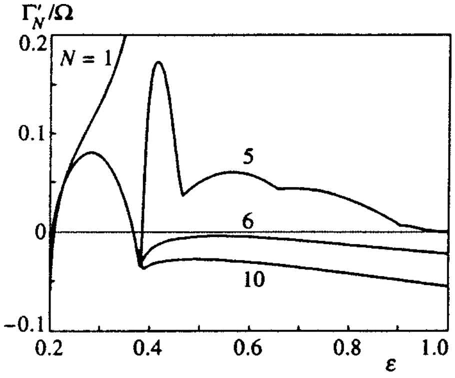


FIG. 5. $\Gamma'_N(\varepsilon)$ at $\kappa=1/2$, $\theta=0$, and $m=1$ for different values of N.

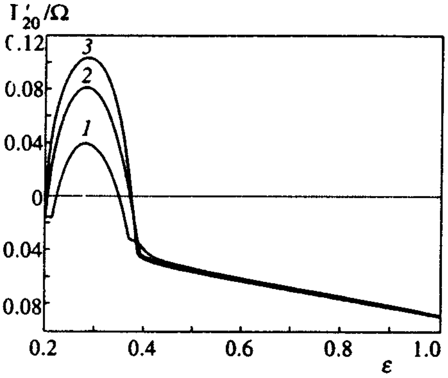


FIG. 6. $\Gamma'_{20}(\varepsilon)$ at $m=1$, $\kappa=1/2$, and $\theta=0$ for different values of $\mu\tau$. Curves *1*, *2*, and *3* correspond, respectively, to $\mu\tau=4\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$, $6\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$, and $8\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$.

assumption of this kind is only justified when $\varepsilon\lesssim0.25$, and leads to the well-known analytic results on parametric excitation of space-charge waves.[10,11]

5. STUDY OF STABILITY

We solved the characteristic equation (10) numerically. For each number N and each set of parameters $\varepsilon=\omega_K/\Omega$, κ, and $\theta=|k_\perp|/K$, we sought the branch of solutions with the maximum possible rate of exponential growth of perturbations, $\Gamma'_N=[\mathrm{Re}\,\Gamma]_{\max}$, i.e., the branch with the highest instability. What follows is the result of our analysis.

We start by investigating instability against generation of the unsplit harmonic $\mathbf{K}/2$, i.e., instability against period doubling. In this case we limit ourselves to the one-dimensional case $(\theta=0)$ and put $\kappa=1/2$. Figure 5 depicts the $\Gamma'_N(\varepsilon)$ dependence at $\mu\tau=6\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$ and $m=1$ for several values of N. As expected, when $\varepsilon\lesssim0.25$ (i.e., $\Omega\gtrsim4\omega_K$), reasonable results are achieved even in the context of the analytic theory $(N=1)$. An increase in N, i.e., the allowance for higher spatial harmonics, yields only negligible corrections to the growth rate.

The situation changes dramatically when $\varepsilon\gtrsim0.28$. Here the analytic model with $N=1$ breaks down and the growth rate $\Gamma'_N(\varepsilon)$ rapidly decreases as the number N' of spatial harmonics taken into account grows. Significantly, when $\varepsilon\gtrsim0.37$ and $N\geqslant6$, the growth rate Γ'_N becomes negative. Thus, there is only one frequency offset range (the range of values of ε) in which excitation of an unsplit subharmonic $\mathbf{K}/2$ is possible. The statement of Pedersen and Johansen that there is a second instability region at large values of ε is a consequence of the unjustified assumptions made in Refs. 21 and 22. Figure 5 clearly shows how the nonphysical region disappears as N increases. Note that the model with $N=4$ provides an extremely accurate description of the growth rate's behavior over the entire subharmonic instability region, $0.22\lesssim\varepsilon\lesssim0.37$.

Figure 6 depicts the dependence of $\Gamma'_{20}(\varepsilon)$ at $m=1$, $\kappa=1/2$, and $\theta=0$ for different values of $\mu\tau$. Clearly, an increase in $\mu\tau$ facilitates instability but does not alter our conclusion that there is only one instability region. Neither do variations of the applied field E_0 and the lattice constant $2\pi/K$ in the vicinity of the values $7\,\mathrm{kV/cm}^{-1}$ and $20\,\mu\mathrm{m}$ considered earlier lead to new qualitative results. A substantial decrease in E_0 or a substantial change in K leads to suppression of instability. At $E_0=7\,\mathrm{kV/cm}^{-1}$ and $2\pi/K=20\,\mu\mathrm{m}$ instability disappears if $\mu\tau\lesssim3\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$.

Note that each curve in Fig. 6 consists of two sections corresponding to the different branches of the solution of the characteristic equation. The transition from one branch to the other occurs in the vicinity of $\varepsilon=0.38$. Actually, the right (stable) sections of the curves are in no way related to subharmonic generation. They describe the damping of perturbations with high spatial frequencies, which are only weakly coupled to the pump wave. However, these sections show that at $\varepsilon\gtrsim0.4$ subharmonic perturbations have even smaller negative values of Γ'.

We now examine the dependence of the instability growth rate on the longitudinal wave number κ. Figure 7 depicts the $\Gamma'_{20}(\kappa)$ dependence at $\theta=0$ for various values of ε. For $\varepsilon\gtrsim0.224$ the curves have a single maximum at $\kappa=1/2$, while for $\varepsilon<0.224$ the curves have two maxima symmetric with respect to $\kappa=1/2$. Such bifurcation agrees with the analytic theory[10,11] and suggests that there may be symmetric longitudinal splitting of the fundamental subharmonic $K/2$ as ε decreases.

Curve *1* in Fig. 8 represents the dependence of the distance between the maxima on the parameter ε and was obtained numerically at $N=20$ and $m=1$, while curve *2* corre-

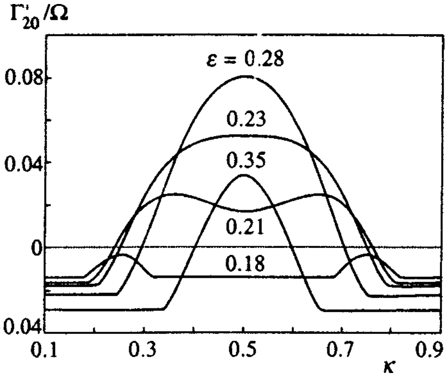


FIG. 7. Γ'_{20} as a function of the longitudinal wave number κ at $m=1$, $\theta=0$, and $\mu\tau=6\times10^{-7}\,\mathrm{cm}^2/\mathrm{V}^{-1}$ for different values of ε.

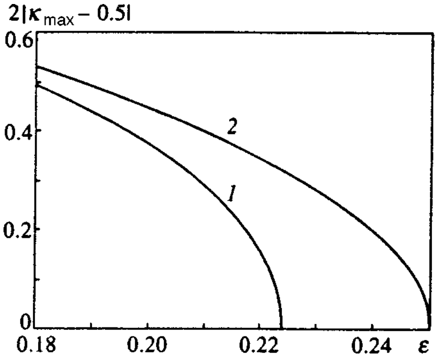


FIG. 8. The ε-dependence of the longitudinal splitting obtained numerically (curve *1*) and analytically (curve *2*).

sponds to the result obtained by the analytic theory, $\sqrt{1-4\varepsilon}$. The basic distinction lies in the shift of the bifurcation point from $\varepsilon=0.25$ to $\varepsilon\simeq0.224$, due to the effect of higher harmonics. The difference between the analytic and numerical values is within the uncertainty of the analytic theory.

Since the maximum in the Γ' (κ) dependence does not generally correspond to $\kappa=1/2$, it would be interesting to calculate the maximum (in κ) value of Γ' as a function of the external parameter ε. The $\Gamma'_{\max}$ (ε) dependence is the most important characteristic of instability in the one-dimensional case. The solid curve in Fig. 9 represents the $\Gamma'_{\max}$ (ε) dependence in the important ε-range. The dot on the curve corresponds to the appearance of longitudinal splitting. The growth rate reaches its maximum value at $\varepsilon\simeq0.265$. The dashed curve represents the Γ' (ε) dependence at $\kappa=1/2$ in the splitting region. Clearly, allowance for splitting substantially shifts the left instability limit toward larger offsets (smaller values of ε).

By comparing the results of our calculations with those of the analytic theory,[10,11] we see that the latter yields qualitatively (or even semiquantitatively) correct results for $\varepsilon\lesssim0.25$. In the $\varepsilon\gtrsim0.25$ range, where the consistent analytic theory breaks down and instability may still be very high, numerical results are important in and of themselves. Recall that allowance for 20 spatial harmonics is not necessary in the instability region: even $N=4$ is sufficient for the above results to be accurate. Note that as the percent modulation m decreases, the role of higher harmonics diminishes significantly, the accuracy of the analytic theory grows accordingly, and instability weakens.

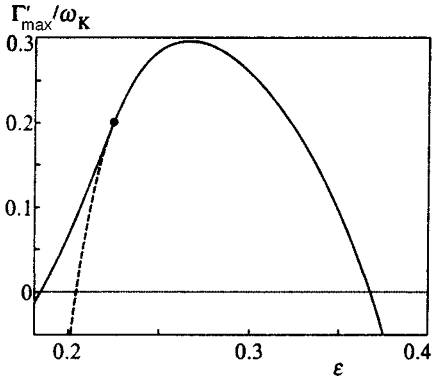


FIG. 9. The solid curve represents the dependence of the maximum (in κ) value of the growth rate on ε at $N=20$ and $\theta=0$. The dashed curve corresponds to $\kappa=1/2$.

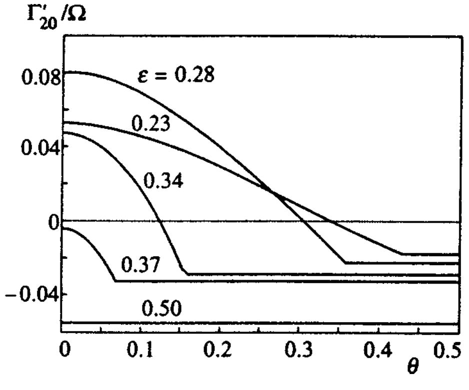


FIG. 10. Γ'_{20} as a function of the transverse wave number θ at $m=1$, $\kappa=1/2$, and $\mu\tau=6\times10^{-7}$ cm^2/V^{-1} for different values of ε.

Finally, let us study the dependence of Γ' on the dimensionless transverse wave number $\theta=|\mathbf{k}_\perp|/K$. This is important if we want to interpret the transverse splitting of the subharmonic **K**/2. Figure 10 depicts the Γ'_{20} (θ) dependence at $m=1$ and $\kappa=1/2$ for different values of ε. The laws that follow from the diagrams are very simple. When $\Gamma'(\theta=0)$ is positive (or even slightly negative), an increase in θ leads to a decrease in Γ', i.e., does not facilitate instability; the rate of decrease of $\Gamma'(\theta)$ grows with ε in the positive growth rate region. When $\theta(\varepsilon)$ is large, i.e., when the growth rate becomes negative, the decrease in $\Gamma'(\theta)$ suddenly becomes saturated. For $\varepsilon\gtrsim0.385$, Γ' is essentially independent of θ. The constant (stable) sections of the curves in Fig. 10 are unrelated to the parametric excitation of subharmonics. As the right (negative) sections of the curves in Fig. 6, these sections represent the damping of perturbations with high spatial frequencies, which are only weakly coupled to the pump wave. The possibility of Γ' increasing with θ on the lower branches of the solution of the characteristic equation (8) cannot be ruled out, but for these branches $\Gamma'<0$. Variation of the longitudinal wave number κ and of $\mu\tau$ does not lead to an increase in $\Gamma'(\theta)$.

6. DISCUSSION

We believe that there are two basic aspects of the present work that may be of interest. First, we have developed a new method for studying the stability of high-contrast photorefractive lattices in semiconducting crystals against three-dimensional perturbations. The method is capable of providing a solution for the problem when a purely analytic or numerical approach is fraught with difficulties. Undoubtedly, it can be used to study the stability of photorefractive lattices excited by an alternating external field. This case is important for applications of the photorefractive effect,[25,26] and the higher spatial harmonics are no less important than when traveling grating techniques are employed.

Second, it has become clear that Pedersen and Johansen's interpretation of transverse splitting[21,22] is incorrect. A consistent linear theory of parametric excitation of space-charge waves (which incorporates the higher harmonics of the space-charge field) suggests that there is no additional instability region and no preferential excitation of transversely split subharmonics.

The inability of the linear theory to explain the transverse splitting of subharmonics does not necessarily mean that this interesting effect cannot be studied. Furthermore, the results shed light on the nature of the effect. As pointed out in Ref. 17, the bias toward transverse splitting over longitudinal splitting ought to show up in the nonlinear stage of instability. On the basis of the results obtained in Ref. 17, it has been established in a recent paper[27] that for the transversely split subharmonic **K**/2 of a finite amplitude there exists a region that is stable against small perturbations, while the unsplit and longitudinally split subharmonics **K**/2 are always modulation-unstable. The results obtained in the present paper indicate that the interpretation of transverse splitting in the context of the nonlinear theory given in Ref. 27 is the only possible one.

In all fairness, there is another aspect of the above problem that must be mentioned. Experiments usually produce values of the higher harmonics of the space-charge field, $E_{2K}, E_{3K}, \ldots$, that are much smaller than the calculated values. An obvious reason for this discrepancy is that the experimental conditions are not perfect; in particular, there is necessarily a broadening of the Fourier spectrum of the light's intensity distribution. From general considerations it is clear that such broadening most directly affects precisely the higher spatial harmonics of the space-charge field. It is still unknown how parametric instability is affected by such broadening.

7. CONCLUSIONS

The higher harmonics of the space-charge field play an important role in the stability analysis of high-contrast photorefractive lattices induced in semiconducting crystals by a traveling interference light pattern in the frequency-offset range $\Omega \lesssim 4\omega_K$. A combination of analytic and numerical methods is capable of providing a complete analysis of the stability of these photorefractive lattices against small three-dimensional perturbations.

The interpretation of the transverse splitting of the subharmonic **K**/2 reported in Refs. 21 and 22 is in error. The consistent theory of parametric excitation of space-charge waves that incorporates the higher spatial harmonics suggests that there is only one instability region. It also provides a quantitative modification of the analytic results of Refs. 10 and 11 in this region.

ACKNOWLEDGMENTS

One of the authors (A.I.Ch.) is grateful to M. G. Stepanov for valuable remarks. This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant No. 96-02-19126).

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

---

[1] R. F. Kazarinov, R. A. Suris, and B. I. Fuks, Fiz. Tekh. Poluprovodn. **6**, 572 (1972) [Sov. Phys. Semicond. **6**, 500 (1972)].

[2] R. A. Suris, and B. I. Fuks, Fiz. Tekh. Poluprovodn. **9**, 1717 (1975) [Sov. Phys. Semicond. **9**, 1130 (1975)].

[3] S. Mallick, B. Imbert, H. Ducollet, J. P. Herrian, and J. P. Huignart, J. Appl. Phys. **63**, 5660 (1988).

[4] D. J. Webb and L. Solymar, Opt. Commun. **74**, 386 (1990).

[5] D. J. Webb, L. B. Au, D. C. Jones, and L. Solymar, Appl. Phys. Lett. **57**, 1602 (1990).

[6] J. Takacs, M. Schaub, and L. Solymar, Opt. Commun. **91**, 252 (1992).

[7] J. Takacs and L. Solymar, Opt. Lett. **17**, 247 (1992).

[8] C. H. Kwak, M. Shamonin, J. Takacs, and L. Solymar, Appl. Phys. Lett. **62**, 328 (1993).

[9] A. Bledovski, J. Otten, K. H. Ringhofer, and B. Sturman, Zh. Éksp. Teor. Fiz. **102**, 406 (1992) [Sov. Phys. JETP **75**, 215 (1992)].

[10] B. I. Sturman, M. Mann, and K. H. Ringhofer, Appl. Phys. A **55**, 325 (1992).

[11] B. I. Sturman, M. Mann, J. Otten, and K. H. Ringhofer, J. Opt. Soc. Am. B **10**, 1919 (1993).

[12] B. I. Sturman, M. Mann, J. Otten, K. H. Ringhofer, and A. Bledowski, Appl. Phys. A **55**, 55 (1992).

[13] T. E. McClelland, D. J. Webb, B. I. Sturman, and K. H. Ringhofer, Phys. Rev. Lett. **73**, 3082 (1994).

[14] B. I. Sturman, T. E. McClelland, D. J. Webb, E. Shamonina, and K. H. Ringhofer, J. Opt. Soc. Am. B **12**, 1621 (1995).

[15] H. C. Pedersen and P. M. Johansen, J. Opt. Soc. Am. B **12**, 1065 (1995).

[16] P. Buchhave, S. Lyuksyutov, M. Vasnetsov, and C. Heyde, J. Opt. Soc. Am. B **13**, 2595 (1996).

[17] B. I. Sturman, M. Aguilar, F. Aguľó-Lòpez, and K. H. Ringhofer, Phys. Rev. E **55**, 6072 (1997).

[18] V. S. L'vov, *Wave Turbulence Under Parametric Excitation. Application to Magnets*, Springer-Verlag, Berlin (1994).

[19] J. Takacs and L. Solymar (private communication, 1992).

[20] H. C. Pedersen and P. M. Johansen, Opt. Lett. **19**, 1418 (1994).

[21] H. C. Pedersen and P. M. Johansen, Phys. Rev. Lett. **77**, 3106 (1996).

[22] H. C. Pedersen and P. M. Johansen, J. Opt. Soc. Am. B **14**, 1418 (1997).

[23] T. E. McClelland, D. J. Webb, B. I. Sturman, E. Shamonina, M. Mann, and K. H. Ringhofer, Opt. Commun. **131**, 315 (1996).

[24] Ph. Refregier, L. Solymar, H. Rajbenbach, and J. P. Huignard, J. Appl. Phys. **58**, 45 (1985).

[25] M. P. Petrov, S. I. Stepanov, and A. V. Khomenko, *Photorefractive Crystals in Coherent Optical Systems*, Springer-Verlag, Berlin (1991).

[26] A. A. Kamshilin, E. Raita, and A. V. Khomenko, J. Opt. Soc. Am. B **13**, 2536 (1996).

[27] E. V. Podivilov, H. C. Pedersen, P. M. Johansen, and B. I. Sturman, Phys. Rev. E **57**, 6112 (1998).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Collective oscillations in two-band superconducting systems with low carrier density

F. G. Kochorbé*) and M. E. Palistrant

Applied Physics Institute, Moldova Academy of Science, 277028 Kishinev, Moldova

(Submitted 5 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 1047–1066 (September 1998)

We study the collective oscillations in two-band superconducting systems with all possible intra- and interband interactions for arbitrary values of carrier density N_0, including low carrier densities, when $\mu\sim\Delta$. Allowance for processes of scattering of an interband Cooper pair into an intraband pair leads to the emergence of a new excitonic mode in addition to the Bogolyubov–Anderson acoustic mode and an excitonic mode of the Leggett type. The presence of a new order parameter Δ_{12}, in addition to Δ_{11} and Δ_{22}, and the asymmetric cutoff of integrals lead to the mixing of fluctuations of the phases and amplitudes of the order parameters of different bands. This mixing and the dependence of the order parameters on N_0 result in a strong dependence of these two excitonic modes on the carrier density N_0. © *1998 American Institute of Physics.* [S1063-7761(98)02009-5]

1. INTRODUCTION

After the discovery of high-T_c superconducting compounds, many models for describing the broad spectrum of magnetic and superconducting properties of such compounds were proposed. Studies have been carried out within the context of the phonon mechanism of superconductivity[1–5] as well as the excitonic,[6] hole,[7] plasmon,[8,9] magnetic,[10] and other mechanisms (see the review in Ref. 11).

In addition to this, the model with two bands overlapping at the Fermi surface has been widely used to describe the superconducting properties of high-T_c superconductors.[12,13] The thermodynamic, electromagnetic, and transport properties of two-band superconductors have been described in two monographs[14,15] and in the review in Ref. 16. This model and its various generalizations to anisotropic systems with singular points in the momentum space, which lead to topological electron transitions, makes it possible to describe a large body of experimental data on high-T_c superconductors.[17–24]

An important feature of the two-band model is the fact that the superconducting transition temperature T_c is independent of the sign of the effective electron–electron interactions V_{ij} $(i\neq j)$. The theory can be applied to the ordinary phonon mechanism of superconductivity and to a nonphonon mechanism. More than that, it has been established that even if there is repulsion between charge carriers, both low-T_c superconductivity (Ref. 14) and high-T_c superconductivity (Refs. 25 and 26) are possible, provided that $V_{11}V_{22}-V_{12}^2<0$.

Numerous calculations (see, e.g., Refs. 27 and 28) corroborate the results of multiband theory and show that several energy bands can indeed cross at the Fermi surface in high-T_c superconductors. For instance, in $\mathrm{YBa_2Cu_3O_{7-\delta}}$ (Ref. 28) the number of energy bands crossing at the Fermi surface increases with oxygen content.

It is important to note that the overlap of energy bands at the Fermi surface leads to results that differ not only quantitatively but also qualitatively from those for a single energy band. For instance, T_c studies done on the basis of the two-band model with allowance for topological transitions[21,23] and on the basis of the three-band model[29] provide an explanation from the stepped dependence of this quantity on oxygen content in $\mathrm{YBa_2Cu_3O_{7-\delta}}$ observed in experiments.

Another interesting phenomenon characteristic only of multiband superconductors is the emergence of collective oscillations of the excitonic type due to fluctuations of the phases of the order parameters of different bands.[30,31] Leggett[30] was the first to detect this phenomenon. He found that in two-band superconductors with the ordinary phonon mechanism of superconductivity there emerges, in addition to a Bogolyubov–Anderson acoustic mode, a collective mode of the excitonic type.

Collective oscillations in the three-photon model with the phonon mechanism of superconductivity have been studied by one of the present authors in Ref. 31. In contrast to Leggett's two-band model,[30] two excitonic modes and one acoustic mode were found to emerge. Collective oscillations in two-band low-dimensional superconductors have been studied in Ref. 32, with both the phonon and the nonphonon mechanisms investigated. These two-band systems also revealed the presence of an excitonic mode in addition to the Bogolyubov–Anderson acoustic mode in quasi-one-dimensional and two-dimensional superconductors.

A nonmagnetic impurity in two-band impurity superconductors[33–35] does not lead to a decay of the collective modes, but it facilitates the increase in the collective oscillation frequency with impurity concentration.

Collective oscillations in a two-band system with local pairs in a Bose-condensate state have been studied in Ref. 36. A single excitonic mode, in addition to a Bogolyubov–Anderson acoustic mode, was detected in the system. However, in contrast to Refs. 30–35, where it is assumed that the collective oscillations are due to fluctuations of the phases of the order parameters of different bands, here these modes are determined by the mixing of fluctuations of the phases and amplitudes of the order parameters of different bands.

We also note that the multilayer model of high-$T_c$superconductivity[37,38] is equivalent to the multiband model. For instance, in studies of collective oscillations in bilayered superconductors[39–41] a Leggett mode of the excitonic type was also detected. This mode is induced by fluctuations of the phases of the order parameters of different layers.

As is known, in high-T_c superconductors the carrier density may be of any value, including small values ($\mu \sim \Delta, T_c$ and $\mu < \omega_D$, where μ is the chemical potential, Δ is the order parameter, and ω_D is the Debye frequency). When the carrier density is low ($\mu \sim \Delta$), in the two-band model we must allow for all possible electron pairings,[42] in contrast to Refs. 12 and 14, where only intraband pairings and transitions of a Cooper pair as a whole from one band into another are taken into account. Here were must discard the diagonal approximation in band indices, since in this model[42] $\Delta_{12} \sim \Delta_{11}, \Delta_{22}$. In such systems all observable physical quantities become strongly dependent on the carrier density. The temperature dependence of the chemical potential exhibits a break at $T = T_c$ (Refs. 43 and 44). Finally, the overlap of the energy bands in such a model lowers the upper limit at which this break disappears by a factor of two to three (~ 5–$6\,\mathrm{meV}$),[42] which facilitates detection of this phenomenon.

In this paper we examine the collective oscillations in a two-band superconductor (with two inequivalent layers; see Sec. 4) with allowance for an arbitrary carrier density, including the case of low carrier density ($\mu \sim \Delta$). This two-band system can be effectively described as a pseudo-three-band model (with the order parameters Δ_{12}, Δ_{11}, and Δ_{22}) with an arbitrary carrier density.

The paper is organized as follows. In Sec. 2 we write the system Hamiltonian, the basic equations for the vertex functions and the order parameters, and the equation for the collective oscillation frequency. In Sec. 3 we study collective oscillations at low frequencies ($\omega^2/4\Delta_i^2 \ll 1$) and $\mathbf{k} = 0$. In Sec. 4 we discuss the main conclusions. Appendices A–D deal with the fine aspects of the calculations.

2. SYSTEM HAMILTONIAN AND BASIC EQUATIONS

The two-band system is described by the Hamiltonian[12–14]

$$H = \sum_{n\mathbf{k}\sigma} [\varepsilon_n(\mathbf{k}) - \mu] a^{\dagger}_{n\mathbf{k}\sigma} a_{n\mathbf{k}\sigma} - \frac{1}{V} \sum_{m_1 \cdots m_4} \sum_{\mathbf{k}\mathbf{k}'} V^{m_3 m_4}_{m_1 m_2} \times (\mathbf{k}, -\mathbf{k}', -\mathbf{k}, \mathbf{k}') a^{\dagger}_{m_1 \mathbf{k}\uparrow} a^{\dagger}_{m_2\, -\mathbf{k}\downarrow} a_{m_3\, -\mathbf{k}'\downarrow} a_{m_4 \mathbf{k}'\uparrow}, \tag{1}$$

where $a^{\dagger}_{n\mathbf{k}\sigma}$ and $a_{n\mathbf{k}\sigma}$ are the creation and annihilation operators for an nth-band electron with spin σ and a qausiwave vector $\mathbf{k}$, and $V^{m_3 m_4}_{m_1 m_2}$ are the intra- and interband interaction constants. The expression (1) is a generalization of the BCS–Bogolyubov Hamiltonian to the two-band case. Account is taken here of all possible methods of electron pairing within each band and of pairing of electrons belonging to different bands. If $m_1 = m_2$ and $m_3 = m_4$, the Hamiltonian (1) is equal to that of the Moskalenko model[12–14] (applicable only to systems with a high carrier density), which considers only intraband pairing and transitions of a Cooper pair as a whole from one band to another.

We introduce the two-particle Green's function describing a density–density correlation:

$$K(\mathbf{x}\alpha\mathbf{y}\beta) = \langle T \psi^{\dagger}_{\alpha}(\mathbf{x}) \psi_{\alpha}(\mathbf{x}) \psi^{\dagger}_{\beta}(\mathbf{y}) \psi_{\beta}(\mathbf{y}) \rangle. \tag{2}$$

The operator $\psi_{\alpha}(\mathbf{x})$ is defined as follows:

$$\psi_{\alpha}(\mathbf{x}) = \sum_{n\mathbf{k}} \psi_{n\mathbf{k}}(\mathbf{x}) a_{n\mathbf{k}\alpha}, \quad \psi_{n\mathbf{k}}(\mathbf{x}) = U_{n\mathbf{k}}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{3}$$

where $\psi_{n\mathbf{k}}(\mathbf{x})$ is a Bloch function in the nth band, and $U_{n\mathbf{k}}(\mathbf{x})$ is the amplitude of that function.

Perturbation-theory reasoning[45] suggests that for the superconducting state the two-particle Green's function $K(\mathbf{x}\alpha\mathbf{y}\beta)$ is defined by sixteen vertex functions, which can be found by solving four independent systems of equations. We start with the system of equations for the four vertex functions

$$\Gamma^{+-}_{+-}, \quad \Gamma^{--}_{--}, \quad \Gamma^{+-}_{--}, \quad \Gamma^{--}_{+-}. \tag{4}$$

The ''plus'' and ''minus'' correspond to situations in which an arrow points away from a vertex and toward a vertex, respectively.

We can show that the functions $\Gamma^{\gamma}(\gamma = {}^{+-}_{--}, {}^{--}_{+-})$ in the $n\mathbf{k}\Omega$-representation enter into the system of equations (A3) (see Appendix A) with a total momentum far removed in value from the Fermi surface. This fact makes it possible to ignore these terms in (A3) in the BCS approximation considered.[46,14]

If we now replace the remaining vertex functions $\Gamma^{\alpha}(\alpha = {}^{+-}_{+-}, {}^{--}_{--})$ with their values on the Fermi surface,

$$\Gamma^{\alpha}_{m_1 m_2 m_3 m_4}(k) = \Gamma^{\alpha}_{m_1 m_2 m_3 m_4}\left(p_F + \frac{k}{2}, -p_F + \frac{k}{2}, p_3, p_4\right), \tag{5}$$

we arrive at the following system of equations:

$$\hat{\Gamma}^{+-}_{+-}(k) = -\hat{V} + \hat{V}\hat{A}(k)\hat{\Gamma}^{+-}_{+-}(k) + \hat{V}\hat{B}(k)\hat{\Gamma}^{--}_{--},$$

$$\Gamma^{--}_{--}(-k) = \hat{V}\hat{A}(-k)\hat{\Gamma}^{--}_{--}(-k) + \hat{V}\hat{B}(-k)\hat{\Gamma}^{+-}_{+-}(k), \tag{6}$$

where

$$\hat{\Gamma}^{\alpha} = \begin{pmatrix} \Gamma^{\alpha}_{1111}(k) & \Gamma^{\alpha}_{1122}(k) & \Gamma^{\alpha}_{1112}(k) \\ \Gamma^{\alpha}_{2211}(k) & \Gamma^{\alpha}_{2222}(k) & \Gamma^{\alpha}_{2212}(k) \\ \Gamma^{\alpha}_{1211}(k) & \Gamma^{\alpha}_{1222}(k) & \Gamma^{\alpha}_{1212}(k) \end{pmatrix},$$

$$\hat{V} = \begin{pmatrix} V_{1111} & V_{1122} & V_{1112} \\ V_{2211} & V_{2222} & V_{2212} \\ V_{1211} & V_{1222} & V_{1212} \end{pmatrix}. \tag{7}$$

Here the matrices $\hat{A}(k)$ and $\hat{B}(k)$ are built in the same way as the matrices $\hat{\Gamma}^{\alpha}(k)$, with

$$A_{m_1 m_2 m_3 m_4}(k) = \frac{1}{\beta} \sum_{q} G_{m_1 m_2}\left(q + \frac{k}{2}\right) G_{m_3 m_4}\left(-q + \frac{k}{2}\right), \tag{8}$$

$$B_{m_1m_2m_3m_4}(k)=\frac{1}{\beta}\sum_q F_{m_1m_2}\left(q+\frac{k}{2}\right)F_{m_3m_4}\left(-q+\frac{k}{2}\right).$$

The definitions of the one-particle Green's functions $G_{m_1m_2}(k)$ and $F_{m_1m_2}(k)(k=(\mathbf{k},\omega))$ are given in Appendix B.

Note that in two-band superconducting systems with low carrier density ($\mu\sim\Delta$) (see Refs. 42–44) one must allow for the mean-field renormalization of the chemical potential:

$$\mu\rightarrow\mu_n=\mu+S_n, \tag{9}$$

where

$$S_n=\sum_{\mathbf{k}m}(2V_{mnnm}-V_{mnmn})\langle a^{\dagger}_{m\mathbf{k}\uparrow}a_{m\mathbf{k}\uparrow}\rangle. \tag{10}$$

Moreover, the fact that such systems possess an additional order parameter $\Delta_{12}\sim\Delta_{11},\Delta_{22}$ gives rise to off-diagonal one-particle Green's functions, G_{12}, $F_{12}\sim G_{11}$, and F_{22} (see Ref. 42).

For the new vertex functions $\hat{\Gamma}^{\mathrm{ph}}$ and $\hat{\Gamma}^{a}$,

$$\begin{aligned}\hat{\Gamma}^{\mathrm{ph}}&=\hat{\Gamma}{}^{\overset{+-}{+-}}(k)+\hat{\Gamma}{}^{\overset{--}{--}}(-k),\\ \hat{\Gamma}^{a}&=\hat{\Gamma}{}^{\overset{+-}{+-}}(k)-\hat{\Gamma}{}^{\overset{--}{--}}(-k),\end{aligned} \tag{11}$$

we have

$$\begin{aligned}\hat{\Gamma}^{\mathrm{ph}}&=-\hat{V}+\hat{V}\hat{\xi}^{+}\hat{\Gamma}^{\mathrm{ph}}+\hat{V}\hat{I}^{-}\hat{\Gamma}^{a},\\ \hat{\Gamma}^{a}&=-\hat{V}+\hat{V}\hat{I}^{+}\hat{\Gamma}^{\mathrm{ph}}+\hat{V}\hat{\xi}^{-}\hat{\Gamma}^{a},\end{aligned} \tag{12}$$

where

$$\begin{aligned}\hat{\xi}^{\pm}_{l}(k)&=\frac{\hat{A}_l(k)-\hat{A}_l(-k)}{2}\pm\frac{\hat{B}_l(k)+\hat{B}_l(-k)}{2},\\ \hat{I}^{\pm}_{l}(k)&=\hat{A}_l(k)-\hat{A}_l(-k)\pm[\hat{B}_l(k)-\hat{B}_l(-k)].\end{aligned} \tag{13}$$

Introducing the matrix $\hat{U}$, the inverse of the interaction matrix $\hat{V}$ defined in (7), we can write

$$\begin{pmatrix}\hat{U}-\hat{\xi}^{+} & -\hat{I}^{-}\\ -\hat{I}^{+} & \hat{U}-\hat{\xi}^{-}\end{pmatrix}\begin{pmatrix}\hat{\Gamma}^{\mathrm{ph}}\\ \hat{\Gamma}^{a}\end{pmatrix}=-\begin{pmatrix}\hat{1}\\ \hat{1}\end{pmatrix}, \tag{14}$$

where $\hat{1}$ is the 3×3 identity matrix. Note that the vertex functions $\hat{\Gamma}^{\mathrm{ph}}$ and $\hat{\Gamma}^{a}$ are divergent when the 6×6 determinant of the system of equations (14) vanishes:

$$\begin{vmatrix}\hat{U}-\hat{\xi}^{+} & -\hat{I}^{-}\\ -\hat{I}^{+} & \hat{U}-\hat{\xi}^{-}\end{vmatrix}=0. \tag{15}$$

Here for an arbitrary matrix $\hat{A}$ we have introduced the notation $|\widehat{A}|\equiv\det\|\widehat{A}\|$.

If we now replace the "imaginary" Matsubara frequencies $i\omega\rightarrow\omega$ with the real frequencies ω, we arrive at an equation for the collective oscillation frequency,

$$\begin{vmatrix}\hat{U}-\hat{\xi}^{+} & -\hat{I}^{-}\\ -\hat{I}^{+} & \hat{U}-\hat{\xi}^{-}\end{vmatrix}=\begin{vmatrix}\hat{R}^{+} & -\omega\hat{J}^{-}\\ -\omega\hat{J}^{+} & \hat{R}^{-}\end{vmatrix}=0. \tag{16}$$

The formulas for the quantities in $\hat{\xi}^{\pm}$ and $\hat{J}^{\pm}$ are given in Appendix C.

Using (14) to calculate the ratios of vertex functions (see Appendix D), we obtain a formula for the collective oscillation frequency:

$$\begin{vmatrix}\hat{U}-\hat{\xi}^{+} & -\omega\hat{J}^{-}\\ -\omega\hat{J}^{+} & \hat{U}-\hat{\xi}^{-}\end{vmatrix}=0, \tag{17}$$

where

$$\hat{\xi}^{\pm}=\begin{pmatrix}\xi^{\pm}_1 & 0 & 0\\ 0 & \xi^{\pm}_2 & 0\\ 0 & 0 & \xi^{\pm}_3\end{pmatrix},\quad \hat{J}^{\pm}=\begin{pmatrix}J^{\pm}_1 & 0 & 0\\ 0 & J^{\pm}_2 & 0\\ 0 & 0 & J^{\pm}_3\end{pmatrix},$$

$$\begin{aligned}\xi^{\pm}_1&=\xi^{\pm}_{1111}+\xi^{\pm}_{1212}\frac{\Gamma^{+}_{2211}}{\Gamma^{+}_{1111}}+2\xi^{\pm}_{1112}\frac{\Gamma^{+}_{1211}}{\Gamma^{+}_{1111}},\\ \xi^{\pm}_2&=\xi^{\pm}_{2222}+\xi^{\pm}_{1212}\frac{\Gamma^{+}_{1111}}{\Gamma^{+}_{2211}}+2\xi^{\pm}_{2212}\frac{\Gamma^{+}_{1211}}{\Gamma^{+}_{2211}},\\ \xi^{\pm}_3&=\xi^{\pm}_{1122}+\xi^{\pm}_{1221}+\xi^{\pm}_{1112}\frac{\Gamma^{+}_{1111}}{\Gamma^{+}_{1211}}+\xi^{\pm}_{2212}\frac{\Gamma^{+}_{2211}}{\Gamma^{+}_{1211}}.\end{aligned} \tag{18}$$

The quantities $J^{\pm}_i$ are constructed from $J^{\pm}_{m_1m_2m_3m_4}$ in the same way as the $\xi^{\pm}_i$.

Equation (17) gives us the frequency of collective oscillations for a two-band system with an arbitrary carrier density, including the case of low carrier density ($\mu\sim\Delta$ and $\mu<\omega_D$) at arbitrary temperatures. Actually, this is an equation for determining the collective oscillation frequency for an ordinary three-band system with order parameters Δ_{11}, Δ_{22}, and Δ_{12} and reduced carrier density ($\mu<\omega_D$).[31,42,47]

In ordinary superconductors ($\mu\gg\omega_D$), $\hat{I}^{\pm}=0$. Equation (17) separates into two independent equations, which determine the collective oscillations caused by fluctuations of the phases, $|\hat{R}^{+}|=0$ (the poles of the vertex functions $\hat{\Gamma}^{\mathrm{ph}}$ defined in Eq. (11)), and amplitudes, $|\hat{R}^{-}|=0$ (the poles of the vertex functions $\hat{\Gamma}^{a}$ defined in Eq. (11)), of the order parameters of the different bands.

In superconductors with low carrier density, $\hat{I}^{\pm}\neq0$. In this case the collective oscillations mix and cannot be separated. This leads to results that differ from those for ordinary superconductors.[30,31]

3. COLLECTIVE OSCILLATIONS

Reducing the rank of the determinant in (17) from six to three by using the theory of partitioned matrices,[48] we can write Eq. (17) for the collective oscillation frequency in the form

$$R^{-}|\hat{R}^{+}-\omega^2\hat{J}^{-}(\hat{R}^{-})^{-1}\hat{J}^{+}|=R^{+}R^{-}-\Delta Z, \tag{19}$$

where

$$\begin{aligned}\Delta Z=\omega^2&\mathrm{Tr}(\hat{J}^{+}\,{}^{t}\hat{Z}^{+}\hat{J}^{-}\,{}^{t}\hat{Z}^{-})\\ &-\omega^4\mathrm{Tr}(\hat{R}^{+}\,{}^{t}\hat{T}^{+}\hat{R}^{-}\,{}^{t}\hat{T}^{-})+\omega^6J^{+}J^{-}.\end{aligned} \tag{20}$$

Here $J^{\pm}=\det\|\hat{J}^{\pm}\|$, $\hat{Z}^{\pm}$ and $\hat{T}^{\pm}$ are the matrices of the cofactors of the elements of the matrices $\hat{R}^{\pm}$ and $\hat{J}^{\pm}$, respectively, and $({}^t\hat{Z}^{\pm})_{ik}=(\hat{Z}^{\pm})_{ki}$.

Equation (19) is symmetric with respect to $\mathbf{k}\to-\mathbf{k}$ and $\omega\to-\omega$ since $\hat{R}^{\pm}(-k)=\hat{R}^{\pm}(k)$ and $\hat{J}^{\pm}(-k)=-\hat{J}^{\pm}(k)$. This fact leads to a situation in which only even powers of ω and $\mathbf{k}$ are present in the series expansion of Eq. (19) in powers of small values of these quantities.

Let us examine Eq. (19) for the collective oscillation frequency in the case where ω is small $(\omega^2/4\Delta_i^2\ll 1)$ and $\mathbf{k}=0$ at $T=0$. To this end we expand Eq. (19) in a power series in this small parameter, keep only terms up to quadratic, and integrate with respect to Ω. We obtain

$$\xi_i^+=\xi_i^0+\omega^2\theta_i^+, \tag{21}$$

where θ_i are the components of the matrix $\hat{\theta}$:

$$\hat{\theta}=\hat{Q}\frac{\Delta_{11}\Delta_{22}}{\Delta_{12}^2}\alpha+\hat{\overline{\theta}}{}^+, \quad \hat{\theta}=\begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\end{pmatrix},$$

$$\hat{\overline{\theta}}{}^+=\begin{pmatrix}\overline{\theta_1^+}\\ \overline{\theta_2^+}\\ \overline{\theta_3^+}\end{pmatrix}, \quad \hat{Q}=\begin{pmatrix}\Delta_{12}^2/\Delta_{11}^2\\ \Delta_{12}^2/\Delta_{22}^2\\ -1\end{pmatrix}. \tag{22}$$

Both $\overline{\theta_-}-i^+$ and α are defined in Appendix D, and ξ_i^0 is defined in (30).

Consider the simple dispersion law

$$\varepsilon_i(\mathbf{k})=\zeta_i+\frac{k_x^2+k_y^2+k_z^2}{2m_i}. \tag{23}$$

We replace the sum with respect to $\mathbf{k}$ with the integral with respect to energy ε and truncate the integrals (22), (30), and (C4)–(C6) (see Appendix C) in accordance with the dispersion law (23), assuming that the carrier density is low $(\mu<\omega_D)$:

$$\frac{1}{V}\sum_{\mathbf{k}}\Phi(\varepsilon_i-\mu_i)=2N_i\int_{-\overline{D}_i}^{\overline{\omega}_{D_i}}d\varepsilon\,\Phi(\varepsilon-S_i)$$
$$=2N_i\int_{-D_i}^{\omega_{D_i}}d\varepsilon\,\Phi(\varepsilon), \tag{24}$$

where $\overline{\omega}_{D_i}$ is the phonon cutoff frequency in the ith band, and $N_i=m_ip_{Fi}/4\pi^2$ is the density of electronic states in the ith band. Here

$$\overline{D}_i=\begin{cases}\overline{\omega}_{D_i} & \text{if } \overline{\omega}_{D_i}<\mu-\zeta_i,\\ \mu-\zeta_i & \text{if } \overline{\omega}_{D_i}>\mu-\zeta_i,\end{cases}$$

and S_i is the mean-field renormalization term in the chemical potential (see Eq. (10)). The need to allow for this renormalization stems from the fact that S_i is of the same order of magnitude as the phonon cutoff frequency $\overline{\omega}_{D_i}$ and the chemical potential (see also Refs. 42–44).

Here we do not write the explicit expressions for the integrals θ_i, $J_i^{\pm}$, and ξ_i^- because of their complexity. Instead we note that $J_i^{\pm}\neq 0$ because the limits of integration in (24) are asymmetric and $\Delta_{12}\neq 0$, in contrast to the case of ordinary superconductors, where $J_i^{\pm}=0$. The presence of the additional order parameter Δ_{12} even when the limits of integration in (24) are symmetric results in a situation in which $J_i^{\pm}\neq 0$.

At low frequencies ω (more precisely, $\omega^2/4\Delta_i^2\ll 1$), $R^-\neq 0$ and Eq. (19) can be written as

$$R^+(\hat{\theta})=\frac{\Delta Z}{R^-}. \tag{25}$$

Reasoning on the basis of the definition (21) of ξ_i^+, we can conclude that when the effective constants of intraband interactions (V_{1111} and V_{2222}) and interactions characterizing the transitions of an interband pair into an interband pair (V_{1212}) are much larger than the constants of interactions characterizing the transitions of an intraband pair into an interband pair (V_{1112} and V_{2212}) and into an intraband pair,

$$V_{1111}, V_{2222}, V_{1212}\gg V_{1112}, V_{2212}, V_{1122}, \tag{26}$$

the equation for the collective oscillation frequency has the form

$$R^+(\hat{\overline{\theta}})=0, \tag{27}$$

where

$$R^+(\hat{\overline{\theta}})=R_0^+-\omega^2\mathrm{Tr}(\hat{\overline{\theta}}\,{}^t\hat{Z}_0^+)+\omega^4\mathrm{Tr}(\hat{R}_0^+\hat{\Theta})-\omega^6\overline{\theta},$$

$$\hat{\overline{\theta}}=\begin{pmatrix}\overline{\theta_1} & 0 & 0\\ 0 & \overline{\theta_2} & 0\\ 0 & 0 & \overline{\theta_3}\end{pmatrix}, \quad \overline{\theta}=\det\|\hat{\overline{\theta}}\|, \quad \overline{\theta_i}=\theta_i+\frac{J_i^+J_i^-}{2B_i}. \tag{28}$$

Here ${}^t\hat{Z}_0^+$ and $\hat{\Theta}$ are the matrices of the cofactors of the elements of the matrices $\hat{R}_0^+$ and $\hat{\overline{\theta}}$, respectively. When the carrier density is low $(\mu\sim\Delta$ and $\mu<\omega_D)$, the presence of the additional order parameter $\Delta_{12}\simeq\Delta_{11},\Delta_{22}$ and the asymmetric cutoff in (24) lead not only to an explicit dependence on the chemical potential via the functions θ_i^+ of (22) and (D4) but also to the mixing of fluctuations of the phases and amplitudes of the order parameters of different bands $(\hat{J}^{\pm},\Delta Z\neq 0)$. With the approximation (26), the effect of such mixing reduces to θ_i being replaced by $\overline{\theta_i}$ (see Eqs. (28)) (or $\widehat{\overline{\theta}}\to\hat{\theta}+\hat{J}^-(\hat{R}^-)^{-1}\hat{J}^+$; see Eq. (19)).

In addition to examining the system of equations (14) for the vertex functions $\hat{\Gamma}^{\mathrm{ph}}$ and $\hat{\Gamma}^a$, we consider the system of equations for the order parameter of a two-band system with low carrier density:[42]

$$\Delta_{ii}=\sum_{j=1}^{2}V_{iijj}\xi_i^0\Delta_{jj}, \tag{29}$$

where

$$\begin{aligned}\xi_1^0=&\left\{\frac{1}{d}\left[\varepsilon_1^2+\Delta_{11}^2-\varepsilon_2^2-\Delta_{22}^2+2\Delta_{12}^2\left(1+\frac{\Delta_{22}}{\Delta_{11}}\right)\right]+1\right\}\\&\times\frac{1}{E_1}-\left\{\frac{1}{d}\left[\varepsilon_1^2+\Delta_{11}^2-\varepsilon_2^2-\Delta_{22}^2+2\Delta_{12}^2\right.\right.\\&\left.\left.\times\left(1+\frac{\Delta_{22}}{\Delta_{11}}\right)\right]-1\right\}\frac{1}{E_2},\quad \xi_2^0=\left\{\frac{1}{d}\left[\varepsilon_2^2+\Delta_{22}^2-\varepsilon_1^2\right.\right.\\&\left.\left.-\Delta_{11}^2+2\Delta_{12}^2\left(1+\frac{\Delta_{11}}{\Delta_{22}}\right)\right]+1\right\}\frac{1}{E_1}\\&-\left\{\frac{1}{d}\left[\varepsilon_2^2+\Delta_{22}^2-\varepsilon_1^2-\Delta_{11}^2+2\Delta_{12}^2\right.\right.\\&\left.\left.\times\left(1+\frac{\Delta_{11}}{\Delta_{22}}\right)\right]-1\right\}\frac{1}{E_2},\quad \xi_3^0=2\left\{\left(\frac{1}{d}[(\varepsilon_1-\varepsilon_2)^2\right.\right.\\&\left.+(\Delta_{11}+\Delta_{22})^2]+1\right)\frac{1}{E_1}-\left(\frac{1}{d}[(\varepsilon_1-\varepsilon_2)^2\right.\\&\left.\left.+(\Delta_{11}+\Delta_{22})^2]-1\right)\frac{1}{E_2}\right\}.\end{aligned}\tag{30}$$

Both E_i and d are defined in Appendix B.

The compatibility condition for this system yields the equation

$$\det\|\hat{U}-\hat{\xi}^0\|=0.\tag{31}$$

In view of this, R_0^+ of (28) vanishes.

Allowing for Eqs. (29) for the order parameters, calculating the traces of the matrices and the determinants in (27) and (28), and doing simple algebraic transformations, we arrive at an equation for the collective oscillation frequency ($\mathbf{k}=0$):

$$\begin{aligned}&\omega^6\bar\theta_1\bar\theta_2\bar\theta_3-\frac{\omega^4}{\det|\hat V|}(\bar\theta_1\bar\theta_2B_{12}+\bar\theta_1\bar\theta_3B_{13}+\bar\theta_2\bar\theta_3B_{23})\\&\qquad+\left(\frac{\omega}{\det|\hat V|}\right)^2\left(\bar\theta_1+\bar\theta_2\frac{\Delta_2^2}{\Delta_1^2}+\bar\theta_3\frac{\Delta_3^2}{\Delta_1^2}\right)P_1=0.\end{aligned}\tag{32}$$

Here

$$\begin{aligned}P_1=&(V_{1212}V_{2211}-V_{2212}V_{1211})\\&\times(V_{2222}V_{1211}-V_{2211}V_{1222})\frac{\Delta_1^2}{\Delta_2\Delta_3}\\&+(V_{1212}V_{2211}-V_{2212}V_{1211})(V_{1111}V_{1222}-V_{1122}V_{1211})\\&\times\frac{\Delta_1}{\Delta_3}+(V_{1111}V_{2212}-V_{1112}V_{2211})\\&\times(V_{2222}V_{1211}-V_{2211}V_{1222})\frac{\Delta_1}{\Delta_2},\\B_{12}=&(V_{2222}V_{1211}-V_{2211}V_{1222})\frac{\Delta_1}{\Delta_3}\\&+(V_{1111}V_{1222}-V_{1122}V_{1211})\frac{\Delta_2}{\Delta_3}.\end{aligned}\tag{33}$$

The coefficient B_{12} can be obtained from B_{12} by subscript interchange 22↔12, while B_{23} can be obtained from B_{12} by subscript interchange 11↔12.

Equation (32) has a trivial solution, $\omega=0$. There are also two nontrivial solutions, which can be found by solving the equation

$$\begin{aligned}&\omega^4\bar\theta_1\bar\theta_2\bar\theta_3-\frac{\omega^2}{\det|\hat V|}(\bar\theta_1\bar\theta_2B_{12}+\bar\theta_1\bar\theta_3B_{13}+\bar\theta_2\bar\theta_3B_{23})\\&\qquad+\left(\frac{1}{\det|\hat V|}\right)^2\left(\bar\theta_1+\bar\theta_2\frac{\Delta_2^2}{\Delta_1^2}+\bar\theta_3\frac{\Delta_3^2}{\Delta_1^2}\right)P_1=0,\end{aligned}\tag{34}$$

whose solutions are

$$\omega_\pm^2=\frac{\bar\theta_1\bar\theta_2B_{12}+\bar\theta_1\bar\theta_3B_{13}+\bar\theta_2\bar\theta_3B_{23}\pm\sqrt{D}}{2\det|\hat V|\bar\theta_1\bar\theta_2\bar\theta_3},\tag{35}$$

where

$$\begin{aligned}D=&[\bar\theta_1\bar\theta_2B_{12}+\bar\theta_1\bar\theta_3B_{13}+\bar\theta_2\bar\theta_3B_{23}]^2-4P_1\bar\theta_1\bar\theta_2\bar\theta_3\\&\times\left(\bar\theta_1+\bar\theta_2\frac{\Delta_2^2}{\Delta_1^2}+\bar\theta_3\frac{\Delta_3^2}{\Delta_1^2}\right).\end{aligned}\tag{36}$$

In the approximation (26), the two solutions (35) of Eq. (34) are positive and real. This suggests that we are dealing with two excitonic modes, ω_+ and ω_-.

In addition to the conditions specified in (26), we assume that transitions of an intraband pair entirely into another band are stronger than transitions into an interband pair ($V_{1122}\gg V_{1112},V_{2212}$). Then the excitonic modes assume the form

$$\omega_+^2=\frac{\bar\theta_1V_{2211}\Delta_{11}/\Delta_{22}+\bar\theta_2V_{1122}\Delta_{22}/\Delta_{11}}{\bar\theta_1\bar\theta_2(V_{1111}V_{2222}-V_{1122}V_{2211})}+\gamma,\tag{37}$$

$$\begin{aligned}\omega_-^2=&\frac{V_{2222}V_{1211}\Delta_{11}/\Delta_{12}+V_{1111}V_{1222}\Delta_{22}\Delta_{12}}{V_{1212}\bar\theta_3(V_{1111}V_{2222}-V_{1122}V_{2211})}\\&\times\frac{\bar\theta_1\Delta_{11}^2+\bar\theta_2\Delta_{22}^2+\bar\theta_3\Delta_{12}^2}{\bar\theta_1\Delta_{11}^2+\bar\theta_2\Delta_{22}^2},\end{aligned}\tag{38}$$

where

$$\begin{aligned}&\bar\theta_i=(\overline{\theta_i^+}+Q_i\overline{\theta_3^+})g_i,\\&g_i=1+\frac{(\hat J^-(\hat R^-)^{-1}\hat J^+)_{ii}+Q_i(\underline{\hat J}^-(\underline{\hat R}^-)^{-1}\underline{\hat J}^+)_3}{\overline{\theta_i^+}+Q_i\overline{\theta_3^+}},\end{aligned}\tag{39}$$

γ contains terms proportional to V_{1112} and V_{2212}, and the definitions of $\underline{\hat J}^\pm$ and $\underline{\hat R}^\pm$ are given in Appendix D.

From (37) it immediately follows that the collective mode ω_+ is a generalized Leggett mode[30] for the case of low carrier density. The mode corresponds primarily to the interference of the scattering of an electron pair as a whole from the first band to the second and the scattering of an interband pair into an interband pair (the first term in (37)), with a small contribution from the scattering of an interband pair into an intraband pair. If we ignore the latter processes ($V_{1112}=V_{2212}=0$), we arrive at a collective Leggett mode[30] for the case where $V_{1212}\neq 0$ $(\Delta_{12}\neq 0)$. In the more general case there are two collective modes, ω_+ of (37) and ω_- of (38).

These excitonic modes depend on the chemical potential (or the carrier density) through the functions $\overline{\theta_i}$ and the coefficients B_{ij} and P_1 of (33).

What is important is that the system of equations (14) for the vertex functions and the order parameters (29) must be augmented by the law of conservation of the number of charge carriers:

$$N_0=\frac{2}{\beta V}\sum_{\mathbf{k}\omega}[G_{11}(\mathbf{k},\omega)+G_{22}(\mathbf{k},\omega)]e^{i\omega 0^+}. \tag{40}$$

The self-consistent system of equations (14), (17), (29), and (40) determines the vertex functions (the collective oscillation frequency), the order parameters Δ_{ij}, and the chemical potential μ for a given carrier density N_0 for the ground state $(T=0)$. A characteristic feature of the ground state of a system with low carrier density is that the position of the Fermi level changes substantially when the superconducting gap is formed. The order parameters are of the same order of magnitude as the chemical potential $(\mu\neq\Delta_{ij})$. This leads to an anomalous temperature dependence of the chemical potential μ (see Refs. 42–44) and the emergence of a break in the μ vs. T curve at $\mu\simeq 6$ meV (Ref. 42). The value of the chemical potential exceeds the values obtained in the single-band BCS model[43] and the Hubbard model[44] by a factor of three. On the other hand, the low carrier density $(\mu\sim\Delta_{ij})$ leads not only to a shift of the generalized Leggett mode[30] (because $\Delta_{12}\neq 0$) and the emergence of an additional excitonic mode (because $V_{1112},V_{2212}\neq 0$) but also to the appearance of a dependence in the acoustic mode and two excitonic collective modes on the carrier density N_0.

4. CONCLUSION

In this paper we have developed a unified approach to the study of collective oscillations in superconducting systems with two energy bands that overlap at the Fermi surface at any (including low) carrier densities ($\mu\sim\Delta$ and $\mu<\omega_D$). Formally this system is equivalent to a three-band superconducting system with order parameters Δ_{11}, Δ_{22}, and Δ_{12} and reduced carrier density N_0 given by Eq. (40). A two-band superconducting system with $\mu\sim\Delta$ $(\mu<\omega_D)$ produces a Bogolyubov–Anderson collective mode corresponding to the acoustic spectrum. This mode has been observed in single-band superconductors and in two- and three-band superconductors.[31,32] Two collective modes of the excitonic type also emerge. One is a modification of a Leggett mode[30] shifted because of the presence of an additional order parameter, Δ_{12}. The reason why such a mode emerges lies in the interference of the scattering of a Cooper pair as a whole from the first energy band to the second and the scattering of an interband pair into an interband pair. The second collective is due to the scattering of an interband Cooper pair into an intraband pair, and vice versa.

Note that when the carrier density is low ($\mu\sim\Delta$ and $\mu<\omega_D$), the frequencies of collective oscillations (acoustic and excitonic) acquire a substantial dependence on the carrier density N_0. Three factors contribute to this:

1) the explicit dependence on N_0 of the functions $\overline{\theta_i}$, which enter into the definition of the collective-oscillation frequencies (Eqs. (35) and (37)) via the functions $\overline{\theta_i^+}+Q_i\overline{\theta_3^+}$ of (39);

2) the mixing of fluctuations of the phases and amplitudes of the order parameters of different bands $(\hat{J},\Delta Z\neq 0)$ via the functions g_i of (39);

3) the strong dependence of the order parameters Δ_i on N_0 (see Refs. 25 and 42).

The first two factors are due to the presence of the additional order parameter Δ_{12} and the asymmetric cutoff of the integrals (24) in systems with low carrier density ($\mu\sim\Delta$ and $\mu<\omega_D$). By analogy with the three-band model,[47] these facts weaken each other in such systems. In systems with reduced carrier density $(\Delta<\mu<\omega_D)$ and in ordinary superconductors $(\mu>\omega_D)$ the two factors balance each other and we have $\overline{\theta_i}\approx N_i/2\Delta_{ii}^2$ $(i=1,2)$. Thus, in systems with low $(\mu\sim\Delta)$ and reduced $(\mu<\omega_D)$ carrier densities, a similar dependence of the order parameters provides the leading contribution to the dependence of the collective-oscillation frequencies on the carrier density N_0. Note that when dealing with systems with low ($\mu\sim\Delta$ and $\mu<\omega_D$) and reduced $(\mu<\omega_D)$ carrier densities, we must take into account the mean-field renormalization of the chemical potential[42–44] and the law of conservation of the number of charge carriers, which lead to a peak in the dependence of the order parameters on N_0. These peaks emerge in the range of reduced (not small) values of N_0, where $\Delta<\mu<\omega_D$. For instance, for a Leggett excitonic mode ω_+ of (37), these facts lead to the emergence of peaks in the ω_+ vs. N_0 dependence and to a strong dependence of the ratios $\omega_+^2/4\Delta_i^2$ ($i=11$, 22, and 12) on the carrier density N_0.

When $\Delta_{12}=0$, only one excitonic mode remains, ω_+. This mode acquires a strong dependence on the carrier density only via a similar dependence of the order parameters of different bands and their ratios.

Thus, an increase in the number of order parameters from two to three $(\Delta_{12}\neq 0)$ in a two-band superconducting system with low carrier density leads to the emergence of an additional excitonic mode and to a strong dependence of the collective modes (acoustic and excitonic) on the carrier density.

Note that in this paper we have studied collective oscillations in a two-band system by using the mean-field approximation (the ladder approximation for the vertex functions), whose applicability depends on the carrier density. For high carrier densities (the radius of an electron pair is much larger than the average interparticle distance), the description is exact. For very low carrier densities (the radius of an electron pair is much smaller than the average interparticle distance) the approximation yields a qualitatively correct picture.[49] The case $\mu\sim\Delta$ is intermediate and requires abandoning the mean-field approximation. The results obtained in the mean-field approximation for this case may be considered a qualitative interpolation of between the above limits.

It is important to note that the two-band system studied in this paper is similar to a system with two inequivalent layers. By diagonalizing the Hamiltonian of such a system we can easily show that

$$V_{1111}=\frac{1}{4}(V_1+V_2-4J+2Y+4W),$$

$$V_{2222}=\frac{1}{4}(V_1+V_2+4J+2Y+4W),$$

$$V_{1122}=\frac{1}{4}(V_1+V_2+4J-2Y),$$

$$V_{2211}=\frac{1}{4}(V_1+V_2-4J-2Y), \tag{41}$$

and

$$V_{1112}=V_{1121}=V_{2212}=V_{2221}=V_{1211}$$

$$=V_{2111}=V_{1222}=V_{2122}=\frac{(V_2-V_1)}{4},$$

$$V_{1212}=V_{2121}=\frac{1}{4}(V_1+V_2-2Y),$$

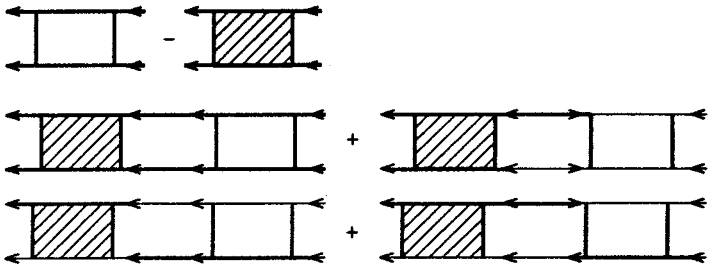
$$V_{1221}=V_{2112}=\frac{1}{4}(V_1+V_2+2Y-4W). \tag{42}$$

Here V_1 and V_2 are the intralayer coupling constants and Y and W are the interlayer coupling constants. Systems with two equivalent layers ($V_1=V_2=V$) have been studied earlier in Ref. 50, 37, and 39.

Thus, the inequivalence of the layers ($V_{1112}\neq 0$) leads to the emergence of a new excitonic mode ω_- of (35) and (38). The strong dependence of the collective-oscillation frequencies on the carrier density N_0 is due to the presence of layer coupling constants, which yields a finite order parameter Δ_{12}, and the asymmetry of the integration limits in (24).

APPENDIX A:

The equation for the vertex function $\Gamma^{\overset{+-}{+-}}$ has the form



(A1)

The other three vertex functions can be constructed in a similar manner without the first absolute term. The hatched part in Eq. (A1) corresponds to the vertex function $\Gamma^{\overset{0+-}{-+}}$ in the zeroth approximation.

Summing over the spin variables in the system of equations (A1) and passing to the $n\mathbf{k}\Omega$-representation for the vertex functions $\Gamma^{\alpha}(n_1\cdots n_4,m_1\cdots m_4=1,2)$, we obtain ($p_3+p_4=k$)

$$\Gamma^{\overset{+-}{+-}}_{m_1m_2m_3m_4}\left(p+\frac{k}{2},-p+\frac{k}{2},p_3,p_4\right)$$

$$=-V_{m_1m_2m_3m_4}+\frac{1}{\beta}\sum_q\sum_{n_1\cdots n_4}V_{m_1m_2n_1n_3}$$

$$\times\left\{G_{n_1n_2}\left(q+\frac{k}{2}\right)G_{n_3n_4}\left(-q+\frac{k}{2}\right)\Gamma^{\overset{+-}{+-}}_{n_2n_4m_3m_4}\right.$$

$$\times\left(q+\frac{k}{2},-q+\frac{k}{2},p_3,p_4\right)+F_{n_1n_2}\left(q+\frac{k}{2}\right)$$

$$\times F_{n_3n_4}\left(-q+\frac{k}{2}\right)\Gamma^{\overset{--}{--}}_{n_2n_4m_3m_4}\left(-q-\frac{k}{2},q-\frac{k}{2},p_3,p_4\right)$$

$$-G_{n_1n_2}\left(q+\frac{k}{2}\right)F_{n_3n_4}\left(-q+\frac{k}{2}\right)\Gamma^{\overset{+-}{--}}_{n_2n_4m_3m_4}$$

$$\times\left(q+\frac{k}{2},q-\frac{k}{2},p_3,p_4\right)+F_{n_1n_2}\left(q+\frac{k}{2}\right)G_{n_3n_4}$$

$$\times\left.\left(-q+\frac{k}{2}\right)\Gamma^{\overset{--}{+-}}_{n_2n_4m_3m_4}\left(-q-\frac{k}{2},-q+\frac{k}{2},p_3,p_4\right)\right\},$$

$$\Gamma^{\overset{--}{--}}_{m_1m_2m_3m_4}\left(p+\frac{k}{2},-p+\frac{k}{2},p_3,p_4\right)$$

$$=\frac{1}{\beta}\sum_q\sum_{n_1\cdots n_4}V_{m_1m_2n_2n_4}\left\{G_{n_1n_2}\left(q+\frac{k}{2}\right)\right.$$

$$\times G_{n_3n_4}\left(-q+\frac{k}{2}\right)\Gamma^{\overset{--}{--}}_{n_1n_3m_3m_4}\left(q+\frac{k}{2},-q\right.$$

$$\left.+\frac{k}{2},p_3,p_4\right)+\overline{F}_{n_1n_2}\left(q+\frac{k}{2}\right)\overline{F}_{n_3n_4}\left(-q+\frac{k}{2}\right)$$

$$\times\Gamma^{\overset{+-}{+-}}_{n_1n_3m_3m_4}\left(-q-\frac{k}{2},q-\frac{k}{2},p_3,p_4\right)-G_{n_1n_2}\left(q+\frac{k}{2}\right)$$

$$\times F_{n_3n_4}\left(-q+\frac{k}{2}\right)\Gamma^{\substack{--\\+-}}_{n_1n_3m_3m_4}\left(q+\frac{k}{2},q-\frac{k}{2},p_3,p_4\right)$$

$$+\overline{F}_{n_1n_2}\left(q+\frac{k}{2}\right)G_{n_3n_4}\left(-q+\frac{k}{2}\right)$$

$$\left.\times\Gamma^{\substack{+-\\--}}_{n_1n_3m_3m_4}\left(-q-\frac{k}{2},-q+\frac{k}{2},p_3,p_4\right)\right\}. \qquad \text{(A2)}$$

In the BCS approximation, for the vertex functions Γ^α of Eq. (11) we arrive at the system of equations (6) in matrix form.

APPENDIX B:

For superconducting systems with low carrier density, the one-particle Green's functions have the following form $(z=i\omega)$:[42]

$$G_{11}(\mathbf{k},z)=\frac{1}{D(z)}[(z+\overline{\varepsilon}_1)(z^2-\xi_2^2)-\Delta_{12}^2(z+\overline{\varepsilon}_2)],$$

$$G_{12}(\mathbf{k},z)=\frac{\Delta_{12}}{D(z)}[\Delta_{11}(z+\overline{\varepsilon}_2)-\Delta_{22}(z+\overline{\varepsilon}_1)],$$

$$F_{11}(\mathbf{k},z)=-\frac{\Delta_{11}}{D(z)}\left(z^2-\xi_2^2+\Delta_{12}^2\frac{\Delta_{22}}{\Delta_{11}}\right),$$

$$F_{12}(\mathbf{k},z)=-\frac{\Delta_{12}}{D(z)}[\Delta_{11}\Delta_{22}-\Delta_{12}^2+(z+\overline{\varepsilon}_1)(z+\overline{\varepsilon}_2)]. \qquad \text{(B1)}$$

The functions G_{22}, G_{21}, F_{22}, and F_{21} can be obtained from G_{11}, G_{12}, F_{11}, and F_{12} by interchanging the subscripts: $1\leftrightarrow 2$. Here

$$\overline{\varepsilon}_n=\varepsilon_n-\mu-S_n,\quad \xi_n^2=\overline{\varepsilon}_n^2+\Delta_{nn}^2,$$

$$D(z)=(z^2-\xi_1^2)(z^2-\xi_2^2)+2\Delta_{12}^2(\overline{\varepsilon}_1\overline{\varepsilon}_2-z^2)$$

$$+(\Delta_{11}\Delta_{22}-\Delta_{12}^2)^2-\Delta_{12}^4, \qquad \text{(B2)}$$

and S_n has been defined in (10). The definitions (B1) and (B2) lead to the following symmetry properties $(k=\mathbf{k},z)$:

$$G_{nm}(k)=G_{mn}(k),\quad F_{nm}(-k)=F_{mn}(k)=\overline{F}_{nm}(k). \qquad \text{(B3)}$$

To integrate the functions $\xi^\pm_{n_1n_2n_3n_4}$ and $I^\pm_{n_1n_2n_3n_4}$ with respect to Ω, we write the normal and anomalous Green's functions in the form

$$G_{ij}=\frac{a_{ij}}{\omega-E_1+i\delta}+\frac{b_{ij}}{\omega-E_2+i\delta}+\frac{c_{ij}}{\omega+E_1-i\delta}+\frac{d_{ij}}{\omega+E_2-i\delta},$$

$$F_{ij}=\frac{l_{ij}}{\omega-E_1+i\delta}+\frac{m_{ij}}{\omega-E_2+i\delta}+\frac{k_{ij}}{\omega+E_1-i\delta}+\frac{n_{ij}}{\omega+E_2-i\delta},$$

$$\overline{F}_{ji}=F_{ij}, \qquad \text{(B4)}$$

where

$$a_{ii}=p_{i1}+q_{i1},\quad b_{ii}=-(p_{i2}+q_{i2}),$$

$$c_{ii}=p_{i1}-q_{i1},\quad d_{ii}=-(p_{i2}-q_{i2}),$$

$$p_{ij}=\frac{E_j^2-\xi_{3-i}^2-\Delta_{12}^2}{2d},\quad q_{ij}=\frac{(-1)^i\Delta_{12}h+\varepsilon_i a_{ij}}{E_j},$$

$$h=-\Delta_{12}\frac{\varepsilon_1-\varepsilon_2}{2d},\quad c=\Delta_{12}\frac{\Delta_{11}+\Delta_{22}}{2d},$$

$$a_{12}=a_{21}=c+r_1,\quad b_{12}=b_{21}=-(c+r_2),$$

$$c_{12}=c_{21}=c-r_1,\quad d_{12}=d_{21}=-(c-r_2),$$

$$r_i=\frac{\varepsilon_1\varepsilon_2-\Delta_{22}h}{E_i},\ \ s_i=-\Delta_{12}\frac{E_i^2-\Delta_{12}^2+\Delta_{11}\Delta_{22}-\varepsilon_1\varepsilon_2}{2E_id},$$

$$l_{ii}=-k_{ii}=t_{i1},\ \ -m_{ii}=n_{ii}=t_{i2},\ \ t_{ij}=-\frac{\Delta_{ii}a_{ii}+\Delta_{12}c}{E_i},$$

$$l_{12}=-k_{21}=s_1+h,\quad -m_{12}=n_{21}=s_2+h,$$

$$k_{12}=-l_{21}=s_1-h,\quad -n_{12}=m_{21}=s_2-h. \qquad \text{(B5)}$$

APPENDIX C:

Integrating $\tilde{\xi}^\pm_{n_1n_2n_3n_4}(k)$ and $\tilde{I}^\pm_{n_1n_2n_3n_4}(k)$ with respect to Ω $(T=0)$, we obtain

$$\tilde{\xi}^+_{lrnm}=\tilde{A}^+_{lrnm}+\tilde{B}^+_{lrnm},\quad \tilde{\xi}^-_{lrnm}=\tilde{A}^+_{lrnm}-\tilde{B}^+_{lrnm},$$

where

$$\tilde{A}^+_{lrnm}(k)$$

$$=\frac{1}{2}\sum_p\left\{\frac{a^+_{lr}a^-_{nm}+a^+_{mn}a^-_{lr}+c^+_{lr}c^-_{nm}+c^+_{nm}c^-_{lr}}{(E_1^++E_1^-)^2-\omega^2}(E_1^++E_1^-)\right.$$

$$+\frac{a^+_{lr}b^-_{nm}+a^+_{mn}b^-_{lr}+c^+_{lr}d^-_{nm}+c^+_{nm}d^-_{lr}}{(E_1^++E_2^-)^2-\omega^2}(E_1^++E_2^-)$$

$$+\frac{b^+_{lr}a^-_{nm}+b^+_{mn}a^-_{lr}+d^+_{lr}c^-_{nm}+d^+_{nm}c^-_{lr}}{(E_2^++E_1^-)^2-\omega^2}(E_2^++E_1^-)$$

$$\left.+\frac{b^+_{lr}b^-_{nm}+b^+_{mn}b^-_{lr}+d^+_{lr}d^-_{nm}+d^+_{nm}d^-_{lr}}{(E_2^++E_2^-)^2-\omega^2}(E_2^++E_2^-)\right\},$$

$$\tilde{B}^+_{lrnm}(k)$$

$$=\frac{1}{2}\sum_p\left\{\frac{l^+_{lr}l^-_{nm}+l^+_{nm}l^-_{lr}+k^+_{lr}k^-_{nm}+k^+_{nm}k^-_{lr}}{(E_1^++E_1^-)^2-\omega^2}(E_1^++E_1^-)\right.$$

$$+\frac{l^+_{lr}m^-_{nm}+l^+_{mn}m^-_{lr}+k^+_{lr}n^-_{nm}+k^+_{nm}n^-_{lr}}{(E_1^++E_2^-)^2-\omega^2}(E_1^++E_2^-)$$

$$+\frac{m^+_{lr}l^-_{nm}+m^+_{mn}l^-_{lr}+n^+_{lr}k^-_{nm}+n^+_{nm}k^-_{lr}}{(E_2^++E_1^-)^2-\omega^2}(E_2^++E_1^-)$$

$$+\frac{m^+_{lr}m^-_{nm}+m^+_{mn}m^-_{lr}+m^+_{lr}m^-_{nm}+m^+_{nm}m^-_{lr}}{(E_2^++E_2^-)^2-\omega^2}$$

$$\left.\times(E_2^++E_2^-)\right\}. \qquad \text{(C1)}$$

The quantities on the right-hand side of Eq. (C1) are functions of $p \pm k/2$, i.e.,

$$a_{lr}^{\pm} = a_{lr}\left(p \pm \frac{k}{2}\right), \quad E_i^{\pm} = E_i\left(p \pm \frac{k}{2}\right). \tag{C2}$$

Similarly, for $\tilde{I}^{\pm}_{n_1 n_2 n_3 n_4}(k)$ we have

$$I^{+}_{lrnm} = \tilde{A}^{-}_{lrnm} + \tilde{B}^{-}_{lrnm} = \omega J^{+}_{lrnm},$$

$$I^{-}_{lrnm} = \tilde{A}^{-}_{lrnm} - \tilde{B}^{-}_{lrnm} = \omega J^{-}_{lrnm},$$

where

$$\begin{aligned}
\tilde{A}^{-}_{lrnm}(k) \\
&= \frac{\omega}{2}\sum_p \left\{ \frac{a^{+}_{lr}a^{-}_{nm} + a^{+}_{mn}a^{-}_{lr} - (c^{+}_{lr}c^{-}_{nm} + c^{+}_{nm}c^{-}_{lr})}{(E^{+}_1 + E^{-}_1)^2 - \omega^2}(E^{+}_1 + E^{-}_1) \right. \\
&\quad + \frac{a^{+}_{lr}b^{-}_{nm} + a^{+}_{mn}b^{-}_{lr} - (c^{+}_{lr}d^{-}_{nm} + c^{+}_{nm}d^{-}_{lr})}{(E^{+}_1 + E^{-}_2)^2 - \omega^2}(E^{+}_1 + E^{-}_2) \\
&\quad + \frac{b^{+}_{lr}a^{-}_{nm} + b^{+}_{mn}a^{-}_{lr} - (d^{+}_{lr}c^{-}_{nm} + d^{+}_{nm}c^{-}_{lr})}{(E^{+}_2 + E^{-}_1)^2 - \omega^2}(E^{+}_2 + E^{-}_1) \\
&\quad + \frac{b^{+}_{lr}b^{-}_{nm} + b^{+}_{mn}b^{-}_{lr} - (d^{+}_{lr}d^{-}_{nm} + d^{+}_{nm}d^{-}_{lr})}{(E^{+}_2 + E^{-}_2)^2 - \omega^2}(E^{+}_2 + E^{-}_2),
\end{aligned}$$

$$\begin{aligned}
\tilde{B}^{-}_{lrnm}(k) \\
&= \frac{\omega}{2}\sum_p \left\{ \frac{l^{+}_{lr}l^{-}_{nm} + l^{+}_{mn}l^{-}_{lr} - (k^{+}_{lr}k^{-}_{nm} + k^{+}_{nm}k^{-}_{lr})}{(E^{+}_1 + E^{-}_1)^2 - \omega^2}(E^{+}_1 + E^{-}_1) \right. \\
&\quad + \frac{l^{+}_{lr}m^{-}_{nm} + l^{+}_{mn}m^{-}_{lr} - (k^{+}_{lr}n^{-}_{nm} + k^{+}_{nm}n^{-}_{lr})}{(E^{+}_1 + E^{-}_2)^2 - \omega^2}(E^{+}_1 + E^{-}_2) \\
&\quad + \frac{m^{+}_{lr}l^{-}_{nm} + m^{+}_{mn}l^{-}_{lr} - (n^{+}_{lr}k^{-}_{nm} + n^{+}_{nm}k^{-}_{lr})}{(E^{+}_2 + E^{-}_1)^2 - \omega^2}(E^{+}_2 + E^{-}_1) \\
&\quad + \frac{m^{+}_{lr}m^{-}_{nm} + m^{+}_{mn}m^{-}_{lr} - (m^{+}_{lr}m^{-}_{nm} + m^{+}_{nm}m^{-}_{lr})}{(E^{+}_2 + E^{-}_2)^2 - \omega^2} \\
&\quad \left. \times (E^{+}_2 + E^{-}_2) \right\}.
\end{aligned} \tag{C3}$$

The quantities on the right-hand side of Eq. (C3) are functions of $p \pm k/2$ (see (B2)).

When $\mathbf{k} = 0$, the expressions (C1) and (C3) for $\tilde{A}^{\pm}$ and $\tilde{B}^{\pm}$ are much simpler:

$$\begin{aligned}
\tilde{A}^{+}_{lrnm} = \sum_p \left\{ 2E_1 \frac{a_{lr}a_{nm} + c_{lr}c_{nm}}{4E_1^2 - \omega^2} + 2E_2 \frac{b_{lr}b_{nm} + d_{lr}d_{nm}}{4E_2^2 - \omega^2} \right. \\
\left. + (E_1 + E_2)\frac{a_{lr}b_{nm} + a_{mn}b_{lr} + c_{lr}d_{nm} + c_{nm}d_{lr}}{(E_1 + E_2)^2 - \omega^2} \right\},
\end{aligned}$$

$$\begin{aligned}
\tilde{B}^{+}_{lrnm} = \sum_p \left\{ 2E_1 \frac{l_{lr}l_{nm} + k_{lr}k_{nm}}{4E_1^2 - \omega^2} + 2E_2 \frac{m_{lr}m_{nm} + n_{lr}n_{nm}}{4E_2^2 - \omega^2} \right. \\
\left. + (E_1 + E_2)\frac{l_{lr}m_{nm} + l_{mn}m_{lr} + k_{lr}n_{nm} + k_{nm}n_{lr}}{(E_1 + E_2)^2 - \omega^2} \right\}.
\end{aligned} \tag{C4}$$

Similarly, for $\tilde{A}^{-}$ and $\tilde{B}^{-}$ we have

$$\tilde{A}^{-}_{lrnm} = \omega A_{lrnm}, \quad \tilde{B}^{-}_{lrnm} = \omega B_{lrnm}, \tag{C5}$$

where

$$\begin{aligned}
A_{lrnm} = \sum_p \left\{ \frac{a_{lr}a_{nm} - c_{lr}c_{nm}}{4E_1^2 - \omega^2} + \frac{b_{lr}b_{nm} - d_{lr}d_{nm}}{4E_2^2 - \omega^2} \right. \\
\left. + \frac{a_{lr}b_{nm} + a_{mn}b_{lr} - (c_{lr}d_{nm} + c_{nm}d_{lr})}{(E_1 + E_2)^2 - \omega^2} \right\},
\end{aligned}$$

$$\begin{aligned}
B_{lrnm} = \sum_p \left\{ \frac{l_{lr}l_{nm} - k_{lr}k_{nm}}{4E_1^2 - \omega^2} + \frac{m_{lr}m_{nm} - n_{lr}n_{nm}}{4E_2^2 - \omega^2} \right. \\
\left. + \frac{l_{lr}m_{nm} + l_{mn}m_{lr} - (k_{lr}n_{nm} + k_{nm}n_{lr})}{(E_1 + E_2)^2 - \omega^2} \right\}.
\end{aligned} \tag{C6}$$

APPENDIX D:

From the system of equations (14) we obtain an equation for finding the ratios of the vertex Green's functions:

$$\begin{pmatrix} \underline{\hat{R}}^{+} & -\omega \underline{\hat{J}}^{-} \\ -\omega \underline{\hat{J}}^{+} + \hat{I}^{+}_1 & \underline{\hat{R}}^{-} \end{pmatrix} \begin{pmatrix} \hat{\Gamma}^{+} \\ \hat{\Gamma}^{-} \end{pmatrix} = \begin{pmatrix} -\hat{R}^{+}_I \\ \omega \hat{J}^{+}_I \end{pmatrix}, \tag{D1}$$

where

$$\underline{\hat{R}}^{+} = \begin{pmatrix} 1 & R^{+}_{12} & R^{+}_{13} \\ 0 & R^{+}_{22} & R^{+}_{23} \\ 0 & R^{+}_{32} & R^{+}_{33} \end{pmatrix}, \quad \underline{\hat{R}}^{-} = \begin{pmatrix} R^{-}_{11} & R^{-}_{12} & R^{-}_{13} \\ R^{-}_{21} & R^{-}_{22} & R^{-}_{23} \\ R^{-}_{31} & R^{-}_{32} & R^{-}_{33} \end{pmatrix},$$

$$\hat{R}^{+}_I = \begin{pmatrix} R^{+}_{11} \\ R^{+}_{21} \\ R^{+}_{31} \end{pmatrix}, \quad \underline{\hat{J}}^{+} = \begin{pmatrix} 0 & J^{-}_{33} & 2J^{-}_{13} \\ 0 & J_{22} & 2J^{+}_{23} \\ 0 & 2J^{-}_{23} & 2J_{12} \end{pmatrix},$$

$$\underline{\hat{J}}^{-} = \begin{pmatrix} J_{11} & J^{+}_{33} & 2J^{+}_{13} \\ J^{-}_{33} & J_{22} & 2J^{-}_{23} \\ 2J^{-}_{13} & 2J^{+}_{23} & 2J_{12} \end{pmatrix}, \quad \hat{J}^{+}_I = \begin{pmatrix} J_{11} \\ J_{33} \\ 2J^{+}_{13} \end{pmatrix},$$

$$\underline{\hat{\Gamma}}^{+} = \begin{pmatrix} 1/\Gamma^{+}_{1111} \\ \Gamma^{+}_{2211}/\Gamma^{+}_{1111} \\ \Gamma^{+}_{1211}/\Gamma^{+}_{1111} \end{pmatrix}, \quad \underline{\hat{\Gamma}}^{-} = \begin{pmatrix} \Gamma^{-}_{1111}/\Gamma^{+}_{1111} \\ \Gamma^{-}_{2211}/\Gamma^{+}_{1111} \\ \Gamma^{-}_{1211}/\Gamma^{+}_{1111} \end{pmatrix},$$

$$\hat{I}^{+}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{D2}$$

To within terms of order ω^2 we have

$$\underline{\hat{\Gamma}}^{+}=\begin{pmatrix}0\\ \Delta_{22}/\Delta_{11}\\ \Delta_{12}/\Delta_{11}\end{pmatrix}+\omega^{2}\hat{\alpha}, \tag{D3}$$

where

$$\hat{\alpha}=(\underline{\hat{R}}^{+})^{-1}\begin{pmatrix}\tilde{\theta}_{1}\\ \tilde{\theta}_{2}\Delta_{22}/\Delta_{11}\\ 2\tilde{\theta}_{3}\Delta_{12}/\Delta_{11}\end{pmatrix},$$

$$(\underline{\hat{R}}^{+})^{-1}=\frac{1}{Z^{+}_{11}}\begin{pmatrix}Z^{+}_{11} & Z^{+}_{21} & Z^{+}_{31}\\ 0 & R_{33} & -R_{23}\\ 0 & -R_{32} & R_{22}\end{pmatrix},$$

$$\underline{\hat{J}}^{+}=\begin{pmatrix}\overline{J}_{11}\\ \overline{J}_{22}\Delta_{22}/\Delta_{11}\\ 2\overline{J}_{12}\Delta_{12}/\Delta_{11}\end{pmatrix},$$

$$\overline{\theta_{i}}=\overline{\theta^{+}_{i}}+(\hat{J}^{-}(\hat{R}^{-})^{-1})_{i1}\overline{J}_{11}+(\hat{J}^{-}(\hat{R}^{-})^{-1})_{i2}\times\frac{\Delta_{22}}{\Delta_{11}}\overline{J}_{22}+2(\hat{J}^{-}(\hat{R}^{-})^{-1})_{i3}\frac{\Delta_{12}}{\Delta_{11}}\overline{J}_{12},$$

$$\overline{\theta^{+}_{1}}=\theta^{+}_{11}+\frac{\Delta_{22}}{\Delta_{11}}\theta^{+}_{33}+2\frac{\Delta_{12}}{\Delta_{11}}\theta^{+}_{13},$$

$$\overline{J}_{11}=J_{11}+\frac{\Delta_{22}}{\Delta_{11}}J^{-}_{33}+2\frac{\Delta_{12}}{\Delta_{11}}J^{-}_{13},$$

$$\overline{\theta^{+}_{2}}=\theta^{+}_{22}+\frac{\Delta_{11}}{\Delta_{22}}\theta^{+}_{33}+2\frac{\Delta_{12}}{\Delta_{22}}\theta^{+}_{23},$$

$$\overline{J}_{22}=J_{22}+\frac{\Delta_{11}}{\Delta_{22}}J^{+}_{33}+2\frac{\Delta_{12}}{\Delta_{22}}J^{+}_{23},$$

$$\overline{\theta^{+}_{3}}=\theta^{+}_{12}+\frac{\Delta_{11}}{\Delta_{12}}\theta^{+}_{13}+\frac{\Delta_{22}}{\Delta_{12}}\theta^{+}_{23},$$

$$\overline{J}_{12}=J_{12}+\frac{\Delta_{11}}{\Delta_{12}}J^{+}_{13}+\frac{\Delta_{22}}{\Delta_{12}}J^{-}_{23}. \tag{D4}$$

In notation (D2)–(D4) we have used the ‘‘pseudoband’’ representation,[42]

$$11\rightarrow 1,\quad 22\rightarrow 2,\quad 12\rightarrow 3. \tag{D5}$$

For instance,

$$J_{11}=J_{1111},\quad J_{33}=J_{1212},\quad J_{12}=J_{1122}+J_{1221},$$

$$R_{11}=U_{11}-\xi^{0}_{1111},\quad R_{12}=U_{12}-\xi^{0}_{1212}.$$

Representing the ratios of the vertex functions as

$$\frac{\Gamma^{+}_{2211}}{\Gamma^{+}_{1111}}=\frac{\Delta_{22}}{\Delta_{11}}(1+\omega^{2}\alpha_{12}),\quad \frac{\Gamma^{+}_{1211}}{\Gamma^{+}_{1111}}=\frac{\Delta_{12}}{\Delta_{11}}(1+\omega^{2}\alpha_{13}), \tag{D6}$$

we arrive at an expression for α of Eq. (22):

$$\alpha=(\alpha_{12}-2\alpha_{13})\xi^{0}_{33}=\frac{\xi^{0}_{33}}{Z^{+}_{11}}\left[\left(R_{32}\frac{\Delta_{22}}{\Delta_{12}}-R_{31}\frac{\Delta_{11}}{\Delta_{12}}\right)\tilde{\theta}_{2}+\left(4R_{21}\frac{\Delta_{11}}{\Delta_{22}}+2R_{23}\frac{\Delta_{12}}{\Delta_{22}}\right)\tilde{\theta}_{3}\right]. \tag{D7}$$

In the approximation (26), for the ratios constructed from the $\hat{\Gamma}^{-}$ of Eqs. (11) we have

$$\frac{\Gamma^{-}_{2211}}{\Gamma^{-}_{1111}}=\frac{Z^{-}_{12}}{Z^{-}_{11}},\quad \frac{\Gamma^{-}_{1211}}{\Gamma^{-}_{1111}}=\frac{Z^{-}_{13}}{Z^{-}_{11}}. \tag{D8}$$

$$J^{-}_{1}=J_{11}+\frac{Z^{-}_{12}}{Z^{-}_{11}}J^{-}_{33}+2\frac{Z^{-}_{13}}{Z^{-}_{11}}J^{-}_{13},\quad J^{-}_{2}=J_{22}+\frac{Z^{-}_{11}}{Z^{-}_{12}}J^{+}_{33}+2\frac{Z^{-}_{13}}{Z^{-}_{12}}J^{+}_{23},$$

$$J^{-}_{3}=J_{12}+\frac{Z^{-}_{11}}{Z^{-}_{13}}J^{+}_{13}\frac{Z^{-}_{12}}{Z^{-}_{13}}J^{-}_{23}. \tag{D9}$$

Similarly, for J^{+}_{i} of Eqs. (18) we have

$$J^{+}_{1}=J_{11}+\frac{\Delta_{22}}{\Delta_{11}}J^{+}_{33}+2\frac{\Delta^{-}_{12}}{\Delta_{11}}J^{+}_{13},$$

$$J^{+}_{2}=J_{22}+\frac{\Delta_{11}}{\Delta_{22}}J^{-}_{33}+2\frac{\Delta_{12}}{\Delta_{22}}J^{-}_{23},$$

$$J^{+}_{3}=J_{12}+\frac{\Delta_{11}}{\Delta_{12}}J^{-}_{13}\frac{\Delta_{22}}{\Delta_{12}}J^{+}_{23}. \tag{D10}$$

The quantities ξ^{-}_{i} of (18) and B_{i} of (28) are constructed in the same manner as J^{-}_{i} of (D9).

*⁾E-mail: statphys@cc.acad.md

---

[1] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **106**, 162 (1957).

[2] N. N. Bogolyubov, D. N. Zubarev, and Yu. A. Tserkovnikov, Dokl. Akad. Nauk SSSR **117**, 788 (1957).

[3] N. N. Bogolyubov, Zh. Éksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys. JETP **7**, 41 (1958)].

[4] N. N. Bogolyubov, V. V. Tolmachev, and D. V. Shirkov, *A New Method in the Theory of Superconductivity*, Consultants Bureau, New York (1959).

[5] A. S. Éliashberg, Zh. Éksp. Teor. Fiz. **38**, 966 (1960) [Sov. Phys. JETP **11**, 696 (1960)].

[6] A. C. Aleksandrov, JETP Lett. **46**, S107 (1987).

[7] J. E. Hirsch and F. Marsiglio, Phys. Rev. B **39**, 11 515 (1989).

[8] V. Z. Kresin, Phys. Rev. B **35**, 8716 (1987).

[9] J. Ruvalds, Phys. Rev. B **35**, 8869 (1987).

[10] P. W. Anderson, Science **235**, 1196 (1987).

[11] V. M. Loktev, Fiz. Nizk. Temp. **22**, 3 (1996) [Low Temp. Phys. **22**, 1 (1996)].

[12] V. A. Moskalenko, Fiz. Met. Metalloved. **8**, 503 (1959).

[13] H. Suhl, B. T. Matthias, and L. R. Walker, Phys. Rev. Lett. **3**, 552 (1959).

[14] V. A. Moskalenko, L. Z. Kon, and M. E. Palistrant, *Low-Temperature Properties of Metals with Band-Spectrum Singularities* [in Russian], Shtiintsa, Kishinev (1989).

[15] V. A. Moskalenko, *Electromagnetic and Kinetic Properties of Superconducting Alloys with Overlapping Energy Bands* [in Russian], Shtiintsa, Kishinev (1976).

[16] B. T. Geĭlikman, Usp. Fiz. Nauk **88**, 327 (1973) [Sov. Phys. Usp. **9**, 142 (1973)].

[17] D. H. Lee and J. Ihm, Solid State Commun. **62**, 811 (1987).

[18] V. A. Moskalenko, M. E. Palistrant, and V. M. Vakalyuk, *Mechanisms of High-Temperature Superconductivity* [in Russian], JINR, Dubna (1988).

[19] V. A. Moskalenko, M. E. Palistrant, and V. M. Vakalyuk, Fiz. Nizk. Temp. **15**, 378 (1989) [Sov. J. Low Temp. Phys. **15**, 213 (1989)].

[20] V. A. Moscalenco, M. E. Palistrant, V. M. Vackalyuk, and I. V. Padure, Solid State Commun. **69**, 747 (1989).

[21] M. E. Palistrant and V. M. Vakalyuk, Sverkhprovodimost': Fiz., Khim., Tekhn. **3**, 1805 (1990) [Supercond., Phys. Chem. Technol. **3**, S215 (1990)].

[22] V. Z. Kresin and S. A. Wolf, Phys. Rev. B **41**, 4278 (1990).

[23] V. A. Moskalenko, M. E. Palistrant, and V. M. Vakalyuk, Usp. Fiz. Nauk **161**, No. 8, 155 (1991) [Sov. Phys. Usp. **34**, 717 (1991)].

[24] V. Z. Kresin and S. A. Wolf, Phys. Rev. B **46**, 6458 (1992).

[25] M. E. Palistrant and F. G. Kochorbé, Physica C **194**, 351 (1992).

[26] M. E. Palistrant and M. G. Kalalb, Izv. Akad. Nauk Moldovy No. 1, 70 (1992).

[27] H. Krakauer and W. E. Pickett, Phys. Rev. Lett. **60**, 1665 (1988).

[28] J. F. Herman, R. V. Kasowski, and W. J. Hsu, Phys. Rev. B **36**, 6904 (1987).

[29] M. G. Kalalb, F. G. Kochorbé, and M. E. Palistrant, Teor. Mat. Fiz. **91**, 483 (1992).

[30] A. J. Leggett, Prog. Theor. Phys. **36**, 901 (1966).

[31] M. E. Palistrant, Teor. Mat. Fiz. **95**, 101 (1993).

[32] M. E. Palistrant, V. M. Vackalyuk, and M. G. Calalb, Physica C **208**, 170 (1993).

[33] M. E. Palistrant, Physica C **235–240**, 2135 (1994).

[34] M. E. Palistrant, Teor. Mat. Fiz. **103**, 312 (1995).

[35] M. E. Palistrant and F. G. Kochorbe, Physica C **241**, 345 (1995).

[36] M. E. Palistrant and M. G. Kalalb, Izv. Akad. Nauk Moldovy No. 1, 30 (1996); Teor. Mat. Fiz. **110**, 162 (1996).

[37] E. V. Gorbar, V. M. Loktev, and S. G. Sharapov, Fiz. Nizk. Temp. **21**, No. 4, 21 (1995) [Low Temp. Phys. **21**, 329 (1995)].

[38] M. Frick and T. Scheider, Z. Phys. B **81**, 337 (1990).

[39] F. Forsthofer, S. Kind, and J. Keller, Phys. Rev. B **53**, 14 481 (1996).

[40] H. A. Fertig and S. Das Sarma, Phys. Rev. Lett. **65**, 1482 (1990).

[41] W.-C. Wu and A. Griffin, Phys. Rev. Lett. **74**, 158 (1995).

[42] F. G. Kochorbé and M. E. Palistrant, Zh. Éksp. Teor. Fiz. **104**, 3084 (1993) [JETP **77**, 442 (1993)]; Teor. Mat. Fiz. **96**, 559 (1993).

[43] D. Van der Marel, Physica C **165**, 35 (1990).

[44] S. Robaszkiewicz, R. Micnas, and K. A. Chao, Phys. Rev. B **26**, 3915 (1982).

[45] A. A. Abrikosov, L. P. Gor'kov, and I. É. Dzyaloshinski, *Quantum Field Theoretical Methods in Statistical Physics*, Pergamon Press, New York (1965).

[46] V. A. Moskalenko and M. E. Palistrant, Zh. Éksp. Teor. Fiz. **49**, 770 (1965) [Sov. Phys. JETP **22**, 536 (1966)].

[47] F. G. Kochorbe and M. E. Palistrant, Czech. J. Phys. **46**, 983 (1996); M. E. Palistrant and F. G. Kochorbe, Czech. J. Phys. **46**, 981 (1996).

[48] F. R. Gantmakher, *The Theory of Matrices*, Chelsea, New York (1959).

[49] A. J. Leggett, in *Modern Trends in the Theory of Condensed Matter* (Proc. XVI Karpacz Winter School of Theoretical Physics, February 19–March 3, 1979, Karpacz, Poland), A. Pekalski and J. Przystawa (eds.), Springer, Berlin (1980), p. 13.

[50] Z. Tesanović and B. I. Halperin, Phys. Rev. B **36**, 4888 (1987).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Radiation breakdown in silicon wafers

V. A. Voĭtenko and S. E. Mal'khanov

St. Petersburg State Technical University, 195251 St. Petersburg, Russia

(Submitted 12 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 1067–1078 (September 1998)

Radiation breakdown in silicon slabs is observed and studied as revealed in anomalous behavior of the dose characteristics of their radiation defects when the radiative intensity is varied. A theory is constructed for reversible radiation breakdown due to the bistability which develops in a gas of radiation vacancies when the gas can be regarded as quasi-two-dimensional. In order to explain the exponential saturation of the dose characteristics as the irradiation intensity is increased, scenarios are proposed in which different forms of the constituent radiation defects develop. Some parameters of the bistable gas of primary vacancies are estimated, including diffusion coefficients, dimensions of inhomogeneity regions, and the rate of movement of the stratification line. On the whole, satisfactory agreement with experiment is obtained. Discrepancies between the diffusion coefficient for neutral vacancies obtained here and in the literature are attributed to the role of interband recombination accompanying radiation defect formation during electron bombardment.
[S1063-7761(98)02109-X]

1. INTRODUCTION

The problem of radiation defect formation, especially in silicon, is currently of high interest, both experimentally and theoretically.[1–5] The well-known simplicity and high symmetry of the silicon crystal structure, the exhaustive studies that have been made of the possible radiation defects in it, and the development of industrial methods for wafer fabrication have made silicon the most appropriate material for studying various models of defect formation. Silicon irradiated by fast electrons is also under study in radiation technology.

Because of the particular way radiation defects are produced during irradiation, they customarily include V vacancies or W divacancies. The latter are also formed independently. Thus, in describing the kinetics of primary radiation defects (interstitial atoms and vacancies), it is of prime importance to include vacancies.

In this paper we obtain the dependence of the concentration N for various types of radiation defects on the irradiation dose Φ for various irradiation intensities I, and an exponential saturation of the dose characteristics is observed. A theory is constructed for the radiation breakdown of semiconductor surfaces and wafers that occurs when the induced crystal lattice vacancies merge. It is shown that the dose dependences found as a result of certain approximations are the same as the theoretical results obtained in a radiation breakdown picture. This breakdown mechanism is based on the bistable kinetics of the nonequilibrium fluctuations which develop in the primary vacancies during irradiation. The subsequent evolution of these fluctuations leads to bifurcations, familiar from hydrodynamics, which can develop, for example, according to the scenario in which Landau turbulence develops.[6] As a result of the formation of secondary radiation defects, the bifurcation pattern is fixed, and this leads to a breakup of the spatial period of the modulation in the resulting structures with the passage of time. Scenarios involving formation and immobilization of secondary radiation defect structures are compared and substantiated.

The accord between theory and experiment on the shape of the dose characteristics $N(\Phi)$ and the good agreement of our estimates of the kinetic parameters of the system of primary vacancies with earlier work[7,8] suggest that radiation breakdown of the wafers, induced by a two-dimensional ordering of certain secondary radiation defects, has been observed. The saturation of the exposure characteristics is also accompanied by the formation of a sulfide film during S-passivation of Ge (Ref. 9) and GaAs during processing in various sulfide solutions,[10] and of a ferrous film on passivated surfaces of these semiconductors.[11] Various geometric patterns are then formed on the surfaces; they can be observed visually on metallic electrodes during electropolishing.[12] Thus, the variety of reconstructed quasi-two-dimensional systems lends credence to the fundamental nature of the saturation in the dose characteristics observed here.

2. KINETICS OF A BISTABLE SYSTEM OF VACANCIES

The first attempts[3,13] to interpret the dependence of the radiation defect concentration on electron irradiation intensity and dose were based on the idea of uniform-rate quasichemical lattice reactions with an exhaustible supply of reagents. Neither the observed form of these dependences nor the correlation between them were explained: the existence of a dependence $N(I)$, of saturation in $N(\Phi)$ for certain radiation defects and their absence for other centers, and the effect of radiation defects of shallow impurity-level type on these and several other[14] characteristics.

In this paper we show that these processes must be treated as Belousov–Zhabotinskiĭ reactions; see, for example, Sec. 10.3 of Ref. 15. They involve an extended catalyst, with the crystal lattice itself acting in this capacity. In this case, lattice inhomogeneities relax self-consistently owing to diffusion and convective field transport.[15] We shall treat single V vacancies as transporters of the diffusion fluxes, filling the same role as holes in the electron theory of semiconductors.

An independent and closed diffusion–kinetic equation, which combines the continuity condition and the diffusion equation, can be formulated for the vacancy concentration $n(t)$:

$$\frac{\partial n}{\partial t}+\operatorname{div}(nb\mathbf{F}-D\nabla n)=G(n)-\frac{n}{\tau_{\mathrm{V}}},\qquad (1)$$

where $G(n)$ is the rate of vacancy generation, τ_V is their lifetime, the field term $j_1=nbF$ in the vacancy flux is determined by the elastic force F and mobility b,[16] and the diffusion term $j_2=-D\nabla n$ is determined by D, the diffusion coefficient. The direction of the vector $\mathbf{F}$ is determined by the preferred direction that emerges within the crystal as a result of irradiation. For example, $\mathbf{F}$ can be directed along the line of intersection of the plane of incidence of the beam with the plane of the wafer surface. Strictly speaking, Eq. (1) alone is insufficient; a contribution to G can arise during the breakup of component radiation defects, such as W divacancies. In addition, the force F can also depend on the concentration of all radiation defects.

However, for the high-intensity and high-energy electron and proton beams that we used (see Sec. 3), the main contribution to $G(n)$ is the external interaction, so that the contribution from the decay of constituent radiation defects can be neglected.

We also assume that F depends only on the concentration n of primary vacancies. The uniform quasichemical lattice reactions studied previously[3,13] obey the balance equation

$$G(n)=n/\tau_{\mathrm{V}}.\qquad (2)$$

By taking the planar geometry of the test wafers into account, the limited depth to which the beam penetrates in some of the radiation experiments, and the known tendency of vacancies to drift toward the surface,[17] it is possible to consider a quasi-two-dimensional model for solving Eq. (1). In this case it turns out to be important that at high enough intensities and radiation doses there is a change in the channeling of the incident particles accompanied by a substantial number of large-angle scattering events on preexisting defects.[18] This is reflected by the fact that the generation rate $G(n)$ becomes a nonlinear function of the concentration, and as will be shown below, this leads to stratification of the vacancy gas. In fact, it is clear from Fig. 1 that Eq. (2) can have three solutions when $G(n)$ is nonlinear. Of the solutions shown in Fig. 1, which correspond to the initial n_l, metastable n_c, and dense n_h phases, only two are stable: n_l and n_h. This means that the system of vacancies can be made bistable.

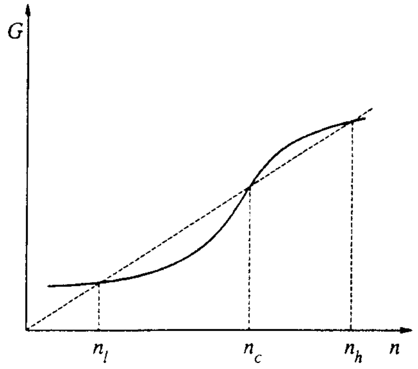


FIG. 1. A schematic plot of the source $G(n)$ (smooth curve) and sink n/τ_V (dashed line) terms of the diffusion kinetic equation (1) as functions of the concentration n. The solutions of the stationary homogeneous equation (2) are indicated on the horizontal axis.

The kinetics of the stratification that takes place for $n_l<n<n_h$ is described by the time-dependent diffusion–kinetic equation (1). In the linear approximation we write the functions $G(n)$ and $F(n)$ as expansions in the small fluctuations δn of the concentration:

$$F(n)=C\,\delta n,\quad G(n)=\left(\frac{1}{\tau_{\mathrm{V}}}+\nu\right)\delta n,\qquad (3)$$

where C is a phenomenological coefficient determined by the projection, parallel to the wafer, of the external force produced by the bombarding beam; the frequency ν, given by

$$\nu=\frac{\partial G}{\partial n}-\frac{1}{\tau_{\mathrm{V}}}.\qquad (4)$$

plays the role of a negative differential conductivity in the linearized equation (1); see Sec. 7.6 of Ref. 15.) Taking the x axis to be in the plane of incidence perpendicular to the wafer surface and taking the Fourier transform, with respect to position y and time, of Eq. (1) as linearized using Eqs. (3) and (4), we obtain

$$[-i(\omega-kV_D)+k^2D-\nu]\delta n=D\,\frac{\partial^2\delta n}{\partial x^2}.\qquad (5)$$

Here k and ω are the two-dimensional wave vector and frequency of the fluctuations, and the product $V_D=nCb$ serves as a drift velocity. According to Eq. (5), for a constant gradient of the concentration of the primary radiation defects inside the wafer, i.e., $dn/dx=\mathrm{const}$, a longitudinal wave

$$\delta n=A(t)\exp\{i(ky-\omega t)\}f(x)\qquad (6)$$

with frequency $\omega=kV_D$ and damping $\gamma=k^2D-\nu$ develops there. For the range of n characterized by a sufficiently steep slope of the $G(n)$ curve (Fig. 1), we have $\nu>k^2D$, so that Eq. (6) describes an exponential rise in the amplitude $A(t)$.

An averaged equation for the square of the absolute value of the amplitude $A(t)$, which describes its saturation, can be obtained by the Landau method[6] if we restrict ourselves to a nondegenerate system. For these systems, the actual stability boundary is determined by the unique form of the perturbations and the frequency $\omega(k)$ that yield zero damping, $\gamma=0$. In this regard, it is possible to expand the time derivative of $|A|^2$ in A, where for the reasons given in Ref. 6 it is sufficient to retain only a few of the nearest terms that fail to vanish upon averaging. A natural average can be taken over the time (and energy) required to produce secondary radiation defects. Terms of odd order in the amplitude A necessarily contain a periodic factor and drop out upon averaging. Even-order terms like the fourth correspond to terms like $A^2A^{*2}=|A|^4$ that do not drop out on averaging.

The time derivatives like $\langle d|A|^2/dt\rangle$ are direct mathematical analogs of the ''transition probabilities per unit time'' that can be calculated using perturbation theory based on the time-dependent Schrödinger equation of quantum mechanics.[19] In this case, we are interested in the square of the absolute value $|A|^2$ because the statistical weight of the vacancy pairs formed per unit volume, which yield W divacancies as they merge, equals the number of combinations $C_n^2=n(n-1)/2$, which is quadratic in n. Thus, the rate of divacancy production

$$\frac{dN_{\mathrm{W}}}{dt}=\frac{1}{2}\left\langle\frac{d|A|^2}{dt}\right\rangle \tag{7}$$

is given by an average derivative of this type. In principle, our expansion can begin with a constant term of zeroth order in A corresponding to a source of divacancies, owing to the decay of other radiation defects. Divacancies, however, have a lower annealing temperature than the K-centers obtained from them (C+O+W; see Ref. 1). This means that this constant source of divacancies is insignificant, so it is sufficient to limit ourselves to terms of second and fourth order in A. Expressing the Landau constant[6] α in terms of the limiting divacancy concentration $N_{0\mathrm{W}}$, we obtain

$$\frac{dN_{\mathrm{W}}}{dt}=\frac{N_{\mathrm{W}}}{\tau_I}\left(1-\frac{N_{\mathrm{W}}}{N_{0\mathrm{W}}}\right), \tag{8}$$

where τ_I is the characteristic time for the evolution of a divacancy structure. Solving Eq. (8) yields a dose characteristic of the form

$$N_{\mathrm{W}}=\frac{N_{0\mathrm{W}}}{1+\exp[(\Phi_0-\Phi)/I\tau_I]}. \tag{9}$$

The only constant that arises as a result of the integration, denoted by $\exp(\Phi_0/I\tau_I)$ in Eq. (9), along with the other parameters τ_I and N_{0W}, can be expressed in terms of the experimental parameters, given the scaling considerations of Sec. 4.

The stability loss regime characterized by plane symmetry corresponds to Poiseuille flow.[6] The other regime, isotropic turbulence, occurs in hydrodynamics during Couette flow (in the space between two cylinders[6]). It is well known that in order to describe it, linear representations must be avoided (Ref. 6, Sec. 30), and that the solution leading to zero damping corresponds to purely imaginary $\omega(k)$ (Ref. 6, Sec. 27), so that $\gamma=0$ when $\omega=0$. This stability loss regime arises in Eq. (1) for a beam normally incident upon a high-symmetry surface of a silicon wafer, whereupon $F(n)$ of Eq. (3) vanishes and $C=0$.

Neglecting elastic forces, we consider stationary axisymmetric solutions of Eq. (1) that are uniform in x, with the concentration $n_0(r)$ taken per unit, area and the nonlinear function $G(n)$ shown in Fig. 1 (r is the radial coordinate in the plane of the wafer). Transforming to cylindrical coordinates in Eq. (1) and integrating with respect to r with weight dn_0/dr instead of taking the Fourier transform, we obtain

$$D\int_0^\infty\left(\frac{dn_0}{dr}\right)^2\frac{dr}{r}=\int_{n_l}^{n_h}\left[G(n)-\frac{n}{\tau_{\mathrm{V}}}\right]dn. \tag{10}$$

An estimate of the thickness Δr of the transition layer between the competing n_l and n_h phases can be obtained from Eq. (10):

$$\sqrt{\frac{D_h}{\nu_h}}\lesssim\Delta r\lesssim\sqrt{\frac{D_l}{\nu_l}}, \tag{11}$$

where the frequencies and diffusion coefficients ν_h, D_h and ν_l, D_l refer to the dense and rarefied phases, respectively. The principal contribution to the integral in Eq. (10) comes from the transition layer: $R_c-\Delta r\leqslant r\leqslant R_c+\Delta r$, where R_c is the critical nucleus radius. The scale length R_c is therefore roughly

$$R_c=\frac{D}{J}\int_0^\infty\left(\frac{dn_0}{dr}\right)^2dr, \tag{12}$$

where the degree of supersaturation in the vacancy gas, J, is

$$J=\int_{n_l}^{n_h}\left[G(n)-\frac{n}{\tau_{\mathrm{V}}}\right]dn. \tag{13}$$

According to Eqs. (12) and (13), in the two-dimensional case R_c can vary continuously over wide limits, depending on the degree of supersaturation J. As a lower bound for the critical nucleus size, we can take a value of Δr that is independent of I and that obeys Eq. (11). According to the experimental data given below, this size is measured in fractions of a micron.

Applying the procedure discussed above for integrating Eq. (1) to nonstationary, two-dimensional, axisymmetric solutions, we find the nucleus rate of growth $v(R)$ of the dense phase. We seek a wave solution of Eq. (1):

$$n=n_0(r-v(R)t). \tag{14}$$

Proceeding as in the derivation of Eq. (12), we obtain from Eq. (1)

$$v(R)=v_\infty(1-R_c/R). \tag{15}$$

Here the quantity

$$v_\infty=J\left[\int_0^\infty\left(\frac{dn_0}{dr}\right)^2dr\right]^{-1} \tag{16}$$

represents the rate of displacement of an interphase wall of small curvature for a given supersaturation J. During a phase transition, attainment of critical size by a nucleus is usually the longest stage.[20] Thus, the characteristic time τ_I in Eq. (9), which will be found below by comparing Eq. (9) with experiment, is approximately

$$\tau_I\sim R_c/v_\infty. \tag{17}$$

Depending on the angle of incidence of the irradiating beam, the secondary radiation defect structures obtained over time τ_I inherit either a plane texture of filaments with $C \neq 0$, or an axisymmetric texture of points of radius R_c from Eq. (12) in the opposite case. The various techniques for processing crystal surfaces[10–12] also provide examples of the practical realization of both possibilities.

3. EXPERIMENTAL RESULTS

We have studied radiation defect formation in $5\times2\times0.2$ mm^3 silicon wafers fabricated from KÉF-7.5 and KÉF-1.0 silicon (with electron concentrations of 5×10^{14} and 4×10^{15} cm^{-3}, respectively) and from KDB-10 silicon (with a hole concentration of 1.3×10^{15} cm^{-3}). Ohmic contacts were made on opposite sides of the wafer surfaces by depositing low-resistance n^+- or p^+-layers by ion implantation (of phosphorus or boron, respectively) to a depth of 0.2–0.3 μm. To recover the amorphous layer the samples were annealed in a neutral medium.

The intensity of the RTÉ-1V accelerator with which the silicon wafers were bombarded with electrons can be varied over wide limits. We used a range $I=10^{13}-10^{14}$ cm^{-2} s^{-1}, dictated by the method for measuring the concentration N available to us. The NG-200U neutron generator was equipped with a special attachment for producing and accelerating a proton beam. The protons penetrated the wafers to a depth $L\sim1$ μm, while the electrons passed right through it. The beams were normally incident upon the $\langle111\rangle$ surface.

The concentration N was measured by a photocapacitance method. The optimum for this method is cancellation of half the original dopant. The lower limit of the concentration measurements (the variation in I and Φ) was directly related to the instrumental sensitivity, while the upper limit was dictated by the large dark currents resulting from cancellation, which increases the resistance of the inner region of the wafer to the that of the reverse-biased $p-n$ junction in the surface layer, which produces the capacitance. Additional details of the experimental procedures and results are available elsewhere.[8,21,22]

Experimental dose plots $N(\Phi)$ obtained by us for two constant values of I differing by a factor of two are shown for several types of radiation defects as different points in Figs. 2 and 3. All the dose characteristics share an almost linear rise, which reflects a constant rate of defect formation, $dN/dt=I(dN/d\Phi)$, at low doses. The essential differences associated with the behavior of the dose characteristics at high doses and variable I make it possible to separate radiation defects into three groups.

K-centers, with $\mathscr{E}_V=+0.35$ eV (C+O+W; see Ref. 1), are characterized by a quadratic dependence[21] $N=\alpha I^2$. The corresponding dose curves *1* and *3* in Fig. 2 have a linear rise followed by a saturation region. The same sort of linear rise followed by saturation in the dose curve $N(\Phi)$ was recorded for radiation defects with $\mathscr{E}_V=+0.37$ eV (curves *2* and *4*), as well as for radiation defects obtained by proton irradiation (curve *5*). The microscopic structure of the defects corresponding to curves 2, 4, and 5 have yet to be fully explained. This structure is discussed in Sec. 4.

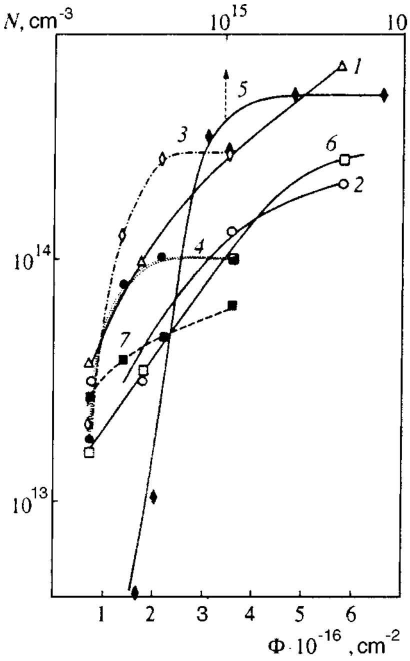


FIG. 2. The dose curves $N(\Phi)$ reflecting the rate of accumulation of various radiation defects in KDB-10 silicon irradiated by 1-MeV electrons and 100-keV protons. The points are experimental data and the smooth curves, theory. Curves *1–5* have been constructed from Eq. (20) and curves *6* and *7*, from Eq. (9): *1, 3* (△, ◇) — K-centers; *2, 4* (○, ●) — radiation defects with $\mathscr{E}_V=+0.37$ eV; *5* (♦) — radiation defects obtained by proton irradiation (upper horizontal scale); *6, 7* (□, ■) — divacancies; *1, 2, 6* (△, ○, □) — intensity $I=6\times10^{13}$ cm^{-2} s^{-1}; *3, 4, 7* (◇, ●, ■) — intensity $I=1.2\times10^{14}$ cm^{-2} s^{-1}.

For the divacancies, either a linear $N=\beta I$ (KÉF-7.5, with $\mathscr{E}_c=-0.4$ eV) or a quadratic $N=\alpha I^2$ (KÉF-1.0, with $\mathscr{E}_c=-0.4$ eV; KDB-10, with $\mathscr{E}_V=+0.29$ eV) variation $N(I)$ is observed. The dose dependence $N(\Phi)$, as for the K-centers at the lower intensity $I=6\times10^{13}$ cm^{-2} s^{-1}, re-

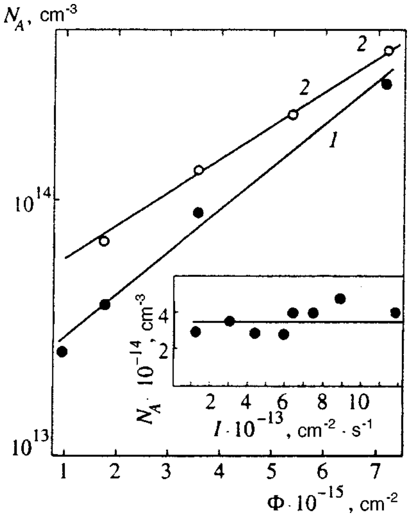


FIG. 3. Dose curves $N_A(\Phi)$ reflecting the buildup of A-centers in KÉF-7.5 silicon bombarded by 1-MeV electrons at intensities (*1*) (●) — of 1.2×10^{14} cm^{-2} s^{-1} and (*2*) (○) — 6.0×10^{13} cm^{-2} s^{-1}. The inset shows the concentration N_A of A-centers as a function of I for a constant dose $\Phi=7.2\times10^{15}$ cm^{-2}.

tains a nearly linear rise over the entire range (curve *6*). However, there is a tendency toward saturation at the higher intensity $I=1.2\times10^{14}$ cm^{-2} s^{-1} (curve *7*).

For A-centers (V+O), the defect concentration N is independent of intensity, i.e., $N(I)=\text{const}$. Here the dose curve $N(\Phi)$ retains an almost linear rise over the entire range of Φ and for all I attainable in our experiments (see Fig. 3, curves *1* and *2*, and the inset). Thus, there is a palpable relationship between the dose characteristics of the radiation defects and the intensity dependence of N: when $N(I)$ is quadratic or linear, the dose curve goes to saturation.

Our measurements also suggest a more or less sharp dependence of the defect formation probability on the type of shallow impurity levels. Thus, K-centers are observed only in p-silicon, while A-centers are observed only in n-silicon. For divacancies in n-type crystals, both linear and quadratic $N(I)$ curves are possible, while in p-type crystals only a quadratic dependence is observed. More subtle differences exist as well.[14]

4. DISCUSSION OF EXPERIMENTAL RESULTS

All these features of the experimental data can be explained in terms of the idea, described in Sec. 2, of an instability of the solutions of Eq. (1) that leads to bistability of the system of primary radiation defects. Bifurcation causes fractionation over time (i.e., as the dose $\Phi=It$ increases) in the spatial modulation period of the primary radiation defects in the medium. The different stages of this process ''freeze out'' as secondary radiation defects of the different generations develop.

The conditions for applicability of the theoretical dose characteristic (9) extend only to the component radiation defects in the first generation, i.e., the divacancies. The constant of integration in Eq. (9) can be expressed in terms of the characteristic delay time $t_0=\Phi_0/I$ for onset of divacancy formation. This time is required to build up enough kinetic energy for the appropriate instability to develop. This energy goes into the phonon excitation that precedes the onset of the breakaway of atoms from lattice sites. In any event, the delay t_0 is determined by details of the preparation of the quasi-two-dimensional unstable system, in this case by the dispersion of the phonon interaction. Thus, t_0 can be regarded as independent of the intensity of the external interaction.

The maximum number $N_{0\mathrm{W}}$ of divacancies is determined by the crystalline structure of silicon and is also independent of I. Ultimately, the only one of the three constants in Eq. (9) that depends on I is the evolution time τ_I. All the dose characteristics shown in Fig. 2, as well as $N(I)$, can be interpreted in terms of Eqs. (9) and (20), using the standard relation for the reciprocal time $\tau_I^{-1}=\xi+\eta I+\zeta I^2$ and choosing the constants ξ, η, and ζ. For example, for curve *7* we obtain $\tau_I=1.7$ min and for curve *6*, which was taken at half the intensity, τ_I is accordingly longer.

The elementary process of forming K-centers ultimately reduces to a reaction of the form[13]

$$[\mathrm{V_2O}]^-+\mathrm{C}_i^+=\mathrm{K}^0, \tag{18}$$

where $[\mathrm{V_2O}]$ is a divacancy–oxygen complex, C_i represents interstitial carbon, and the signs indicate charge states. The crystal lattice participates in reaction (18) as a catalyst because of the Debye correlation that builds up during the divacancy formation stage.[23] If, given this correlation, the entire surface of the wafer is covered by ordered Debye spheres, then Eq. (18) leads to random filling of some of these spheres with K-centers. The resulting fluctuation aggregations of K-centers then become mesoscopic nucleation centers. They accelerate the stratification of the vacancy gas into n_l and n_h phases in accordance with Sec. 2.

The probability of such aggregations can be estimated using an Ising model.[24] It is considerably greater than the probability of spontaneous formation of a critical nucleus.[20] This process is sustained by localization of the aggregations along dislocation lines, in accordance with electron microscopy data.[25] The emerging nuclei of the dense phase vacancies, in turn, begin to sustain the formation of new K-centers within themselves, so that their formation in other places becomes unlikely. This leads to the chaining of K-centers in the direction of the radiation flux. Ultimately, stable linear structures of K-centers, extending from one surface of the wafer to the other, in effect cause it to undergo radiation breakdown, which becomes irreversible and can be regarded as a structural phase transition. Because the appearance of new K-centers outside the structure becomes essentially impossible, the expansion[6] for their rate of formation $I(dN_\mathrm{K}/d\Phi)$ must be in terms of the number of remaining free sites in the already formed structure, $N_{0\mathrm{K}}-N_\mathrm{K}$, i.e.,

$$I\,\frac{dN_\mathrm{K}}{d\Phi}=\frac{N_{0\mathrm{K}}-N_\mathrm{K}}{\tau_{I\mathrm{K}}}, \tag{19}$$

where the time $\tau_{I\mathrm{K}}$ corresponds to a K-center. The dose characteristic

$$N_\mathrm{K}=N_{0\mathrm{K}}\left[1-\exp\left(-\frac{\Phi}{I\tau_{I\mathrm{K}}}\right)\right], \tag{20}$$

which is confirmed experimentally by curves *1* and *3* of Fig. 2, is a solution of Eq. (19).

Note that Eq. (19) results from the expansion of the radiation defect formation rate, including a zeroth-order term for a constant source owing to synthesis from other centers. The dose curves *2*, *4*, and *5* constructed using Eq. (20) suggest that the corresponding radiation defects (with $\mathscr{E}_\mathrm{V}=+0.37$ eV) and the defects induced by proton bombardment also belong to an older generation, after the divacancies. Measurements of N at different intensities[21] show that they can be obtained from divacancies. The $\mathscr{E}_\mathrm{V}=0.37$ eV level may turn out to be another component of a multiplet of K-centers, while hydrogen dimers probably predominate among the defects produced by proton irradiation.[26]

A lower-bound on the distance along the wafer between chains of radiation defects can be estimated from their average two-dimensional concentration $n=NL$, where $N\sim10^{14}$ cm^{-3} (see Fig. 2). As an example, for proton bombardment $L\sim1$ μm. This implies a lower bound for the inhomogeneity of $R_c\sim1/\sqrt{n}$, which yields a lower limit on the size of the two-dimensional nuclei of fractions of a micron, with $R_c \sim 0.5$ μm. The rate of growth of the radiation defect-enriched two-dimensional region is then of the order of $n_\infty = R_c/\tau_I \sim 0.1$ μm/min. Note that this rate is of the same order for both proton and electron bombardment.

In order for two-dimensional structures of this type to develop, the vacancies must be able to traverse the major transverse dimension of the problem, the distance L to the wafer surface, within the time (17). The time τ_I obtained by integrating curve *5* of Fig. 2, which coincides with Eq. (17), is of the order of one hour. Given this and the required escape of the vacancies to the surface, for the diffusion coefficient we obtain the estimate $D = L^2/\tau_I$, whereupon $D \sim 10^{-12}$ cm^2/s. This value is the same, to within the measurement errors, as the $D = 1.2 \times 10^{-12}$ cm^2/s previously obtained[8] for positively charged divacancies. This agreement in the values of D found by different methods tends to confirm the existence of a structural phase transition in an atmosphere of radiation defects.

On the other hand, the 1 MeV electrons we have used go right through the wafer. In this case, the L for estimating D must be the full width of the wafer, i.e., $L = 0.2$ mm. This value is several orders of magnitude greater than for protons, while the time τ_I, on the other hand, is an order of magnitude shorter. This leads to huge values $D \sim 10^{-7}$ cm^2/s for the diffusion coefficient of the neutral vacancies obtained in this case. This strong predominance of diffusion processes in electron-irradiated silicon compared to proton-irradiated silicon appears to be due to a larger contribution from interband recombination, which induces momentum transfer to the lattice.

Silicon wafers grown by the Czochralski method were used in our experiments. These crystals contain oxygen atoms at overwhelming concentrations $N_0 \sim (6-7) \times 10^{17}$ cm^{-3} compared to other impurities. Thus, the oxygen impurity is present in sufficient amounts and cannot limit the probability of A-center formation processes, whose rate is therefore determined by the deficit of vacancies. Any scenario for the formation of A-centers must, above all, therefore address the question of why their dose characteristics $N(\Phi)$ (see Fig. 3 and the inset) differ radically from the analogous dependences for divacancies and K-centers (Fig. 2), whose formation probabilities are also limited by a deficit of vacancies.

These differences are related to the fact that oxygen is not an isoelectronic impurity in silicon, so it interacts only with charged (but not neutral) vacancies. The number of such vacancies is determined by the donor concentration, which does not depend on the external irradiation. Ultimately, the concentration of A-centers is independent of the total vacancy concentration n, i.e., it is also independent of I, which is determined solely by the rate of a reaction of the form V+O=A. This scenario is confirmed by the fact that the concentration of A-centers is always rigidly tied to (on the order of, but slightly smaller than) the donor concentration (see Fig. 3). Accordingly, the linear rise in their dose characteristic does not saturate, owing to the maintenance of a large diffusion coefficient D throughout irradiation.

In conclusion, we note that the observed coupling of the saturation in the dose characteristics of radiation defects with the formation of self-organized structures of the latter, as expressed by Eqs. (9) and (20), is a promising result. Thus, in the future, radiation defect formation can likely become a test model for studying the principles of self-organization in discrete two-dimensional systems. This model can possibly the creation and use of mathematical lattice techniques for solving the corresponding class of nonlinear kinetic equations.

Efforts in this area offer promise, as they are capable of supplementing, at the level of microscopic scale lengths, existing work on self-organization of lateral structures in the physics of superlattices.[27]

On the other hand, the formation of self-organized defect structures during proton, electron, and possibly neutron irradiation is an indication of the colossal radiation vulnerability of planar semiconductor devices. In p-Si, with the mediation (capture) of mobile holes, divacancy formation predominates, forming a correlated extended ‘‘lattice getter.’’ This structure entrains the predominant oxygen and then, according to Eq. (18), carbon. *A*-centers lose out in terms of their rate of formation and are present in negligible amounts. In n-Si, A-centers are formed and compete with the divacancies. The latter are chaotic; they do not have a Debye-correlated structure. Thus, an extended getter does not exist, and there are no K-centers, as they lose out to the A-centers.

[1] Y. H. Lee, J. W. Corbett, and K. L. Brower, Phys. Status Solidi A **41**, 637 (1977).

[2] V. V. Emtsev, T. V. Mashovets, and V. V. Mikhnovich, Fiz. Tekh. Poluprovodn. **27**, 708 (1993) [Semiconductors **27**, 390 (1993)].

[3] A. I. Baranov, N. I. Boyarkina, and A. V. Vasil'ev, Fiz. Tekh. Poluprovodn. **29**, 1570 (1995) [Semiconductors **29**, 817 (1995)].

[4] E. M. Verbitskaya, V. K. Eremin, A. M. Ivanov *et al.*, Fiz. Tekh. Poluprovodn. **31**, 235 (1997) [Semiconductors **31**, 127 (1997)].

[5] S. K. Estreicher, J. Weber, A. Derecskei-Kovacs *et al.*, Phys. Rev. B **55**, 5037 (1997).

[6] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* 2nd ed., Pergamon, New York (1987).

[7] G. D. Watkins, in *Deep Centers in Semiconductors*, S. T. Pantelides (Ed.), Gordon and Breach, New York (1986), p. 147.

[8] S. E. Mal'khanov, Fiz. Tekh. Poluprovodn. **29**, 725 (1995) [Semiconductors **29**, 377 (1995)].

[9] G. W. Anderson, M. C. Hanf, P. R. Norton *et al.*, Appl. Phys. Lett. **66**, 1123 (1995).

[10] V. N. Bessolov, E. V. Konenkova, M. V. Lebedev *et al.*, in *Third All-Russian Conference on the Physics of Semiconductors. Abstracts* [in Russian], FIAN, Moscow (1997), p. 304.

[11] G. W. Anderson, P. Ma, and P. R. Norton, J. Appl. Phys. **79**, 5641 (1996).

[12] V. V. Yuzhakov, Hsueh-Chia Chang, and A. E. Miller, Phys. Rev. B **56**, 12608 (1997).

[13] N. V. Kolesnikov, V. N. Lomasov, and S. E. Mal'khanov, Fiz. Tekh. Poluprovodn. **23**, 1921 (1989) [Sov. Phys. Semicond. **23**, 1191 (1989)]; Fiz. Tekh. Poluprovodn. **24**, 372 (1990) [Sov. Phys. Semicond. **24**, 230 (1990)].

[14] N. T. Bagraev, D. E. Onopko, A. I. Ryskin *et al.*, Fiz. Tekh. Poluprovodn. **30**, 1855 (1996) [Semiconductors **30**, 970 (1996)].

[15] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes*, Springer-Verlag, New York (1987).

[16] I. P. Ipatova, V. G. Malyshkin, A. N. Starodubtsev *et al.*, in *Third All-Russian Conference on the Physics of Semiconductors. Abstracts* [in Russian], FIAN, Moscow (1997), p. 272.

[17] B. S. Bokshteĭn, *Atoms Wander through a Crystal* [in Russian], Nauka, Moscow (1984).

[18] N. P. Kalashnikov, *Coherent Interactions of Charged Particles in Single Crystals* [in Russian], Atomizdat, Moscow (1981), Ch. 2.

[19] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, 3rd ed., Pergamon, New York (1977), Ch. 6.

[20] L. D. Landau and E. M. Lifshitz, *Statistical Physics. Part I* [in Russian], Nauka, Moscow (1976).
[21] N. V. Kolesnikov, V. N. Lomasov, S. E. Mal'khanov *et al.*, Fiz. Tekh. Poluprovodn. **18**, 1496 (1984) [Sov. Phys. Semicond. **18**, 936 (1984)].
[22] S. E. Mal'khanov, Fiz. Tekh. Poluprovodn. **28**, 1431 (1994) [Semiconductors **28**, 804 (1994)].
[23] V. A. Voĭtenko and S. E. Mal'khanov, Zh. Éksp. Teor. Fiz. **112**, 707 (1997) [JETP **85**, 386 (1997)].
[24] J. Ziman, *Principles of the Theory of Solids*, Cambridge Univ. Press, Cambridge (1974).
[25] A. K. Gutakovskiĭ, L. I. Fedina, A. L. Aseev *et al.*, in *Third All-Russian Conference on the Physics of Semiconductors. Abstracts* [in Russian], FIAN, Moscow (1997), p. 267.
[26] A. N. Safonov, E. C. Lightowlers, and G. Davies, Phys. Rev. B **56**, 15517 (1997).
[27] N. N. Ledentsov, in *Proceedings of 23rd International Conference on the Physics of Semiconductors*, M. Scheffler and R. Zimmermann (eds.), Vol. **1**, World Scientific, Singapore (1996), p. 19.

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Charge states of atoms in the lattices of the high-temperature superconductors $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ and $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$

V. F. Masterov, F. S. Nasredinov, N. P. Seregin,*) and P. P. Seregin

St. Petersburg State Technical University 195251 St. Petersburg, Russia

(Submitted 12 November 1997; resubmitted 4 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 1079–1088 (September 1998)

To determine the charges of atoms in the lattices of the compounds $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ and $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$ ($n=1,2,3$), the parameters of the electric field gradient tensor at the copper sites of the indicated lattices were found by emission Mössbauer spectroscopy on the isotopes $^{61}Cu(^{61}Ni)$ and $^{67}Cu(^{67}Zn)$, and a calculation of these parameters was performed in the point-charge approximation. A comparison of the resulting values and the published data on ^{63}Cu nuclear quadrupole resonance showed that agreement between the experimental and computed values of the parameters obtains for models in which the holes resulting from a reduction in the valence of some of the thallium (bismuth) atoms are localized predominantly at oxygen sites located in the same plane as the copper atoms (for the compounds $Tl_2Ba_2Ca_2Cu_3O_{10}$ and $Bi_{1.6}Pb_{0.4}Sr_2Ca_2Cu_3O_{10}$—at oxygen sites in the same plane as the Cu(2) atoms). © *1998 American Institute of Physics.* [S1063-7761(98)02209-4]

1. INTRODUCTION

The compounds $Tl_2Ba_2Ca_{n-1}Cu_nO_{2n+4}$ (TlBaCaCuO) and $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$ (BiSrCaCuO) ($n=1,2,3$) have high superconducting transition temperatures T_c. This explains the interest in the investigation of their properties. A pressing problem for the TlBaCaCuO and BiSrCaCuO lattices is to determine the charge states of the oxygen atoms, which are responsible for superconductivity in these compounds.

For this reason, to us it seemed sensible to employ emission Mössbauer spectroscopy on $^{67}Cu(^{67}Zn)$ to determine the charge state of the atoms in TlBaCaCuO and BiSrCaCuO compounds. The essence of the method, which we proposed in Ref. 1, consists in using a $^{67}Zn^{2+}$ probe to measure the parameters of the electric field gradient tensor produced at the copper sites by the lattice ions, calculating these parameters in the point-charge approximation, and determining the effective charges of the atomic centers by comparing the experimental and computed values of the parameters of the tensor of the crystal gradient of the electric field. The effective charges give a good picture of the valence states of the ions at the lattice sites and substantial deviations from the standard valence states. We have demonstrated the effectiveness of this method for typical high-temperature superconductors (HTSCs)—the compounds $RBa_2Cu_3O_7$ (where R is a rare-earth metal).[2]

However, the number of available experimental parameters for TlBaCaCuO and BiSrCaCuO compounds is much smaller than the number of charges to be determined. For this reason, in the present work the correlation between the quadrupole interaction constants $C(Cu)$ and $C(Zn)$ of $^{63}Cu^{2+}$ centers (^{63}Cu nuclear quadrupole resonance (NQR) and nuclear magnetic resonance (NMR) data) and $^{67}Zn^{2+}$ centers ($^{67}Cu(^{67}Zn)$ emission Mössbauer spectroscopy data) in copper metal oxides (here $C=eQU_{zz}/h$, where Q is the quadrupole moment of the probe nucleus, U_{zz} is the principal component of the tensor of the total electric field gradient at the probe, and h is Planck's constant) was used to limit the number of possible combinations of these charges.[3]

Moreover, a high defect rate is typical of TlBaCaCuO and BiSrCaCuO compounds, even for materials with high values of T_c. In consequence, the experimental results obtained for nominally identical compounds by different groups of investigators are not reproducible. For example, obvious inconsistencies in the values of the ^{63}Cu NQR frequencies are found for the compounds $Tl_2Ba_2Ca_2Cu_3O_{10}$[4,5] and $Bi_{1.6}Pb_{0.4}Sr_2Ca_2Cu_3O_{10}$.[5–9] The correlation indicated above makes it possible not only to advance reasons for the discrepancy in the ^{63}Cu NQR data but also to assess the validity of the models proposed for the charge distribution. An investigation of TlBaCaCuO and BiSrCaCuO compounds via $^{61}Cu(^{61}Ni)$ emission Mössbauer spectroscopy confirms the conclusions drawn on the basis of the data obtained by this method on $^{67}Cu(^{67}Zn)$.

2. EXPERIMENTAL PROCEDURE AND RESULTS

The Mössbauer sources were prepared by diffusion doping of the compounds $Tl_2Ba_2Ca_2Cu_3O_{10}$ [Tl(2223)] ($T_c \approx 120$ K), $Bi_{1.6}Pb_{0.4}Sr_2Ca_2Cu_3O_{10}$ [Bi(2223)] ($T_c \approx 108$ K), $Tl_2Ba_2CaCu_2O_8$ [Tl(2212)] ($T_c \approx 60$ K), $Bi_2Sr_2CaCu_2O_8$ [Bi(2212)] ($T_c \approx 80$ K), $Tl_2Ba_2CuO_6$ [Tl(2201)] and $Bi_2Sr_2CuO_6$ [Bi(2201)] (for the latter two $T_c < 4.2$ K) with radioactive ^{61}Cu and ^{67}Cu isotopes using the procedure described in Ref. 10.

The Mössbauer spectra of $^{67}Cu(^{67}Zn)$ were measured at 4.2 K with ^{67}ZnS as the absorber (surface density 1000 mg/cm^2 of ^{67}Zn), while the $^{61}Cu(^{61}Ni)$ spectra were measured at 80 K with $Ni_{0.86}V_{0.14}$ as the absorber (surface density 1500 mg/cm^2 of nickel). The spectra of the ^{67}Cu and ^{61}Cu sources (in metallic copper) with the indicated absorb-

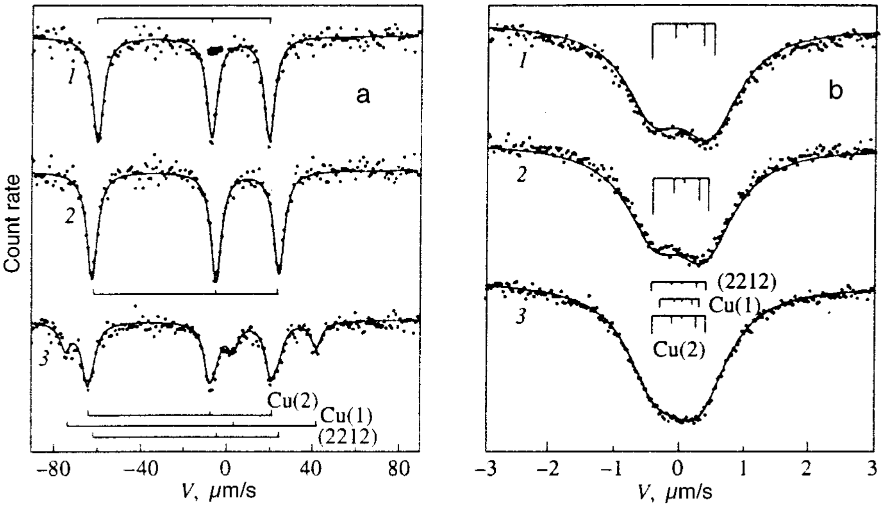


FIG. 1. ^{67}Cu (^{67}Zn) (a) and ^{61}Cu (^{61}Ni) (b) Mössbauer spectra of the compounds Tl(2201) (*1*), Tl(2212) (*2*), and Tl(2223) (*3*). The position of the components of the quadrupole triplets (a) and multiplets (b) corresponding to $^{67}Zn^{2+}$ (a) and $^{61}Ni^{2+}$ centers at Cu(1) and Cu(2) sites of the compound Tl(2223) and at the copper sites of the compounds Tl(2201) and Tl(2212) is shown.

ers had widths at half-height 2.7(4) μm/s and 0.95(5) mm/s, respectively. Typical spectra of all TlBaCaCuO compounds are presented in Fig. 1. The spectra of the compounds BiSrCaCuO have a similar structure. The results of the analysis of all spectra are summarized in Table I.

Since the copper atoms in the (2201) and (2212) lattices occupy a unique position,[11–13] it was expected that the ^{61}Cu (^{61}Ni) and ^{67}Cu (^{67}Zn) Mössbauer spectra of these compounds would correspond to a unique state of the Mössbauer probes $^{61}Ni^{2+}$ and $^{67}Zn^{2+}$. In actual fact, as one can see from Fig. 1 (curves *1* and *2*), the ^{67}Cu (^{67}Zn) and ^{61}Cu (^{61}Ni) spectra for both ceramics, Tl(2201) and Tl(2212), are quadrupole multiplets corresponding to a single state of the probes $^{67}Zn^{2+}$ and $^{61}Ni^{2+}$.

More complicated spectra consisting of three quadrupole triplets were obtained for the (2223) compounds. As an example, the ^{67}Cu (^{67}Zn) spectrum of the compound Tl(2223) is shown in Fig. 1a (curve 3). Since the copper atoms occupy two crystallographically inequivalent positions in the (2223) lattices,[14,15] one of the three quadrupole triplets in the spectrum should correspond to $^{67}Zn^{2+}$ centers at copper sites of an additional phase. The parameters of the spectra of this phase are close to those of the spectrum of the compound Tl(2212). Quadrupole triplets corresponding to $^{67}Zn^{2+}$ centers at Cu(1) and Cu(2) sites in the Tl(2223) lattice (their intensity ratio is close to 1:2), and a quadrupole triplet corresponding to $^{67}Zn^{2+}$ centers at copper sites of the Tl(2212) phase (its relative intensity is approximately 0.8), are shown in Fig. 1a (curve *3*).

TABLE I. Parameters of ^{61}Cu (^{61}Ni) and ^{67}Cu (^{67}Zn) emission Mössbauer spectra.

		TlBaCaCuO		BiSrCaCuO	
Compound	Site	*C*(Ni)	*C*(Zn)	*C*(Ni)	*C*(Zn)
(2201)	Cu	−48(3)	+13.3(3)	−51(3)	+12.4(3)
(2212)	Cu	−43(3)	+14.5(3)	−46(3)	+13.4(3)
(2223)	Cu(1)	−30(4)	+19.5(3)	−35(3)	+19.0(3)
	Cu(2)	−43(3)	+14.3(3)	−44(3)	+13.8(3)

Note. *C*(Ni) and *C*(Zn)—quadrupole interaction constants (in MHz) for $^{61}Ni^{2+}$ and $^{67}Zn^{2+}$ probes, respectively; for all probes the asymmetry parameter of the electric field gradient tensor $\eta < 0.2$; the widths of the components of the quadrupole multiplets were 5.8(2) μm/s for ^{67}Zn and 1.15(5) mm/s for ^{61}Ni.

Obviously, the ^{61}Cu (^{61}Ni) spectra of the (2223) compounds should also consist of a superposition of three multiplets. The ^{61}Cu (^{61}Ni) spectra of the (2223) compounds were analyzed taking this into account. Quadrupole multiplets corresponding to $^{61}Ni^{2+}$ centers at Cu(1) and Cu(2) sites in the Tl(2223) lattice and a quadrupole multiplet corresponding to $^{61}Ni^{2+}$ centers at the copper sites of the Tl(2212) phase are shown in Fig. 1b (curve *3*). The position of the line of the latter multiplet was based on the spectrum of the (2212) compound, while the ratio of the areas under the spectral curves was held at 1:2:0.8. As one can see from the table, the components of the quadrupole multiplets of ^{67}Zn and ^{67}Ni are broadened, compared with the natural width Γ_{nat} of the nuclear levels ($2\Gamma_{\mathrm{nat}} = 0.32$ μm/s for ^{67}Zn and $2\Gamma_{\mathrm{nat}} = 0.77$ mm/s for ^{61}Ni), and compared with the spectra of the ^{67}Cu and ^{61}Cu sources in metallic copper. The broadening of the ^{67}Cu (^{67}Zn) and ^{61}Cu (^{61}Ni) spectra, though not so great, has also been observed in other HTSCs.[16] The broadening in TlBaCaCuO and BiSrCaCuO compounds can probably be explained by the inhomogeneity of the experimental samples, which is also observed in ^{63}Cu NQR spectra[4–9] of these compounds. The broadening of the ^{67}Zn spectra is more pronounced due to the narrower spectral line (for ^{67}Zn substantial broadening is typical even for the absorption spectra[17]).

3. EXPERIMENTAL RESULTS

3.1. $^{67}Zn^{2+}$ probe

In the general case the measured quadrupole interaction constant *C* is a sum of two terms

$$C = eQ(1-\gamma)V_{zz}/h + eQ(1-R_0)W_{zz}/h, \tag{1}$$

where V_{zz} and W_{zz} are the principal components of the tensors of the crystal and valence electric field gradients, and γ and R_0 are the Sternheimer coefficients of the probe atom.

The contribution of the valence electrons to the total electric field gradient tensor can be neglected for the probe $^{67}\mathrm{Zn}^{2+}$. Therefore

$$C(\mathrm{Zn})\approx eQ(1-\gamma)V_{zz}/h. \tag{2}$$

Thus, the experimental data obtained on the parameters of the electric field gradient tensor using the $^{67}\mathrm{Zn}^{2+}$ probe can be correlated with calculations of these parameters based on an ionic model of the crystal lattice (point charge model).

The applicability of this model to copper metal oxides has frequently been discussed in the literature. Numerous attempts to compare the parameters of the electric field gradient tensor, calculated for oxygen and copper lattice sites on the basis of the point-charge model, with the values of the parameters determined experimentally by ^{17}O and ^{63}Cu NQR and NMR, must probably be deemed unsuccessful (see, for example, the review in Ref.16). In no case was satisfactory agreement obtained between the computed and measured values. This can be explained by the fact that the valence contribution to the electric field gradient at $^{17}\mathrm{O}^{2-}$ and $^{63}\mathrm{Cu}^{2+}$ probe nuclei is substantial. However, such agreement is achieved using probes at whose nuclei the electric field gradient is produced predominantly by the lattice ions, for example, $^{139}\mathrm{La}^{3+}$ and $^{137}\mathrm{Ba}^{2+}$ in $\mathrm{La}_{2-x}\mathrm{Ba}_x\mathrm{CuO}_4$ or $\mathrm{RBa}_2\mathrm{Cu}_3\mathrm{O}_{7-x}$,[16] as well as the $^{67}\mathrm{Zn}^{2+}$ probe, which we used earlier, at copper sites in a number of HTSCs.[1,2] On this basis, the point-charge model can be expected to be suitable for the TlBaCaCuO and BiSrCaCuO compounds that we investigated.

We calculated the tensors of the crystal electric field gradient at the copper sites of the TlBaCaCuO and BiSr–CaCuO lattices based on the point-charge model. In so doing, following the x-ray crystallographic and neutron-diffraction data,[11–15] the lattices were represented as a superposition of the following sublattices:

$\mathrm{Tl_2Ba_2CuO(1)_2O(2)_2O(3)_2}$,
$\mathrm{Tl_2Ba_2CaCu_2O(1)_4O(2)_2O(3)_2}$,
$\mathrm{Tl_2Ba_2Ca_2Cu(1)Cu(2)_2O(1)_2O(2)_4O(3)_2O(4)_2}$,
$\mathrm{Bi_2Sr_2CuO(1)_2O(2)_2O(3)_2}$,
$\mathrm{Bi_2Sr_2CaCu_2O(1)_2O(2)_2O(3)_2O(4)_2}$,
$\mathrm{(Bi_{1.6}Pb_{0.4})Sr_2Ca_2Cu(1)Cu(2)_2O(1)_2O(2)_2O(3)_2}$
$\mathrm{-O(4)_2O(5)_2}$.

The components of the electric field gradient tensors were calculated as sums of products of fixed lattice sums with the charges of the sublattices.[1]

For subsequent developments, it is significant that the O(1) sites for Tl(2201), Bi(2201), and Tl(2212) compounds are located in the copper–oxygen plane, for Bi(2212) the O(3) sites are located in the strontium–oxygen plane, for Tl(2223) the O(2) sites are located in the Cu(2)–O(2) plane, and for Bi(2223) the O(2) and O(3) sites are located in the Cu(2)–O(2,3) plane. Structural data from Refs. 11–15 were used in the calculations.

If the values $Q=0.17$ b (Ref. 17) and $\gamma=-12.2$ (Ref. 18) are taken for the $^{67}\mathrm{Zn}^{2+}$ centers, then model A, corresponding to the standard valence states of atoms in the Tl(2223) lattice (Tl^{3+}, Ba^{2+}, Ca^{2+}, Cu^{2+}, O^{2-}), yields $eQ(1-\gamma)V_{z1}=78$ MHz for Cu(1) sites and $eQ(1-\gamma)V_{z2}=68$ MHz for the Cu(2) sites.

These values differ considerably (by a factor of 4–4.5) from the experimental values of $C(\mathrm{Zn})$ at Cu(1) ($C(\mathrm{Zn1})$) and Cu(2) ($C(\mathrm{Zn2})$) sites; see Table I. A similar situation is also observed for other compounds. Such discrepancies can be due to erroneous values of γ and Q, an incorrect choice of model for the charge distribution, or a strong dependence of the crystal structure of the samples on their past history. The latter reason can probably be ruled out, since despite the many structural investigations of TlBaCaCuO and BiSrCaCuO compounds, all experimental data for specific compounds agree satisfactorily with one another.[19] The existing neglible discrepancies in the structural parameters cannot be explained by the observed differences in the values of $C(\mathrm{Zn})$.

To eliminate inaccurate values of γ and Q (or more accurately, their product) from the analysis, instead of studying the magnitude of the quadrupole interaction constants for Tl(2223) and Bi(2223) compounds, the experimental ratio $S=C(\mathrm{Zn1})/\mathrm{C(Zn2)}$ can be compared with the calculated value of $s=V_{zz1}/V_{zz2}$. Specifically, for Tl(2223) the values $S=1.36(4)$ and $s=1.15$ were obtained. The large difference between s and S shows that at the very least, the charge distribution was incorrectly modeled, i.e., the charges of the ions at the lattice sites deviate from their standard values.

To simplify the problem we took advantage of the correlation between the quadrupole interaction constants $C(\mathrm{Cu})$ obtained by ^{63}Cu NQR and $C(\mathrm{Zn})$ obtained by ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy for the same copper sites. As shown in Ref. 3, for a series of divalent copper metal oxides the experimental data for $C(\mathrm{Cu})$ and $C(\mathrm{Zn})$ fall on the curve

$$C(\mathrm{Cu})=197-11.3C(\mathrm{Zn}), \tag{3}$$

where $C(\mathrm{Cu})$ and $C(\mathrm{Zn})$ are given in MHz.

The linear dependence (3) is a consequence of the fact that in the copper metal oxides that were investigated, the valence component of the electric field gradient for Cu^{2+} is the same, while the decrease in $C(\mathrm{Cu})$ with increasing $C(\mathrm{Zn})$ suggests that the valence and crystal contributions to $C(\mathrm{Cu})$ have opposite signs and that $|(1-R_0)W_{zz}|>|(1-\gamma)V_{zz}|$ for the $^{63}\mathrm{Cu}^{2+}$ probe. The points for monovalent copper ($\mathrm{Cu_2O}$, Cu(1) state in $\mathrm{YBa_2Cu_3O_6}$) were found not to lie on the straight line (3). This can be explained by the loack of a valence contribution to the electric field gradient at the nucleus for the Cu^{+} probe. Thus, the $C(\mathrm{Cu})-C(\mathrm{Zn})$ diagram makes it possible to select copper positions where the copper is not divalent: the deviation of the $C(\mathrm{Cu})$ and $C(\mathrm{Zn})$ values from the straight line (3) for a specific compound indicates that the copper valence in that compound is other than $+2$.

As shown in Ref. 3, information about the charges of the lattice atoms can be obtained from the $C(\mathrm{Cu})-V_{zz}$ diagram, where V_{zz} is the principal component of the crystal electric field gradient tensor at the copper sites, calculated in the point-charge model for the specific copper compound, while $C(\mathrm{Cu})$ is the quadrupole interaction constant determined for this compound by the ^{63}Cu NQR method. For divalent-copper metal oxides this dependence has the form (Fig. 2b)[3]

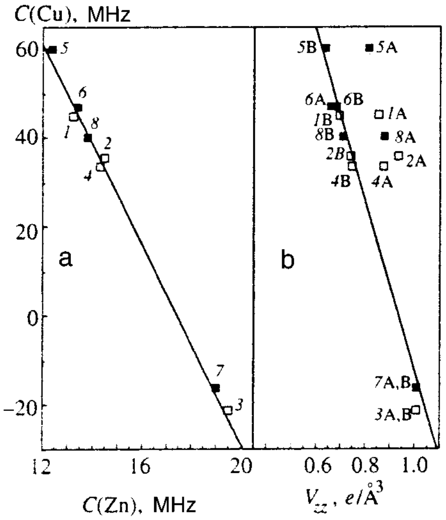


FIG. 2. a) $C(\mathrm{Cu})-C(\mathrm{Zn})$ diagram for divalent-copper compounds (solid line). b) $C(\mathrm{Cu})-V_{zz}$ diagram for divalent-copper compounds (solid straight line). The open squares show data for TlBaCaCuO: *1*—Cu in Tl(2201);[20] *2*—Cu in Tl(2212);[21] *3*—Cu(1) in Tl(2223);[4] *4*—Cu(2) in Tl(2223).[4] The filled squares show data for BiSrCaCuO: *5*—Cu in Bi(2201);[22] *6*—Cu in Bi(2212);[5,7,23] *7*—Cu(1) in Bi(2223);[6] *8*—Cu(2) in Bi(2223).[6] The indices A and B denote the model used to calculate V_{zz} (the calculation and the models themselves are described in the text). The citations refer to $C(\mathrm{Cu})$ data for TlBaCaCuO and BiSrCaCuO compounds. The values of $C(\mathrm{Zn})$ for these compounds were taken from Table I.

$$C(\mathrm{Cu})=179-191.4\,V_{zz}\,, \tag{4}$$

where $C(\mathrm{Cu})$ is given in MHz and V_{zz} in $e/\text{Å}^3$.

Just as in Fig. 2a, in Fig. 2b the points for the monovalent copper compounds do not lie on the straight line (4). However, if the copper is divalent, i.e., the corresponding point lies on the straight line (3) in the $C(\mathrm{Cu})-C(\mathrm{Zn})$ diagram, then a deviation from the straight line (4) in the $C(\mathrm{Cu})-V_{zz}$ diagram indicates that the ionic charge distribution was incorrectly modeled in calculating V_{zz}. Such an analysis makes it possible to choose possible alternative charge distributions in the lattices.

The ^{63}Cu NQR data for the compounds Tl(2201),[20] Tl(2212),[21] Tl(2223),[4] Bi(2201),[22] Bi(2212),[5,7,23] and Bi(2223)[6], together with our ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy data, are presented in the $C(\mathrm{Cu})-C(\mathrm{Zn})$ diagram (Fig. 2a). One can see that all points satisfactorily correspond to the relation (3), i.e., the copper is divalent in TlBaCaCuO and BiSrCaCuO compounds.

It is clear from Fig. 2b, however, that the data from Refs. 4–7 and 20–23 are inconsistent with the linear dependence (4) in the $C(\mathrm{Cu})-V_{zz}$ diagram if V_{zz} is calculated for model A, again confirming the shortcomings of this model. To reconcile the data with the dependence (4) for the Tl(2201), Bi(2201), and Tl(2212) compounds, lesser charges of the oxygen ions located in the Cu–O planes must be used, i.e., holes must be inserted into the corresponding sublattices. For the Bi(2212) compound, the holes must be inserted into the O(3) sublattice. Finally, for the Tl(2223) and Bi(2223) compounds, the holes must be positioned at the oxygen ions located in the same plane of the Cu(2) atoms. The corresponding hole concentrations are 0.20(2), 0.10(1), 0.10(1), 0.15(2), 0.12(1), and 0.20(2) per oxygen site.

The hole distribution indicated here is not, strictly speaking, the only possible one. However, the substantial deviations of the points from the straight line (4) in the $C(\mathrm{Cu})-V_{zz}$ diagram require that corrections to the standard charges in model A be made for ions located in the nearest-neighbor environment of the copper sites. The contributions made by the rest of the sublattices to the electric field gradient are at least an order of magnitude smaller than the contributions of the designated oxygen sublattices, and an order of magnitude larger deviations of the charges from the standard values would be required to reconcile the computed values of V_{zz} with Eq. (4). Such deviations are scarcely physically meaningful. Thus, the foregoing hole distribution in the oxygen sublattices is probably necessary in the model adopted.

Holes in the oxygen sublattices require introducing acceptor centers into the model in concentrations corresponding to neutrality. For the reason discussed above, their charges have virtually no effect on the value of V_{zz} at the copper sites. For TlBaCaCuO compounds, holes can result from a transition of some of the thallium atoms to a monovalent state. This is supported, specifically, by ^{205}Tl NMR data for Tl(2212) and Tl(2223).[24] One can see from Fig. 2b that satisfactory agreement with the linear dependence (4) is observed for models B, which take account of the appearance of holes at the oxygen sites, as well as a transition of 10(1)% of the thallium atoms in the compounds (2210) and (2223) and 12(1)% in the compound Tl(2212) to an monovalent state.

Direct experimental data on the nature of the acceptors do not exist for BiSrCaCuO compounds. The origin of holes in their oxygen sublattices can be explained, for example, by the defect rate in the material. Figure 2b shows agreement with the dependence (4) of models B, according to which the charge of holes at oxygen sites is balanced (to make the discussion specific) by a decrease in the charge of the bismuth sublattice. For Bi(2223) such a charge decrease does occur, due to the substitution of Pb for Bi.

For Tl(2223) and Bi(2223) compounds, there exist other experimental ^{63}Cu NQR data (in Ref. 5 (Tl), and Refs. 5 and 7–9 (Bi)) that differ appreciably from the data in Refs. 4 and 6. We see from Fig. 3a that the data from Refs. 5 and 7–9 leave the points for the Cu(2) sites on the straight line (3), confirming the divalence of Cu(2) in both cases, but lead to a considerable deviation of the Cu(1) points from the straight line (3). Such a deviation can be explained by a reduced valence contribution in $C(\mathrm{Cu})$ due to the partial filling of holes in the $3d$ shell of Cu(1). On the basis of Ref. 3, the observed deviation from the straight line (3) corresponds to a Cu(1) charge of $+1.8e$ in Tl(2223) and between $+1.85e$ and $+1.7e$ in Bi(2223). This could mean that (2223) samples with partially reduced copper in the Cu(1) positions were used in Refs. 5 and 7–9, and the degree of reduction depended on uncontrolled material fabrication conditions.

One can see from Fig. 3b that when the models A and the data of Refs. 5 and 7–9 are used, the Cu(1) and Cu(2) points for both compounds deviate from the straight line (4).

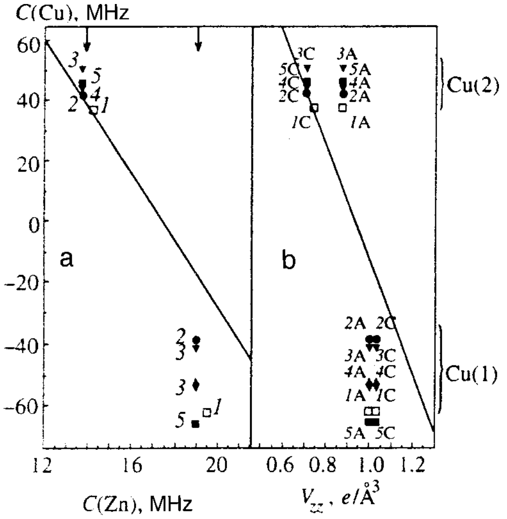


FIG. 3. a) C(Cu)–C(Zn) diagram for divalent-copper compounds (solid line). b) C(Cu)–V_{zz} diagram for divalent-copper compounds (solid line). The open squares *1* show data for Cu(1) and Cu(2) in the compound Tl(2223);[5] the filled symbols show data for Cu(1) and Cu(2) in the compound Bi(2223): *2*—Ref. 8; *3*—Ref. 7; *4*—Ref. 9; *5*—Ref. 5. The letters A and C denote the models used to calculate V_{zz} (the calculation and the models themselves are described in the text). The citations refer to the values of C(Cu) for the compounds Tl(2223) and Bi(2223). The values of C(Zn) for these compounds are taken from Table I.

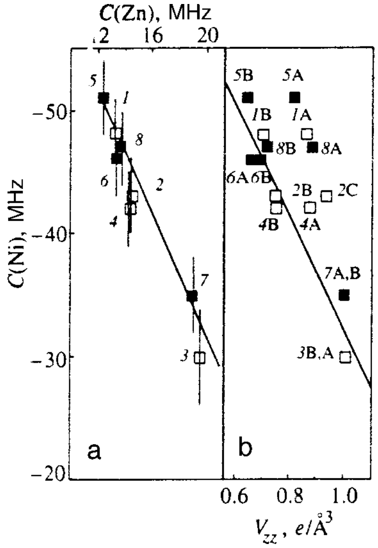


FIG. 4. a) C(Ni)–C(Zn) diagram for divalent-copper compounds (solid line). b) C(Ni)–V_{zz} diagrams for divalent-copper compounds (solid line). The open squares show the data for TlBaCaCuO: *1*—Cu in Tl(2201); *2*—Cu in Tl(2213), *3*—Cu(1) in Tl(2223); *4*—Cu(2) in Tl(2223). The filled squares show the data for BiSrCaCuO: *5*—Cu in Bi(2201); *6*—Cu in Bi(2212); *7*—Cu(1) in Bi(2221); *8*—Cu(2) in Bi(2223). The letters A and B designate the model to calculate V_{zz} (the calculation and the models themselves are described in the text).

The deviation of the Cu(1) points can be explained by the aforementioned deviation of the copper valence from $+2$ at these sites. The deviation of the Cu(2) points must be attributed to the appearance of holes at the oxygen sites, located in the same plane as the Cu(2) atoms. To explain their presence, it is necessary to take account of the additional source of holes due to reduction of the Cu(1) charge. Specifically, if it is assumed that the Cu(1) charge in Tl(2232) is $+1.8e$, then the Cu(2) point falls on the straight line (4) with 0.11 holes in the O(2) sublattice and 6% monovalent thallium. Similarly, if it is assumed that the Cu(1) charge in Bi(2232) is $+1.7e$, then with due account of the divalence of the lead, the presence of 0.16 holes at each O(2) and O(3) site, which is necessary in order that the Cu(2) point fall on the straight line (4), does not require additional acceptor centers. These models are marked with the letter C in Fig. 3b.

The models constructed for the charge distribution in (2223) lattices describe the ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy data in conjunction with a variety of ^{63}Cu NQR data. The improved charge distribution models not only reconcile the measured ^{63}Cu quadrupole interaction constants with the computed crystal electric field gradients, they also reconcile the values of S and s (for Tl(2223) compound, $s=1.34$ for the B model and $s=1.38$ for the C model; for a Bi(2223) compound, $s=1.41$ for the B model and $s=1.43$ for the C model). The B and C models presuppose the presence of holes at oxygen sites in the Cu(2)–O plane, but they differ with respect to the degree of filling of the hole in the $3d$ shell of copper at the Cu(1) sites. The choice between alternative models can be made only by eliminating the discrepancies in the experimental ^{63}Cu NQR data. It cannot be ruled out that the difference in the values of C(Cu) for the Cu(1) sites might be due to the difference in the technology used to prepare the (2223) samples, i.e., the charges of the Tl, Bi Cu(1), O(2), and O(3) atoms in the (2223) lattices can be controlled.

3.2. ^{61}Ni^{2+} probe

To obtain additional information about the charge distribution in the lattices of copper metal oxides and to check the accuracy of the ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy data, data obtained by this method on ^{61}Cu (^{61}Ni) can be used. Figure 4 displays the C(Cu)$-C$(Zn) diagram constructed from the ^{61}Cu (^{61}Ni) and ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy measurements of the C(Ni) and C(Zn) quadrupole interaction constants for ^{61}Ni^{2+} and ^{67}Zn^{2+} probes, respectively, at the copper sites of the same copper metal oxides.[25] For ^{61}Ni^{2+} the electric field gradient is produced both by the lattice ions and the valence electrons of the probe itself. Since the C(Ni)$-C$(Zn) diagram in Fig. 4a is a straight line, according to the relation (1) this means that the valence contribution to C(Ni) is constant. Points for TlBaCaCuO and BiSrCaCuO compounds are also plotted in Fig. 4a, and one can see that good agreement obtains with the existing data for other copper metal oxides.

Just as in the C(Cu)$-V_{zz}$ diagram, the validity of the charge-distribution models can be checked using the C(Ni)$-V_{zz}$ diagram (Fig. 4b) (here V_{zz} are the computed principal components of the tensor of the crystal gradient of the electric field for copper positions at which C(Ni) was measured).[25] The C(Cu)$-V_{zz}$ diagram is a straight line. The deviation from this straight line migiht be due to an incorrect calculation of the electric field gradient tensor because of an incorrect choice of atmoic charges. Figure 4b shows our results for two models for calculating the tensor of the crystal gradient of the electric field in TlBaCaCuO and BiSrCaCuO lattices. The results are inconsistent with the linear dependence in the diagram $C(\mathrm{Ni})-V_{zz}$, if V_{zz} is calculated for the A model. Satisfactory agreement with the linear dependence obtains for the B models, which take account of the appearance of holes at the oxygen atoms.

4. CONCLUSIONS

The parameters of the electric field gradient tensor at copper sites in the lattices of TlBaCaCuO and BiSrCaCuO compounds were determined by ^{61}Cu (^{61}Ni) and ^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy, and these parameters were calculated in the point-charge approximation. The quadrupole interaction constants for $^{67}\mathrm{Zn}^{2+}$ centers (^{67}Cu (^{67}Zn) emission Mössbauer spectroscopy data), $^{61}\mathrm{Ni}^{2+}$ centers (^{61}Cu (^{61}Ni) emission Mössbauer spectroscopy data), and $^{63}\mathrm{Cu}^{2+}$ centers (published ^{63}Cu NQR data) as well as the computed values of the principal component of the crystal electric field gradient tensor at the copper sites in these compounds were analyzed. To reconcile the experimental and computed values of the parameters of the electric field gradient tensor, it must be assumed that holes are present in the oxygen sublattices of all TlBaCaCuO and BiSrCaCuO compounds. These holes are located predominantly at oxygen sites located in the same plane as the Cu atoms (Tl(2201), Tl(2212), and Bi(2201)), Cu(2) atoms (Tl(2223) and Bi(2223)), or strontium atoms (Bi(2212)). The presence of holes in the oxygen sublattices is probably a typical feature of HTSCs and related compounds. The sources of these holes, at least for the compounds (2223), might be intrinsic atoms (Tl, Bi, Cu(1)) whose charge states depend on the technology used to fabricate the (2223) samples.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 97-02-16216).

*)E-mail: seregin@tuexph.stu.neva.ru

---

[1] F. S. Nasredinov, V. F. Masterov, N. P. Seregin, and P. P. Seregin, Zh. Éksp. Teor. Fiz. **99**, 1027 (1991) [Sov. Phys. JETP **72**, 570 (1991)].

[2] V. F. Masterov, P. P. Seregin, F. S. Nasredinov, and N. P. Seregin, Phys. Status Solidi B **196**, 11 (1996).

[3] P. P. Seregin, V. F. Masterov, F. S. Nasredinov, and N. P. Seregin, Phys. Status Solidi B **201**, 269 (1997).

[4] Yu. I. Zhdanov, A. M. Bogdanovich, K. N. Mikhalev *et al.*, Sverkhprovodimost': Fiz., Khim., Tekhn. **6**, 750 (1993).

[5] T. Oashi, K. Kumagai, H. Nakajima *et al.*, Physica C **161**, 367 (1989).

[6] B. W. Statt and L. M. Song, Physica C **183**, 372 (1991).

[7] H. Lutgemeir, Hyperfine Interact. **61**, 1051 (1990).

[8] K. Fujiwara, Y. Kitaoka, K. Asayama *et al.*, J. Phys. Soc. Jpn. **58**, 380 (1989).

[9] H. Reisemeier, G. Stadermann, H. Kamphausen *et al.*, J. Less-Common Met. **164/165**, 1106 (1990).

[10] V. F. Masterov, F. S. Nasredinov, Ch. S. Saidov, and P. P. Seregin, Fiz. Tverd. Tela (Leningrad) **34**, 2294 (1992) [Sov. Phys. Solid State **34**, 1228 (1992)].

[11] A. W. Hewat, E. A. Hewat, Y. Beyhestad *et al.*, Physica C **152**, 438 (1988).

[12] C. C. Torardi, M. A. Subramanian, J. C. Calabrese *et al.*, Phys. Rev. B **38**, 225 (1988).

[13] H. Borelet, Physica C **156**, 189 (1988).

[14] J. K. Liang, Y. L. Zhang, J. Q. Huang *et al.*, Physica C **156**, 616 (1988).

[15] W. Carrilo-Cabrera and M. T. Welker, Physica C **161**, 373 (1989).

[16] V. F. Masterov, F. S. Nasredinov, and P. P. Seregin, Fiz. Tverd. Tela (St. Petersburg) **37**, 1265 (1995) [Phys. Solid State **37**, 687 (1995)].

[17] A. Forster, W. Potzel, and G. M. Kalvius, Z. Phys. B **37**, 209 (1980).

[18] R. Sternheimer, Phys. Rev. **146**, 140 (1966).

[19] K. Yvon and M. Francois, Z. Phys. B **76**, 413 (1989).

[20] N. E. Alekseevskii, E. G. Nikolaev, A. V. Mitin *et al.*, Physica C **192**, 147 (1992).

[21] Yu. I. Zhdanov, B. A. Aleksashin, K. N. Mikhaljov *et al.*, Physica C **185/189**, 1187 (1991).

[22] Y. Kohori, K.-I. Ueda, and T. Kohara, Physica C **185/189**, 1187 (1991).

[23] A. K. Rajarajan, V. R. Palkar, N. C. Mishra *et al.*, Solid State Commun. **71**, 835 (1989).

[24] Yu. I. Zhdanov, K. N. Mikhalev, B. A. Aleksashin *et al.*, SFKhT **3**, 194 (1990).

[25] F, S. Nasredinov, P. P. Seregin, V. F. Masterov, and N. P. Seregin, J. Phys.: Condens. Matter **7**, 2339 (1995).

Translated by M. E. Alferieff

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Electronic superstructures in doped superlattices

I. P. Zvyagin*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 16 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 1089–1100 (September 1998)

Using an approach based on the density functional, we show that the exchange-correlation contribution to the system energy can be bigger than the sum of the kinetic energy and the Hartree contribution due to redistribution of carriers over the quantum wells in doped composite superlattices at low temperatures and moderate impurity densities. As a result, the ground state of the system can correspond to an inhomogeneous electron distribution over the quantum wells. Conditions when the homogeneous state is stable against small and finite density fluctuations are determined, and a phase diagram is plotted. A nonlinear theory of the inhomogeneous state is considered. © *1998 American Institute of Physics.* [S1063-7761(98)02309-9]

1. INTRODUCTION

Direct measurements of the conductivity in the vertical direction in GaAs/GaAlAs superlattices uniformly doped with silicon have revealed some unexpected features.[1] In particular, structures without artificially introduced disorder but with a relatively low doping level demonstrated at higher temperatures a transition from the quasi-metallic to activation conductivity with an activation energy comparable to or even higher then the minigap width. It was suggested[1] that this anomaly, which could not be interpreted in terms of the standard theory, might be caused by effects of Coulomb interaction.

It is well known that the role of the Coulomb interaction in structures of low dimensionality (in particular, both the exchange and correlation contribution to the total energy of a system) is especially important. It seems that these energies are responsible for the observed band gap reduction at high carrier densities in quantum wells. It has been noted[1] that the exchange-correlation effects can lead to distributions of the electron density varying in the vertical direction, which may be a cause of the observed anomalies in the conductivity of doped superlattices.

The fact that the ground state of a system may be inhomogeneous when the interaction between electrons is important (Wigner crystallization[2]) has been actively discussed in the literature. Accurate calculations of the Wigner crystallization parameters, however, are difficult because the difference between the important energies to be compared is small; therefore calculations of the threshold electron density differ considerably. For example in the three-dimensional case numerical Monte Carlo calculations produced a Wigner sphere with a radius measured in units of the Bohr radius, $a_0 = \varepsilon h^2/me^2$, equal to $r_s^{(3)} = 100 \pm 20$ [$r_s^{(3)} = (4\pi n_0^{(3)}/3)^{-1/3}/a_0$, $n_0^{(3)}$ is the average electron density, and ε is a permittivity of the material], and for two-dimensional systems $r_s \simeq 26$, where $r_s = (\pi\nu_0)^{-1/2}/a_0$ and ν_0 is the two-dimensional electron density.[3,4] In recent years, Wigner crystallization has been described using the density functional approach,[5,6] which was previously successfully applied to analyzing crystallization of simple liquids. In particular, a modified density functional theory[6] has yielded threshold values of r_s, in fair agreement with numerical calculations.

Experimental observation of Wigner crystallization is also very difficult because conditions for this process are quite restrictive. It seems that no Wigner crystallization has ever been detected in three dimensions. Only crystallization of two-dimensional electrons over a liquid helium surface was observed in experiments.[7,8]

Quasi-two-dimensional systems offer new opportunities for spontaneous symmetry breaking owing to electron density distributions that vary in the vertical direction because of specific features of the electron spectrum size quantization conditions.[1,9] Earlier[9] a mechanism giving rise to states with inhomogeneous carrier distributions among quantum wells in superlattices with a specially designed potential profile owing to delocalization of carriers in the second size quantization subband was discussed.

The present study shows that, generally speaking, a specially designed potential is unnecessary for generating inhomogeneous electron distributions in doped compositional superlattices. They can be created by a different mechanism in the presence of exchange-correlation interaction between quasi-two-dimensional electrons in quantum wells even if only the lowest subband of size quantization is important. We will see that the conditions for creating an electronic superstructure are less restrictive in this case than the conditions for Wigner crystallization, and they can be realized in conventional doped superlattices.

2. PROBLEM STATEMENT

Here we study the possibility of spontaneous symmetry breaking in a system of interacting electrons in doped compositional superlattices (note that this model can be applied to modulation-doped superlattices, which are composed of periodic p–n junctions).

It is convenient to use an approach based on the density functional theory. Let us consider electrons in a homogeneously doped compositional superlattice characterized by the Hamiltonian

$$H=T+V_{ee}+V_{ei}+V_{\text{conf}}+V_{\text{ext}}, \tag{1}$$

where T is the kinetic energy operator, V_{ee} is the operator of electron–electron interaction, V_{ei} is the operator of interaction between electrons and positively charged ions, V_{conf} is the operator describing interaction between electrons and confining potential of the superlattice, which determines the spatial profile of the conduction band edge, and V_{ext} is the operator of interaction with an external electric field.

The theory of the density functional is based on the Hohenberg–Kohn theorem, which asserts that a thermodynamic potential (such as free energy) of a system of interacting particles can be expressed in terms of a universal functional of one-particle density, which is independent of the external field and has a maximum corresponding to the ground state for an equilibrium density distribution $n(\mathbf{r})$.[10,11] In what follows, we limit our discussion to the case of the ground state at $T=0$ and consider extremal points of the total system energy $E[n]$. In a general case, the expression for energy, $E[n]$, considered as a functional of the density, can be written as follows:

$$E[n]=T[n]+E_H[n]+E_{xc}[n]+E_{\text{conf}}[n]+E_{\text{ext}}[n], \tag{2}$$

where $T[n]$ is the kinetic energy of a system of noninteracting electrons, $E_H[n]$ is the energy of interaction among electrons calculated in the classical Hartree approximation, which also includes the energy of interaction with the positive background compensating for the average negative electron charge, $E_{\text{conf}}[n]$ is the energy of interaction with the confining potential, and $E_{\text{ext}}[n]$ is the energy of interaction with external fields; the remaining part of the energy is denoted by $E_{xc}[n]$ and is called the exchange-correlation contribution (note that all the energies are measured per unit area of the superlattice). The major difficulty of the density functional theory is that the exact form of the functional $E_{xc}[n]$ is unknown, and it is calculated using various approximations, such as the local density approximation based on the gradient expansion of the energy.[10–12]

A remarkable feature of the system under investigation is that the potential V_{conf} is not a small parameter, but is largely responsible for forming electron states in the superlattice. We consider the case of sufficiently narrow quantum wells, when the populations of all subbands, except the lowest one, is negligible, and the overlap between electron wave functions of neighboring wells is small. In this case, the tight-binding approximation applies, and the problem can be formulated in terms of electron states localized in isolated quantum wells. In this approximation, the component $E_{\text{conf}}[n]$ is taken into account when the basis functions are defined. Neglecting the overlap between electron wave functions of neighboring quantum wells (this condition will be discussed in Sec. 6), we can consider the electron density as a function of the discrete argument described as a set of points i corresponding to isolated quantum wells, and the energy (2) as a functional defined on set $\{\boldsymbol{\rho},i\}$, where $\boldsymbol{\rho}$ is the continuous radius-vector in the quantum well plane, and i is the well number. Since the functional (2) is known, the problem of the electron density is solved by minimizing this functional.

3. ENERGY AS A FUNCTIONAL OF THE ELECTRON DENSITY IN A SUPERLATTICE

Our goal is to study the possibility of spontaneous symmetry breaking associated with different average electron densities in different quantum wells. Since, as will be shown below, conditions for the existence of such states are less restrictive than the conditions for creating inhomogeneous continuous distributions of the electron density, like that in a two-dimensional Wigner crystal, we limit our discussion to inhomogeneous distributions of electrons with a constant density in each quantum well. In other words, we assume that the electron density distribution function is independent of the coordinate $\boldsymbol{\rho}$, i.e., the function is defined on a discrete set of points corresponding to different quantum wells in the superlattice. For simplicity, we consider the case when the electron density is concentrated in narrow layers intersecting with the z-axis, which is aligned with the vertical direction, at $z_i=id$, where i is the quantum well number and d is the superlattice period. Hence the density can be expressed as

$$n(\mathbf{r})=\sum_n \nu_i\delta(z-z_i),$$

where ν_i is the two-dimensional electron density in the i-th well, so that the three-dimensional distribution of the electron density is determined by the set of electron densities $\nu=\{\nu_i\}$ in all quantum wells, and the density functional is equivalent to a simple function of many variables $\{\nu_i\}$.

Let us write expressions for different components of the energy (2). It is obvious that the kinetic energy is additive:

$$T[\nu]=\sum_i T^{(2)}(\nu_i), \tag{3}$$

where $T^{(2)}(\nu_i)$ is the kinetic energy of the two-dimensional electron gas in the i-th well. In the case when only the lowest subband is important and tunneling through barriers is negligible, we have

$$T^{(2)}(\nu_i)=\nu_i^2/2\rho_0, \tag{4}$$

where ρ_0 is the two-dimensional density of states in the quantum well. The Hartree energy $E_H[\nu]$ for narrow quantum wells can be expressed in the form

$$E_H[\nu]=\sum_{i,j} V_{ij}\nu_i\nu_j-\sum_i \nu_iN_dV^*, \tag{5}$$

where V_{ij} is the potential of the Coulomb interaction between electrons of the i-th and j-th wells, V^* is the Coulomb potential integrated over the entire volume, and N_d is the density of positively charged dopants. It is known that the last term on the right of Eq. (5) exactly compensates the interaction with the background charge of negatively charged carriers. For simplicity, we use the ''jellium'' model, assuming that the positive charge is uniformly spread over the space. In this case, the Hartree energy can be expressed as

$$E_H[\nu]=\frac{1}{2}\sum_{\substack{i,j\\(i\neq j)}} V_{ij}(\nu_i-\nu_0)(\nu_j-\nu_0), \tag{5a}$$

where

$$V_{ij}=-(2\pi e^2/\varepsilon)|z_i-z_j|, \tag{6}$$

ε is the permittivity, and z_i and z_j are the coordinates of the i-th and j-th wells. Given the homogeneous distribution of electrons within one well, the average electron density per unit area ν_0 is related to the three-dimensional density N_{d} of dopants by the formula $\nu_0=N_d d$.

Let us analyze separately the exchange and correlation contributions to the energy $E_{xc}[\nu]$. The exchange contribution $E_x[\nu]$ is additive if the overlap between wave functions of neighboring wells is negligible, i.e.,

$$E_x[\nu]=\sum_i E_x^{(2)}(\nu_i), \tag{7}$$

where $E_x^{(2)}(\nu_i)$ is the exchange energy of the two-dimensional electron gas in the ith quantum well. We have[13,14]

$$E_x^{(2)}(\nu_i)=-C_x(e^2/\varepsilon)\nu_i^{3/2}, \tag{8}$$

where $C_x\simeq 0.81$. In addition, corrections due to the finite quantum well width and finite temperature[15] can be easily included.

The correlation contribution $E_c[\nu]$ is due to a decrease in the spin-independent pair correlation function at small separations between electrons in the presence of Coulomb correlations (formation of a Coulomb hole around an electron). It is clear that the correlation contribution $E_c[\nu]$ is also additive if the Coulomb hole size is smaller than the separation between quantum wells:

$$E_c[\nu]=\sum_i E_c^{(2)}(\nu_i), \tag{9}$$

where $E_c^{(2)}(\nu_i)$ is the correlation energy of the two-dimensional electron gas with density ν_i. With due account of ring diagrams for $E_c^{(2)}(\nu)$ at $\nu<2a_0^{-2}$, we have[13]

$$E_c^{(2)}(\nu)=\frac{e^2}{2\varepsilon a_0^3}\left\{-1.29(\nu a_0^2)^{4/3}+1.20(\nu a_0^2)^{3/2}\right.$$
$$\left.-0.14(\nu a_0^2)^{5/3}+O\left((\nu a_0^2)^2\ln\frac{1}{\nu a_0^2}\right)\right\}. \tag{10}$$

Note, however, that additivity fails when the Coulomb hole radius is larger than the separation between neighboring wells. If the Coulomb hole radius is much larger than the superlattice period (i.e., $\nu_0^{-1/2}\gg d$), the effect of the correlation contribution on the redistribution of electrons among quantum wells is negligible in comparison with that of the exchange interaction. Hence, choosing the expression for the exchange-correlation energy in the form

$$E_{xc}[\nu]=\sum_i E_{xc}^{(2)}(\nu_i)$$

and setting $E_{xc}^{(2)}(\nu)\simeq E_x^{(2)}(\nu)$, we obtain a lower bound for the absolute value of the exchange-correlation energy. A more accurate estimate is $E_{xc}^{(2)}(\nu)\simeq E_x^{(2)}(\nu)+\lambda E_c^{(2)}(\nu)$, where $0<\lambda<1$. A more accurate account of the correlation energy leads to some (insignificant for $\nu_0^{1/2}d\ll 1$) weakening of conditions for an electronic superstructure, so in our estimates we use the approximation $E_{xc}^{(2)}(\nu)\simeq E_x^{(2)}(\nu)$.

Finally, in the presence of an external field the energy of its interaction with electrons in this approximation has the form

$$E_{\mathrm{ext}}[\nu]=\sum_i V_i^{(\mathrm{ext})}\nu_i, \tag{11}$$

where $V_i^{(\mathrm{ext})}$ is the energy of electron interaction with an external field (which is assumed to depend only on z) at $z=z_i$.

4. STABILITY OF THE HOMOGENOUS STATE AGAINST SMALL VARIATIONS IN THE ELECTRON DENSITY

It is known that Wigner crystallization in a homogeneous electron gas proceeds via a spin-polarized state corresponding to Stoner ferromagnetic ordering of electron spins.[4,5] An analysis of conditions for ferromagnetic ordering of band electrons at low electron densities similar to that performed by Shimizu[16] can be based on the expressions for the energy functional given above. Concentrations of electrons with spins up and down can be expressed as $\nu_\uparrow=(\nu/2)(1+x)$ and $\nu_\downarrow=(\nu/2)(1-x)$, where ν is the two-dimensional electron density in the layer and $|x|\leqslant 1$. Thus, we can derive from Eqs. (4) and (8) the energy difference between the polarized and nonpolarized states in a two-dimensional layer:

$$\Delta_{\mathrm{sp}}E(x)=\frac{\nu^2}{4\rho_0}x^2-C_x\frac{e^2\nu^{3/2}}{2\sqrt{2}\,\varepsilon}\{(1+x)^{3/2}+(1-x)^{3/2}-2\}. \tag{12}$$

It is clear that, in addition to the extremum at $x=0$ at sufficiently small ν $(\sqrt{\nu}<2\sqrt{2}(3/2)C_x e^2\rho_0/\varepsilon)$, the function $\Delta_{\mathrm{sp}}E(x)$ has a local minimum at $|x|=1$; when

$$\sqrt{\nu}<\frac{2\sqrt{2}(\sqrt{2}-1)C_x e^2\rho_0}{\varepsilon}, \tag{13}$$

the difference $\Delta_{\mathrm{sp}}E(1)$ is negative, i.e., the ground state corresponds to a fully spin-polarized electron gas in the layer. Finally, at $\sqrt{\nu}<(3\sqrt{2}/2)C_x e^2\rho_0/\varepsilon$ the nonpolarized state is unstable against small fluctuations in the spin polarization.

In the region of ferromagnetic ordering in isolated quantum wells, a quasi-one-dimensional system similar to a one-dimensional chain of one-electron atoms with a small overlap between wave functions of nearest neighbors is considered. For homogeneous doping, if the finite overlap integral between the electron wave functions in nearest quantum wells (interlayer exchange) is taken into account, the spin-dependent interaction between electrons in the superlattice should lead to ferromagnetic ordering of electron spins in neighboring quantum wells. Note the difference between the superlattice and a chain of one-electron atoms, because the analogue of the Hubbard energy in the superlattice (the change in the Coulomb energy when one additional electron is introduced to a quantum well) vanishes, so the Hubbard gap also goes to zero.

Next we prove that the spin-ordered state with equal electron densities in the layers is unstable against z-dependent variations of the electron density in a certain part of the density range defined by Eq. (13). Consider the conditions under which a transition from a spin-ordered state to an inhomogeneous state with different electron densities in different wells lowers energy. To this end, we study the expansion of the functional $E[\nu]$ in terms of $\delta\nu_i=\{\nu_i-\nu_0\}$ in the absence of an external field ($V_i^{(\mathrm{ext})}=0$) about the homogeneous state $\nu_i=\nu_0$, which always corresponds to an extremum of $E[\nu]$. With terms of up to the second order in $\delta\nu$, we have

$$\Delta E[\nu]=\sum_i\left(\frac{1}{\rho_0}-C_x\frac{3e^2}{4\varepsilon\sqrt{\nu_0}}\right)\delta\nu_i^2+\frac{1}{2}\sum_{\substack{i,j\\(j\neq i)}}V_{ij}\delta\nu_i\delta\nu_j. \tag{14}$$

Let us denote by N the total number of quantum wells in the superlattice and introduce normal coordinates using the expansion

$$\delta\nu_i=\frac{1}{\sqrt{N}}\sum_k[Q_k\exp(ikz_i)+Q_k^*\exp(-ikz_i)], \tag{15}$$

where the sum is performed over all wave numbers k. Assuming periodic boundary conditions on $\delta\nu_i$, the parameter k runs through the values $k=(2\pi/Nd)n$ within the first Brillouin zone: $-\pi/d<k\leqslant\pi/d$, so the integers n belong to the interval $-N/2<n\leqslant N/2$. By substituting Eq. (15) in expansion (14), we obtain

$$\Delta E[\nu]=\frac{1}{2}\sum_k\left[\frac{1}{\rho_0}-\frac{3C_xe^2}{4\varepsilon\sqrt{\nu_0}}+V(k)\right]|Q_k|^2. \tag{16}$$

Here

$$V(k)=\sum_i V(z_i)\exp(-ikz_i), \tag{17}$$

and we have taken into account that the parameters $V_{ij}=V(z_i-z_j)$ depend only on the differences between coordinates z_i-z_j. Functions $V(k)$ are periodic in the reciprocal one-dimensional lattice with period $b=2\pi/d$.

It follows from Eq. (16) that the homogeneous state of the system is stable against small fluctuations only if all factors

$$W(k)=\frac{1}{\rho_0}-\frac{3C_xe^2}{4\varepsilon\sqrt{\nu_0}}+V(k)$$

are positive. If at least one of them vanishes or becomes negative at a certain k_0, the system becomes unstable against a wave with wave number k_0. For the system under investigation, the function $V(k)$ can be calculated explicitly. We have

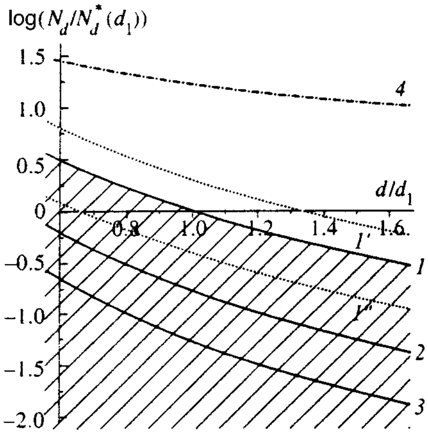


FIG. 1. Domain of electronic superstructure (hatched). Curves *1–3* define domains of ground states with periods $2d$, $3d$, and $4d$, respectively. The dotted line *1′* is the boundary of the instability region for the homogeneous distribution of electrons over the quantum wells of a superlattice, and the curve *1″* limits the domain of the metastable inhomogeneous state. The curve *4* defines the region in which the spin-ordered state is the ground state.

$$W(k)=\frac{1}{\rho_0}+\frac{e^2d}{\varepsilon\sin^2(kd/2)}-\frac{3C_xe^2}{4\varepsilon\sqrt{\nu_0}}. \tag{18}$$

It is clear that the function $W(k)$ has a minimum at $k=k_0=\pi/d$, i.e., at the Brillouin zone boundary. If

$$\frac{1}{\rho_0}+\frac{2\pi e^2d}{\varepsilon}-\frac{3C_xe^2}{4\varepsilon\sqrt{\nu_0}}<0, \tag{19}$$

a uniform distribution of electrons over the quantum wells becomes unstable, and the instability first appears for small harmonic variations in the electron density with $k=k_0$, i.e., for a wavelength double the superlattice period.

Condition (19) can be transformed to

$$N_d<N_{d1}^*, \tag{20}$$

where

$$N_{d1}^*=\frac{A_1}{d(d+d^*)^2},\quad A_1=\left(\frac{3C_x}{8\pi}\right)^2,\quad d^*=\frac{\varepsilon}{2\pi e^2\rho_0}.$$

The boundary of the region of instability of the spin-ordered homogeneous state against small spatial fluctuations is shown schematically in Fig. 1 by the dotted line 1′. The dash-dotted line 4 in Fig. 1 defines the region in which the spin-ordered state is the ground state.

Thus, Eqs. (13) and (19) demonstrate that the transition to the electronic superstructure at lower electron densities proceeds through a spin-ordered state. The reason is that the Coulomb interaction between charges in different layers described by the positive term $E_H[\nu]$ makes a transition to a state with an inhomogeneous distribution of electrons over the quantum wells more difficult. Moreover, when a spin-ordered state is formed in a layer, the two-dimensional electron density is half the density at which all electrons transfer to one of the two neighboring quantum wells.

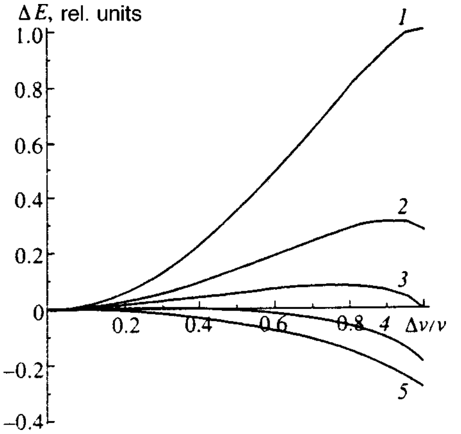


FIG. 2. Energies of inhomogeneous states versus $\Delta\nu/\nu_0$ at different values of $\nu_0=N_d d$ (the shapes of the curves are shown schematically). The parameter ν_0 decreases when the number labeling the curves increases.

5. PROPERTIES OF THE ELECTRONIC SUPERSTRUCTURE

One can investigate the inhomogeneous state of the system by minimizing the functional $E[\nu]$ in the absence of an external field under the additional condition

$$\sum_i \nu_i = N_d L,$$

where L is the system size in the vertical direction. It has been shown in the previous section that, at the point where stability is lost, the system becomes unstable against electron density fluctuations with the period which is double the superlattice constant. Therefore, let us first consider as an illustration the nonlinear theory of inhomogeneous states with a period twice as large as the superlattice period. Denote the electron density in even and odd quantum wells as $\nu_{1,2}=\nu_0 \pm \Delta\nu$, respectively. After introducing the parameter $u=\Delta\nu/\nu_0$, we can express the energy change per one superlattice period with respect to that in the homogeneous state as

$$\Delta E(u)=\frac{1}{2}\left(\frac{1}{\rho_0}+\frac{2\pi e^2 d}{\varepsilon}\right)\nu_0^2 u^2 - C_x\frac{e^2\nu_0^{3/2}}{2\varepsilon}\{(1+u)^{3/2} + (1-u)^{3/2}-2\}. \tag{21}$$

It is obvious that the homogeneous solution $u=0$ to equation $d\Delta E(u)/du=0$ exists for all ν_0, but, if condition (19) holds, it is unstable against small variations in the electron density. Curves of $\Delta E(u)$ for various ν_0 are plotted in Fig. 2. One can see that, at small electron densities ν_0, the energy has minima at $u=\Delta\nu/\nu_0=\pm 1$, i.e., when the density distribution is such that all electrons go to even (odd) quantum wells. The range of electron densities corresponding to an inhomogeneous state stable against small fluctuations is determined by the condition

$$\left.\frac{d\Delta E(u)}{du}\right|_{u=1}<0. \tag{22}$$

It can be transformed to

$$\frac{1}{\rho_0}+\frac{2\pi e^2 d}{\varepsilon}-\frac{3C_x(\sqrt{2}-1)}{4}\frac{e^2}{\varepsilon\sqrt{\nu_0}}<0$$

or

$$N_d<N_{d2}^*, \tag{23}$$

where

$$N_{d2}^*=\frac{A_2}{d(d+d^*)^2},\quad A_2=2\left(\frac{3C_x}{8\pi}\right)^2.$$

The region of the inhomogeneous state stable against small fluctuations is bounded in Fig. 1 by dotted curve *1″*.

The energy of the inhomogeneous distribution becomes lower than that of the homogeneous distribution when $\Delta E(u)|_{u=1}<0$. This condition can be expressed as

$$N_d<N_{d3}^*, \tag{24}$$

where

$$N_{d3}^*=\frac{A_3}{d(d+d^*)^2},\quad A_3=\frac{(3-2\sqrt{2})C_x^2}{\pi^2}.$$

Equation (24) determines the range of the dopant density in which the ground state is spatially nonuniform, i.e., an electronic superstructure with a period that is double the superlattice constant exists. This region (it is hatched in Fig. 1) is bounded by curve *1*. An important point is that conditions (21), (23), and (24) are not identical. In the region

$$N_{d1}^*<N_d<N_{d3}^* \tag{25}$$

the homogeneous state is stable against small fluctuations, but unstable against relatively large fluctuations. In the region

$$N_{d3}^*<N_d<N_{d2}^*$$

the inhomogeneous state is stable against small fluctuations, but is metastable.

Now let us address a more general case and suppose that a periodic structure with a period that is s-fold of the superlattice constant is formed, such that all electrons from the well labeled by $(n-1)$ go the sth well, so $\nu_1=s\nu_0$ and $\nu_i=0$ for $i-2,\dots,s$. Then the change in the energy per one superlattice period with respect to the homogeneous state can be expressed as

$$\Delta E(s)=\frac{\pi e^2 d}{2\varepsilon}\nu_0^{3/2}\left\{\left[(s-1)d^*+\frac{s^2-1}{3}d\right]\nu_0^{1/2} - \frac{C_x}{\pi}(s^{1/2}-1)\right\}. \tag{26}$$

It is clear that for $N_d<N_d^{(s)}$, where

$$N_d^{(s)}=\{s+1-2s^{1/2}\}\frac{C_x}{\pi}\frac{A^{(s)}}{d[(s-1)d^*+(s^2-1)d/3]^2},$$

and $A^{(s)}=(s+1-2s^{1/2})(C_x/\pi)^2$, the ground state corresponds to the period $s_{\min}d$, where $s_{\min}$ is the value of s corresponding to the minimum of energy (26). Thus, at low electron densities inhomogeneous states with periods larger than twice the superlattice constant can form. The regions in which the ground state periods are $3d$ and $4d$ are defined in Fig. 1 by curves *2* and *3*. Note, however, that such states have lower stability than those with the double superlattice period at intermediate densities because the gain in energy $\Delta E(s)$ is smaller at lower ν_0.

6. DISCUSSION

Thus, in superlattices with sufficiently small periods at low densities of charge carriers, states with nonuniform carrier distributions over the quantum wells or electronic superstructures can exist. Note that the mechanism for spontaneous symmetry breaking in the vertical direction is different from that of Wigner crystallization. Although the problem is formally one-dimensional, multidimensional aspects are also important in this problem, because the resulting inhomogeneous distribution essentially depends on the exchange and correlation interactions within quasi-two-dimensional quantum wells. The conditions for the spontaneous symmetry breaking in this case are less restrictive for the following reasons. It is known that the local density approximation yields considerably lower critical densities than more accurate methods, such as those based on numerical calculations.[3] The reason is that corrections deriving from the gradient expansion of the density functional are important. The critical density in the system under investigation is lower than in a homogeneous electron gas because overlap integrals, which play the role of gradient corrections, are small. They can be easily taken into account, but they produce only an additional term in equations that determine the domains of the superstructure existence and stability. Note that the overlap integrals depend on positions of energy levels in the wells, and they become smaller in the inhomogeneous state when the energy levels are shifted by the Coulomb fields due to the carrier redistribution between quantum wells.

A transition to an inhomogeneous state leads to shifts of size quantization subbands in neighboring wells. Hence a gap can appear in the spectrum of elementary excitations associated with electronic transitions between neighboring wells. These are the excitations that determine the vertical conductivity in superlattices at finite temperatures.

One can easily estimate the dopant density at which electronic superstructures are possible. By taking for an estimate $\rho_0 = 3\times 10^{13}$ eV^{-1} cm^{-2}, $\varepsilon = 12.7$, and $d = 9$ nm, we obtain $d^* \approx 4.7$ nm and $N^*_{d1} \approx 8.2\times 10^{16}\,\mathrm{cm}^{-3}$ (this corresponds to an average two-dimensional density $\nu_0 \approx 7\times 10^{10}\,\mathrm{cm}^{-2}$). As was stated above, inclusion of the correlation energy leads to an increase in constant C_{xc}, hence in the critical density N^*_{di}. This estimate indicates a possibility of inhomogeneous electronic states under conditions which are less restrictive than those for Wigner crystallization [at the parameters given above, $r_s \simeq 26$ (see Ref. 4) corresponds to a critical two-dimensional density $\nu_0 \approx 3\times 10^{8}\,\mathrm{cm}^{-2}$ for Wigner crystallization]. On the other hand, it follows from our estimate that formation of an electronic superstructure (and, consequently, of a gap in the spectrum of electronic excitations) could be expected in the range of dopant densities where anomalies in the superlattice vertical conductivity have been detected.[1]

I am indebted to A. G. Mironov for helpful discussions.

The work was supported by the Russian Fund for Fundamental Research (Grant 97-02-17334) and INTAS (Grant 94-4435).

*⁾E-mail: zvyagin@sc.phys.msu.su

---

[1] G. Richter, W. Stolz, P. Thomas, S. Koch, K. Maschke, and I. P. Zvyagin, Superlattices Microstruct. **22**, 475 (1997).
[2] E. P. Wigner, Phys. Rev. **46**, 1002 (1934).
[3] D. M. Ceperley and B. J. Adler, Phys. Rev. Lett. **45**, 567 (1980).
[4] D. M. Ceperley, Phys. Rev. B **18**, 3126 (1978).
[5] G. Senatore and N. H. March, Rev. Mod. Phys. **66**, 445 (1994).
[6] G. Senatore and G. Pastore, Phys. Rev. Lett. **64**, 303 (1990).
[7] C. C. Grimes and G. Adams, Phys. Rev. Lett. **42**, 795 (1979).
[8] T. Ando, A. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 1 (1982).
[9] V. V. Tugushev and E. A. Zhukovskii, Solid State Commun. **96**, 203 (1995).
[10] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[11] W. Kohn and P. Vashishta, in *Theory of the Inhomogeneous Electron Gas*, ed. by S. Lundquist and N. H. March, Plenum, New York (1983), p. 79.
[12] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).
[13] A. Isihara, *Quantum Liquids*, Springer, New York (1995).
[14] W. Koch and H. Haug, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 2nd ed., World Scientific, Singapore (1993).
[15] H. Totsuji, H. Tachibana, C. Totsuji, and S. Nara, Phys. Rev. B **51**, 11148 (1995).
[16] M. Shimizu, Proc. Roy. Soc. **84**, 397 (1964).

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

New intensity magneto-optical effect in materials exhibiting giant magnetoresistance

N. F. Kubrakov,*) A. K. Zvezdin, K. A. Zvezdin, and V. A. Kotov

General Physics Institute, Russian Academy of Sciences, 117942 Moscow, Russia

R. Atkinson†)

The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, United Kingdom

(Submitted 16 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1101–1114 (September 1998)

The change in the reflectivity of a metallic magnetic multilayer that exhibits giant magnetoresistance for a monochromatic electromagnetic plane wave with polarization along the magnetization (s polarization) in response to a change from the antiferromagnetic magnetic configuration of the multilayer to the ferromagnetic configuration is investigated. This magneto-optical effect is treated in the effective-medium approximation, in which the dielectric constant needed is found analytically with consideration of the interface roughness scattering of electrons. It is shown in the example of an Fe/Cu multilayer that the effect amounts to $\sim 0.7\%$. The representation found for the complex conductivity is convenient in a special case for investigating the magnetoresistive effect.
[S1063-7761(98)02409-3]

1. INTRODUCTION

Since the discovery of giant magnetoresistance in metallic magnetic multilayer films (multilayers), significant advances have been made in explaining its nature.[1–13] The model most frequently used for it is associated with roughness of the interfaces between the layers, on which the electrons comprising the macroscopic current are scattered differently, depending on which of the two possible orientations of the spin along a fixed direction of the magnetization in the plane of the multilayer they have. An external magnetic field **H** can change the configuration of the magnetization between neighboring magnetic layers separated by a layer of a nonmagnetic metal from an antiparallel, or, as it may be termed, antiferromagnetic, configuration (when $\mathbf{H}=0$) to a parallel, or ferromagnetic, configuration. As a result, the symmetry in the scattering of electrons with different spin orientations changes, and a pronounced magnetoresistive effect consequently appears. It can exceed 50%, as, for example, in the case of Co/Cu and Fe/Cr multilayers with a specific combination of parameters characterizing the transport properties of electrons in the multilayers.[7] The theory developed not so long ago on the basis of this idea[7,8] is a major generalization of the Fuchs–Sondheimer theory for describing the size effect of a single metallic layer[14] and yields results that are close to experiment. Other approaches to the problem based on the Kubo–Greenwood formula and Chambers and Pipard's method were considered, for example, in Refs. 12 and 13 and the references cited therein. Apart from the giant magnetoresistance of the multilayers, their optical properties, which vary in response to changes in their magnetic configuration, i.e., the magneto-optical effects, can be of considerable interest. Sufficient attention has not heretofore been focused on their investigation.

In this paper we shall examine the change in the reflectivity for a wave polarized along the magnetization (s polarization). This change is, in fact, a new magneto-optical effect, which is caused by the spin-dependent scattering of electrons (see Ref. 15).[1)] A method for obtaining a microscopic description of the optical properties of a metallic multilayer in the approximation of an effective homogeneous anisotropic medium will be demonstrated in the example of the determination of the diagonal element ε_{xx} of the dielectric tensor $\hat{\varepsilon}$, which is needed to find the dependence of the new intensity magneto-optical effect on the parameters of the multilayer.

2. EFFECTIVE MEDIUM AND REFLECTIVITY

When magneto-optical effects are considered in layered structures, it is usually assumed that each layer is characterized by its own dielectric tensor, which is the same as that of the corresponding bulk medium. Then the characteristic matrix of each layer and the parameters of the reflected or transmitted wave of interest are found from Maxwell's equations and the boundary conditions.[19,20] In metallic multilayers exhibiting giant magnetoresistance the electron mean free paths should exceed the thicknesses of the individual layers. For this reason, the use of the tensor of the corresponding bulk medium for each layer becomes unacceptable.

However, a multilayer can be treated as a homogeneous anisotropic medium[2)], for which the dielectric tensor $\hat{\varepsilon}$ must be determined. Moreover, such an effective medium can be regarded as a semi-infinite medium filling the half-space $z>0$ in order to demonstrate the occurrence of a magneto-optical effect consisting of a change in the reflectivity R for an s-wave in response to a change in the magnetic configuration of the multilayer.

It follows from the wave equation

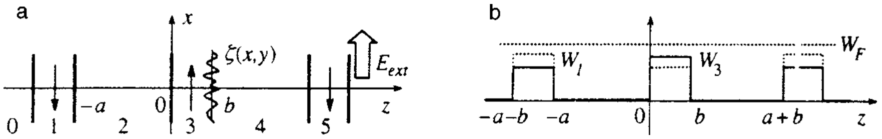


FIG. 1. a) Antiferromagnetic superlattice configuration; b) potential energy of electrons with spin up (solid line) and spin down (dashed line), $W_2=0$; W_F is the Fermi energy.

$$\nabla^2\mathbf{E}+k_0^2(1+i\mu_0c^2\omega^{-1}\hat{\sigma})\mathbf{E}=0 \tag{2.1}$$

that the element of the dielectric tensor needed to determine R is

$$\varepsilon_{xx}=1+i\mu_0c^2\omega^{-1}\sigma_{xx}, \tag{2.2}$$

where c is the speed of light, ω is the frequency of the wave, μ_0 is the magnetic constant, and σ_{xx} is the conductivity (a diagonal element of the tensor $\hat{\sigma}$), which relates the longitudinal current J_x averaged over a period of the structure to the external wave field inducing it. If the value of $\varepsilon_{xx}=\mathrm{Re}\,\varepsilon_{xx}+i\,\mathrm{Im}\,\varepsilon_{xx}$ is known, the solution of the corresponding boundary-value problem for Eq. (2.1) gives[20]

$$R=\left|\frac{n_0\cos\varphi-\sqrt{n^2-n_0^2\sin^2\varphi}}{n_0\cos\varphi+\sqrt{n^2-n_0^2\sin^2\varphi}}\right|^2, \tag{2.3}$$

where

$$\mathrm{Re}\,n,\quad \mathrm{Im}\,n=\frac{1}{\sqrt{2}}\sqrt{\sqrt{(\mathrm{Re}\,\varepsilon_{xx})^2+(\mathrm{Im}\,\varepsilon_{xx})^2}\pm\mathrm{Re}\,\varepsilon_{xx}}$$

are the real and imaginary parts of the refractive index of the effective medium, φ is the angle of incidence, and n_0 is the refractive index in the region $z<0$. It is henceforth assumed that $n_0=1$. Thus, the problem reduces to determining σ_{xx} in (2.2) with consideration of the features of the behavior of the conduction electrons in the multilayer.

3. DISTRIBUTION FUNCTION OF ELECTRONS AND CONDUCTIVITY

In the effective-medium approximation the value of ε_{xx} must be found under the following obvious condition: the multilayer (Fig. 1a) is an infinite periodic (along the z axis) structure (a one-dimensional magnetic superlattice). The periods of its antiferromagnetic and ferromagnetic configurations consist of four and two layers, respectively, and are specified by the mean coordinates of the rough interfaces.[3)] The wave

$$\mathbf{E}_{\mathrm{ext}}=\mathbf{E}_0\exp[-i\omega t+ik_0(x\sin\varphi+z\cos\varphi)],$$

which is polarized along the magnetization in the layers (s polarization), leads to the existence of a longitudinal current $J_{xj}(z)$ in each jth layer of the superlattice, and the value of σ_{xx} in (2.2) is found from the condition

$$J_x=\sigma_{xx}E_0. \tag{3.1}$$

The conductivity σ_{xx} can be determined using the kinetic theory, which is usually used to investigate the magnetoresistive effect in multilayers.[7,8,23–25] In this theory electrons are regarded as classical particles having a coordinate $\mathbf{r}$ and a quasimomentum $\mathbf{p}$. The nonequilibrium distribution function f_j of the electrons in each jth layer must satisfy the Boltzmann equation, which, in the relaxation-time approximation, has the form

$$\frac{\partial f_j}{\partial t}+(\mathbf{v}_j,\nabla_{\mathbf{r}}f_j)+e(\mathbf{E}_{\mathrm{ext}},\nabla_{\mathbf{p}}f_j)=-\frac{f_j-f_0}{\tau_{0j}}, \tag{3.2}$$

where $\mathbf{v}_j$ is the mean velocity of electrons characterized by the relaxation time τ_{0j}, $e=1.6\times10^{-19}$ C is the charge of an electron, and f_0 is the Fermi–Dirac distribution function. If the usual representation $f_j=f_0+\psi(\mathbf{v},t)$ is used, the deviation from Ohm's law is neglected, allowance is made for the fact that f_0 depends only on the energy of the electrons W and that $\nabla_{\mathbf{p}}f_0=\mathbf{v}_j\partial f_0/\partial W$, it is assumed that $\psi_j\propto\exp(-i\omega t+ik_0x\sin\varphi)$, and the relation $|\mathbf{v}|\tau_jk_0\ll1$ (the electron mean free path is much shorter than the wavelength of the light) is taken into account, then Eq. (3.2) is simplified:

$$\frac{\partial\psi_j}{\partial z}+\frac{1}{\tau_jv_{zj}}\psi_j=-eE_0\frac{v_{xj}\exp(ik_0z)}{v_{zj}}\frac{\partial f_0}{\partial W}. \tag{3.3}$$

Here v_{xj} and v_{zj} are the projections of the velocity $\mathbf{v}_j$ onto the x and z axes, and the parameter $\tau_j=\tau_{0j}(1-i\omega\tau_{0j})^{-1}$ can be called the frequency-dependent relaxation time. The function

$$\psi_j(z)=eE_0\frac{\partial f_0}{\partial W}\left[C_j\exp\left(-\frac{z}{\tau_j|v_{zj}|}\right)-\frac{\tau_jv_{xj}\exp(ik_0z)}{1+ik_0\tau_jv_{zj}}\right],$$

which is a general solution of Eq. (3.3), can be separated into two equations:

$$\psi_j^+(z)=eE_0\frac{\partial f_0}{\partial W}\left[C_j^+\exp\left(-\frac{z}{\tau_j|v_{zj}|}\right)-\tau_jv_{xj}\right],\quad v_{zj}>0,$$

$$\psi_j^-(z)=eE_0\frac{\partial f_0}{\partial W}\left[C_j^-\exp\left(\frac{z}{\tau_j|v_{zj}|}\right)-\tau_jv_{xj}\right],\quad v_{zj}<0, \tag{3.4}$$

with consideration of the approximation $\tau_j^{-1}+ik_0v_{zj}=\tau_{0j}^{-1}-i\omega(1-2\pi v_{zj}c^{-1})\approx\tau_j^{-1}$, which can be made in view of the condition $|v_{zj}|<v_{Fj}\ll c$, where v_{Fj} is the Fermi velocity. The functions ψ_j^+ and ψ_j^- have the same form as they would have, if there were a uniform electric field along the x axis in the superlattice instead of a plane wave (Fig. 1a). Therefore, there should be periodicity in the electron distribution for such symmetry, i.e., it is sufficient to determine only the functions $\psi_1^\pm,\dots,\psi_4^\pm$ for the antiferromagnetic configuration.

The arbitrary constants C_j^+ and C_j^- in (3.4) can be found from the boundary conditions:[7]

$$\psi_1^+(-a-b)=P_{10}\psi_1^-(-a-b)+Q_{01}\psi_0^+(-a-b),$$

$$\psi_1^-(-a)=P_{12}\psi_1^+(-a)+Q_{21}\psi_2^-(-a),$$

$$\begin{aligned}
&\psi_2^+(-a)=P_{21}\psi_2^-(-a)+Q_{12}\psi_1^+(-a),\\
&\psi_2^-(0)=P_{23}\psi_2^+(0)+Q_{32}\psi_3^-(0),\\
&\psi_3^+(0)=P_{32}\psi_3^-(0)+Q_{23}\psi_2^+(0),\\
&\psi_3^-(b)=P_{34}\psi_3^+(b)+Q_{43}\psi_4^-(b),\\
&\psi_4^+(b)=P_{43}\psi_4^-(b)+Q_{34}\psi_3^+(b),\\
&\psi_4^-(a+b)=P_{45}\psi_4^+(a+b)+Q_{54}\psi_5^-(a+b),
\end{aligned}\tag{3.5}$$

where P_{jk} is the transparency factor, which is equal to the probability of the specular reflection of an electron from the interface between layers j and k (Fig. 1a), and Q_{jk} is the transmission coefficient, which characterizes the passage of electrons through this interface. The roughness of each interface is assumed to be a random function $\zeta_{jk}(x,y)$, but it is such that its average value in a plane is equal to zero: $\langle\zeta_{jk}(x,y)\rangle=0$. The roughness can be characterized by the parameter $\eta=\sqrt{\langle\zeta_{jk}^2(x,y)\rangle}$, which will henceforth be assumed to be identical for all the interfaces. The explicit representations of P_{jk} and Q_{jk} in terms of η and the electron parameters are known as the generalized Ziman–Soffer formulas:[23]

$$\begin{aligned}
&P_{jk}=R_{jk}\exp\left[-\left(\frac{2\eta}{\hbar}m_jv_j\cos\theta_j\right)^2\right],\\
&Q_{jk}=(1-R_{jk})\exp\left[-\frac{\eta^2}{\hbar^2}(m_jv_j\cos\theta_j-m_kv_k\cos\theta_k)^2\right],\\
&R_{jk}=\left(\frac{v_j\cos\theta_j-v_k\cos\theta_k}{v_j\cos\theta_j+v_k\cos\theta_k}\right)^2,
\end{aligned}\tag{3.6}$$

where m_j is the effective mass and v_j is the absolute value of the velocity.

When electrons pass through an interface, which acts as a potential barrier, the angles θ_j and θ_k must satisfy the refraction law

$$\frac{\sin\theta_j}{\sin\theta_k}=\frac{m_kv_k}{m_jv_j}.\tag{3.7}$$

For the possible total internal reflection of electrons, where their angle of incidence θ_j exceeds a certain critical value, the reflectivity $R_{jk}=1$, and the transmission coefficient $Q_{jk}=0$.

The boundary conditions (3.5) become periodic, if

$$\begin{aligned}
&P_{10}=P_{54},\quad Q_{01}=Q_{45},\quad \psi_0^+(-a-b)=\psi_4^+(a+b),\\
&\psi_5^-(a+b)=\psi_1^-(-a-b).
\end{aligned}$$

This gives (in the case of the antiferromagnetic configuration) a matrix equation for $C_1^\pm,\ldots,C_4^\pm$. The dimensionality of the matrix (8×8), however, is halved, if the following relations for the P_{jk} and Q_{jk} and the limiting values of the ψ_j within a period of the superlattice, which are obvious in view of the symmetry, are used:

$$\begin{aligned}
&\psi_1^+(-a-b)=\psi_1^-(-a),\quad P_{10}=P_{12},\quad Q_{01}=Q_{21},\\
&\psi_0^+(-a-b)=\psi_2^+(-a),\\
&\psi_1^+(-a)=\psi_1^-(-a-b),\quad \psi_1^+(-a)=\psi_1^-(-a-b),\\
&\psi_3^-(0)=\psi_3^+(b).
\end{aligned}$$

Thus, it is sufficient to have only four boundary conditions, viz.,

$$\begin{aligned}
&\psi_1^-(-a)=P_{12}\psi_1^-(-a-b)+Q_{21}\psi_2^-(-a),\\
&\psi_2^+(-a)=P_{21}\psi_2^-(-a)+Q_{12}\psi_1^-(-a-b),\\
&\psi_2^-(0)=P_{23}\psi_2^+(0)+Q_{32}\psi_3^+(b),\\
&\psi_3^+(0)=P_{32}\psi_3^+(b)+Q_{23}\psi_2^+(0),
\end{aligned}\tag{3.8}$$

in order to determine the distribution functions needed. The substitution of (3.4) into (3.8) leads to the equation

$$\begin{bmatrix}
\exp\left(-\frac{a}{\xi_1}\right)\left[1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right] & 0 & -Q_{21}\exp\left(-\frac{a}{\xi_2}\right) & 0\\
-Q_{12}\exp\left(-\frac{a+b}{\xi_1}\right) & \exp\left(\frac{a}{\xi_2}\right) & -P_{21}\exp\left(-\frac{a}{\xi_2}\right) & 0\\
0 & -P_{23} & 1 & -Q_{32}\exp\left(-\frac{b}{\xi_3}\right)\\
0 & -Q_{23} & 0 & 1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)
\end{bmatrix}$$

$$\times\begin{bmatrix}C_1^-\\ C_2^+\\ C_2^-\\ C_3^+\end{bmatrix}=\begin{bmatrix}\tau_1v_{x1}(1-P_{12})-\tau_2v_{x2}Q_{21}\\ \tau_2v_{x2}(1-P_{21})-\tau_1v_{x1}Q_{12}\\ \tau_2v_{x2}(1-P_{23})-\tau_3v_{x3}Q_{32}\\ \tau_3v_{x3}(1-P_{32})-\tau_2v_{x2}Q_{23}\end{bmatrix},\tag{3.9}$$

where $\xi_j=\tau_j|v_{zj}|$. A direct calculation of the inverse matrix for (3.9) gives

$$\begin{aligned}
C_1^- ={}& D^{-1}\exp\left(\frac{a}{\xi_1}\right)\Bigg\{\tau_1 v_{x1}\Bigg[(1-P_{12})\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)\\
&-\exp\left(-\frac{2a}{\xi_2}\right)((1-P_{12})P_{21}+Q_{12}Q_{21})\Bigg(P_{23}-(P_{23}P_{32}\\
&-Q_{23}Q_{32})\exp\left(-\frac{b}{\xi_3}\right)\Bigg)\Bigg]-\tau_2 v_{x2}Q_{21}\left(1-\exp\left(-\frac{a}{\xi_2}\right)\right)\\
&\times\left[\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)\left(1+P_{23}\exp\left(-\frac{a}{\xi_2}\right)\right)\right.\\
&\left.+Q_{23}Q_{32}\exp\left(-\frac{b}{\xi_3}-\frac{a}{\xi_2}\right)\right]-\tau_3 v_{x3}Q_{21}Q_{32}\\
&\times\exp\left(-\frac{a}{\xi_2}\right)\left(1-\exp\left(-\frac{b}{\xi_3}\right)\right)\Bigg\},
\end{aligned}$$

$$\begin{aligned}
C_2^- ={}& D^{-1}\Bigg\{-\tau_1 v_{x1}Q_{12}\exp\left(-\frac{a}{\xi_2}\right)\left(1-\exp\left(-\frac{b}{\xi_1}\right)\right)\\
&\times\left(P_{23}-(P_{23}P_{32}-Q_{23}Q_{32})\exp\left(-\frac{b}{\xi_3}\right)\right)+\tau_2 v_{x2}\\
&\times\Bigg[\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)\Bigg((1-P_{23})\Bigg(1-P_{32}\\
&\times\exp\left(-\frac{b}{\xi_3}\right)\Bigg)-Q_{23}Q_{32}\exp\left(-\frac{b}{\xi_3}\right)\Bigg)\\
&+\left(P_{23}-(P_{23}P_{32}-Q_{23}Q_{32})\exp\left(-\frac{b}{\xi_3}\right)\right)\Bigg((1-P_{21})\\
&\times\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)-Q_{12}Q_{21}\exp\left(-\frac{b}{\xi_1}\right)\Bigg)\\
&\times\exp\left(-\frac{a}{\xi_2}\right)\Bigg]-\tau_3 v_{x3}Q_{32}\left(1-\exp\left(-\frac{b}{\xi_3}\right)\right)\\
&\times\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)\Bigg\},
\end{aligned}$$

$$\begin{aligned}
C_2^+ ={}& D^{-1}\exp\left(-\frac{a}{\xi_2}\right)\Bigg\{-\tau_1 v_{x1}Q_{12}\left(1-\exp\left(-\frac{b}{\xi_1}\right)\right)\\
&\times\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)+\tau_2 v_{x2}\Bigg[\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)\\
&\times\Bigg((1-P_{21})\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)-Q_{12}Q_{21}\\
&\times\exp\left(-\frac{b}{\xi_1}\right)\Bigg)+\Bigg(P_{21}-(P_{12}P_{21}-Q_{12}Q_{21})\\
&\times\exp\left(-\frac{b}{\xi_1}\right)\Bigg)\Bigg((1-P_{23})\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)\\
&-Q_{23}Q_{32}\exp\left(-\frac{b}{\xi_3}\right)\Bigg)\exp\left(-\frac{a}{\xi_2}\right)\Bigg]-\tau_3 v_{x3}Q_{32}\\
&\times\exp\left(-\frac{a}{\xi_2}\right)\left(1-\exp\left(-\frac{b}{\xi_3}\right)\right)\Bigg(P_{21}-(P_{12}P_{21}\\
&-Q_{12}Q_{21})\exp\left(-\frac{b}{\xi_1}\right)\Bigg)\Bigg\},
\end{aligned}$$

$$\begin{aligned}
C_3^+ ={}& D^{-1}\Bigg\{-\tau_1 v_{x1}Q_{12}Q_{23}\exp\left(-\frac{a}{\xi_2}\right)\left(1-\exp\left(-\frac{b}{\xi_1}\right)\right)\\
&-\tau_2 v_{x2}Q_{23}\left(1-\exp\left(-\frac{a}{\xi_2}\right)\right)\Bigg[\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)\\
&\times\left(1+P_{21}\exp\left(-\frac{a}{\xi_2}\right)\right)+Q_{12}Q_{21}\exp\left(-\frac{b}{\xi_1}-\frac{a}{\xi_2}\right)\Bigg]\\
&+\tau_3 v_{x3}\Bigg[(1-P_{32})\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)-\exp\left(-\frac{2a}{\xi_2}\right)\\
&\times((1-P_{32})P_{23}+Q_{23}Q_{32})\Bigg(P_{21}-(P_{12}P_{21}\\
&-Q_{12}Q_{21})\exp\left(-\frac{b}{\xi_1}\right)\Bigg)\Bigg]\Bigg\},
\end{aligned}$$

$$\begin{aligned}
D={}&\left(1-P_{12}\exp\left(-\frac{b}{\xi_1}\right)\right)\left(1-P_{32}\exp\left(-\frac{b}{\xi_3}\right)\right)\\
&-\exp\left(-\frac{2a}{\xi_2}\right)\left(P_{21}-(P_{12}P_{21}-Q_{12}Q_{21})\exp\left(-\frac{b}{\xi_1}\right)\right)\\
&\times\left(P_{23}-(P_{23}P_{32}-Q_{23}Q_{32})\exp\left(-\frac{b}{\xi_3}\right)\right). \qquad (3.10)
\end{aligned}$$

The remaining constants are related to the constants already found in the following manner:

$$C_1^+=C_1^-\exp\left(-\frac{2a+b}{\xi_1}\right),\quad C_3^-=C_3^+\exp\left(-\frac{b}{\xi_3}\right),$$

$$C_4^+=C_2^-\exp\left(-\frac{b}{\xi_2}\right),\quad C_4^-=C_2^+\exp\left(-\frac{b}{\xi_2}\right).$$

After substituting (3.10) into (3.4), we obtain the distribution functions.

If it is assumed that the Fermi surface is spherical, the velocity

$$v_j=\sqrt{\frac{2(W-W_j)}{m_j}}, \qquad (3.11)$$

where W_j is the potential energy of the electrons in layer j. The potential energy of the electrons (Fig. 1b) depends on the orientation of their spin.[7] Therefore, the distribution functions also depend on spin. Then, the current J_x induced by the wave can be regarded as the sum of the two currents caused by electrons with spins whose orientations along the magnetization can be called by convention up (↑) and down (↓), respectively, as is shown in Fig. 1b. The distribution functions found can clearly be used to calculate each of these two currents in all the layers of a superlattice within a period. For this purpose, the known values of the potential energy of electrons with an assigned spin orientation must be substituted into (3.11). In spherical coordinates the velocity $\mathbf{v}_j = \mathbf{v}_j(W)(\cos\varphi\sin\theta_j,\ \sin\varphi\sin\theta_j,\ \cos\theta_j)$. Then the definition of the current in layer j has the form[24,25]

$$J_{xj}(z)=\frac{e}{4\pi}\int_0^\infty \rho_j v_j\,dW\int_0^{2\pi}\cos\varphi\,d\varphi \times\left[\int_0^{\pi/2}\psi_j^+\sin^2\theta_j\,d\theta_j+\int_{\pi/2}^{\pi}\psi_j^-\sin^2\theta_j\,d\theta_j\right], \tag{3.12}$$

where ρ_j is the density of states. The currents J_{x1}, J_{x2}, J_{x3}, and $J_{x4}=J_{x2}$ are found from (3.12) with consideration of the approximation $\partial f_0/\partial W=-\delta(W-W_F)$, where W_F is the Fermi energy. After averaging the current within a period, we obtain the following necessary result:

$$\sigma_{xx}=\frac{1}{2(a+b)}[b\sigma_1\gamma_1+2a\sigma_2\gamma_2+b\sigma_3\gamma_3], \tag{3.13}$$

where

$$\sigma_j=\frac{1}{3}e^2\rho_{Fj}\tau_j v_{Fj}^2=\frac{e^2m_j^2\tau_j v_{Fj}^3}{6\pi^2\hbar^3}$$

is the bulk conductivity of the jth layer.[7] The Fermi velocity $v_{Fj}=\sqrt{2(W_F-W_j)/m_j}$ must be substituted for v_j in (3.6), (3.7), and (3.10), which are needed to calculate the parameters

$$\begin{aligned}
\gamma_1=1-\frac{3}{2b}\int_0^{\pi/2}&\Bigl(1-\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\Bigl\{l_1\sin\theta_1\\
&\times\Bigl[(1-P_{12})\Bigl(1-P_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\\
&-\exp\Bigl(-\frac{2a}{l_2\cos\theta_2}\Bigr)\Bigl(P_{23}\Bigl(1-P_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\\
&+Q_{23}Q_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)((1-P_{12})P_{21}+Q_{12}Q_{21})\Bigr]\\
&-l_2\sin\theta_2Q_{21}\Bigl(1-\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\Bigl[\Bigl(1-P_{32}\\
&\times\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\Bigl(1+P_{23}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\\
&+Q_{23}Q_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr]\\
&-l_3\sin\theta_3Q_{21}Q_{32}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\\
&\times\Bigl(1-\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\Bigr\}D^{-1}\sin^2\theta_1\cos\theta_1\,d\theta_1,
\end{aligned}$$

$$\begin{aligned}
\gamma_2=1-\frac{3}{4a}\int_0^{\pi/2}&\Bigl(1-\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\Bigl\{\Bigl[l_2\sin\theta_2\\
&\times\Bigl((1-P_{21})\Bigl(1-P_{12}\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)-Q_{12}Q_{21}\\
&\times\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)-l_1\sin\theta_1Q_{12}\\
&\times\Bigl(1-\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\Bigr]\Bigl[\Bigl(1-P_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\\
&\times\Bigl(1+P_{23}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)+Q_{23}Q_{32}\\
&\times\exp\Bigl(-\frac{a}{l_2\cos\theta_2}-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr]+\Bigl[l_2\sin\theta_2\Bigl((1-P_{23})\\
&\times\Bigl(1-P_{32}\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)-Q_{23}Q_{32}\\
&\times\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)-l_3\sin\theta_3Q_{32}\\
&\times\Bigl(1-\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\Bigr]\Bigl[\Bigl(1-P_{12}\\
&\times\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\Bigl(1+P_{21}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\\
&+Q_{12}Q_{21}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr]\Bigr\}\\
&\times D^{-1}\sin^2\theta_2\cos\theta_2\,d\theta_2,
\end{aligned}$$

$$\begin{aligned}
\gamma_3=1-\frac{3}{4a}\int_0^{\pi/2}&\Bigl(1-\exp\Bigl(-\frac{b}{l_3\cos\theta_3}\Bigr)\Bigr)\Bigl\{-l_1\sin\theta_1Q_{12}\\
&\times Q_{23}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigl(1-\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\\
&-l_2\sin\theta_2Q_{23}\Bigl(1-\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\Bigl[\Bigl(1-P_{12}\\
&\times\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\Bigl(1+P_{21}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}\Bigr)\Bigr)\\
&+Q_{21}Q_{12}\exp\Bigl(-\frac{a}{l_2\cos\theta_2}-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr]+l_3\sin\theta_3\\
&\times\Bigl[(1-P_{32})\Bigl(1-P_{12}\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\\
&-\exp\Bigl(-\frac{2a}{l_2\cos\theta_2}\Bigr)\Bigl(P_{21}\Bigl(1-P_{12}\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)\\
&+Q_{12}Q_{21}\exp\Bigl(-\frac{b}{l_1\cos\theta_1}\Bigr)\Bigr)((1-P_{32})P_{23}\\
&+Q_{32}Q_{23})\Bigr]\Bigr\}D^{-1}\sin^2\theta_3\cos\theta_3\,d\theta_3,
\end{aligned} \tag{3.14}$$

which take into account the interface roughness and the size effect, or, more specifically, the fact that the thicknesses of the layers can be much smaller than the electron mean free path $l_j=\tau_j v_F$. The angles θ_k in (3.6) are related to the angle over which the integration in (3.14) is carried out by Eq. (3.7), from which it follows that the angle of total internal reflection of the electrons

$$\theta_{j\,\max}=\arcsin\sqrt{\frac{m_k(W_F-W_k)}{m_j(W_F-W_j)}},\quad \text{if}\quad W_k>W_j.$$

It is important to note that since the spatial dispersion could be neglected, the difference between $\sigma_{xx}=\sigma_{xx}^{\uparrow}+\sigma_{xx}^{\downarrow}$ in (3.1) and the conductivity in the static case is related only to $\tau_j(\omega)$. Therefore, the magnetoresistive effect of a magnetic superlattice can be investigated using (3.13) as $\omega\to 0$.

For the antiferromagnetic configuration, we clearly have $\sigma_{xx}^{\uparrow}=\sigma_{xx}^{\downarrow}$, and the conductivity σ_{xx}, which is denoted by σ_{AF}, is

$$\sigma_{\mathrm{AF}}=\frac{1}{a+b}[b\sigma_1\gamma_1+2a\sigma_2\gamma_2+b\sigma_3\gamma_3], \tag{3.15}$$

where the parameters σ and γ are calculated using (3.14) for electrons with spin up or spin down.

Reversal of the direction of the magnetization in the third layer of the superlattice period (Fig. 1a) gives a ferromagnetic configuration, for which (3.13) remains valid, since the potential energy W_3 was assumed to be an arbitrary parameter during its derivation. In this case $W_3=W_1$ for electrons with a certain spin orientation. Then, in (3.14) $v_{F3}=v_{F1}$, $m_3=m_1$, and, according to (3.7), $\theta_3=\theta_1$. It follows from (3.6) that $P_{32}=P_{12}$, $P_{23}=P_{21}$, and $Q_{32}=Q_{12}$, and in (3.13) $\sigma_3=\sigma_1$ and $\gamma_3=\gamma_1$. However, $\sigma_{xx}^{\uparrow}\neq\sigma_{xx}^{\downarrow}$. When these relations are taken into account, the conductivity for the ferromagnetic configuration has the form

$$\sigma_{\mathrm{FM}}=\frac{1}{a+b}\sum_{\uparrow\downarrow}(b\sigma_1\gamma_1+a\sigma_2\gamma_2). \tag{3.16}$$

If the conductivities of the second and fourth layers (Fig. 1a) within a period are set equal to zero, it follows from (3.14)–(3.16) that

$$\sigma_{\mathrm{AF}}=\sigma_{\mathrm{FM}}=\frac{b}{a+b}(\sigma_1\gamma_1+\sigma_3\gamma_3).$$

The expressions for γ_1 and γ_3 are similar to the expression for the parameter

$$\gamma=1-\frac{3l}{2b}\times\int_0^{\pi/2}\frac{(1-P)[1-\exp(-b/l\cos\theta)]}{1-P\exp(-b/l\cos\theta)}\sin^3\theta\cos\theta\,d\theta,$$

which takes into account the size effect (for a layer of thickness b) in the Fuchs–Sondheimer theory.[14] If all the layers are identical ($W_j=0$, $\sigma_j=\sigma$) and there is no roughness ($\eta=0$), it follows from (3.14)–(3.16) that $\sigma_{\mathrm{AF}}=\sigma_{\mathrm{FM}}=\sigma$.

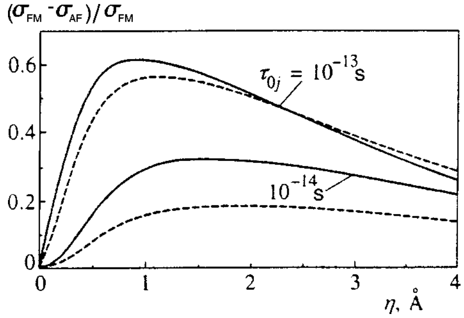


FIG. 2. Dependence of the magnetoresistive effect on η for $a=b=10$ Å (solid lines) and $a=b=20$ Å (dashed lines).

4. MAGNETORESISTIVE EFFECT

The expressions obtained for the conductivities σ_{AF} and σ_{FM} in the special case where $\omega=0$ clearly enable us to see how the magnetoresistive effect, which is characterized by the quantity $(\sigma_{\mathrm{FM}}-\sigma_{\mathrm{AF}})/\sigma_{\mathrm{FM}}$, depends on the thicknesses of the layers (a and b), the interface roughness parameter η, and the relaxation time.

As an example, we shall examine an Fe/Cu superlattice, for which the potential-energy distribution of the electrons is known (Fig. 1b):[7] $W_F-W_j=8.23$, 8.54, and 5.73 eV ($j=1,2,3$). The effective masses are assumed to be identical: $m_j=4m_0$, where m_0 is the rest mass of the electron. If $\tau=10^{-13}$ s, then the Fermi velocities for electrons with spin up (Fig. 1b) are $v_{Fj}=8.51\times10^5$, 8.67×10^5, and 7.1×10^5 m/s, and the corresponding mean free paths are much greater than the thicknesses of the layers.

It follows from (3.14)–(3.16) that the conductivities σ_{AF} and σ_{FM} decrease with increasing η and that the magnetoresistive effect increases to a certain maximum value and then decreases (Fig. 2). If all the relaxation times τ_{0j} are identical, the effect is not observed for ideal superlattice interfaces ($\eta=0$). Differences between the relaxation times in the layers produce a relatively small effect when $\eta=0$. In particular, if $\tau_1=3\times10^{-14}$ s, $\tau_2=10^{-14}$ s, and $\tau_3=5\times10^{-14}$ s, for $a=b=10$ Å the effect amounts to 3.2% when $\eta=0$, whereas it amounts to 37% in the case of $\eta=1$ Å. The greater is the relaxation time τ_{0j}, the stronger is the effect, but it decreases as the thicknesses of the layers decrease (Fig. 2). A similar numerical analysis for an Fe/Cr multilayer gives a dependence of the effect on η that coincides exactly with one of the results in Ref. 7.

5. NEW INTENSITY MAGNETO-OPTICAL EFFECT

As follows from (3.14) and (3.15), the dielectric constant (2.2) of the effective medium for the antiferromagnetic superlattice configuration has the form

$$\varepsilon_{\mathrm{AF}}=1-\frac{1}{a+b}\left[\frac{b\omega_{p1}^2\gamma_1}{\omega(\omega+i\omega_{\tau1})}+\frac{2a\omega_{p2}^2\gamma_2}{\omega(\omega+i\omega_{\tau2})}+\frac{b\omega_{p3}^2\gamma_3}{\omega(\omega+i\omega_{\tau3})}\right], \tag{5.1}$$

where

$$\omega_{pj}=\frac{ecm_j}{\pi}\sqrt{\frac{\mu_0 v_{Fj}^3}{6\hbar^3}}$$

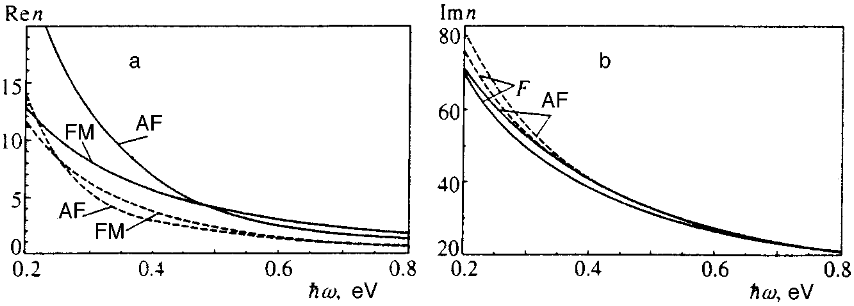


FIG. 3. Real (a) and imaginary (b) parts of the refractive indices (for the antiferromagnetic and ferromagnetic superlattice configurations) as functions of $\hbar\omega$; $\tau_{0j}=10^{-13}$ s, $\eta=2$ Å, $a=b=10$ Å (solid lines) and $a=b=20$ Å (dashed lines).

is the plasma frequency of layer j and $\omega_{\tau j}=\tau_{0j}^{-1}$. The parameters γ_j in (5.1), as in (3.15), are defined for electrons with the same spin orientation. In the case of the ferromagnetic configuration, the dielectric constant is given by the expression

$$\varepsilon_{\mathrm{FM}}=1-\frac{1}{a+b}\sum_{\uparrow\downarrow}\left[\frac{b\,\omega_{p1}^2\gamma_1}{\omega(\omega+i\omega_{\tau1})}+\frac{a\,\omega_{p2}^2\gamma_2}{\omega(\omega+i\omega_{\tau2})}\right]. \tag{5.2}$$

For a homogeneous nonmagnetic metal ($\xi_j=0$, $\eta=0$, $\gamma_j=1$, and $2\omega_{p1}^2=2\omega_{p2}^2=\omega_p^2$) both formulas transform into the familiar expression

$$\varepsilon=1-\frac{\omega_p^2}{\omega(\omega+i\omega_\tau)}.$$

The parameters γ_j in (5.1) and (5.2) are integrals, which depend on all the parameters of the superlattice and contain the frequency ω of the wave, the thicknesses of the magnetic (b) and nonmagnetic (a) layers, the interface roughness η, the effective masses of the electrons (m_1, m_2, and m_3), their mean free paths (l_1, l_2, and l_3), the relaxation times (τ_1, τ_2, and τ_3), and the potential-energy distribution for spin-up or spin-down electrons (Fig. 1b).

The difference between γ_j for the antiferromagnetic superlattice configuration and γ_j for the ferromagnetic configuration is the only reason why $\varepsilon_{\mathrm{AF}}\neq\varepsilon_{\mathrm{FM}}$. Thus, according to (2.3), the reflectivities for a plane wave with polarization along the magnetization of the multilayer (the intensity of the reflected wave) will also differ. This can be called a magneto-optical effect, which was apparently not previously considered. It is convenient to characterize this new effect by the change in reflectivity $(R_{\mathrm{FM}}-R_{\mathrm{AF}})/R_{\mathrm{FM}}$. The effect is caused mainly by the interface roughness of the multilayer. If there is no roughness, purely specular reflection of the electrons takes place, all the relaxation times are identical, the effect is not observed.

The dependence of the real and imaginary parts of the refractive indices of an Fe/Cu superlattice (for the antiferromagnetic and ferromagnetic configurations) on the energy $\hbar\omega$ is shown in Fig. 3. Plots of the dependence of the magneto-optical effect on $\hbar\omega$ and the roughness parameter η are shown in Fig. 4.

The maximum of the magneto-optical effect for different values of the parameters characterizing the multilayer is achieved at normal incidence. The magneto-optical effect is considered here in reference to the configuration which exhibits the Kerr magneto-optical effect, which consists of a change in the intensity of the reflected wave upon magnetization of the medium and is governed by off-diagonal elements of the magnetic permeability tensor. The off-diagonal elements of the dielectric tensor do not make any contribution. As was demonstrated experimentally in Refs. 16 and 18, the Kerr magneto-optical effect has a value of $\sim10^{-5}$, which is at least two orders of magnitude smaller than the new magneto-optical effect in magnetic multilayers exhibiting giant magnetoresistance. As the angle of incidence decreases, the Kerr effect decreases, and it vanishes at normal incidence, while the angular dependence of the new magneto-optical effect has the opposite character.

6. CONCLUSION

The change in the reflectivity of a metallic multilayer that exhibits giant magnetoresistance for a monochromatic plane wave of light at normal incidence and with polarization along the magnetization in the plane in response to a change

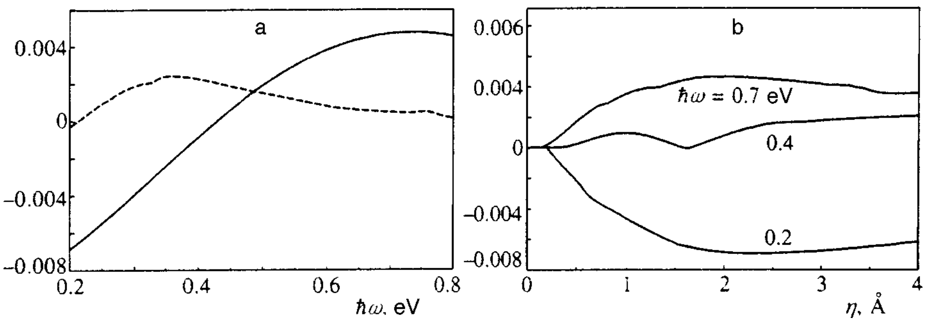


FIG. 4. Dependence of the magneto-optical effect on $\hbar\omega$ (a) for $\tau_{0j}=10^{-13}$ s, $\eta=2$ Å, and $a=b=10$ Å (solid line) and $a=b=20$ Å (dashed line) and on η (b) for $a=b=10$ Å and $\tau_{0j}=10^{-13}$ s.

from the antiferromagnetic magnetic configuration of the multilayer to the ferromagnetic configuration under the action of an external field has been investigated. This reflectivity change is a new magneto-optical effect since the definition of a ‘‘magneto-optical effect’’ includes any change in any characteristic of a wave in response to a change in the magnetization of the medium from which the wave is reflected. It has been shown in the example of an Fe/Cu multilayer that the effect (which is defined as the relative change in the reflectivity of a semi-infinite medium) amounts to $\sim 0.7\%$.

Because the mean free path significantly exceeds the thicknesses of the individual layers, the conventional approach, in which the dielectric tensor of each metal comprising the multilayer is assumed to be known, cannot be employed to study the effect. The possible, more correct analysis performed in the present work involves the use of the approximation of an effective medium, whose dielectric tensor is determined with consideration of the transport properties of the conduction electrons already in a magnetic superlattice, i.e., an infinite periodic medium consisting of alternating magnetic and nonmagnetic metallic layers. Then, to study the magneto-optical effect under consideration, it is sufficient to assume that the medium is semi-infinite.

Since the Fermi velocity is much smaller than the velocity of light, in the present work it was possible to determine the complex conductivity of the effective medium (corresponding to an assigned polarization of the wave) according to almost the same scheme which was previously used in the theory of giant magnetoresistance. The important factors here are: 1) the refraction law for electrons passing through an interface acting as a potential barrier and the existence of angles for total internal reflection; 2) the means for taking into account the interface roughness (modified Ziman–Soffer formulas) in the boundary conditions for the nonequilibrium distribution functions of the electrons in the layers; 3) the dependence of the potential energy of electrons on the orientation of their spins relative to the magnetization of the layers.

Utilization of the symmetry properties of the distribution functions within a superlattice period has enabled us to find analytical expressions for these functions, as well as for the currents induced by the wave field, which are convenient for numerical experiments. The analytical representations of the complex conductivity for the antiferromagnetic and ferromagnetic superlattice configurations (besides being needed to calculate the corresponding dielectric constants) can be used in a special case to study giant magnetoresistance. In particular, testing of the computer program developed demonstrated complete agreement with the known theoretical results for an Fe/Cr multilayer.[7]

It has been shown within the theoretical model used that the interface roughness is mainly responsible for both the intensity magneto-optical effect and giant magnetoresistance. If the interfaces are ideal (if there is purely specular reflection of electrons) and the relaxation times in all the layers are identical, the effects are not observed.

This work was supported by Ministry for the Science and Technical Policy of Russia under the Physics of Solid-State Nanostructures Program (Project No. 97-1071) and by the Russian Fund for Fundamental Research (Grants 96-02-16520 and 96-02-18253).

---

*)E-mail: kubrakov@orc.ru

†)E-mail: Ronald.Atkinson@qub.ac.uk

1)It differs from the known Kerr magneto-optical effects, including the intensity effects,[16] in that the conventional effects (see, for example, Refs. 16–18) are determined by the magneto-optical parameter Q (for an optically isotropic medium $Q=\varepsilon_{xy}/\varepsilon_{xx}$, where ε_{xx} and ε_{xy} are, respectively, diagonal and off-diagonal elements of the dielectric tensor $\hat{\varepsilon}$), while the effect discussed in this paper does not depend on Q (it is manifested in the zeroth order with respect to Q). It can also be stated that the conventional Kerr magneto-optical effects are a consequence of the magnetic gyrotropy of the medium, whereas the effect considered in this paper is not gyrotropic, i.e., it is determined by the diagonal elements of $\hat{\varepsilon}$. It also clearly differs from the quadratic (anisotropic) magneto-optical effects manifested in the second order with respect to Q.

2)The spatial dispersion will be ignored. This is justifiable, since the wavelength of the electromagnetic wave significantly exceeds the thicknesses of the individual layers.

3)Here we shall disregard the effects associated with the existence of surface magnetic structures of the fan type.[9,21,22]

[1] M. N. Babich, J. M. Bruno, A. Fert *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).

[2] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. B **39**, 4828 (1989).

[3] S. S. P. Parkin, Appl. Phys. Lett. **61**, 1358 (1992).

[4] S. Zhang and P. M. Levy, Phys. Rev. B **43**, 11 048 (1991).

[5] P. F. Garcia and A. Suna, J. Appl. Phys. **54**, 20 000 (1983).

[6] R. E. Camly and J. Barnas, Phys. Rev. Lett. **63**, 664 (1989).

[7] R. Q. Hood and L. M. Falikov, Phys. Rev. B **46**, 8287 (1992).

[8] J. Barnas, A. Fuss, R. E. Camly *et al.*, Phys. Rev. B **42**, 8110 (1990).

[9] A. K. Zvezdin, S. N. Utochkin, JETP Lett. **57**, 433 (1993).

[10] A. Vedyaev, C. Cowache, N. Ryzhanova, and B. Dieny, J. Phys.: Condens. Matter **5**, 8289 (1993).

[11] V. V. Ustinov and E. A. Kravtsov, J. Phys.: Condens. Matter **7**, 3471 (1995).

[12] X.-G. Zhang and W. H. Butler, Phys. Rev. B **51**, 10 085 (1995).

[13] P. M. Levy, in *Solid State Physics, Vol. 47*, Academic Press, New York (1994).

[14] E. H. Sondheimer, Adv. Phys. **1**, 1 (1952).

[15] R. Atkinson, N. F. Kubrakov, A. K. Zvezdin, and K. A. Zvezdin, J. Magn. Magn. Mater. **156**, 169 (1996).

[16] G. S. Krinchik, *Physics of Magnetic Phenomena* [in Russian], Mosk. Gos. Univ., Moscow (1985).

[17] A. K. Zvezdin and V. A. Kotov, *Magneto-optics of Thin Films* [in Russian], Nauka, Moscow (1988).

[18] A. Zvezdin and V. Kotov, *Modern Magnetooptics and Magnetooptical Materials*, Institute of Physics, Bristol–Philadelphia (1997).

[19] R. Atkinson and N. F. Kubrakov, Proc. R. Soc. London, Ser. A **449**, 205 (1995).

[20] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam (1977).

[21] A. K. Zvezdin and S. N. Utochkin, J. Magn. Magn. Mater. **140–144**, 787 (1995).

[22] R. Atkinson, N. F. Kubrakov, S. N. Utochkin, and A. K. Zvezdin, J. Appl. Phys. **75**, 6786 (1994).

[23] *The Physics of Metals, Vol. 1: Electrons*, J. M. Ziman (ed.), Cambridge University Press, Cambridge (1969) [Russ. transl., Mir, Moscow (1972)].

[24] A. A. Abrikosov, *Introduction to the Theory of Normal Metals (Solid State Physics, Suppl. 12)*, Academic Press, New York (1972) [Russ. transl., Nauka, Moscow (1972)].

[25] V. Bezak, M. Kedro, and A. Pevala, Thin Solid Films **23**, 305 (1974).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Energy relaxation and transport of indirect excitons in AlAs/GaAs coupled quantum wells in magnetic field

L. V. Butov*) and A. I. Filin

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

(Submitted 16 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 1115–1120 (September 1998)

The evolution of indirect exciton luminescence in AlAs/GaAs coupled quantum wells after excitation by pulsed laser radiation has been studied in strong magnetic fields ($B \leqslant 12$ T) at low temperatures ($T \geqslant 1.3$ K), both in the normal regime and under conditions of anomalously fast exciton transport, which is an indication of the onset of exciton superfluidity. The energy relaxation rate of indirect excitons measured in the range of relaxation times between several and several hundreds of nanoseconds is found to be controlled by the properties of the exciton transport, specifically, this parameter increases with the coefficient of excitonic diffusion. This behavior is qualitatively explained in terms of migration of excitons between local minima of the random potential in the plane of the quantum well. *© 1998 American Institute of Physics.* [S1063-7761(98)02509-8]

Owing to their long lifetime, indirect excitons in coupled quantum wells (CQWs) can cool down to temperatures close to that of the crystal lattice. Therefore, condensation of excitons similar to the Bose–Einstein condensation is possible at low lattice temperatures.[1] It has been shown in some theoretical investigations that the critical conditions for exciton condensation can be greatly improved by applying a magnetic field perpendicular to the quantum well plane, mainly because of the full quantization of the electron and hole energy spectra[2] and lifting of the spin degeneracy. In recent investigations of indirect excitons contained in AlAs/GaAs CQWs, effects indicating condensation of excitons in strong magnetic fields have been detected, namely, an anomalous increase in the diffusion coefficient[3] and radiative decay rate[4] of excitons, which have been interpreted in terms of excitonic superfluidity and superluminescence of the excitonic condensate, together with anomalously large fluctuations in the total intensity of exciton photoluminescence interpreted as critical fluctuations in the region of the phase transition associated with instability of condensate domains.[5]

In the reported work, we have investigated the evolution of indirect exciton luminescence in AlAs/GaAs CQWs after termination of a laser pulse in both the normal regime and the regime of anomalously fast transport and high radiative decay rate of excitons, i.e., in the regime of the suggested condensation of excitons. We have studied the energy relaxation of indirect excitons in the time interval from several nanoseconds to several hundreds of nanoseconds and its relation to exciton transport.

The n^+-i-n^+ heterostructures with AlAs/GaAs CQWs tuned by a gate voltage V_g are similar to those studied in earlier experiments.[3–5] The *i*-layer consists of two quantum wells: an AlAs well with thickness 40 Å and a GaAs well 30 Å thick between $Al_{0.48}Ga_{0.52}As$ barriers. In the indirect regime ($V_g \lesssim 0.5$ V), electrons are confined in the AlAs quantum well and holes in the GaAs quantum well (Fig. 1a). The electron state in AlAs is constructed from states of the X_z minima in the conduction band, which ensures that the indirect excitons have long lifetimes. Carriers were generated in the GaAs well by a semiconductor laser ($\hbar\omega = 1.85$ eV) operated in the pulsed mode. The laser pulse had an approximately rectangular shape with a duration of 50 ns and rise and decay times of about 1 ns. The time resolution of the light detecting system was 0.3 ns.

Photoluminescence decay in magnetic fields $B=0$, 4, 6, and 12 T at $V_g=0$ and $T=1.3$ K is illustrated by Fig. 2a (the signal was detected in a spectral range with a width of 3 meV centered at the photoluminescence line peak). These curves are similar to those measured in previous experiments.[3–5] In the indirect regime, the radiative lifetime of excitons is much longer than the nonradiative time τ_{nr}, and the total lifetime $\tau \approx \tau_{nr}$.[5] In narrow CQWs $\tau_{\rm nr}$ is determined by the exciton transport toward the centers of nonradiative recombination.[6] Direct time-of-flight measurements of exciton transport in the AlAs/GaAs CQWs studied[3,4] indicate that an increase (decrease) in τ really corresponds to a decrease (increase) in indirect exciton diffusion coefficient.

Thus, in magnetic fields $B \lesssim 7$ T the diffusion coefficient monotonically increases with the temperature and decreases with the increasing delay time and magnetic field (Fig. 2a and 2c). This behavior can be well described in terms of one-exciton transport in random potential (the random potential in narrow quantum wells is largely controlled by interface roughness): (i) the increase in the diffusion coefficient with the temperature is due to thermal activation of excitons from local potential minima; (ii) the diffusion coefficient drops with the increasing delay since more and more strongly localized excitons dominate in the luminescence spectra (those which have not had enough time to travel to centers of nonradiative recombination and recombine there); (iii) the decrease in the diffusion coefficient with the increas-

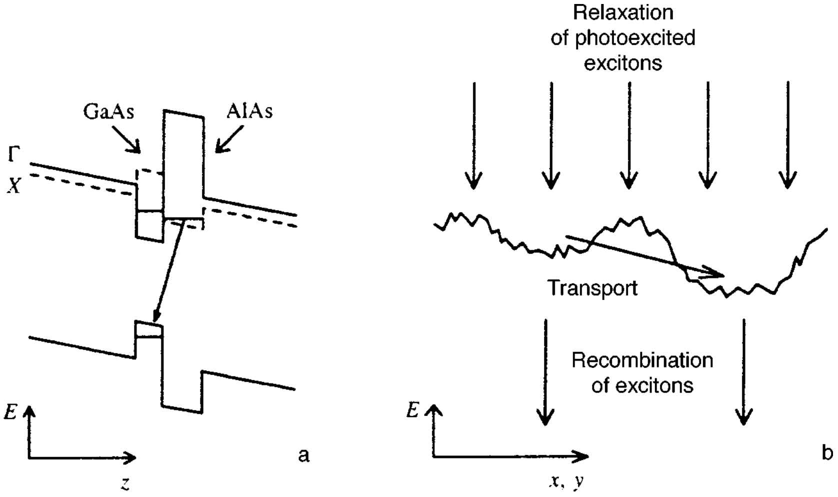


FIG. 1. (a) Band diagram of AlAs/GaAs CQWs; the solid (dashed) line represents energies of Γ (X) extrema in the Brillouin zone. (b) Scheme of a photogenerated exciton energy relaxation in the presence of random potential in the CQW plane.

ing field can be qualitatively explained by the increase in the magneto-exciton mass.[7]

For $B \lesssim 7$ T and $T \lesssim 5$ K, we have detected anomalous increase in the diffusion coefficient with the increasing magnetic field and a decrease in the diffusion coefficient with the increasing temperature at initial decay times (Fig. 2a and 2c). This behavior can not be explained in terms of one-exciton transport and is interpreted as the onset of exciton superfluidity owing to their condensation.[3] The fast decay of the exciton photoluminescence, which corresponds to fast exciton transport, is observed until the exciton density drops severalfold; the subsequent decay is slow and corresponds to slow exciton transport (Fig. 2a). The transition from the initially fast to the subsequent slower transport corresponds to elimination of excitonic superfluidity when the exciton density drops below the critical value, which is determined by the temperature, magnetic field, and random potential. In the general case of excitonic condensation in the presence of a random potential, domains of condensed (superfluid) excitons alternate with normal regions in a random manner, and the boundaries between these regions are determined by the potential relief. A superfluid domain may include several microdomains connected by weak bonds, which should lead to a coherence across the entire domain (it is an analogue of a network of the Josephson junctions in superconductors). Measurements of the photoluminescence decay time yield the parameters of the exciton transport averaged over normal and superfluid regions in the laser excitation spot.

Evolution of indirect exciton photoluminescence spectra with time and spectra integrated with respect to time at typical values of temperature and magnetic field are shown in Fig. 3. In integrated spectra, one can see a line of lower

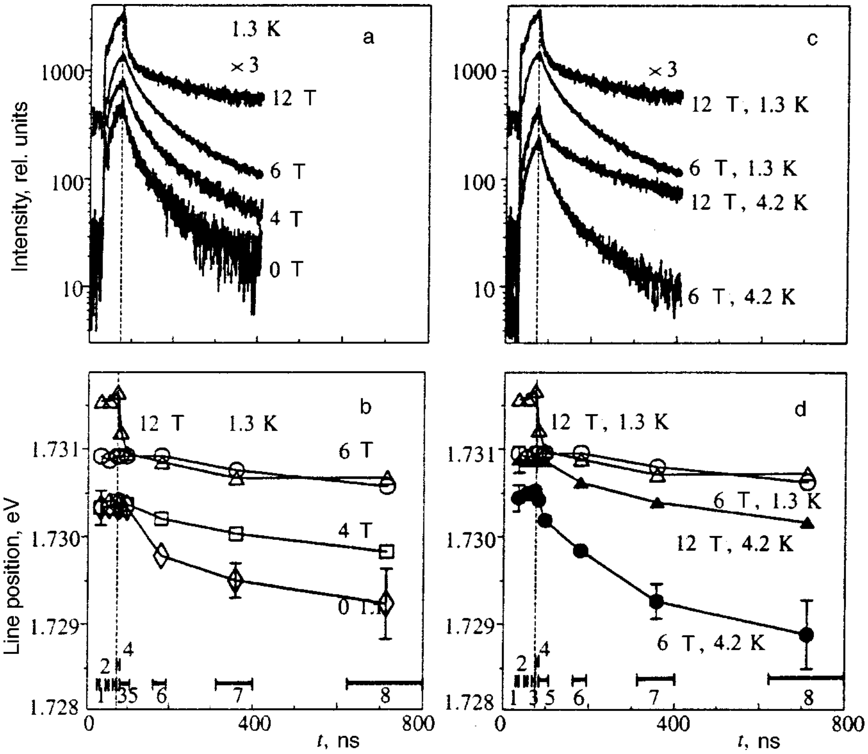


FIG. 2. (a,c) Decay and (b,d) photoluminescence line positions of indirect excitons at $V_g=0$, $B=0$, 4, 6, and 12 T, $T=1.3$ and 4.2 K. The dashed vertical line shows the position of the 50-ns laser pulse end. The positions of the luminescence line (b,d) were derived from spectra measured in time intervals 1–8 shown at the bottom of the graph.

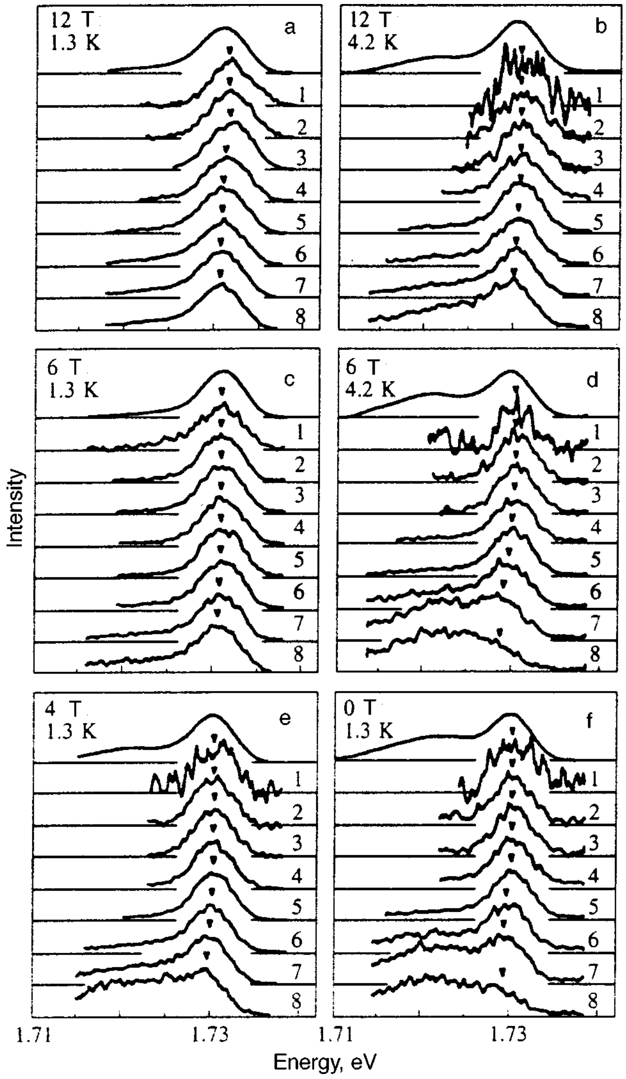


FIG. 3. Evolution of photoluminescence spectra of indirect excitons with time. Spectra 1–8 were recorded in time intervals shown in Fig. 2b and 2d. The positions of the photoluminescence line of mobile excitons are marked by ticks. Spectra integrated over time are shown at the tops. The spectra are normalized so that have almost the same peak intensities.

intensity on the low-energy side of the dominant line (Fig. 3b and 3d–f). The magnitudes of the shifts of these lines toward lower energy with the gate voltage V_g indicate that both of them are due to recombination of indirect excitons. The relative intensity of the low-energy line increases with the delay time (Fig. 3). Therefore, we associate this line with the recombination of strongly localized indirect excitons, whose nonradiative recombination rate is small owing to their lower mobility (it is likely to be smaller than their radiative recombination rate). The high-energy line is due to recombination of indirect excitons with a higher mobility (even though the localization radius of such excitons is finite, we dub them mobile for definiteness). These are the excitons which demonstrate the variations in the diffusion coefficient and radiative decay rate[4] with the magnetic field and temperature discussed above.

The observed changes in the ratio I_m/I_l between the intensities of the mobile and strongly localized exciton luminescence lines with the magnetic field and temperature are related to variations in both the diffusion coefficient and radiative decay rate of mobile excitons. For example, in fields below 7 T the radiative decay rate of mobile excitons drops with the increasing temperature,[4] whereas the diffusion coefficient grows (Fig. 2c). Both these effects lead to a decrease in I_m/I_l (Fig. 3c and 3d). This effect of temperature indicates that the distribution of excitons between localized and mobile states is nonequilibrium. The effect would have an opposite sign if there were equilibrium between these states, namely, the relative intensity of higher energy mobile excitons would increase with temperature. The time of experiment (about 1 μs), which is limited by the necessity of detecting the decaying luminescence signal, is insufficient for establishing equilibrium between the mobile and localized indirect excitons.

As the delay time increases, one can see a shift of the mobile exciton luminescence line, which reflects energy relaxation of excitons (Fig. 2b and 2d and Fig. 3). In order to determine the spectral position the line shape was approximated by a Gaussian. In magnetic fields $B \lesssim 7$ T, the energy relaxation rate of mobile excitons gradually drops with the delay; in addition, it decreases with the field and grows with the temperature at all delay times (Fig. 2b and 2d). In magnetic fields $B \gtrsim 7$ T the energy relaxation rate of mobile excitons is considerably higher at smaller delays; in addition, it increases with the magnetic field and drops with the temperature at small delays, whereas it decreases with the field and rises with the temperature at longer delays (Fig. 2b and 2d). Thus, the energy relaxation of mobile excitons is faster, the higher the diffusion coefficient throughout the entire range of experimental parameters studied (magnetic field, temperature, and delay).

After photoexcitation, electron–hole pairs are rapidly bound in excitons and lose their kinetic energy by emitting phonons. In this ''fast'' relaxation stage, $\Gamma - X$ electron transfer from GaAs to AlAs takes place and indirect excitons are formed. The times of these processes are considerably shorter than the lifetime of indirect excitons, and they are not revealed in our measurements. The ''slow'' energy relaxation, which is observed in the range between several nanoseconds and hundreds of nanoseconds (Fig. 2b and 2d), is controlled by excitons migrating between local minima of the random potential in the plane of the CQWs (Fig. 1b). This migration of excitons in the random potential is characterized by a large spread of relaxation times,[8] and the long lifetime of indirect excitons allows us to trace the transport and energy relaxation of excitons in the range of long delay times. The speed of exciton transport to deeper local minima increases with the their mobility, which leads to a faster energy relaxation of excitons. The relationship between the energy relaxation and exciton transport persists with the onset of exciton superfluidity, when superfluid domains of excitonic condensate are formed at some local minima.

The diffusion coefficient and radiative decay rate of indirect excitons are sensitive to the bath temperature down to 1 K, which indicates that excitons in local minima thermalize down to ~1 K. The recombination line of indirect excitons of each local minimum cannot be much wider than the exciton temperature ($\sim$1 K), so the measured line width of about 50 K (Fig. 3) indicates that the laser excitation spot covers a lot of local minima, and the distribution of excitons among them is nonequilibrium. The measurement time (about one microsecond) is insufficient for establishing equilibrium among all local minima of the random potential. The condensation of excitons should result in a narrowing of luminescence line from the condensate domain owing to the macroscopic filling of state with the lowest energy. In our experiments, we have not detected a narrowing of the exciton luminescence line in strong magnetic fields and at low temperatures when the diffusion coefficient and radiative decay rate of indirect excitons increased anomalously (Fig. 3). This can be explained by the presence of a large number of condensate domains in the laser excitation spot with a nonequilibrium distribution of excitons among them.

We are grateful to G. Abstreiter, G. Böhm, G. Weimann, M. Hagn, and A. Zrenner for their contribution on the earlier stage of studies of indirect excitons in CQWs, and to V. D. Kulakovskiĭ and S. G. Tikhodeev for helpful discussions. The financial support from the Russian Fund for Fundamental Research and *Physics of Solid-State Nanostructures* program is gratefully acknowledged.

---

1)E-mail: butov@issp.ac.ru

[1] Yu. E. Lozovik and V. I. Yudson, Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)]; S. I. Shevchenko, Fiz. Nizk. Temp. **2**, 505 (1976) [Sov. J. Low Temp. Phys. **2**, 251 (1976)]; T. Fukuzawa, S. S. Kano, T. K. Gustafson, and T. Ogawa, Surf. Sci. **228**, 482 (1990); X. Zhu, P. B. Littlewood, M. S. Hybersten, and T. M. Rice, Phys. Rev. Lett. **74**, 1633 (1995).

[2] Y. Kuramoto and C. Horie, Solid State Commun. **25**, 713 (1978); I. V. Lerner and Yu. E. Lozovik, J. Low Temp. Phys. **38**, 333 (1980); I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)].

[3] L. V. Butov, A. Zrenner, M. Hagn, G. Abstreiter, G. Böhm, and G. Weimann, Surf. Sci. **361/362**, 2434 (1996).

[4] L. V. Butov, in *Proceedings of 23th Int. Conf. on the Physics of Semiconductors*, ed. by M. Scheffler and R. Zimmermann, World Scientific, Singapore (1996), p. 1927; L. V. Butov, Usp. Fiz. Nauk **168**(2), 127 (1998).

[5] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).

[6] F. Minami, K. Hirata, K. Era, T. Yao, and Y. Masumoto, Phys. Rev. B **36**, 2875 (1987); M. Maaref, F. F. Charfi, D. Scalbert, C. Benoit a la Guillaume, and R. Planel, Phys. Status Solidi B **170**, 637 (1992); G. D. Gilliland, A. Antonelli, D. J. Wolford, K. K. Bajaj, J. Klem, and J. A. Bradley, Phys. Rev. Lett. **71**, 3717 (1993).

[7] A. B. Dzyubenko and G. E. W. Bauer, Phys. Rev. B **51**, 14524 (1995).

[8] T. Takagahara, Phys. Rev. B **31**, 6552 (1985).

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Electrical characteristics of three-component dielectric media

Yu. P. Emets

Institute of Electrodynamics, National Academy of Sciences of Ukraine, 252680 Kiev, Ukraine

(Submitted 29 January 1997)
Zh. Éksp. Teor. Fiz. **114**, 1121–1136 (September 1998)

We study the effective parameters of a three-component plane dielectric medium with a doubly periodic arrangement of circular inclusions. The problem is solved in the one-dipole approximation. The computational results are compared with a two-component Rayleigh model. The general structure of the formulas for the effective parameters is discussed and the reciprocity relations for a three-component matrix system are determined. Explicit expressions are given for the dielectric permittivity when the concentration of circular inclusions is low, and their domain of applicability is determined. Under certain polarization conditions, the effective conductivity in a three-component medium is exactly equal to the dielectric permittivity of the matrix. © *1998 American Institute of Physics.* [S1063-7761(98)02609-2]

1. INTRODUCTION

Of the many problems that arise during the study of physical transport processes in inhomogeneous media, determining the macroscopic (effective) characteristics and parameters occupies a central place; it is simple to formulate and is often invoked to account for various phenomena in composite materials. In solving this problem, however, one encounters serious mathematical difficulties that must be overcome when calculating and averaging inhomogeneous physical fields. The situation is especially complicated when calculating three-dimensional fields in multicomponent systems. two-dimensional systems are in a more favorable position, especially two-component matrix media, which can be studied using the powerful mathematical machinery of complex variable theory. A number of important results have thus been obtained for two-component systems, including the Keller–Mendelson theorem for the reciprocity relations,[1,2] the Dykhne symmetry transform,[3] and exact solutions of boundary value problems for several matrix systems with a doubly periodic arrangement of regularly shaped inclusions.[4–6]

Multicomponent systems in which the main phase contains, not one, but several forms of inclusions with differing physical properties and dimensions have yet to be studied adequately. Multiphase systems can have rather complicated structure, and their macroscopic characteristics are more diverse. The difference between the structures of two-component and multicomponent systems is in some respects similar to the difference between black-and-white and color images of the textures of real multicolored materials.

In this paper we determine the effective parameters of a three-component dielectric medium. A matrix system with a doubly periodic arrangement of two types of parallel cylinders is considered. In transverse cross section, the system is two-dimensional and can be subdivided arbitrarily into square cells, with periodically alternating circular inclusions of two different types and sizes at their centers.

Periodic lattice models were introduced by Rayleigh to calculate the refractive index of highly nonuniform media.[7] In his model, circular cylinders of one kind were arranged doubly periodically at the centers of rectangular cells; in the three-dimensional analog of this system, spherical inclusions were placed in cubic cells. For systems with this sort of structure it is possible to formulate a boundary value problem for a harmonic equation and completely calculate the field in a single cell with a single isolated inclusion, taking the effect of all surrounding inclusions into account. The method proposed by Rayleigh yields formulas for the effective parameters of the system he examined to any required accuracy, and this is important for evaluating the various approximations. Unfortunately, this method cannot be applied directly to multicomponent systems. Calculating the fields in these systems involves the formulation of complicated matching boundary value problems in multiply-connected regions, requiring new approaches for their solution.

The solution for two dielectric cylinders located in an external uniform electric field is invoked below to calculate the characteristics of a three-component matrix system. This problem has an exact analytic solution under general assumptions about the radii of the cylinders, their respective locations in the external field, and the relationship between the dielectric permittivities of the matrix and the inclusion materials.[8] These results show that the mutual effect of the inclusions in a multicomponent system can be taken into account through dipole–dipole interactions. In the following sections, we investigate an inhomogeneous medium with a low concentration of inclusions; the calculations can be simplified by restricting attention to the one-dipole approximation. In this case, of the infinite number of dipoles determining the electric field of the interacting inclusions, only the first dipoles located at the center of circular inclusions are taken into account. The accuracy of the calculations is then evaluated in terms of the small parameters characterizing the system and by comparison with calculations based on the Rayleigh formula under equivalent conditions.

The expressions for the dielectric permittivity of the system under study can be generalized naturally to a quasistationary electric field by introducing the complex dielectric permittivity. The physical interpretation of the results becomes more convincing in this case upon examination of equivalent electrical circuits with periodically alternating complex impedances.

The general properties of three-component systems are similar in many respects to the properties inherent in two-component systems. They also obey reciprocity relations, which now take a form reflecting their dependence on two additional phases. In addition, three-component dielectric media can acquire characteristics that do not exist in two-component materials.

Of the many possible ways of arranging circular inclusions in the plane, an isotropic system with a square structure is studied below. The method for calculating the fields used here, however, is entirely applicable to systems with other structures, such as hexagonal.

The solution is presented in terms of the dielectric permittivity. In light of the well known analogies, these results can also be used to examine the macroscopic characteristics of other physical media.

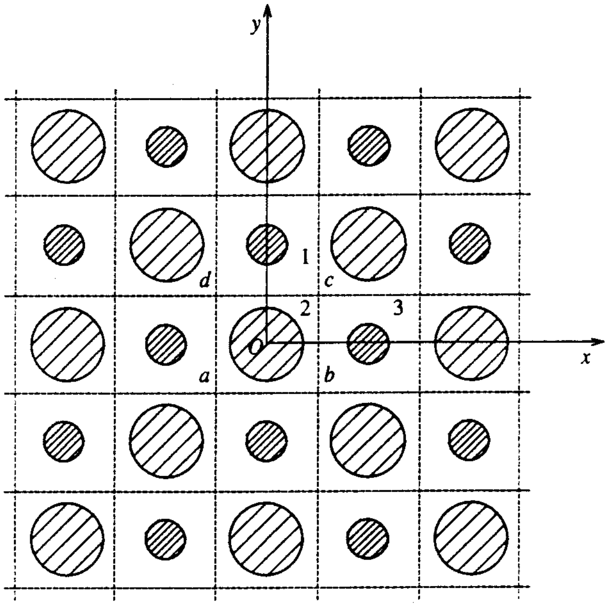


FIG. 1.

2. THE ELECTRIC FIELD IN THE SYSTEM

Let two types of circular dielectric cylinders with permittivities ε_2 and ε_3 and radii r_1 and r_2, respectively, be located in doubly periodic alternating order in an unbounded dielectric medium with permittivity ε_1. In transverse cross section, the long unidirectional cylinders form a plane system with circular inclusions located at the centers of square cells with sides of length l (Fig. 1). Systems of this type serve as a convenient theoretical model for the macroscopic characteristics of various multicomponent heterogeneous structures, for example, periodic lattices, composite materials, thin films with topological structure, and so on. Above all, systems with a low concentration of heterogeneous inclusions ($\pi r_1^2, \pi r_2^2 \ll l^2$) are of practical interest. In terms of computation, they are simpler to study, but they do preserve many of the basic properties inherent in three-component media.

To determine the effective dielectric constant of the system under study, it is first necessary to calculate the field inside the inclusions and in the matrix medium. Because of the regular structure of the system, the field pattern repeats doubly periodically; thus, it suffices to calculate the field in one cell with two individual, dissimilar inclusions. Here it must be borne in mind that the formation of an electric field in each inclusion depends on the presence of all other inclusions in the system. This interdependence of the inclusions can be represented by the sum of the pairwise interaction of an individual inclusion with every other inclusion in the system. To determine the electric field in the system, it is therefore necessary to solve for the interaction of two dissimilar dielectric cylinders in an external electric field. This is the key problem in calculating the field. The solution is known,[8] and is given in the Appendix for reference. According to Eq. (A1), the electric field is given by an infinite sum of linear dipoles, whose moments and positions depend on the radii of the cylinders, the distance between them, their properties, and the external electric field in the system.

Given the above assumption of a low concentration of inclusions, we can use an approximate expression for the electric field that accounts only for the first, principal dipoles located at the inclusion centers. These dipoles have the largest moments and are the main contributors to the electric field at low concentrations of the inclusions ($\pi r_1^2, \pi r_2^2 \ll l^2$). If a square cell $abcd$ containing an inclusion with dielectric constant ε_2 are radius r_1 is singled out in Fig. 1, then according to Eqs. (A1), in the approximation taken here, the electric field in the cell will be

$$E_1(z)=E_0-\bar{E}_0\left\{\Delta_{12}r_1^2z^{-2}+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}[\Delta_{12}r_1^2 \times(z-a_{mn})^{-2}+\Delta_{13}r_2^2(z-b_{mn})^{-2}]\right\} \tag{1}$$

outside the inclusion and

$$E_2(z)=(1+\Delta_{12})\left\{E_0-\bar{E}_0\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}[\Delta_{12}r_1^2(z-a_{mn})^{-2} +\Delta_{13}r_2^2(z-b_{mn})^{-2}]\right\} \tag{2}$$

inside. Here $E_k(z)=E_{kx}-iE_{ky}$ ($k=1,2$) is the complex electric field function, $z=x+iy$ is a complex variable, $E_0=E_{0x}-iE_{0y}$ is the external uniform electric field, the bar over E_0 denotes the complex conjugate, Δ_{1p} is the relative dielectric constant,

$$\Delta_{1p}=\frac{\varepsilon_1-\varepsilon_p}{\varepsilon_1+\varepsilon_p}\quad(-1\leqslant\Delta_{1p}\leqslant 1)\quad p=2,3, \tag{3}$$

a_{mn} are the coordinates of the dipoles in all inclusions of the same type with parameters ε_2 and r_1 except the one located in the cell $abcd$ centered at the origin, and b_{mn} are the coordinates of the dipoles in all the inclusions of the other type, with parameters ε_3 and r_2 (see Fig. 1). The coordinates a_{mn} and b_{mn} are also the coordinates of the centers of the circular inclusions in the system. We have

$$a_{mn}=\pm 2ml, \quad b_{mn}=\pm l(2m-1)$$

on the x axis,

$$a_{mn}=\pm 2inl, \quad b_{mn}=\pm il(2n-1)$$

on the y axis, and

$$\begin{gathered}
a_{mn}=\pm 2l(m+in), \quad b_{mn}=\pm l[2m+i(2n-1)],\\
a_{mn}=\pm 2l(m-in), \quad b_{mn}=\pm l[2m-i(2n-1)],\\
a_{mn}=\pm l[2m-1+i(2n-1)], \quad b_{mn}=\pm l[2m-1+2in],\\
a_{mn}=\pm l[2m-1-i(2n-1)], \quad b_{mn}=\pm l[2m-1-2in],\\
m,n=1,2,... \qquad (4)
\end{gathered}$$

off the axes.

If the electric field is to be determined in a cell with an inclusion that has parameters ε_3 and r_2, then the same Eqs. (1) and (2) can be used with the replacements $E_2(z)\rightarrow E_3(z)$, $\Delta_{12}\leftrightarrow\Delta_{13}$, $r_1\leftrightarrow r_2$, and $a_{mn}\leftrightarrow b_{mn}$. Here the coordinate origin has been shifted to the center of the inclusion in the newly selected cell.

Equations (1) and (2) yield a first approximation for the interaction of inclusions in the system. If there is no interaction among them, then the double sums in Eqs. (1) and (2) must be dropped, and the electric field at the inclusions will then be uniform and, for each type

$$E_2(z)=(1+\Delta_{12})E_0, \quad E_3(z)=(1+\Delta_{13})E_0,$$

as appropriate for isolated cylinders in an external, uniform field.

When the concentration of the additional phases increases, the mutual effect of inclusions becomes greater and the expressions for the electric field must account for the next dipoles, refining the interaction intensity of the inclusions. The number of dipoles to be additionally included depends on the required accuracy. Fundamentally, this approach to refining the calculations does not pose any problems, but the expressions for the electric field become messier, and the computational load increases significantly.

Since the interaction among inclusions is proportional to the inverse square of their separation, in practical calculations the infinite sums in Eqs. (1) and (2) can be replaced by finite, low-order sums. The interaction among widely separated inclusions can thereby be neglected.

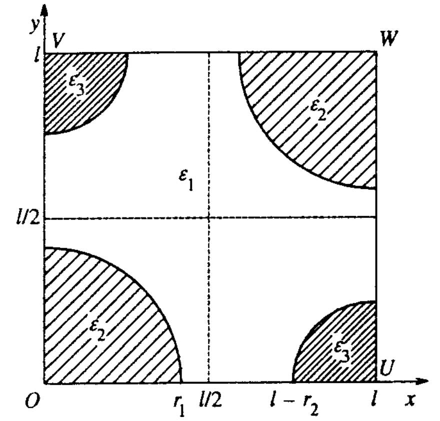


FIG. 2.

3. EFFECTIVE DIELECTRIC CONSTANT

By definition, the effective dielectric constant relates the displacement and electric field vectors averaged over the region, i.e.,

$$\langle\mathbf{D}\rangle=\varepsilon_{\text{eff}}\langle\mathbf{E}\rangle. \tag{5}$$

The averages are taken over a region whose characteristic dimensions are equal to or greater than those of the system.

For the medium with a periodic structure being considered here, it is sufficient to calculate the average fields in a single square cell of size l. It is convenient, for example, to choose the unit cell $OUWV$ shown in Fig. 2. If the external electric field vector is directed along the x axis, i.e., $E_0=E_{0x}$, then the segments OV and UW coincide with equipotentials and the segments OU and VW lie on lines of force. This property of the cell $OUWV$ makes it possible to reduce the averaging procedure to a determination of the average fields by calculating the corresponding contour integrals on the segments OU and OV:

$$\begin{aligned}
\langle E_x\rangle&=\frac{1}{l}\left[\int_0^{r_1}E_{x2}(x)dx+\int_{r_1}^{l/2}E_{x1}(x)dx+\int_{l/2}^{r_2}E_{x1}(x)dx\right.\\
&\quad\left.+\int_{r_2}^{l}E_{x3}(x)dx\right],\\
\langle D_x\rangle&=\frac{\varepsilon_1}{l}\left[\frac{\varepsilon_2}{\varepsilon_1}\int_0^{r_1}E_{x2}(y)dy+\int_{r_1}^{l/2}E_{x1}(y)dy\right.\\
&\quad\left.+\int_{l/2}^{r_2}E_{x1}(y)dy+\frac{\varepsilon_3}{\varepsilon_1}\int_{r_2}^{l}E_{x3}(y)dy\right].
\end{aligned} \tag{6}$$

In calculating the integrals of Eqs. (6), recall that the electric field $E_1(z)$ in the matrix is given by different expressions in regions abutting dissimilar cells; this was discussed in Sec. 2.

Finally, after calculating the integrals of Eq. (6) using Eqs. (1), (2), and (4), we obtain the following expression for the effective dielectric constant of the three-component system under consideration:

$$\varepsilon_{\text{eff}}=\varepsilon_1\frac{1-\Delta_{12}s_1/2-\Delta_{13}s_2/2+\Delta_{12}^2A_1+\Delta_{13}^2A_2+\Delta_{12}\Delta_{13}(B_1+B_2)}{1+\Delta_{12}s_1/2+\Delta_{13}s_2/2+\Delta_{12}^2A_1+\Delta_{13}^2A_2+\Delta_{12}\Delta_{13}(B_1+B_2)}. \tag{7}$$

Here

$$s_k=\frac{\pi r_k^2}{l^2}=\pi r_{k*}^2,\quad r_{k*}=\frac{r_k}{l}\quad (k=1,2) \tag{8}$$

are the concentrations of the inclusions with dielectric constants ε_2 and ε_3, respectively. The parameters A_k and B_k $(k=1,2)$ are functions of the radii of the inclusions,

$$\begin{aligned}A_k=2r_k^2\Bigg\{&2r_k^3\sum_{m=1}^{\infty}\frac{1}{r_k^4-16m^4}+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\Bigg[\frac{r_k-2m}{(r_k-2m)^2+4n^2}\\&+\frac{r_k+2m}{(r_k+2m)^2+4n^2}+\frac{r_k-2m+1}{(r_k-2m+1)^2+(2n-1)^2}\\&+\frac{r_k+2m-1}{(r_k+2m-1)^2+(2n-1)^2}\Bigg]\Bigg\}.\\B_k=2r_{3-k}^2\Bigg\{&2r_k^3\sum_{m=1}^{\infty}\frac{1}{r_k^4-(2m-1)^4}\\&+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\Bigg[\frac{r_k-2m}{(r_k-2m)^2+(2n-1)^2}\\&+\frac{r_k+2m}{(r_k+2m)^2+(2n-1)^2}+\frac{r_k-2m+1}{(r_k-2m+1)^2+4n^2}\\&+\frac{r_k+2m-1}{(r_k+2m-1)^2+4n^2}\Bigg]\Bigg\}.\end{aligned} \tag{9}$$

In Eqs. (9) the radii of the inclusions can be written in relative units, $r_{r*}=r_k/l$, with the asterisks omitted for brevity. Using Eqs. (8), A_k and B_k can be represented as functions of the concentrations s_1 and s_2 of the inclusions. In actual calculations the infinite sums can be replaced by finite sums, and sufficient accuracy for practical calculations is ensured by including a relatively small number of terms $(m,n=10-100)$. The coefficients of the first-order parameters Δ_{12} and Δ_{13} in Eq. (7) can be determined when calculating the corresponding integrals in Eq. (6), which yields

$$A_{0k}=r_k^2(2+A_{11}+B_{11}),\quad k=1,2, \tag{10}$$

where

$$\begin{aligned}A_{11}=4\Bigg\{&2\sum_{m=1}^{\infty}\frac{1}{1-256m^4}+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\Bigg[\frac{1-4m}{(1-4m)^2+16n^2}\\&+\frac{1+4m}{(1+4m)^2+16n^2}+\frac{1-2(2m-1)}{[1-2(2m-1)]^2+4(2n-1)^2}\\&+\frac{1+2(2m-1)}{[1+2(2m-1)]^2+4(2n-1)^2}\Bigg]\Bigg\}=0.09644,\\B_{11}=4\Bigg\{&2\sum_{m=1}^{\infty}\frac{1}{1-16(2m-1)^4}\\&+\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\Bigg[\frac{1-4m}{(1-4m)^2+4(2n-1)^2}\\&+\frac{1+4m}{(1+4m)^2+4(2n-1)^2}+\frac{1-2(2m-1)}{[1-2(2m-1)]^2+16n^2}\\&+\frac{1+2(2m-1)}{[1+2(2m-1)]^2+16n^2}\Bigg]\Bigg\}=-0.52562.\end{aligned} \tag{11}$$

In Eq. (10) the expression in parentheses is exactly equal to $\pi/2$, so that $A_{01}=s_1/2$ and $A_{02}=s_2/2$, as shown in Eq. (7).

4. PROPERTIES OF THE EFFECTIVE DIELECTRIC CONSTANT

The effective dielectric constant defined by Eq. (7) depends on the concentrations s_1 and s_2 of the inclusions (or, equivalently, on the radii of the differing inclusions, r_1 and r_2) and on the parameters Δ_{12} and Δ_{13}, which characterize the relationship between the dielectric constants of the matrix and inclusions.

For weakly inhomogeneous media, in which the quadratic terms Δ_{12}^2 and Δ_{13}^2 can be neglected, Eq. (7) takes the especially simple form

$$\varepsilon_{\text{eff}}=\varepsilon_1\frac{1-(\Delta_{12}s_1+\Delta_{13}s_2)/2}{1+(\Delta_{12}s_1+\Delta_{13}s_2)/2}. \tag{12}$$

Equation (12) describes exactly the macroscopic characteristics of a system with inclusions whose dielectric constants differ little from that of the matrix. It also holds for systems with very low concentrations of inclusions, i.e., $s_1,s_2\ll 1$, where the parameters Δ_{12} and Δ_{13} can then take arbitrary values within their applicable domains, $-1\leqslant\Delta_{12},\ \Delta_{13}\leqslant 1$. Equation (12) is, in fact, valid for media characterized by small values of the parameter combinations $\Delta_{12}s_1/2$ and $\Delta_{13}s_2/2$.

If the concentrations of the additional phases are equal $(s_1=s_2=s)$, i.e., all inclusions have the same radius $(r_1=r_2=r)$, then Eq. (12) takes the form

$$\varepsilon_{\text{eff}}=\varepsilon_1\frac{1-s(\Delta_{12}+\Delta_{13})/2+A(\Delta_{12}^2+\Delta_{13}^2)+2B\Delta_{12}\Delta_{13}}{1+s(\Delta_{12}+\Delta_{13})/2+A(\Delta_{12}^2+\Delta_{13}^2)+2B\Delta_{12}\Delta_{13}}, \tag{13}$$

where $A_1=A_2=A$ and $B_1=B_2=B$. If all inclusions also have the same dielectric constant $(\varepsilon_2=\varepsilon_3$, or $\Delta_{12}=\Delta_{13})$, then the expression for the effective dielectric constant simplifies further to

$$\varepsilon_{\text{eff}}=\varepsilon_1\frac{1-\Delta_{12}s+2\Delta_{12}^2(A+B)}{1+\Delta_{12}s+2\Delta_{12}^2(A+B)}. \tag{14}$$

Equation (14) now defines the effective dielectric constant of a two-component system that is the same as the Rayleigh model,[7] in which unidirectional circular cylinders with identical characteristics form a doubly periodic lattice. This agreement makes it possible to compare the present results with the more accurate calculations obtained by by Rayleigh in a different manner,[7] and therefore to establish the domain of applicability of Eq. (14) and, indirectly, the more general Eqs. (12) and (13).

For weakly inhomogeneous media, where the parameters Δ_{12} and s are very small $(\Delta_{12},s_1\ll 1)$, the effective dielectric constant for the Rayleigh model in the first approximation is (in the notation of the present paper)

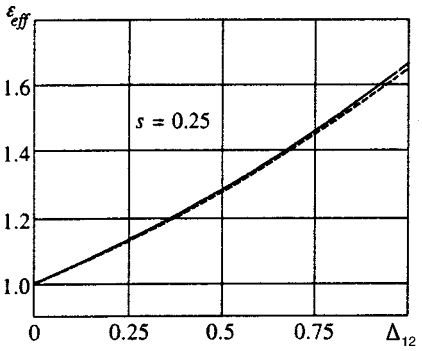


FIG. 3.

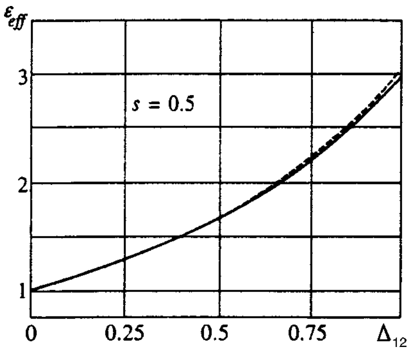


FIG. 4.

$$\varepsilon_{\mathrm{eff}}=\varepsilon_1\frac{1-\Delta_{12}s}{1+\Delta_{12}s}. \tag{15}$$

This corresponds to Eq. (14) in the linear approximation in Δ_{12}.

In the general case, if we restrict attention to the first three approximations, the expression for the effective dielectric constant corresponding to the Rayleigh model can be written in the form

$$\varepsilon_{\mathrm{eff}}=\varepsilon_1\frac{1-\Delta_{12}s-f_j(\Delta_{12},s)}{1+\Delta_{12}s-f_j(\Delta_{12},s)}\quad (j=1,2,...), \tag{16}$$

where for the first approximation

$$f_1=0,$$

which leads to Eq. (15), for the second approximation

$$f_2=0.306\Delta_{12}^2s^4 \tag{17}$$

for the second approximation, and to

$$f_3=\frac{f_2}{1-1.403\Delta_{12}^2s^8}+0.0134\Delta_{12}^2s^8 \tag{18}$$

for the third.

A comparison of Eqs. (14) and (16) (the latter for the case of f_3), obtained by two independent methods, shows that they yield the same results for concentrations of the inclusions $s<0.4$, for all values of the parameter Δ_{12}, and for arbitrary concentrations of the inclusions $0\leqslant s\leqslant\pi/4$ if $\Delta_{12}<0.5$. This can be seen from Figs. 3–5 for $\varepsilon_{\mathrm{eff}}(\Delta_{12})$ with $s=0.25$, 0.5, and 0.75, respectively. In these figures the dashed curves correspond to the Rayleigh model and were plotted using Eqs. (16) and (18), while the smooth curves correspond to Eq. (14) derived here.

A number of interesting special solutions can be derived from the general equation (13) for the effective dielectric constant.

For example, let one of the additional phases—the second, with permittivity ε_3—be eliminated from the system; then, setting $\varepsilon_3=\varepsilon_1$ ($\Delta_{13}=0$), we have a two-component medium for which

$$\varepsilon_{\mathrm{eff}}=\varepsilon_1\frac{1-\Delta_{12}s/2+A\Delta_{12}^2}{1+\Delta_{12}s/2+A\Delta_{12}^2}. \tag{19}$$

If one of the two types of inclusion, for example the second, is a metal ($\varepsilon_3\to\infty$, $\Delta_{13}=-1$), then Eq. (13) becomes

$$\varepsilon_{\mathrm{eff}}=\varepsilon_1\frac{1+s/2+A-\Delta_{12}^2(s/2+2B)+A\Delta_{12}^2}{1-s/2+A+\Delta_{12}^2(s/2+2B)+A\Delta_{12}^2}. \tag{20}$$

The latter two cases can also be examined using the general expression (7) for $\varepsilon_{\mathrm{eff}}$ with unequal concentrations of the inclusions.

5. COMPOSITE MEDIUM WITH CHARACTERISTIC PROPERTIES

Of all the possible relationships among the geometric parameters and physical characteristics of a three-component medium, it is necessary to distinguish the case in which the dielectric constant of the matrix has some average value between those of the inclusions.

Let us first consider a system with equal concentrations of the additional phases, and therefore equal radii for all inclusions, i.e., $s_1=s_2=s$ ($r_a=r_2=r$). Then, for

$$\Delta_{12}=-\Delta_{13}, \tag{21}$$

Eq. (13) implies that

$$\varepsilon_{\mathrm{eff}}=\varepsilon_1. \tag{22}$$

Using Eq. (3), Eq. (21) can be rewritten in the form

$$\varepsilon_1=\sqrt{\varepsilon_2\varepsilon_3}. \tag{23}$$

Thus, if the dielectric constant ε of the matrix equals the geometric mean of the constants of the two dissimilar inclusions, ε_2 and ε_3, the effective dielectric constant of this

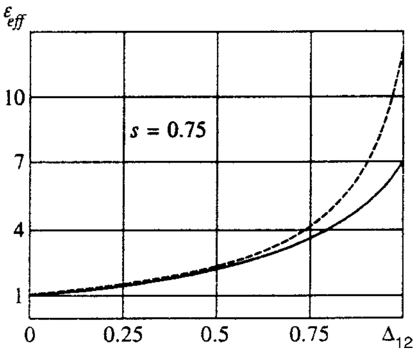


FIG. 5.

inhomogeneous medium will equal that of the matrix. Physically, this characteristic of the medium is explained by the fact that the dielectric polarization vectors in the dissimilar inclusions are equal in magnitude and and opposite in direction. Here the electric field in the composite material is highly nonuniform: the displacement vector is squeezed out of the inclusions whose permittivity is below ε_1 and drawn into the others, whose permittivity exceeds ε_1.

If a system with these properties is interpreted in terms of equivalent electrical circuits with RLC components, then Eq. (23) means that the circuit has the characteristic resistance

$$R=\sqrt{LC}. \tag{24}$$

It must be borne in mind that Eqs. (21)–(23) were derived for a system with equal concentrations of the additional phases. However, systems in which the dissimilar cylindrical inclusions have different radii, and therefore unequal concentrations of the two phases in the matrix, can be studied entirely analogously. In this case, it is necessary to turn to the general formula (7) for the effective dielectric constant. Setting

$$\Delta_{12}s_1=-\Delta_{13}s_2 \tag{25}$$

in Eq. (7), we again obtain $\varepsilon_{\text{eff}}=\varepsilon_1$. In explicit form, Eq. (25) becomes

$$\varepsilon_1=\frac{\Delta_s}{2}(\varepsilon_2-\varepsilon_3)+\sqrt{\frac{\Delta_s^2}{2}(\varepsilon_2-\varepsilon_3)^2+\varepsilon_2\varepsilon_3}, \tag{26}$$

where $\Delta_s=(s_1-s_2)/(s_1+s_2)$.

For equal concentrations of the additional phases, i.e., $s_1=s_2$, Eq. (25) transforms to Eq. (21), and Eq. (26) to Eq. (23).

In a system with unequal concentrations of the phases, the dielectric polarization vectors are also oppositely directed in the dissimilar inclusions, but unlike the previous case, the magnitudes of the dielectric polarization vectors in the phases are not equal—they depend on the transverse cross-sectional area of the cylinders. In order for the condition $\varepsilon_{\text{eff}}=\varepsilon_1$ to be met, the polarization of the dielectric in the cylinders with smaller radii must be larger, and *vice versa*. The quantitative relationships among the parameters of the medium are then established by Eq. (25). As before, the dielectric constant of the matrix, ε_1, has a value intermediate between those of the inclusions, ε_2 and ε_3. But now, according to Eq. (26), ε_1 also depends on the ratio of the concentrations of the additional phases.

The approach proposed here to calculate the effective parameters can be extended to matrix systems with more than three components, and therefore the conditions that determine the compensation of polarization phenomena in multicomponent systems can be obtained.

6. RECIPROCITY RELATIONS

Keller's theorem occupies an important place in the theory of two-dimensional, two-component matrix systems.[1] This theory establishes a relationship among the components of the effective dielectric permittivity tensor:

$$\varepsilon_{\text{eff}\,xx}(\varepsilon_1,\varepsilon_2)\varepsilon_{\text{eff}\,yy}(\varepsilon_2,\varepsilon_1)=\varepsilon_1\varepsilon_2, \tag{27}$$

where the x component is determined for a medium in which the matrix and inclusions have the parameters ε_1 and ε_2, respectively, while the y component corresponds to the parameters of a medium in which the dielectric constant of the matrix is ε_2 and that of the inclusions is ε_1. Equation (27) establishes the most general property characterizing two-component, two-dimensional systems, independent of their specific structure. At a practical level, it often makes it possible to simplify the analysis of systems by reducing the volume of analytic and numerical calculations.

To prove Eq. (27), Keller considered a medium with circular cylindrical inclusions, which are parallel to one another and form a doubly periodic rectangular lattice in their transverse cross section. This is the same system as that studied by Rayleigh,[7] and it can be shown that Eqs. (15)–(18), obtained by him for the effective dielectric constant, satisfy Eqs. (27).

Subsequently, Eqs. (27) were extended to more general structures for two-component inhomogeneous media. Dykhne found a symmetry relation for media of this type.[3] Developing Dykhne's method, Balagurov showed that Eqs. (27) are valid for arbitrary concentrations, shapes, and distributions of the inclusions in a medium.[9] To prove the reciprocity relations, Mendelson used tensor analysis and obtained extremely general results.[2] Schulgasser proposed a persuasive interpretation of the reciprocity relations employing a combination of geometry, algebra, and mathematical analysis.[10] Note that in the papers cited above, as well as in many others, the reciprocity relations are usually discussed in terms of electrical conductivity (sometimes thermal conductivity, as in Ref. 10); this is in fact of no fundamental significance and, in accordance with the familiar analogy, the results obtained here are completely applicable to the study of magnetic, diffusion, electric, and other processes in inhomogeneous systems.

The reciprocity relation takes a slightly different form for three-component systems. In accordance with the preceding investigation, it is convenient to write it in a form using the parameters Δ_{12} and Δ_{13}, which arise naturally in solving the field problem (1) and (2) and are present in the averaged formulas. The main result can be formulated as a theorem: if a dielectric medium with permittivity ε_1 contains parallel dielectric cylindrical inclusions of two varieties with constants ε_2 and ε_3, and the inclusions are distributed with doubly periodic ordering and a step size equal to the side of the square cell, then the effective dielectric constant of this medium will satisfy

$$\varepsilon_{\text{eff}}(\Delta_{12},\Delta_{13})\varepsilon_{\text{eff}}(\Delta_{21},\Delta_{31})=\varepsilon_1^2, \tag{28}$$

where, according to the definition (3), $\Delta_{1p}=-\Delta_{p1}$ $(p=2,3)$. If rectangular cells are considered, the system as a whole will be anisotropic, and

$$\varepsilon_{\text{eff}\,xx}(\Delta_{12},\Delta_{13})\varepsilon_{\text{eff}\,yy}(\Delta_{21},\Delta_{31})=\varepsilon_1^2. \tag{29}$$

The concentrations of the inclusions can differ. As a special case, Eqs. (28) and (29) yield the reciprocity relation for a two-component medium upon setting $\varepsilon_3=\varepsilon_1$ $(\Delta_{13}=0)$. In the form taken here, it can be written

$$\varepsilon_{\text{eff}}(\Delta_{12})\varepsilon_{\text{eff}}(\Delta_{21})=\varepsilon_1^2, \tag{30}$$

for an isotropic system, and

$$\varepsilon_{\text{eff}\,xx}(\Delta_{12})\varepsilon_{\text{eff}\,yy}(\Delta_{21})=\varepsilon_1^2 \tag{31}$$

for an anisotropic system.

It can be shown that the reciprocity relation for two-component systems in the form (30) is indeed satisfied by the known exact solutions. This is the case, for example, in the Rayleigh model (15)–(18). It is also valid for a checkerboard system, for which[3,5]

$$\varepsilon_{\text{eff}}=\sqrt{\varepsilon_1\varepsilon_2}=\varepsilon_1\sqrt{\frac{1-\Delta_{12}}{1+\Delta_{12}}}. \tag{32}$$

It also corresponds to the exact solution for a doubly periodic distribution in a matrix of rectangular inclusions with concentration $s=0.25$.[4] In the special case of the square inclusions from Ref. 4, some simple calculations yield (in terms of the dielectric constant)

$$\varepsilon_{\text{eff}}=\varepsilon_1\sqrt{\frac{2-\Delta_{12}}{2+\Delta_{12}}}. \tag{33}$$

The reciprocity relation (31) is also applicable to one-dimensional structures and stratified media, for which, when the concentrations of the phases are equal,

$$\begin{aligned}\varepsilon_{\text{eff}\,xx}&=\frac{2\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}=\varepsilon_1(1-\Delta_{12}),\\ \varepsilon_{\text{eff}\,yy}&=\frac{1}{2}\,(\varepsilon_1+\varepsilon_2)=\frac{\varepsilon_1}{1+\Delta_{12}}.\end{aligned} \tag{34}$$

The reciprocity relations for three-component media can be proven by directly solving the field problem in the given periodic system and averaging the electric field in a unit cell with two dissimilar inclusions. We have taken this approach here.

In the general case, the formula for the parameter has the structure

$$\varepsilon_{\text{eff}}=\varepsilon_1\,\frac{F(\Delta_{12},\Delta_{13},s_1,s_2)}{F(\Delta_{21},\Delta_{31},s_1,s_2)}, \tag{35}$$

where we recall that $\Delta_{21}=-\Delta_{12}$ and $\Delta_{31}=-\Delta_{13}$ (in accordance with Eq. (3)) and $F(\dots)$ is a function of its parameters. This is an implicit expression for the dielectric constant; it holds for arbitrary concentrations of inclusions from two additional phases.

For high concentrations of the inclusions, multidipole interactions must be taken into account in solving the field problem. In the special case of low concentrations, where we can limit ourselves to one-dipole interactions, an explicit expression for $F(\Delta_{12},\Delta_{13},s_1,s_2)$ has been obtained in the present paper (Eqs. (7)–(11)). In Eq. (35), $F(\Delta_{21},\Delta_{31},s_1,s_2)$ determines the electric field averaged over the length of the unit cell along the x axis, $\langle E_x\rangle$, and coincides with the direction of the external field $E_0=E_{0x}$ in the system. Thus, the function $\varepsilon_1F(\Delta_{12},\Delta_{13},s_1,s_2)$ gives the average value of the electric displacement $\langle D_x\rangle$ in the cell. Here the average fields $\langle E_x\rangle$ and $\langle D_x\rangle$ are given by the integrals (6).

Equations (35) yield the reciprocity relations (28) for square cells, and (29) for rectangular cells.

Clearly the reciprocity relation for a three-component system is satisfied by stratified media with equal concentrations of the phases (one-dimensional periodic structures), for which the effective parameters are given by formulas that generalize Eq. (34):

$$\begin{aligned}\varepsilon_{\text{eff}\,xx}&=\frac{3\varepsilon_1\varepsilon_2\varepsilon_3}{\varepsilon_1\varepsilon_2+\varepsilon_1\varepsilon_3+\varepsilon_2\varepsilon_3}=3\varepsilon_1\,\frac{(1-\Delta_{12})(1-\Delta_{13})}{(1-\Delta_{12})(1+\Delta_{13})+(1+\Delta_{12})(1-\Delta_{13})+(1-\Delta_{12})(1-\Delta_{13})},\\ \varepsilon_{\text{eff}\,yy}&=\frac{1}{3}\,(\varepsilon_1+\varepsilon_2+\varepsilon_3)=\frac{\varepsilon_1}{3}\,\frac{(1+\Delta_{12})(1-\Delta_{13})+(1-\Delta_{12})(1+\Delta_{13})+(1+\Delta_{12})(1+\Delta_{13})}{(1+\Delta_{12})(1+\Delta_{13})}.\end{aligned} \tag{36}$$

The reciprocity relation (29) has been written in the form of a general algebraic expression: it no longer has the simple physical appeal characteristic of the two-component case. The main advantage of writing the reciprocity relation in a form containing the product of functions of the parameters Δ_{12} and Δ_{13}, however, is that it admits of a natural generalization to systems with an arbitrary number of components.

Thus, if the matrix of a composite dielectric contains n dissimilar inclusions, then the reciprocity relation takes the form

$$\varepsilon_{\text{eff}\,xx}(\Delta_{12},\Delta_{13},\dots,\Delta_{1n})\varepsilon_{\text{eff}\,yy}(\Delta_{21},\Delta_{31},\dots,\Delta_{n1})=\varepsilon_1^2. \tag{37}$$

The validity of Eq. (37) can be tested by examining periodic structures with the aid of the computational technique used here. In particular, it is clear that Eq. (37) is satisfied by multicomponent one-dimensional structures with formulas that are generalizations of Eq. (36).

7. COMPLEX EFFECTIVE DIELECTRIC CONSTANT

In the quasistationary approximation with a wavelength that exceeds the characteristic dimensions of the system, all three components of the dielectric constant of a given medium can be regarded as complex:

$$\hat{\varepsilon}_j=\varepsilon_j'-i\varepsilon_j'' \quad (j=1,2,3). \tag{38}$$

An imaginary part of the complex dielectric constant signifies losses in the dielectric medium.

In this case the complex effective dielectric constant of an inhomogeneous system, $\hat{\varepsilon}_{\mathrm{eff}}=\varepsilon'_{\mathrm{eff}}-i\varepsilon''_{\mathrm{eff}}$, is given by Eq. (7), in which scalar quantities must be replaced by complex values of the dielectric constant (38). Here we use the complex parameters

$$\hat{\Delta}_{1k}=\frac{\hat{\varepsilon}_1-\hat{\varepsilon}_p}{\hat{\varepsilon}_1+\hat{\varepsilon}_p}\quad (p=2,3). \tag{39}$$

Losses in an inhomogeneous dielectric medium are in general due to displacement currents and conduction currents; they depend in a complicated fashion on the relationships among the components of the system and on the frequency of the external electric field.

In this case, in the original expression (15) and Eqs. (13), (19), and (20) it is necessary to make the substitutions $\varepsilon_{\mathrm{eff}}\to\hat{\varepsilon}_{\mathrm{eff}}$, $\varepsilon_1\to\hat{\varepsilon}_1$, and $\Delta_{1p}\to\hat{\Delta}_{1p}$. Then, for

$$\hat{\Delta}_{12}=-\hat{\Delta}_{12} \tag{40}$$

we have

$$\hat{\varepsilon}_{\mathrm{eff}}=\hat{\varepsilon}_1, \tag{41}$$

with

$$\hat{\varepsilon}_1=\sqrt{\hat{\varepsilon}_2\hat{\varepsilon}_3}. \tag{42}$$

Equations (40)–(42) are generalizations of the corresponding Eqs. (21)–(23). Equations (24), (32), and (33) can also be written in complex form.

If the inclusions have low dielectric losses, $\tan\,\delta_2,\tan\,\delta_3\ll 1$ (δ_2 and δ_3 are the dielectric loss angles in the differing cylinders), then using the approximate expansion of the root, Eq. (42) can be written

$$\hat{\varepsilon}_1=\sqrt{\varepsilon'_2\varepsilon'_3}\left[1-\frac{i}{2}\,(\tan\,\delta_1+\tan\,\delta_2)\right].$$

In low-loss media, the dielectric constant of the matrix is obviously close to $\sqrt{\varepsilon'_2\varepsilon'_3}$.

8. CONCLUSION

Three-component systems have heretofore not been adequately studied due to difficulties in calculating the physical fields in such systems. Techniques for boundary value problems in the theory of analytic functions can be invoked for two-dimensional systems with an ordered distribution of dissimilar inclusions; this has been done here for a medium with circular inclusions, where we took advantage of the solution for a pair of inclusions.

Three-component matrix systems share many of the properties of two-component systems. They also obey reciprocity relations expressible as a product of functions of the parameters that determine the relative values of the dielectric constants of the inclusions. In addition, polarization phenomena are more varied in three-component systems. In some cases, it has been shown that the dielectric polarization vectors in dissimilar inclusions can have opposite directions and, under certain conditions, they can cancel one another. The effective dielectric constant of such a system will be equal to the permittivity of the matrix. For two-component systems these conditions cannot be realized, even in principle.

APPENDIX

Here we give the exact analytic solution for two circular dielectric inclusions in a uniform external electric field; this is then used to study the given three-component system.

Let two circular dielectric cylinders that are parallel to one another and have radii r_1 and r_2 and dielectric constants ε_2 and ε_3, respectively, lie in an dielectric medium with permittivity ε_1. The axes of the cylinders are separated by a distance h $(h>r_1+r_2)$. The external electric field E_0 is uniform and perpendicular to the axes of the cylinders, which lie along the x axis. The coordinate origin lies on the axis of the cylinder with radius r_1. Under these conditions, the solution of the problem is

$$E_1(z)=E_0-(x_1-x_2)^2\sum_{k=1}^{\infty}\left\{\Delta^k\left\{\bar{E}_0\left[\frac{1}{\Delta_{13}}\left(\frac{g_{1k}}{z-z_{1k}}\right)^2\right.\right.\right.$$
$$\left.\left.\left.+\frac{1}{\Delta_{12}}\left(\frac{g_{2k}}{z-z_{2k}}\right)^2\right]-E_0\left[\left(\frac{g_{3k}}{z-z_{3k}}\right)^2+\left(\frac{g_{4k}}{z-z_{4k}}\right)^2\right]\right\}\right\} \tag{A.1}$$

outside the cylinders,

$$E_2(z)=(1+\Delta_{12})\left\{E_0-(x_1-x_2)^2\right.$$
$$\left.\times\sum_{k=1}^{\infty}\left\{\Delta^k\left[\frac{\bar{E}_0}{\Delta_{12}}\left(\frac{g_{2k}}{z-z_{2k}}\right)^2-E_0\left(\frac{g_{3k}}{z-z_{3k}}\right)^2\right]\right\}\right\}$$

in the first cylinder, and

$$E_3(z)=(1+\Delta_{13})\left\{\bar{E}_0-(x_1-x_2)^2\right.$$
$$\left.\times\sum_{k=1}^{\infty}\left\{\Delta^k\left[\frac{\bar{E}_0}{\Delta_{13}}\left(\frac{g_{1k}}{z-z_{1k}}\right)^2-E_0\left(\frac{g_{4k}}{z-z_{4k}}\right)^2\right]\right\}\right\}$$

in the second.

Here $E_j(z)=E_{xj}-iE_{yj}$ $(j=1,2,3)$ are the complex electric field functions; $z=x+iy$ is the complex variable, x_1 and x_2 are symmetry points on the x axis, with coordinates

$$x_1=br_1,\quad x_2=\frac{r_1}{b},$$

$$b=\frac{1}{2r_1h}\,[h^2+r_1^2+r_2^2-\sqrt{(h^2+r_1^2+r_2^2)^2-4r_1^2h^2}],$$

the g_{vk} $(v=1,\ldots 4)$ are parameters given by

$$g_{vk}=\frac{a_v}{1-a_v^2},\quad g_{4k}=-g_{3k},$$

$$a_1=a_4\left(\frac{h-x_1}{h-x_2}\right)^{1/2},\quad a_2=\frac{1}{a_4}\left(\frac{x_1}{x_2}\right)^{1/2},$$

$$a_3=\frac{1}{a_4},\quad a_4=\left(\frac{x_1h-r_1^2}{x_2h-r_2^2}\right)^{k/2};$$

the z_{vk} are coordinates of the dipoles on the x axis,

$$z_{vk}=\frac{x_1-x_2a_v^2}{1-a_v^2};$$

Δ is a parameter given by

$$\Delta=\Delta_{12}\Delta_{13},\quad -1\leqslant\Delta\leqslant 1;$$

the bar over E_0 denotes the complex conjugate.

These expressions have been obtained by solving the corresponding boundary value problem. This solution is discussed in detail in Ref. 8, where a constructive solution is also obtained using the inversion method.

[1] J. B. Keller, J. Math. Phys. **5**, 548 (1964).

[2] K. S. Mendelson, J. Appl. Phys. **46**, 4740 (1975).

[3] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1970)].

[4] Yu. V. Obnosov, Dokl. Akad. Nauk SSSR **319**, 1125 (1991) [Sov. Phys. Dokl. **36**(8), 573 (1991)].

[5] Yu. P. Emets, Zh. Éksp. Teor. Fiz. **96**, 701 (1989) [Sov. Phys. JETP **69**, 397 (1989)].

[6] Yu. P. Emets, *Electrical Characteristics of Composite Materials with Regular Structure* [in Russian], Naukova Dumka, Kiev (1986).

[7] Lord Rayleigh, Phil. Mag. **34**, 481 (1882).

[8] Yu. P. Emets and Yu. P. Onofrichuk, IEEE Trans. Dielectr. Electr. Insul. **3**, 87 (1996).

[9] B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **81**, 665 (1981) [Sov. Phys. JETP **54**, 355 (1981)].

[10] K. A. Schulgasser, Int. Comm. Heat Mass Transfer **19**, 639 (1992).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 87, NUMBER 3 SEPTEMBER 1998

Quantum and semiclassical oscillations in the organic metal $(\mathrm{BEDO\text{-}TTF})_2\mathrm{Cl}_x(\mathrm{H_2O})_y$

S. I. Pesotskiĭ*)

Institute of Chemical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia;
International High Magnetic Field and Low temperature Laboratory, 53-529 Wroclaw, Poland;
Walther–Meissner Institute, D-85748 Garching, Germany

R. B. Lyubovskiĭ

Institute of Chemical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia;
International High Magnetic Field and Low Temperature Laboratory, 53-529 Wroclaw, Poland

P. Behm, W. Biberacher, and K. Andres

Walther–Meissner Institute, D-85748 Garching, Germany

M. V. Kartsovnik

Walther–Meissner Institute, D-85748 Garching, Germany;
Institute of Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

A. E. Kovalev and T. G. Togonidze

Institute of Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russian

J. A. A. J. Perenboom

High Field Magnetic Laboratory, Catholic University, NL 6525, ED Nijmegen, Netherlands

N. D. Kushch, R. N. Lyubovskaya, and E. I. Zhilyaeva

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

(Submitted 3 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 1137–1146 (September 1998)

We study quantum (Shubnikov–de Haas and de Haas–van Alphen) oscillations and angular oscillations of the reluctance in the organic quasi-two-dimensional metal $(\mathrm{BO})_2\mathrm{Cl}_x(\mathrm{H_2O})_y$. We show that the Fermi surface in this compound consists of a slightly corrugated cylinder with its axis perpendicular to the conducting plane. The cross section of the cylinder in this plane is a perfect circle of radius $k_F \simeq 3\times10^7\ \mathrm{cm}^{-1}$. The effective carrier mass associated with this cylinder is $m^* = (1.65\text{–}2.0)\,m_0$ in the conducting plane, while the Dingle temperature is $T_D = 3\text{–}4$ K. *© 1998 American Institute of Physics.* [S1063-7761(98)02709-7]

1. INTRODUCTION

The compound $(\mathrm{BO})_2\mathrm{Cl}_x(\mathrm{H_2O})_y$ belongs to the class of quasi-two-dimensional organic metals. The major structural material in this compound, which forms the conducting layers, is the organic molecule BO=(BEDO-TTF) (bis(ethylenedioxy)-tetrathiofulvalene), an analog of the well known ET molecule used as a basis for synthesizing the overwhelming majority of quasi-two-dimensional organic metals and superconductors.[1] Unlike the ET molecule, in the BO molecule the peripheral atoms of sulfur are replaced by oxygen atoms. Since overlap of the orbitals of these sulfur atoms is responsible for the high conductivity in a layer made up of ET molecules, one might expect substantial changes in the conducting properties on going to metals based on BO.

The first report on the synthesis, structure, and properties of the metallic chloride of BO contained the chemical formula of this compound, $(\mathrm{BO})\mathrm{Cl}(\mathrm{H_2O})$, a simple salt.[2] The simple salt is characterized by a 1:1 ratio of the BO cation to the $\mathrm{Cl}(\mathrm{H_2O})$ anion, which corresponds to the transfer of an electron from each BO molecule to an anion and implies a half-filled metal band. This report was of great interest, since before then no organic metals with a half-filled conduction band were known.

The next paper[3] showed that the chemical formula of the BO chloride was more properly written in the form of the complex salt $(\mathrm{BO})_2\mathrm{Cl}(\mathrm{H_2O})_3$. In this case, one electron arrives at an ion from every two BO molecules, and the conduction band is then one-quarter full. A comparison of theoretical calculations of the structure of BO chloride with the

1063-7761/98/87(9)/7/$15.00

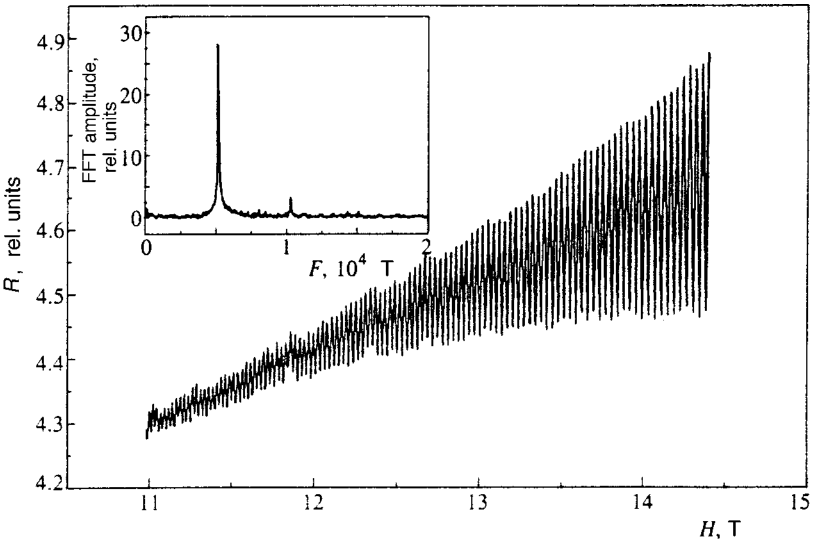


FIG. 1. Shubnikov–de Haas oscillations: $\theta=0°$, $T=1.45$ K, $\mathbf{I}\perp\mathbf{ab}$. The inset shows the FFT amplitude of these oscillations as a function of the frequency F of the quantum oscillations.

parameters of the quantum mechanical oscillations first observed in this compound[3] confirms this level of filling of the band.

Data from an analysis of the composition and structure corresponding to the formula $(BO)_2Cl(H_2O)_3$ were later presented,[4] along with some results from a study of the reluctance of this complex that were consistent with the proposed formula.

Nevertheless, the chemical composition of BO chloride is still under discussion. A chemical formula of the form $(BO)_2Cl_{1.28}(H_3O)_{0.28}(H_2O)_{2.44}$ has been proposed,[5] while fully retaining the structural motif proposed before.[3,4] This difference in the description of the chemical composition does not, however, lead to a change in the degree of filling of the band, since the excess negative charge on the anion (associated with the chlorine) beyond that implied by the formula $(BO)_2Cl(H_2O)_3$, is balanced by the positive charge on the H_3O.

In this paper we offer a more detailed investigation of the Shubnikov–de Haas and de Haas–van Alphen oscillations and semiclassical angular oscillations in the reluctance. These studies provide a more detailed idea of the size and shape of the Fermi surface in BO chloride, and make it possible to estimate the parameters of its electron system. The possible effect of the chemical composition of the anion on the behavior of these oscillations is not discussed here.

2. EXPERIMENT

Single crystal samples of $(BO)_2Cl_x(H_2O)_y$ obtained in two chemical groups under differing synthesis conditions[4,5] were used in the experiments. Nevertheless, the samples were shown to be identical by x-ray structural analysis. The samples consisted of irregular slabs with average dimensions of $1\times1\times0.1\,\mathrm{mm}^3$. The plane of the samples coincided with the crystallographic **ab** plane, which contains the conductive layers. These layers alternate along the short dimension of the samples, which coincides with the **c*** axis.

The reluctance was measured by the standard four contact method using a 330 Hz alternating current. Here the measurement current **I** could be directed either in the **ab** plane or perpendicular to it. A magnetic field was created either by a superconducting solenoid with a maximum field of 15 T or by a resistive magnet with a field of up to 20 T, and the minimum temperature was 0.45 K. In the reluctance measurements the orientation of the sample was varied using a two-axis gimbal mount, which made it possible to rotate the sample in declination and in the azimuthal angle.

The de Haas–van Alphen effect was studied in terms of the dependence of the torque on the magnetic field.[6] Here the maximum magnetic field was 14 T and the minimum temperature lay below 0.48 K.

3. RESULTS OF MEASUREMENTS

The average conductivity of the test samples at room temperature in the conducting plane is $50\ \Omega^{-1}\,\mathrm{cm}^{-1}$. All the samples are characterized by a metallic variation in the reluctance as a function of temperature: as the temperature is reduced from room temperature to liquid nitrogen temperature, the reluctance decreases on average by a factor of a few dozen.

Figures 1 and 2 show the Shubnikov–de Haas and de Haas–van Alphen oscillations for a magnetic field directed almost perpendicular to the conducting plane. For this field direction there is a single fundamental frequency for the quantum oscillations $F_0\simeq4900$ T (see insets to Figs. 1 and 2). When the field deviates significantly from this direction, beats become appreciable in the fundamental frequency (see Figs. 3 and 4). Analysis of the quantum oscillations and fast Fourier transforms reveal the complex nature of these beats, which result from the mixing of three oscillations with similar frequencies (insets to Figs. 3 and 4), at least for some directions of the field.

The angular dependence of these frequencies for the de Haas–van Alphen oscillations is shown in Fig. 5. There, for each direction of the field, different symbols correspond to different frequencies. This curve fit well by the formula

$$F(\theta)=\frac{4900}{\cos\theta}[\mathrm{T}], \tag{1}$$

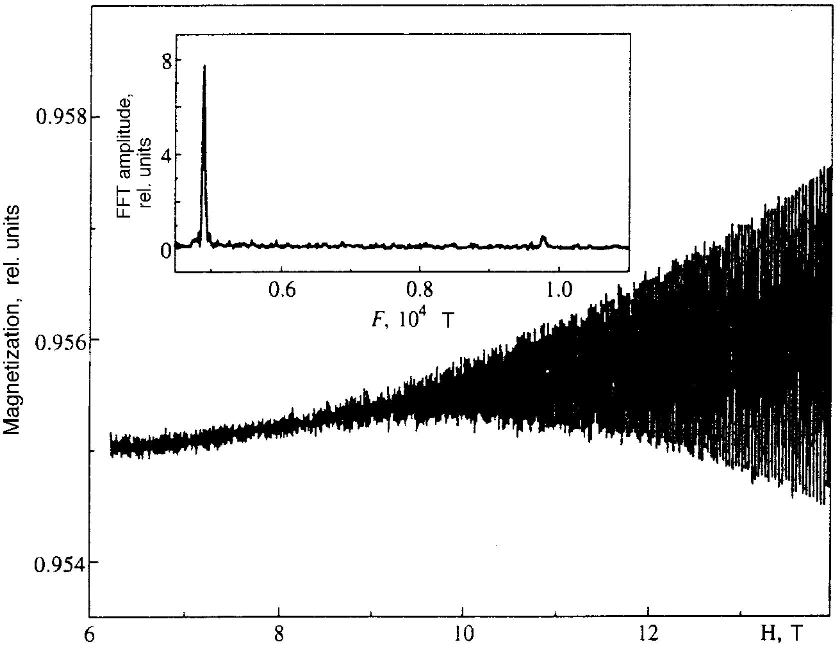


FIG. 2. De Haas–van Alphen oscillations: $\theta = 4.7°$, $T = 0.5$ K. The inset shows the FFT of these oscillations.

where θ is the angle between the field and the normal to the conducting plane. Here, recording several frequencies at each angle has essentially no effect on the curve, since the difference between the frequencies is never greater than a few percent and does not exceed the average scatter. Thus, we can assume that Eq. (1) yields the fundamental frequency to high accuracy.

Figure 6 illustrates the angular dependence of the amplitude of the de Haas–van Alphen oscillations. As in the case of the angular variation of the frequency, the amplitudes corresponding to oscillations at differing frequencies are indicated by different symbols. Despite the complicated angular dependence of the amplitude of the oscillations at different frequencies, the common zeroes in the amplitude at $\theta = \pm 41$ and $\pm 57°$ are evident. The amplitude null at $\theta = 0°$ is nonphysical and is related to the method for observing the de Haas–van Alphen oscillations.[6] The amplitude nulls at $\theta = \pm 41°$ and $\theta = \pm 57°$ are probably related to the fact that the decreasing spin factor in the Lifshitz–Kosevich formula[7] vanishes in those directions, i.e.,

$$R_s = \cos(p\pi g\mu/2) = 0, \tag{2}$$

where p is the harmonic number, g is the g-factor, $\mu = m^*/m_0$, m^* is the effective electron mass, and m_0 is the free electron mass.

The effective carrier mass was estimated on the basis of the temperature dependence of the amplitude of the quantum oscillations. Here the decreasing temperature factor was taken to have the form

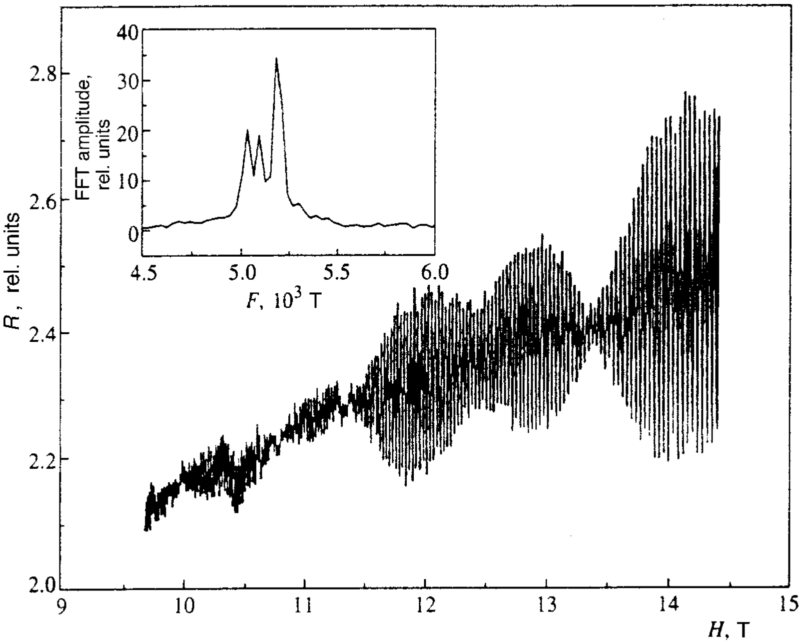


FIG. 3. Shubnikov–de Haas oscillations: $\theta = 23°$, $T = 1.45$ K, $\mathbf{I} \perp \mathbf{ab}$. The inset shows the FFT of these oscillations.

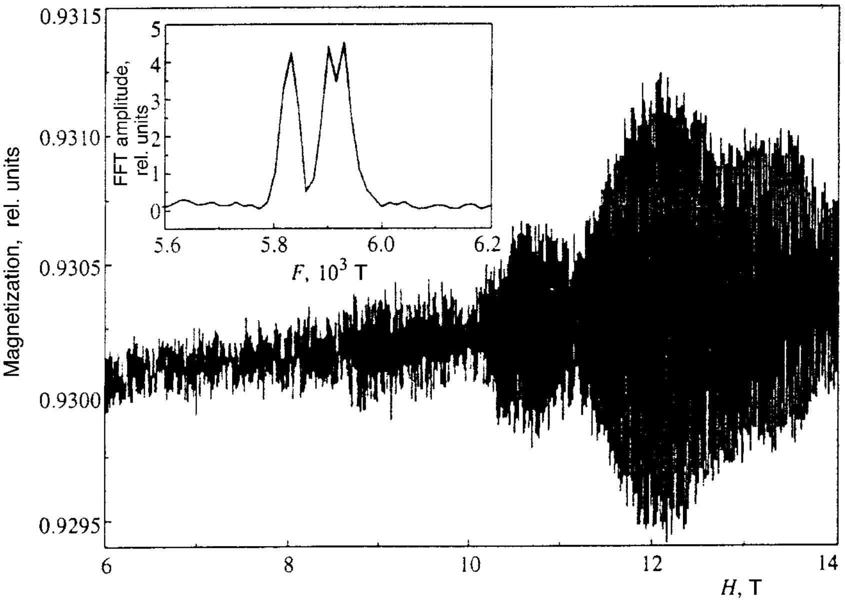


FIG. 4. De Haas–van Alphen oscillations: $\theta=34.7°$, $T=0.5$ K. The inset shows the FFT of these oscillations.

$$R_t=\frac{\alpha p\mu T/H}{\sinh(\alpha\mathbf{p}\mu T/H)},$$

where $\alpha=2\pi^2k_Bm_0/e\hbar=14.7$ T/K, T is the temperature, and H is the magnetic field. The mass in the conducting plane determined in this way for several of the samples had a substantial scatter and lay within the interval $m^*=(1.65–2.0)\,m_0$.

The Dingle temperature was determined from the field-induced variation in the amplitude of the quantum oscillations using a Dingle reducing factor of the form

$$R_D=\exp(-\alpha p\mu T_D/H).$$

The Dingle temperature lies in the range $T_D=3–4$ K for all samples measured.

Figure 7 shows the angular variation in the reluctance for a constant field of 14.3 T. The maxima of the pronounced angular oscillations are periodic in $\tan\theta$. In Fig. 8, the period of these oscillations is plotted in polar coordinates as a function of the azimuthal angle φ in the conducting plane. Clearly the period is essentially independent of the azimuthal angle, and is $\Delta\approx0.5$.

4. DISCUSSION

The crystal lattice parameters of BO chloride were first given in Ref. 2 and confirmed in Refs. 4 and 5. The band structure of this material was calculated on the basis of these data and found to be in agreement with preliminary observations of quantum oscillations.[3] According to this calculation,

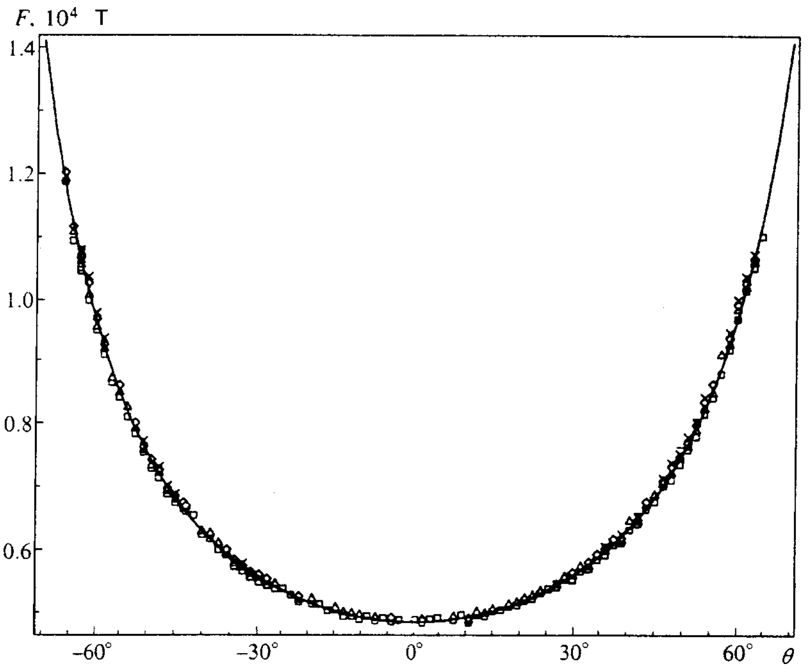


FIG. 5. Angular dependence of the frequency of de Haas–van Alphen oscillations; $T=0.5$ K. The smooth curve corresponds to $F(\theta)=4900/\cos\theta$ [T].

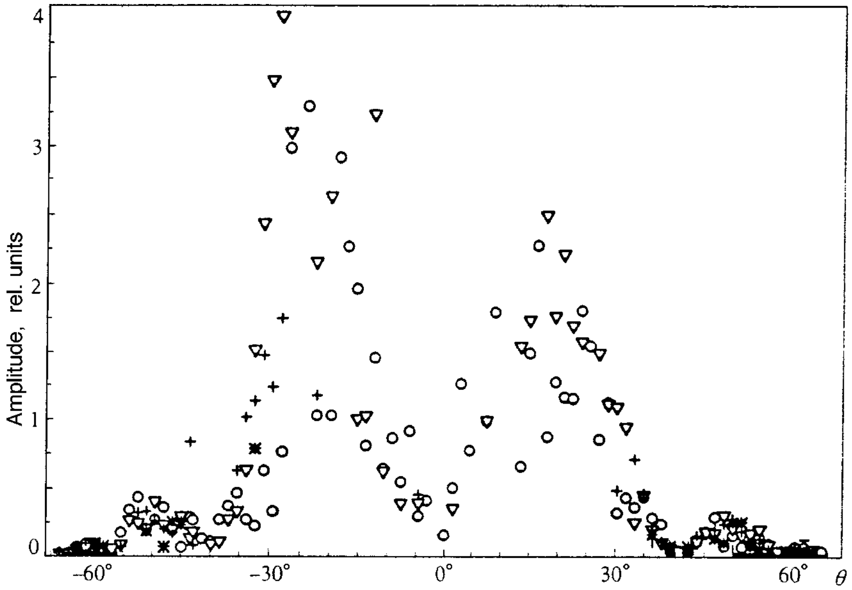


FIG. 6. Angular dependence of the amplitude of the de Haas–van Alphen oscillations; $T=0.5$ K.

the Fermi surface (Fig. 9) in $(BO)_2Cl_x(H_2O)_y$ is a cylinder with its axis oriented in the **c*** direction. The cross section of this cylinder in the **ab** plane coincident with the plane of best conductivity is a perfect circle whose area is roughly half that of the corresponding cross section of the Brillouin zone.

For all the BO chloride crystals studied here, the frequency of the quantum oscillations is $F_0 \approx 4900$ T for a field **H**⊥**ab**, which corresponds to 50% of the area transected in the first Brillouin zone, and agrees well with theoretical calculations. The angular dependence of the frequency of these oscillations is fit well by Eq. (1) and corresponds to a cylindrical Fermi surface oriented along **c***.

The angular oscillations in the reluctance, with peaks periodic in $\tan\theta$ (Fig. 7), are also related to the motion of carriers over the cylindrical Fermi surface, and the cylinder must be weakly corrugated along its axis.[8,9] The period of oscillations in the reluctance is independent of azimuthal angle in BO chloride crystals (Fig. 8), so the cross section of the cylinder can be represented as a perfect circle in the **ab** plane. In this case, the Fermi momentum k_F is related to the period Δ of angular oscillations in the reluctance by the simple formula[8]

$$\Delta=\pi/k_F d,$$

where d is the distance between conducting planes. The estimated cross section $S=\pi k_F^2 \simeq 3\times10^{15}\,\mathrm{cm}^{-2}$ is in good agreement with theoretical calculations and the quantum oscillation data. Thus, the quantum and semiclassical oscillations in the organic metal $(BO)_2Cl_x(H_2O)_y$ correspond to a weakly corrugated cylindrical Fermi surface with a perfectly circular cross section of radius $\simeq 3\times10^7\,\mathrm{cm}^{-1}$ in the **ab** plane.

The slight corrugation in the cylindrical sheet of the Fermi surface must result in the existence of several similar

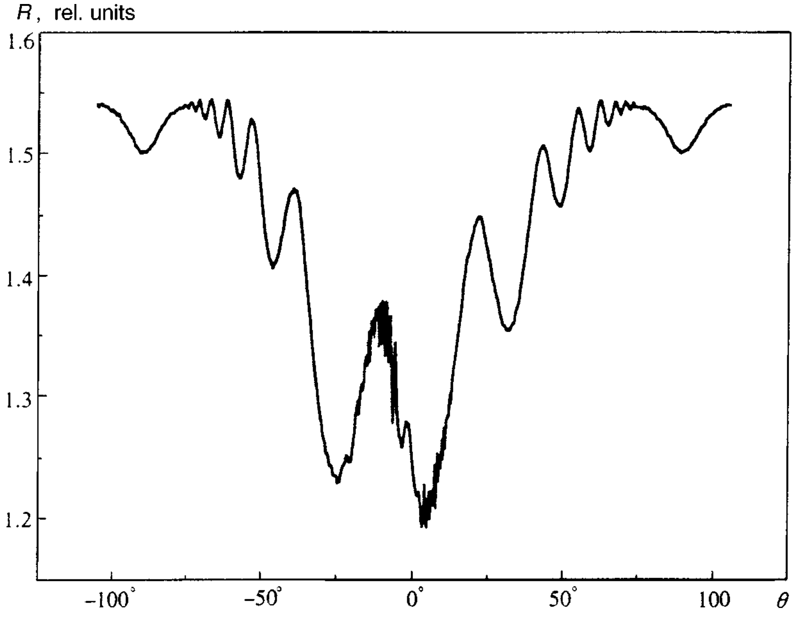


FIG. 7. Angular dependence of the reluctance; $T=0.5$ K, $H=14.3$ T, **I**⊥**ab**.

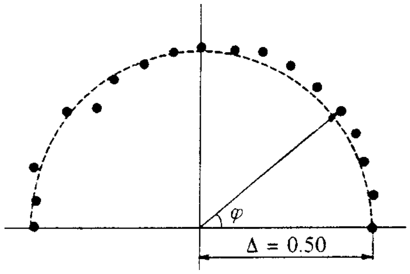


FIG. 8. The period of angular oscillations in the reluctance as a function of the azimuthal angle φ in polar coordinates.

extremal cross sections (in the case of simple corrugation, there should be two), and this may show up as beats in the fundamental frequency. In the compound studied here, beats are observed for field directions not coincident with $\mathbf{H}\|\mathbf{c}^*$ (Figs. 3 and 4). These beats, however, have a complex shape, since they result from combining more than two oscillations at differing frequencies. We might assume a complicated corrugation with several differing extremal cross sections, but one would then scarcely expect the pronounced angular oscillations in the reluctance observed in the sample crystals. In addition, there is still no explanation for the lack of any beats whatsoever for a field with $\mathbf{H}\|\mathbf{c}^*$.

The beats might be related to crystal imperfections of the samples, such as twinning, intergrowth, etc. This explanation, however, can be ruled out by direct x-ray analysis of the test samples, which confirms that they are high-quality single crystals.

The distorted form of the quantum oscillations (including beats) might arise through magnetic interaction,[7] but the latter is only significant when the absolute amplitude of oscillations in the magnetic moment $\mu_B\mathbf{M}$ is comparable to the period H^2/F of the oscillations. In the test samples,

$$\mu_B M \approx 2.5\,\mathrm{G} \ll 200\,\mathrm{G} \approx H^2/F,$$

where $\mu_B M \approx 2.5\,\mathrm{G}$ is the maximum attainable absolute amplitude in the experiments with a 10 T field, so magnetic interaction can be neglected.

Finally, there is yet another possible reason for the beats. The calculated Fermi surface in Ref. 5 differs somewhat from that in Fig. 9. The cylindrical sheet intersects the boundary of the Brillouin zone near the point X, rather than being tangent to it, forming very small pockets at this site. This can lead to the coexistence of a small orbit with the main orbit, which corresponds to the fundamental frequency. We can then expect sum and difference frequencies to appear, including some near the fundamental. This latter scenario would seem to be the most plausible, but it is obviously very vulnerable at present and requires more convincing confirmation and additional study.

The condition (2) for the appearance of "spin nulls" at $\theta \approx \pm 41$ and $\pm 57°$ is satisfied by the first harmonic, assuming that the effective mass and angle of inclination are related by the characteristic formula for a cylindrical Fermi surface, $\mu(\theta)=\mu(0)/\cos\theta$. Here the splitting factor is $S_s = g\mu(0)/2 \approx 1.85$. Given the spread in the estimate of the effective mass obtained from the temperature dependences of the amplitude of the quantum oscillations, $m=(1.65\text{–}2.0)\,m_0$, the g factor becomes $g=1.85\text{–}2.23$. The g-factor determined in this way usually includes corrections owing to many-particle interactions, which makes it different from the free electron g-factor, $g_0=2$.[7] The closeness of g and g_0 argues for the weakness of these interactions. In the test samples, any assumptions regarding the influence of many-particle interactions would be incorrect because of the excessive error in determining the g-factor due to the spread in the value of the effective mass. More detailed experiments will be required to make the latter more accurate.

5. CONCLUSION

We have studied the behavior of the magnetic moment and reluctance of a group of samples of the organic metal $(BO)_2Cl_x(H_2O)_y$. Shubnikov–de Haas and de Haas–van Alphen quantum oscillations and semiclassical angular oscillations in the reluctance were observed. An analysis of these data indicate that the Fermi surface of this compound is a cylinder weakly corrugated along its axis, and is a perfect circle of radius $k_F \simeq 3\times10^7\,\mathrm{cm}^{-1}$ perpendicular to the axis. These results agree with theoretical calculations. Complicated beats involving the fundamental frequency of the quantum oscillations were observed, but their nature is as yet unclear.

We thank L. P. Rozenberg for the x-ray structural data, and E. B. Yagubskiĭ and R. P. Shibaeva for useful discussions. This work was supported by the Russian Fund for

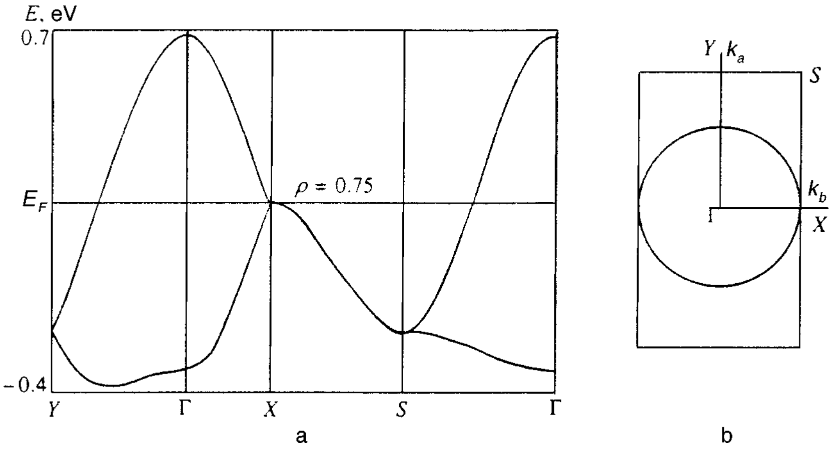


FIG. 9. The energy band structure (a) and Fermi surface (b) in the organic metal $(BO)_2Cl(H_2O)_3$.[3]

Fundamental Research (Grants No. 96-02-18957 and No. 96-02-17475), the State Scientific Program on Statistical Physics, Grant 1/70 206 from the Volkswagen–Stiftung, a Grant from NWO, and Grant No. INTAS-93 2400 EXT.

*⁾E-mail: pesot@icp.ac.ru

---

[1] J. M. Williams, J. R. Ferraro, R. J. Thorn, K. D. Karlson, U. Geiser, H. H. Wang, A. A. Kini, and M. H. Whangbo, *Organic Superconductors: Synthesis, Structure, and Theory*, Prentice-Hall, Englewood Cliffs, NJ (1992).

[2] D. Schweitzer, S. Kahlich, I. Heiman, S. E. Lan, B. Nuber, H. J. Keller, K. Winzer, and H. W. Helberg, Synth. Met. **56**, 2827 (1993).

[3] T. Mori, S. Oshima, H. Okuno, K. Kato, H. Mon, and S. Tanaka, Phys. Rev. B **51**, 11110 (1995).

[4] E. I. Zhilyaeva, S. A. Torunova, R. N. Lyubovskaya, S. V. Konovalikhin, O. A. Dyachenko, R. B. Lyubovskii and S. I. Pesotskii, Synth. Met. **83**, 7 (1996).

[5] R. P. Shibaeva, S. S. Khasanov, B. Zh. Narymbetov, L. V. Zorina, L. P. Rozenberg, A. V. Bazhenov, N. D. Kusch, E. B. Yagubskii, C. Rovira, and E. Canadell, submitted to J. Mat. Chem.

[6] P. Christ, W. Biberacher, H. Muller, and K. Andres, Solid State Commun. **91**, 451 (1994).

[7] D. Shoenberg, *Magnetic Oscillations in Metals*, Cambridge Univ. Press, Cambridge (1984).

[8] Y. K. Yamaji, J. Phys. Soc. Jpn. **58**, 1520 (1989).

[9] M. V. Kartsovnik, V. N. Laukhin, S. I. Pesotskii, I. F. Schegolev, and V. M. Yakovenko, J. de Phys. I **2**, 89 (1992).

Translated by D. H. McNeill