JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Propagation of circularly polarized light in media with large-scale inhomogeneities

E. E. Gorodnichev,*) A. I. Kuzovlev, and D. B. Rogozkin

Moscow Institute of Engineering Physics, 115409 Moscow, Russia

(Submitted 28 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 769–790 (March 1999)

Small-angle multiple scattering of circularly polarized waves in disordered systems composed of large (larger than the light wavelength) spherical particles is discussed. The equation for Stokes's fourth parameter V — the difference between the intensities of the left- and right-hand polarized light — is shown to have the form similar to that of the scalar transport equation for intensity I, the only difference being the presence of an additional ''non-small-angle'' term responsible for depolarization. In the case of small-angle scattering, depolarizing collisions are relatively rare and, in contrast to the scalar case, the problem contains an additional spatial scale, namely the depolarization depth. The polarization degree and helicity of the scattered light are calculated for the case of purely elastic scattering and in the presence of absorption in the medium. For strong absorption, depolarization is shown to follow the transition to the asymptotic regime of wave propagation. The features appearing in strong (non-Born) single scattering are also discussed. © *1999 American Institute of Physics.* [S1063-7761(99)00103-1]

1. INTRODUCTION

In recent times, a lot of attention has been focused on polarization properties of light multiply scattered in a medium with large-scale inhomogeneities.[1–4] This interest is stimulated, in particular, by various applications to the diffusion spectroscopy of scattering media (suspensions, colloid solutions, etc.)[5,6] and diagnostics of biological tissues.[7,8] Experiments[1,4] and computer simulations[4] detected a number of new features in propagation of polarized light through disordered media. The theoretical analysis of some aspects of this problem, however, has remained incomplete. As was noted in our earlier publication,[9] there is no theory of light depolarization in small-angle multiple scattering.

This paper considers multiple scattering of circularly polarized waves with a view to extending our earlier study[9] of polarization effects in propagation of unpolarized light. We assume that the medium consists of large-scale (larger than the light wavelength) spherical particles. The Stokes parameters of scattered light are calculated for both purely elastic scattering and in the presence of absorption in the medium. The difference between the intensities of the left-hand and right-hand polarized light, or Stokes's fourth parameter V, is shown to be governed by an equation similar to the scalar transport equation for intensity I, but taking into account both conventional and ''depolarizing'' scattering events. We have found that, in small-angle scattering, the depolarization length $l_{\rm dep}$ is larger than the transport length $l_{\rm tr}$ of elastic scattering. Therefore, in the presence of strong absorption ($l_a<l_{\rm tr}$, where l_a is the absorption length), when the distribution of multiply scattered light remains highly anisotropic at all penetration depths z,[10,11] the polarization is lost at very large depths z, where an asymptotic regime of light propagation is established.[9–11] Thus, unlike the scalar case, the problem contains another characteristic length, namely, the depolarization depth. We also discuss features of strong scattering, which appear beyond the Born approximation for single scattering by an isolated inhomogeneity.

It follows from our results that, alongside measurements of the angular distribution of light, the polarization degree allows one to determine further characteristic parameters of a medium and can be employed in optical studies of inhomogeneous media.

2. GENERAL RELATIONS

Let a wide beam of circularly polarized light be incident along the normal to a layer of a medium composed of large (of radius $a\gg\lambda$) spherical scatterers. The refraction index n of an isolated scatterer is assumed to be close to unity ($|n-1|\ll1$), and single scattering occurs predominantly forward.[13] Owing to the azimuthal symmetry of the configuration, the four equations for Stokes's parameters of scattered light are decoupled into two independent equation systems[12,13]:

$$\left\{\mu\frac{\partial}{\partial z}+n_0\sigma_{\rm tot}\right\}\begin{pmatrix}I(z,\mu)\\ Q(z,\mu)\end{pmatrix}=n_0\int d\mathbf{n}'\begin{pmatrix}a_1 & b_1\cos\beta\\ b_1\cos\beta' & a_1\cos\beta\cos\beta'-a_2\sin\beta\sin\beta'\end{pmatrix}\times\begin{pmatrix}I(z,\mu')\\ Q(z,\mu')\end{pmatrix},\qquad(1)$$

$$\left\{\mu\frac{\partial}{\partial z}+n_0\sigma_{\rm tot}\right\}\begin{pmatrix}U(z,\mu)\\ V(z,\mu)\end{pmatrix}=n_0\int d\mathbf{n}'\begin{pmatrix}a_2\cos\beta\cos\beta'-a_1\sin\beta\sin\beta' & -b_2\cos\beta'\\ b_2\cos\beta & a_2\end{pmatrix}\times\begin{pmatrix}U(z,\mu')\\ V(z,\mu')\end{pmatrix},\qquad(2)$$

1063-7761/99/88(3)/12/$15.00

where n_0 is the number of scatterers per unit volume, $\sigma_{\mathrm{tot}} = \sigma + \sigma_a$ is the total cross section of light attenuation, σ and σ_a are the scattering and absorption cross sections, respectively,

$$a_1 = \frac{1}{2}(|A_\parallel|^2 + |A_\perp|^2), \quad a_2 = \mathrm{Re}\, A_\parallel A_\perp^*,$$

$$b_1 = \frac{1}{2}(|A_\parallel|^2 - |A_\perp|^2), \quad b_2 = \mathrm{Im}\, A_\parallel A_\perp^*,$$

$A_\parallel$ and $A_\perp$ are the single scattering amplitudes for waves polarized in the scattering plane and normally to this plane,

$$\cos\beta = 1 - \frac{2(1-\mu^2)(1-\cos^2\psi)}{1-\cos^2\xi},$$

$$\sin\beta = \frac{2\sqrt{1-\mu^2}(\mu\sqrt{1-\mu'^2} - \mu'\sqrt{1-\mu^2}\cos\psi)\sin\psi}{1-\cos^2\xi},$$

$$\cos\xi = \mu\mu' + \sqrt{(1-\mu^2)(1-\mu'^2)}\cos\psi,$$

$$\mu = \cos\theta, \quad \mu' = \cos\theta', \quad \psi = \varphi - \varphi'.$$

Quantities $\cos\beta'$ and $\sin\beta'$ are obtained from $\cos\beta$ and $\sin\beta$ by replacing μ and φ with μ' and φ'. The unit vectors $\mathbf{n}' = \{\sin\theta'\cos\varphi', \sin\theta'\sin\varphi', \cos\theta'\}$ and $\mathbf{n} = \{\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta\}$ define the directions of photon propagation before and after scattering.

The boundary condition for Eqs. (1) and (2) for the unit flux density of incident light has the form

$$\left.\begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix}\right|_{z=0,\mu>0} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \delta(1 - \mathbf{n}\cdot\mathbf{n}_0), \tag{3}$$

where $\mathbf{n}_0$ is the vector of the interior normal to the medium surface.

Within the small-angle approximation, a solution of Eq. (1) was studied in detail previously.[9] In this work we focus our attention on the second pair of equations (2). As in our earlier study,[9] it seems interesting to analyze separately the cases of single scattering in the Born approximation ($a|n-1| \ll \lambda$) and in the approximation of strong (non-Born) scattering ($a|n-1| > \lambda$). In both these cases, functions a_i and b_i ($i = 1,2$) are known and given in monographs.[13,14]

3. BORN SCATTERERS ($a|n-1| \ll \lambda$)

In the Born approximation, the scattering amplitudes $A_\parallel(\cos\gamma)$ and $A_\perp(\cos\gamma)$ are related by the equation[13,14]

$$A_\parallel(\cos\gamma) = A_\perp(\cos\gamma)\cos\gamma. \tag{4}$$

Therefore, the following formulas apply to a_i and b_i:

$$a_1 = |A_\perp|^2 \frac{1+\cos^2\gamma}{2}, \quad a_2 |A_\perp|^2 \cos\gamma,$$

$$b_1 = |A_\perp|^2 \frac{\cos^2\gamma - 1}{2}, \quad b_2 = 0. \tag{5}$$

It follows from the latter equation and boundary condition (3) that Stokes's third parameter U is identically zero, $U \equiv 0$, therefore equation system (2) is reduced to a single equation:

$$\left\{\mu\frac{\partial}{\partial z} + n_0\sigma_{\mathrm{tot}}\right\} V(z,\mu) = n_0 \int d\mathbf{n}' |A_\perp(\mathbf{n}\cdot\mathbf{n}')|^2 \times (\mathbf{n}\cdot\mathbf{n}') V(z,\mu'). \tag{6}$$

According to Eq. (5), the differential scattering cross sections in the equations for the intensity and Stokes's fourth parameter in the range of small angles are identical for terms of orders up to γ^2 ($a_1 \approx a_2$). Therefore, an equation derived from Eq. (6) after the conventional small-angle expansion (i.e., when only terms of order of θ^2 and θ'^2 are retained[9]) will be identical to the small-angle scalar transport equation for I_{scal}.

Our analysis has demonstrated that this approximation is inadequate. If we assume that $V \approx I_{\mathrm{scal}}$, the polarization degree

$$P = \sqrt{V^2 + Q^2}/I \tag{7}$$

may be larger than unity, which is absurd.

In order to avoid this error, one should take account of the difference between V and I_{scal}. Let us write Eq. (6) in the form

$$\left\{\mu\frac{\partial}{\partial z} + n_0\sigma_{\mathrm{tot}}\right\} V(z,\mu) = n_0 \int d\mathbf{n}' a_1(\mathbf{n}\cdot\mathbf{n}') V(z,\mu') - n_0 \int d\mathbf{n}' \frac{d\sigma_{\mathrm{dep}}(\mathbf{n}\cdot\mathbf{n}')}{d\mathbf{n}'} V(z,\mu'). \tag{8}$$

Equation (8) is identical to the scalar transport equation for intensity of light in the presence of two types of scattering in a medium, namely, the conventional scattering with the differential cross section

$$\frac{d\sigma}{d\mathbf{n}} = a_1 = \frac{1}{2}|A_\perp(\mathbf{n}\cdot\mathbf{n}')|^2[1 + (\mathbf{n}\cdot\mathbf{n}')^2]$$

and the ''depolarizing'' scattering with the cross section

$$\frac{d\sigma_{\mathrm{dep}}}{d\mathbf{n}} = \frac{1}{2}|A_\perp(\mathbf{n}\cdot\mathbf{n}')|^2(1 - \mathbf{n}\cdot\mathbf{n}')^2. \tag{9}$$

The cross section given by Eq. (9) is proportional to that of transformation between the left-hand and right-hand polarized light. One can easily check out this statement using the definitions $I = I_+ + I_-$ and $V = I_+ - I_-$, where $I_\pm$ are the intensities of left- and right-hand polarized light, and deriving equations for $I_\pm$ by combining the equations for I and V. In the equations for $I_\pm$ the scattering differential cross section is $d\sigma - (1/2)d\sigma_{\mathrm{dep}}$, and that of transformation between the left- and right-hand polarized light is $(1/2)d\sigma_{\mathrm{dep}}$.

The form factor $|A_\perp|^2$ for weakly scattering spheres ($a|n-1| \ll \lambda$) averaged over the angular scale $\Delta\gamma > \lambda/a$ drops in the range of small angles following the power law[13]

$$|A_\perp|^2 \propto 1/\gamma^4. \tag{10}$$

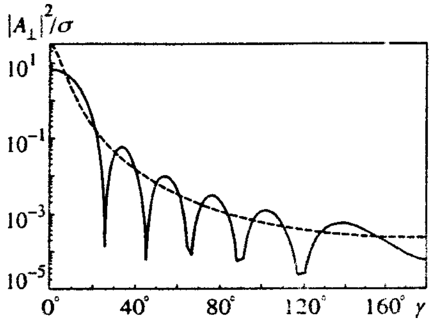

FIG. 1. Form factor $|A_\perp|^2$ for transparent scatterers. The solid curve shows calculations with the Mie formulas ($n-1=0.01$, $a=10\lambda/2\pi$), the dashed line is the approximation given by Eq. (11) ($\gamma_0=0.1$).

Therefore one can describe $|A_\perp|^2$ by the model function[15]

$$|A_\perp(\cos\gamma)|^2=\frac{\sigma\gamma_0^2}{\pi[\gamma_0^2+2(1-\cos\gamma)]^2}, \tag{11}$$

where $\sigma=(2\pi)^3a^4|n-1|^2/\lambda^2$ (Ref. 13) and $\gamma_0=\lambda/2\pi a$ are the cross section and characteristic angle of a single scattering event, respectively.[1)] Direct comparison (Fig. 1) clearly shows that function (11) adequately describes the behavior of the form factor calculated by Mie formulas[13] over the entire range of scattering angles.

By substituting expression (11) in Eq. (9), we obtain the depolarization cross section

$$\frac{d\sigma_{\text{dep}}}{d\mathbf{n}}=\frac{\sigma\gamma_0^2(1-\cos\gamma)^2}{2\pi[\gamma_0^2+2(1-\cos\gamma)]^2}. \tag{12}$$

In the most interesting case of multiple scattering, the scattering angle θ is larger than the characteristic angle of single scattering, $\theta>\gamma_0$. Under these conditions, the depolarization cross section is independent of the angular variables:

$$\frac{d\sigma_{\text{dep}}}{d\mathbf{n}}\approx\frac{\sigma\gamma_0^2}{8\pi}=\frac{\sigma_{\text{tr}}}{8\pi\ln(2/\gamma_0)}, \tag{13}$$

where $\sigma_{\text{tr}}=\sigma\gamma_0^2\ln(2/\gamma_0)$ is the transport cross section of elastic scattering.[9] As a result, we have for the mean free path l_{dep} with respect to a depolarizing collision

$$l_{\text{dep}}=\left(n_0\int d\sigma_{\text{dep}}\right)^{-1}=l_{\text{tr}}\ln\frac{4}{\gamma_0^2}, \tag{14}$$

where $l_{\text{tr}}=(n_0\sigma_{\text{tr}})^{-1}$ is the transport mean free path. According to Eqs. (12)–(14), the depolarization length due to small-angle scattering is inversely proportional to the fourth power of the single-scattering angle, $l_{\text{dep}}\sim l/\langle\gamma^4\rangle$ ($l=(n_0\sigma)^{-1}$ is the mean free path) and is greater than the transport mean free path: $l_{\text{dep}}>l_{\text{tr}}\sim l/\langle\gamma^2\rangle$. This means that depolarizing collisions are relatively rare, and the second term on the right-hand side of Eq. (8) can be treated as a small correction, in comparison with the other terms in this equation. This circumstance allows one to solve Eq. (8) by an iterative technique, treating the solution of the conventional scalar transport equation as an initial approximation.

3.1. Weakly absorbing medium

Let us analyze transmission through a layer of a weakly absorbing medium $[l_{\text{tr}}<l_a=(n_0\sigma_a)^{-1}]$. In this case, the multiple scattering angle is small as long as the layer thickness is limited by the condition $L<l_{\text{tr}}$.

Treating the last term on the right-hand side of Eq. (8) as an inhomogeneous term, let us rewrite Eq. (8) in the integral form

$$\begin{aligned}\tilde{V}(z,\mu)&=\tilde{I}_{\text{scal}}(z,\mu)\\&\quad-n_0\int_0^z dz'\int d\mathbf{n}'\tilde{G}(z-z'|\mathbf{n},\mathbf{n}')\\&\quad\times\int d\mathbf{n}''\frac{d\sigma_{\text{dep}}(\mathbf{n}'\cdot\mathbf{n}'')}{d\mathbf{n}''}\tilde{V}(z',\mu''),\end{aligned} \tag{15}$$

where

$$\tilde{V}(z,\mu)=\exp(n_0\sigma_a z)V(z,\mu),$$
$$\tilde{I}_{\text{scal}}(z,\mu)=\exp(n_0\sigma_a z)I_{\text{scal}}(z,\mu),$$
$$\tilde{G}(z|\mathbf{n},\mathbf{n}')=\exp(n_0\sigma_a z)G(z|\mathbf{n},\mathbf{n}'),$$

I_{scal} is the solution of the scalar transport equation (i.e., Eq. (8) at $\sigma_{\text{dep}}=0$) for a normally incident light beam, and G is the Green's function of the same equation.

As follows from Eq. (15), in calculating Stokes's fourth parameter V, one must above all know the intensity I_{scal} and the Green's function of the scalar transport equation. One cannot, however, calculate these quantities in the general case even in the small-angle approximation, so we have to use results that apply only under certain conditions.

As concerns small-angle scattering in a relatively thin layer of a weakly absorbing medium ($L<l_{\text{tr}}<l_a$), we can neglect photon path-length fluctuations due to multiple scattering and use the standard version of the small-angle approximation.[9,14] In this approximation, the intensity $\tilde{I}_{\text{scal}}$ and Green's function $\tilde{G}$ of the scalar transport equation are well known:[14,17]

$$\begin{aligned}\tilde{I}_{\text{scal}}(z,\theta)&=\frac{1}{2\pi}\int_0^\infty\omega d\omega J_0(\omega\theta)\\&\quad\times\exp[-n_0\sigma(1-\chi(\omega))z],\end{aligned} \tag{16}$$

$$\begin{aligned}\tilde{G}(z|\mathbf{n},\mathbf{n}')\approx\tilde{G}(z|\boldsymbol{\theta}-\boldsymbol{\theta}')&=\frac{1}{2\pi}\int_0^\infty\omega d\omega J_0\\&\quad\times(\omega|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\exp\{-n_0\sigma[1-\chi(\omega)]z\},\end{aligned} \tag{17}$$

where $\boldsymbol{\theta}$ is the component of vector $\mathbf{n}$ parallel to the interface, $J_0(x)$ is the zero-order Bessel function,[18] and

$$\begin{aligned}\chi(\omega)&=\frac{2\pi}{\sigma}\int_0^\infty\theta d\theta J_0(\omega\theta)a_1(\theta)\\&\approx1-\frac{1}{2}(\omega\gamma_0)^2\ln\frac{1}{\omega\gamma_0}.\end{aligned} \tag{18}$$

In the limiting cases of relatively small and large angles, the intensity $\tilde{I}_{\text{scal}}$ in Eq. (16) is[9,19]

$$\tilde{I}_{\rm scal}(z,\theta)=\begin{cases}\dfrac{l_{\rm tr}\ln(2/\gamma_0)}{\pi z\ln(z/l)}, & \theta<\theta_z,\\[2ex] \dfrac{z}{\pi\theta^4 l_{\rm tr}\ln(2/\gamma_0)}\left(1+\dfrac{8z}{\theta^2 l_{\rm tr}\ln(2/\gamma_0)}\ln\dfrac{\theta}{\gamma_0}\right), & \theta>\theta_z,\end{cases} \tag{19}$$

where $\theta_z=\gamma_0\sqrt{(z/l)\ln(z/l)}$ is the characteristic scattering angle in a layer of thickness $z\gg l$.

By substituting expressions (13) and (16)–(18) into Eq. (15) and performing one iteration in Eq. (15), we obtain Stokes's fourth parameter $\tilde{V}$:

$$\tilde{V}(z,\theta)\approx\tilde{I}_{\rm scal}(z,\theta)-\frac{z}{4\pi l_{\rm dep}}+\ldots \tag{20}$$

The applicability of Eq. (20) is governed by the condition of small-angle scattering ($\theta<1$).

Expression (7) for the polarization degree can be conveniently transformed to

$$P=\sqrt{S^2+P_{\rm un}^2}, \tag{21}$$

where $S=\tilde{V}/\tilde{I}$ is the photon helicity[20,21] and $P_{\rm un}=Q/I$ is the polarization degree of the initially unpolarized light beam. Our analysis indicates that, in calculating P, one can neglect polarization corrections to the intensity and set $I\approx I_{\rm scal}$. In this approximation, $P_{\rm un}$ is given by the expression[9]

$$P_{\rm un}=\begin{cases}-\dfrac{\theta^2}{4\ln(z/l)}, & \theta<\theta_z,\\[2ex] -\dfrac{\theta^2}{2}\left(1-\dfrac{8z}{\theta^2 l_{\rm tr}\ln(2/\gamma_0)}\ln\dfrac{\theta}{\gamma_0}\right), & \theta>\theta_z.\end{cases} \tag{22}$$

Now, by substituting (19), (20), and (22) into Eq. (21), we find that in the lowest order in the small parameter $l_{\rm tr}/l_{\rm dep}\ll 1$, the polarization degree around the peak ($\theta<\theta_z$) of the angular distribution is constant with θ and determined by the expression

$$P=1-\frac{\theta_z^4}{8\ln(z/l)}, \tag{23}$$

whereas in the wings ($\theta\gg\theta_z$) of the distribution

$$P=1-\theta^2\theta_z^2\,\frac{\ln(\theta/\gamma_0)}{\ln(z/l)}. \tag{24}$$

It follows from Eqs. (23) and (24) that, in transmission through a relatively thin layer ($z<l_{\rm tr}$), the depolarization is low and the polarization degree is close to unity ($\delta P=1-P\ll 1$). The depolarization is the greatest at relatively large ($\theta>\theta_z$) scattering angles: $\delta P(\theta<\theta_z)/\delta P(\theta>\theta_z)\propto\theta_z^2/\theta^2\ll 1$.

The process of light depolarization can be described qualitatively as follows. At a small depth ($z<l$), where the dominant process is single scattering, the polarization degree is unity.[13] With increasing z ($z>l$), P drops. In the region of small angles, $\theta<\theta_z$, the drop in P is fully determined by the transformation of the circularly polarized component to unpolarized light due to multiple scattering. The relative contribution of the linearly polarized component, which is generated in scattering of the depolarized component,[9] $P_{\rm un}$, is negligible. At $\theta>\theta_z$ the angular distribution is determined by light waves that undergo multiple scattering through small ($\theta\ll\theta_z$) angles and a single deviation through a large angle ($\theta>\theta_z$).[9] In the angular range in question, the drop in the helicity in the first order in the small ratio $\theta_z/\theta\ll 1$, $S\approx 1-\theta^4/8$, as in the case of single scattering, is offset by the linearly polarized component generated by depolarized radiation, $P_{\rm un}\approx-\theta^2/2$. The difference between P and unity at $\theta>\theta_z$ is determined by corrections of higher orders in θ_z/θ and is proportional to $\theta^2\theta_z^2$.

3.2. Medium with high absorption

Now consider propagation of circularly polarized light in a medium with high absorption ($l_a<l_{\rm tr}$), when the small-angle approximation holds at all penetration depths.[10,11] The evolution of the angular distribution with the increase in z in such a medium proceeds as follows.[9,19]

At relatively small depths ($l<z<l_a$) the effect of absorption on the intensity distribution is negligible, and the intensity is described by Eq. (19). With increasing z ($z>l_a$) owing to the absorption of photons scattered through large angles, the "Rutherford" law for the intensity decrease, $\tilde{I}\propto\theta^{-4}$ [Eq. (19)], is replaced in the region $\theta>\sqrt{l_a/z}$ by a faster function[9,19]:

$$\tilde{I}_{\rm scal}(z,\theta)=\frac{2\sigma_{\rm tr}\tilde{E}(z)}{\pi\sigma_a\ln(2/\gamma_0)\,\theta^6}\left(1+5\frac{\langle\theta^2\rangle_z}{\theta^2}+\cdots\right), \tag{25}$$

where $\tilde{E}(z)=\exp(z/l_a)E(z)$ and $\langle\theta^2\rangle_z$ are the "reduced" flux and mean square of multiple scattering angle at range z, respectively.

Prior to the transition to the asymptotic propagation regime ($z<l_d$, where l_d is the length over which the "reduced" flux decays), the dependence $\tilde{E}(z)$ can be neglected, and $\tilde{E}(z)\approx 1$. Quantity $\langle\theta^2\rangle_z$ in the region of z under consideration is given by

$$\langle\theta^2\rangle_z=\frac{2\pi}{\tilde{E}(z)}\int_0^\infty\theta^3 d\theta\,\tilde{I}_{\rm scal}(z,\theta)\approx\frac{2z}{l_{\rm tr}}\,\frac{\ln(\sqrt{l_a/z}/\theta_z)}{\ln(2/\gamma_0)}, \tag{26}$$

At depths $z>l_d$ the "Rutherford" region in the angular distribution vanishes and the intensity is[9]

$$\tilde{I}_{\rm scal}(z,\theta)\approx\exp\left(-\frac{z}{l_d}\right)\times\begin{cases}\dfrac{2}{\pi\langle\theta^2\rangle_\infty}\left(1-\dfrac{\theta^2}{\langle\theta^2\rangle_\infty}+\cdots\right), & \theta<\sqrt{\langle\theta^2\rangle_\infty},\\[2ex] \dfrac{4\sigma_{\rm tr}\gamma_0^2}{\pi\sigma_a\theta^6\ln(2/\gamma_0)}\left(1+5\dfrac{\langle\theta^2\rangle_\infty}{\theta^2}\right), & \theta>\sqrt{\langle\theta^2\rangle_\infty},\end{cases} \tag{27}$$

where $\langle\theta^2\rangle_\infty=2l_a/l_d$ is the mean square multiple scattering angle in the asymptotic regime. According to earlier calculations,[9] the length l_d is given by

$$l_d=\sqrt{2l_a l_{\rm tr}\ln(2/\gamma_0)/\ln\sqrt{2l_a/l\gamma_0^2}}. \tag{28}$$

It follows from the equation relating $\tilde V$ to $\tilde I_{\rm scal}$ [Eq. (15)] and from the above statements concerning $\tilde I_{\rm scal}$ that Stokes's fourth parameter in a highly absorbing medium at a depth $z<l_d$ at relatively small angles $\theta<\sqrt{l_a/z}$ should be the same as in a weakly absorbing medium ($l_a<l_{\rm tr}$), and therefore it should be described by Eqs. (23) and (24).

The effect of absorption on $\tilde V$ begins to show up at $z>l_a$ in the region of angles $\theta>\sqrt{l_a/z}$ [Eq. (25)]. At larger z the absorption radically changes the angular dependence of Stokes's fourth parameter $\tilde V$.

At large scattering angles ($\theta>\sqrt{l_a/z}$ for $z<l_d$ and $\theta>\sqrt{\langle\theta^2\rangle_\infty}$ for $z>l_d$), the correction to $\tilde V$ due to the last term on the right-hand side of Eq. (8) drops following the law $\delta\tilde V\propto 1/\theta^2$ [compare with Eq. (20)]. This follows directly from Eq. (8).

By transferring $n_0\sigma V$ from the left to right-hand side and estimating separately the contributions of the elastic scattering integral[9] and the term responsible for depolarization at relatively large angles, we obtain for $\delta\tilde V(z,\mu)$

$$\delta\tilde V(z,\mu)\approx-\frac{n_0\sigma_{\rm dep}}{4\pi\mu}\int_0^z dz'\times\exp\left(-\frac{n_0\sigma_a(z-z')(1-\mu)}{\mu}\right)\tilde E^{(V)}(z')\approx-\frac{\sigma_{\rm dep}}{2\pi\sigma_a\theta^2}\tilde E^{(V)}(z), \tag{29}$$

where

$$\tilde E^{(V)}(z)=\int d\mathbf{n}\tilde V(z,\mu). \tag{30}$$

Consequently, like the intensity in our previous calculations,[9,19], $\delta\tilde V$ at large θ can be expressed in terms of the integral of $\tilde V$ over angles, i.e., in terms of the spatial ''density'' of $\tilde E^{(V)}$.

The correction $\delta\tilde V$ is the component of $\tilde V$ that most slowly decreases with θ.

Separating out the contribution given by Eq. (29) from $\tilde V(z,\mu)$, one can develop an iterative procedure for calculating $\tilde V(z,\mu)$ at relatively large angles θ. With this end in view, let us express $\tilde V$ in the form ($\mu>0$)

$$\tilde V(z,\mu)=-\frac{n_0\sigma_{\rm dep}}{4\pi\mu}\int_0^z dz'\times\exp\left(-\frac{n_0\sigma_a(z-z')(1-\mu)}{\mu}\right)\times\tilde E^{(V)}(z')+\tilde v(z,\mu) \tag{31}$$

and substitute Eq. (31) into (8). As a result, we obtain for $\tilde v(z,\mu)$ an equation which, unlike Eq. (8), can be treated in the small-angle approximation (whereas in Eq. (8) this cannot be done owing to the last term on the right-hand side, which is constant in angle). Using the conventional small-angle expansion in the equation for $\tilde v(z,\mu)$, i.e., setting $1-\mu\approx\theta^2/2$ and retaining nonvanishing terms of the lowest orders,[9,11] we obtain

$$\left\{\frac{\partial}{\partial z}+n_0\sigma_a\frac{\theta^2}{2}\right\}\tilde v(z,\theta)=n_0\times\int d\boldsymbol{\theta}'a_1(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)(\tilde v(z,\theta')-\tilde v(z,\theta))-n_0\times\int d\boldsymbol{\theta}'a_1(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\frac{n_0\sigma_{\rm dep}}{4\pi}\int_0^z dz'\tilde E^{(V)}(z')\times\left[\exp\left(-\frac{n_0\sigma_a(z-z')\theta'^2}{2}\right)-\exp\left(-\frac{n_0\sigma_a(z-z')\theta^2}{2}\right)\right]. \tag{32}$$

The boundary condition for Eq. (32) has the form

$$\tilde v(z=0,\theta)=\delta(\theta)/2\pi\theta.$$

Now, using the power expansion in terms of reciprocal θ, which was developed earlier for the intensity $\tilde I_{\rm scal}$, [9,19] we can calculate an asymptotic expansion of $\tilde v(z,\theta)$ (see Appendix A).

In the approximation of the lowest order in the small parameter $\sigma_{\rm dep}/\sigma_a\ll 1$, Stokes's fourth parameter is determined in the ''wings'' of the angular distribution by the expression

$$\tilde V(z,\theta)\approx\tilde E^{(V)}(z)\times\left[-\frac{\sigma_{\rm dep}}{2\pi\sigma_a\theta^2}\left(1+\frac{\langle\theta^2\rangle_z}{\theta^2}+\cdots\right)+\frac{\tilde I_{\rm scal}(z,\theta)}{\tilde E(z)}\right]=\tilde E^{(V)}(z)\left[-\frac{\sigma_{\rm dep}}{2\pi\sigma_a\theta^2}\left(1+\frac{\langle\theta^2\rangle_z}{\theta^2}+\cdots\right)+\frac{4\sigma_{\rm dep}}{\pi\sigma_a\theta^6}\left(1+5\frac{\langle\theta^2\rangle_z}{\theta^2}+\cdots\right)\right], \tag{33}$$

where $\tilde I_{\rm scal}(z,\theta)$ is the intensity (25). Using the earlier results[9,19] for

$$P_{\rm un}\approx-\frac{\theta^2}{2}\left(1-4\frac{\langle\theta^2\rangle_z}{\theta^2}\right), \tag{34}$$

we can derive from Eqs. (25) and (33) the polarization degree of circularly polarized light in the region of relatively large scattering angles:

$$P\approx\sqrt{\left(1-\frac{\theta^4}{4}+\theta^2\langle\theta^2\rangle_z\right)\left(\frac{\tilde{E}^{(V)}(z)}{\tilde{E}(z)}\right)^2+\frac{\theta^4}{4}\left(1-8\frac{\langle\theta^2\rangle_z}{\theta^2}\right)}\,. \tag{35}$$

The first summand under the radical sign in Eq. (35) describes the drop in the helicity of circularly polarized light, the second one is due to the linear polarization generated in scattering of depolarized light. At relatively small depths, when $\tilde{E}^{(V)}(z)\approx\tilde{E}(z)\approx1$, we derive from Eq. (35)

$$P\approx1-\frac{1}{2}\theta^2\langle\theta^2\rangle_z\,. \tag{36}$$

The comparison between Eqs. (36) and (24) shows that absorption leads to slowing down of the polarization degree decay. The cause is the drop in the number of the ''most depolarized'' photons scattered through relatively large angles.

Now let us analyze the polarization degree in the asymptotic ($z>l_d$) regime, when the effect of absorption on light propagation is significant throughout the entire range of scattering angles. In this case, the solution of Eq. (8) should have the form[22]

$$\tilde{V}(z,\mu)=\tilde{V}(\mu)\exp(-\varepsilon_0^{(V)}z), \tag{37}$$

where

$$\tilde{V}(\mu)=\psi_0^{(V)}(\mu=1)\,\psi_0^{(V)}(\mu),$$

and $\psi_0^{(V)}(\mu)$ and $\varepsilon_0^{(V)}$ are the first eigenfunction and the first eigenvalue of the spectral problem corresponding to Eq. (8).

The angular function $\tilde{V}(\mu)$ and damping constant $\varepsilon_0^{(V)}$ can be easily calculated by the perturbation theory on the basis of our earlier results[9] concerning the intensity at large depths (Appendix B). In the approximation of the first order in small parameter $\sigma_{\mathrm{dep}}/\sigma_a\ll1$, we obtain [see Eq. (B3)]

$$\varepsilon_0^{(V)}=\frac{1}{l_d}\left(1+\frac{\sigma_{\mathrm{dep}}}{\sigma_a}\right). \tag{38}$$

Within the same accuracy as in Eq. (38), function $\tilde{V}(\theta)$ around the peak of the angular distribution ($\theta<\sqrt{\langle\theta\rangle_\infty}$) is determined by the following expression [see Eqs. (27) and (B5)]:

$$\tilde{V}(\theta)=\tilde{I}_{\mathrm{scal}}(\theta)+\delta\tilde{V}(\theta)\approx\frac{2}{\pi\langle\theta\rangle_\infty}+\frac{\sigma_{\mathrm{dep}}}{\pi\sigma_a\langle\theta\rangle_\infty}\ln2. \tag{39}$$

It follows from Eq. (39) that at $\theta<\sqrt{\langle\theta^2\rangle_\infty}$ the contributions of $\delta\tilde{V}(\theta)$ to the helicity S and ''density'' $\tilde{E}^{(V)}(z)$ can be neglected in the first approximation at $z/l_d>1$. As a result, the polarization degree in this angular region is given by the expression

$$P\approx\sqrt{\exp\left[-\frac{2}{l_d}\left(\frac{l_a}{l_{\mathrm{dep}}}\right)z\right]+\left(\frac{\ln2}{2}\,\frac{l_a\gamma_0^2\theta^2}{l\langle\theta^2\rangle_\infty^2}\right)^2}\,. \tag{40}$$

In deriving Eq. (40) we have taken into account the relation[9,19]

$$P_{\mathrm{un}}=-\frac{\ln2}{2}\,\frac{l_a\gamma_0^2\theta^2}{l\langle\theta^2\rangle_\infty^2},\qquad\theta<\sqrt{\langle\theta^2\rangle_\infty}. \tag{41}$$

A formula similar to Eq. (40) can be obtained for the ''wings'' of the angular distribution. With this end in view, we should use the formula[9,19]

$$P_{\mathrm{un}}=-\frac{\theta^2}{2}\left[1-\frac{4\langle\theta^2\rangle_\infty}{\theta^2}\right],\qquad\theta>\sqrt{\langle\theta^2\rangle_\infty} \tag{42}$$

and the expression for $\tilde{E}^{(V)}(z)$ at large penetration depths:

$$\tilde{E}^{(V)}(z)\approx\tilde{E}(z)\exp\left[-\frac{1}{l_d}\left(\frac{l_a}{l_{\mathrm{dep}}}\right)z\right]. \tag{43}$$

The accuracy of this result is the same as that of Eq. (40).

Substituting Eqs. (42) and (43) into expression (35) for the polarization degree of circularly polarized light in the ''wings'' of the angular distribution, we finally obtain in the asymptotic regime ($z>l_d$)

$$P\approx\sqrt{\left(1-\frac{\theta^4}{4}+\theta^2\langle\theta^2\rangle_\infty\right)\exp\left[-\frac{2}{l_d}\left(\frac{l_a}{l_{\mathrm{dep}}}\right)z\right]+\frac{\theta^4}{4}-2\theta^2\langle\theta^2\rangle_\infty}\,. \tag{44}$$

It follows from Eqs. (40) and (44) that the length of polarization decay

$$l_{\mathrm{circ}}=l_dl_{\mathrm{dep}}/l_a \tag{45}$$

is much larger than the depth l_d, at which a transition to the asymptotic regime occurs.

According to Eqs. (40) and (44), around the angular distribution peak for $z\approx l_{\mathrm{circ}}$, the process transforming the circularly polarized component to unpolarized light is dominant. The polarization degree drops to

$$P\approx1-z/l_{\mathrm{circ}}\,. \tag{46}$$

in the ''wings'' of the angular distribution, the depolarization of the circular component is partially compensated by generation of linearly polarized light, and the polarization degree is determined by the following expression:

$$P\approx1-\frac{z}{l_{\mathrm{circ}}}-\frac{1}{2}\theta^2\langle\theta^2\rangle_\infty\,. \tag{47}$$

As z increases ($z>l_{\mathrm{circ}}$), the circularly polarized component vanishes, and the polarization degree coincides with that for an initially unpolarized beam, $P\approx P_{\mathrm{un}}$.

4. STRONG SCATTERERS $(a|n-1|>\lambda)$

Consider multiple scattering of circularly polarized light in a medium containing transparent $(a\,\mathrm{Im}\,n\ll\lambda)$ spheres of large radius $(a(n-1)>\lambda)$. Such a configuration can be easily realized in an experiment.[3,4,23–25]

In the region of angles larger than the diffraction angle $(\gamma\gg\lambda/a)$, the differential cross sections $a_i(\cos\gamma)$ and $b_i(\cos\gamma)$ can be expressed as expansions in terms of the number of ray collisions with a scatterer surface[13]:

$$a_1(\cos\gamma)=\frac{1}{2}\sum_{j=1}^{3}\left[|A_\parallel^{(j)}(\cos\gamma)|^2+|A_\perp^{(j)}(\cos\gamma)|^2\right],\qquad(48)$$

$$b_1(\cos\gamma)=\frac{1}{2}\sum_{j=1}^{3}\left[|A_\parallel^{(j)}(\cos\gamma)|^2-|A_\perp^{(j)}(\cos\gamma)|^2\right],$$

$$a_2(\cos\gamma)=\mathrm{Re}\sum_{j=1}^{3}A_\parallel^{(j)}(\cos\gamma)A_\perp^{(j)*}(\cos\gamma)$$

$$=a_1(\cos\gamma)-\frac{1}{2}\sum_{j=1}^{3}|A_\parallel^{(j)}(\cos\gamma)-A_\perp^{(j)}(\cos\gamma)|^2,\qquad(49)$$

$$b_2(\cos\gamma)=\mathrm{Im}\sum_{j=1}^{3}A_\parallel^{(j)}(\cos\gamma)A_\perp^{(j)*}(\cos\gamma).\qquad(50)$$

These expansions (48)–(50) are derived under the assumption that interference between rays that have undergone different numbers of reflections can be neglected.

Amplitudes $A_{\parallel,\perp}^{(j)}$ characterize rays deviating as a result of specular reflection $(j=1)$, those passing through a scatterer $(j=2)$, and those which are additionally reflected within a scatterer $(j=3)$. At scattering angles $\gamma\ll\sqrt{n-1}$, the dominant terms in a_1, a_2, and b_1 are the terms with $j=2$, whereas at $\gamma\gg 2\sqrt{2(n-1)}$ these terms are identically zero. Terms with $j=1$, 3 are important when $\gamma>\gamma_{\mathrm{cr}}=2\sqrt{2(n-1)}$. The expressions for a_1 and b_2 in these regions of angles are given in Ref. 9. The depolarization cross section [compare with Eq. (9)] is determined in the general case by the formula

$$\frac{d\sigma_{\mathrm{dep}}}{d\mathbf{n}}=\frac{1}{2}|A_\parallel-A_\perp|^2\qquad(51)$$

and, with due account of Eqs. (48)–(50), can be expressed as

$$\frac{d\sigma_{\mathrm{dep}}}{d\mathbf{n}}=\frac{\sigma(n-1)^2}{16\pi}\begin{cases}1, & \gamma<\sqrt{n-1},\\ 2, & \gamma>\sqrt{n-1}.\end{cases}\qquad(52)$$

The latter case in Eq. (52) applies to the angles $\gamma\le\pi$. Using Eq. (52), we calculate the total depolarization cross section σ_{dep}:

$$\sigma_{\mathrm{dep}}=\frac{\sigma_{\mathrm{tr}}}{3\ln[1/(n-1)]},\qquad(53)$$

where

$$\sigma_{\mathrm{tr}}=\frac{3}{2}\sigma(n-1)^2\ln\frac{1}{n-1}\qquad(54)$$

is the transport cross section in the case of strong scattering.[9] The mean free path with respect to a depolarizing collision, as in the case of the Born approximation [Eq. (14)], is greater than the transport mean free path: $l_{\mathrm{dep}}/l_{\mathrm{tr}}\sim\ln[1/(n-1)^3]\gg 1$.

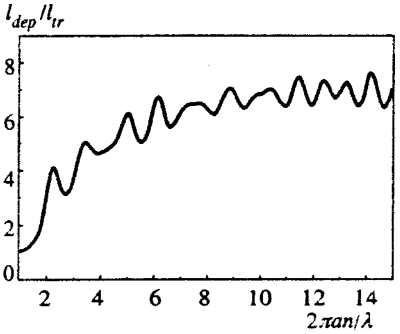

FIG. 2. Curve of $l_{\mathrm{dep}}/l_{\mathrm{tr}}$ as a function of scatterer radius (latex particles in water, $n-1=0.2$, λ is the wavelength of light in vacuum).

As an illustration, Fig. 2 plots calculations of the ratio $l_{\mathrm{dep}}/l_{\mathrm{tr}}$ as a function of the scatterer radius based on the exact Mie formulas.[13] These calculations were performed for particles of latex in water. This is a disordered medium extensively used in experiments on multiple light scattering.[4,26]

As concerns the cross section b_2 defined by Eq. (50), it is known[27] that for $j=2$,3 the following equality holds in the absence of absorption:

$$\mathrm{Im}\,A_\parallel^{(2,3)}A_\perp^{(2,3)*}=0,\qquad(55)$$

whereas the contribution of $\mathrm{Im}\,A_\parallel^{(1)}A_\perp^{(1)*}$, which corresponds to specularly reflected rays,[13,27] can be expressed as

$$\mathrm{Im}\,A_\parallel^{(1)}A_\perp^{(1)*}=\frac{a^2}{4}\sin(\delta_\parallel-\delta_\perp),\qquad(56)$$

where $\delta_{\parallel,\perp}$ are phase shifts of light waves with polarization vectors parallel and perpendicular to the incidence plane in the Fresnel reflection from a particle surface. In the absence of absorption, the phase difference $\delta_\perp-\delta_\parallel$ is zero if the optical density of scatterers is higher $(n>1)$ than that of ambient medium. Therefore the cross section b_2 also turns to zero in this case, $b_2=0$.

Let us consider propagation of light through a disordered medium containing transparent scatterers immersed in a highly absorbing material $(l_a<l_{\mathrm{tr}})$. Suppose that $n>1$. In these conditions, Stokes's third parameter, as in the case of the Born approximation, is identically zero, $U\equiv 0$. Parameter V is governed by Eq. (8) with a depolarization cross section given by Eq. (52). The routine for solving Eq. (8) remains the same (see the previous section). The details of the multiple scattering process, however, are more complicated.

In the range of angles where absorption does not affect the shape of the angular distribution $(\theta<\sqrt{l_a/z})$[9] the first iteration in Eq. (15) yields

$$\tilde{V}(z,\mu)\approx\tilde{I}_{\mathrm{scal}}(z,\mu)-n_0\int_0^z dz'\int d\mathbf{n}'\tilde{G}(z-z'|\mathbf{n},\mathbf{n}')\times\int d\mathbf{n}''\frac{d\sigma_{\mathrm{dep}}(\mathbf{n}'\cdot\mathbf{n}'')}{d\mathbf{n}''}\tilde{I}_{\mathrm{scal}}(z',\mu''). \tag{57}$$

At $z<l/(n-1)\ln[1/(n-1)]$ the characteristic angle of multiple scattering,

$$\theta_z=\sqrt{(z/l)(n-1)^2\ln(z/l)}, \tag{58}$$

is less than the angle γ_{cr} of total internal reflection: $\theta_z<\gamma_{\mathrm{cr}}=2\sqrt{2(n-1)}$.[27] Taking into account the characteristic range of angles in which integration over $\mathbf{n}'$ and $\mathbf{n}''$ is performed being of order of θ_z and using Eq. (52), we obtain the following formula for $\tilde{V}$:

$$\tilde{V}(z,\theta)=\begin{cases}\tilde{I}_{\mathrm{scal}}(z,\theta)-z(n-1)^2/16\pi l, & \theta<\gamma_{\mathrm{cr}},\\ \tilde{I}_{\mathrm{scal}}(z,\theta)-z(n-1)^2/8\pi l, & \gamma_{\mathrm{cr}}<\theta<\sqrt{l_a/z}.\end{cases} \tag{59}$$

The function $\tilde{I}_{\mathrm{scal}}(z,\theta)$ in Eq. (59) is given by[9]

$$\tilde{I}_{\mathrm{scal}}(z,\theta)=\begin{cases}\dfrac{l}{2\pi z(n-1)^2\ln(z/l)}, & \theta<\theta_z,\\[2ex] \dfrac{2z(n-1)^2}{\pi\theta^4 l}\left[1+\dfrac{8z(n-1)^2}{\theta^2 l}\ln\dfrac{\theta^2}{(n-1)^2}\right], & \theta_z<\theta<\gamma_{\mathrm{cr}},\\[2ex] \dfrac{z(n-1)^2}{\pi\theta^4 l}\left[1+\dfrac{4z(n-1)^2}{\theta^2 l}\ln\dfrac{\theta^2}{(n-1)^3}\right], & \gamma_{\mathrm{cr}}<\theta<\sqrt{\dfrac{l_a}{z}}.\end{cases} \tag{60}$$

In the case $z>l/(n-1)\ln[1/(n-1)]$, we have the expression for θ_z[9]

$$\theta_z=\sqrt{\frac{z}{l}(n-1)^2\ln\left(\frac{z}{l}\frac{1}{n-1}\right)}. \tag{61}$$

In this case, the characteristic range of angles $\Delta\theta\sim\theta_z$, which makes the main contribution to the integrals with respect to $\mathbf{n}'$ and $\mathbf{n}''$ in Eq. (57), extends beyond the critical angle γ_{cr}. Therefore we can use instead of the cross section $d\sigma_{\mathrm{dep}}/d\mathbf{n}$ its asymptotic expression for large ($\gamma>\gamma_{\mathrm{cr}}$) angles [see Eq. (52)]. As a result, we obtain

$$\tilde{V}(z,\theta)=\tilde{I}_{\mathrm{scal}}(z,\theta)-\frac{z(n-1)^2}{8\pi l}, \tag{62}$$

where

$$\tilde{I}_{\mathrm{scal}}(z,\theta)=\begin{cases}\dfrac{l}{\pi z(n-1)^2\ln[(z/l)/(n-1)]}, & \theta<\theta_z,\\[2ex] \dfrac{z(n-1)^2}{\pi\theta^4 l}\left[1+\dfrac{4z(n-1)^2}{\theta^2 l}\ln\dfrac{\theta^2}{(n-1)^3}\right], & \theta_z<\theta<\sqrt{\dfrac{l_a}{z}}.\end{cases} \tag{63}$$

In calculating the polarization degree (21), let us use our earlier results[9] for the polarization degree of an initially unpolarized beam, P_{un}. At relatively small depths, $z<l/(n-1)\ln[1/(n-1)]$

$$P_{\mathrm{un}}=\begin{cases}\dfrac{\theta^2}{8\ln(z/l)}, & \theta<\theta_z,\\[2ex] \dfrac{\theta^2}{4}\left[1-\dfrac{8z(n-1)^2}{\theta^2 l}\ln\dfrac{\theta^2}{(n-1)^2}\right], & \theta_z<\theta<\gamma_{\mathrm{cr}},\\[2ex] -\dfrac{\theta^2}{2}\left[1-\dfrac{4z(n-1)^2}{\theta^2 l}\ln\dfrac{\theta^2}{(n-1)^3}\right], & \gamma_{\mathrm{cr}}<\theta<\sqrt{\dfrac{l_a}{z}}.\end{cases} \tag{64}$$

At larger penetration depths z ($z>l/(n-1)\ln[1/(n-1)]$) the polarization degree P_{un} around the angular distribution peak ($\theta<\theta_z$) becomes[9]

$$P_{\mathrm{un}}=-\frac{\theta^2}{4\ln[(z/l)/(n-1)]}. \tag{65}$$

In the "wings" ($\theta>\theta_z$) P_{un} is given by the last line of Eq. (64).

Given these results, we find that a change in the light polarization in the region of very small angles ($\theta<\theta_z$) is controlled, as in the case of the Born approximation, by the depolarization of the circularly polarized component (i.e., by the decay of helicity). For $\theta<\theta_z$ we have

$$P=\begin{cases}1-\dfrac{\theta_z^4}{2\ln(z/l)}, & z<\dfrac{l}{(n-1)\ln[1/(n-1)]},\\[2ex] 1-\dfrac{\theta_z^4}{2\ln[(z/l)/(n-1)]}, & z>\dfrac{l}{(n-1)\ln[1/(n-1)]}.\end{cases} \tag{66}$$

In the approximation of the lowest order in small parameter θ_z/θ, the angular dependence of Stokes's parameters in the "wings" ($\theta>\theta_z$) of the angular distribution of scattered light coincides with the law of single scattering.[13] By retaining terms of higher orders in the reciprocal scattering angle in expansions of Stokes's parameters, we find that the polarization degree in the range of angles $\theta>\theta_z$ drops following the law

$$P\approx\begin{cases}1-\dfrac{z(n-1)}{2l}\theta^2\ln\dfrac{\theta}{n-1}, & \theta<\gamma_{\mathrm{cr}},\\[2ex] 1-\dfrac{z(n-1)^2}{2l}\theta^2\ln\dfrac{\theta^2}{(n-1)^3}, & \gamma_{\mathrm{cr}}<\theta<\sqrt{\dfrac{l_a}{z}}.\end{cases} \tag{67}$$

The effect of absorption on $\tilde{V}$ at relatively large scattering angles, $\theta>\sqrt{l_a/z}$, can be conveniently analyzed using the same method as in the case of Born scatterers. Let us express $\tilde{V}$ in the form similar to Eq. (31):

$$\tilde{V}(z,\mu) = -\frac{n_0}{\mu}\int_0^z dz' \exp\left(-\frac{n_0\sigma_a(z-z')(1-\mu)}{\mu}\right)$$
$$\times \int d\mathbf{n}' \frac{d\sigma_{\mathrm{dep}}(\mathbf{n}\cdot\mathbf{n}')}{d\mathbf{n}'}\tilde{V}(z',\mu') + \tilde{v}(z,\mu). \tag{68}$$

Substituting expression (68) in Eq. (8) and taking into account the fact that in the "wings" ($\theta > \sqrt{l_a/z} > \gamma_{\mathrm{cr}}$) [see also Eq. (52)]

$$\int d\mathbf{n}' \frac{d\sigma_{\mathrm{dep}}(\mathbf{n}\cdot\mathbf{n}')}{d\mathbf{n}'}\, \tilde{V}(z,\mu') \approx \frac{\sigma(n-1)^2}{8\pi}\tilde{E}^{(V)}(z), \tag{69}$$

we consider the resulting equation for $\tilde{v}(z,\mu)$ in the small-angle approximation.$^{2)}$ As a result, we obtain for the function $\tilde{v}(z,\theta)$ an equation identical to Eq. (32) but with the substitution $\gamma_0 = n-1$ in its coefficients. Therefore Eqs. (33) and (35) for $\tilde{V}(z,\theta)$ and $P(z,\theta)$ still hold in the approximation of the lowest order in small parameter $\sigma(n-1)^2/\sigma_a \ll 1$.

At small penetration depth $z < l_d$, where in the case of strong scatterers

$$l_d = \left(\sqrt{\frac{(n-1)^2}{2l_a l}\ln\sqrt{\frac{l_a}{l(n-1)^4}}}\right)^{-1}, \tag{70}$$

we have $\tilde{E}^{(V)}(z) \approx \tilde{E}(z) \approx 1$ and the polarization degree drops with z following Eq. (36), where the mean square multiple scattering angle $\langle\theta^2\rangle_z$ must be set to

$$\langle\theta^2\rangle_z = 2\frac{z}{l}(n-1)^2$$
$$\times \begin{cases} \ln\dfrac{\gamma_{\mathrm{cr}}\sqrt{l_a/z}}{\theta_z^2}, & z < \dfrac{l}{(n-1)\ln[1/(n-1)]}, \\[2ex] \ln\dfrac{\sqrt{l_a/z}}{\theta_z}, & z > \dfrac{l}{(n-1)\ln[1/(n-1)]}. \end{cases} \tag{71}$$

The characteristic angle θ_z in Eq. (71) is given by Eqs. (58) and (61).

At large penetration depths ($z > l_d$) and if the condition $l < l_a < l_{\mathrm{tr}}$ holds, the width of the radiation angular distribution is always larger than γ_{cr}. Therefore, qualitatively, the evolution of polarization parameters of multiply scattered light at these depths is not sensitive to the singularity in the single scattering cross section at $\gamma = \gamma_{\mathrm{cr}}$ and remains the same as in the Born approximation. In particular, the polarization degree is still determined by Eqs. (40) and (42). All changes affect only the parameters l_d [Eq. (70)] and $\langle\theta^2\rangle_\infty = 2l_a/l_d$, which contain additional contributions from rays twice crossing the boundaries of an isolated scatterer.

5. CONCLUSIONS

Let us summarize the results of our work.

As was shown above, the propagation of circularly polarized light in a disordered medium can be reduced to an equation for Stokes's parameter *V*, which has the form similar to the scalar transport equation for the light intensity. The only difference is the presence of an additional term responsible for depolarization.

The presence of the depolarization term in the equation for *V*, on the one hand, results in the faster decay of parameter *V* than that of the intensity, and on the other, it modifies the shape of *V* as a function of scattering angle. Since Stokes's parameter *V* decays with *z* faster than the intensity, circularly polarized light loses its polarization. In the case of small-angle light scattering in disordered media with large particles (larger than the light wavelength), the additional term is small, so the light depolarization should be very slow. Under conditions of high absorption, when multiple scattering at all *z* occurs through small angles, the depolarization occurs at anomalously large penetration depths ($z \sim l_{\mathrm{circ}} \gg l_a, l_{\mathrm{tr}}$) when the asymptotic regime of propagation is already set.

All said is in agreement with the recent discovery of the effect of circular polarization "conservation" in diffusion of light in media with large-scale inhomogeneities.[4,28] A degree of circular polarization can persist even after isotropization of the intensity angular distribution. The depolarization rate depends only on how smooth are the trajectories of wave propagation, and in the case of scattering by large-scale inhomogeneities it is proportional to the fourth power of the single scattering angle (unlike the second power of the angle in the isotropization rate $1/l_{\mathrm{tr}}$). Therefore, measurements of the depolarization of circularly polarized light can yield additional information on the medium, different from that derived from measurements of an angular distribution of light intensity.

In accordance with the results of this paper, the polarization degree of multiply scattered light is governed, on the one hand, by depolarization of circularly polarized light, and on the other, by the appearance of a linearly polarized component generated in scattering of previously depolarized light. In an single scattering, these two processes fully compensate for one another, therefore the polarization degree remains unity ($P=1$). A drop in *P* can occur only as a result of multiple scattering.

As long as the circularly polarized component persists ($z < l_{\mathrm{circ}}$), the contribution of the linear component in the region of small angles ($\theta < \theta_z$) is small and *P* is fully determined by the drop in the polarization helicity [see Eqs. (23), (46), and (66)]. In the "wings" of the angular distribution ($\theta > \theta_z$), the contribution of linearly polarized light is, on the contrary, quite significant. In this range of angles the difference between *P* and unity is due to the difference between the polarization degree of the initially unpolarized light, P_{un}, and its value in the single scattering approximation, $P_{\mathrm{un}}^{(s)} = -\theta^2/2$ [see Eqs. (24), (36), (47), and (67)].

The circularly polarized component decays with *z* ($z > l_{\mathrm{circ}}$), and the polarization degree *P* becomes identical to P_{un} throughout the entire range of angles.

If light propagates through a medium with sufficiently large particles ($n-1 > \lambda/a$), effects due to the "non-Born" single scattering turn up. These effects are most significant at relatively small depths, where the angular distribution width is still within the critical angle of single scattering [γ_{cr}

$=2\sqrt{2(n-1)}$]. The angular scale $\gamma_{\rm cr}$ in this case affects the angular dependence of both intensity and polarization degree.

The effects discussed in this paper may be interesting for optical research of structural inhomogeneities in multiply scattering media.

This work was supported by the Russian Fund for Fundamental Research (Grant 96-02-17518).

APPENDIX A: Appendix A

For the function

$$\tilde{v}(p,\omega)=2\pi\int_0^\infty dz\exp(-pz)\int_0^\infty \theta d\theta J_0(\omega\theta)\tilde{v}(z,\theta)$$

one can easily derive from Eq. (32) the equation

$$\left[p+n_0\sigma(1-\chi(\omega))-\frac{1}{2}n_0\sigma_a\Delta_\omega\right]\tilde{v}(p,\omega)$$
$$=1+\left(\frac{\sigma_{\rm dep}}{\sigma_a}\right)n_0\sigma(1-\chi(\omega))K_0\left(\omega\sqrt{\frac{2p}{n_0\sigma_a}}\right)\tilde{E}^{(V)}(p),\tag{A1}$$

where $J_0(x)$ and $K_0(x)$ are the Bessel and modified Bessel functions of order zero,[18] and $\chi(\omega)$ is determined by Eq. (18),

$$\Delta_\omega=\frac{1}{\omega}\frac{\partial}{\partial\omega}\omega\frac{\partial}{\partial\omega},$$

$$\tilde{E}^{(V)}(p)=\int_0^\infty\exp(-pz)\tilde{E}^{(V)}(z).$$

We are interested in values of function $\tilde{v}(x,\theta)$ at relatively large angles. The ''wings'' of the angular dependence $\tilde{v}(z,\theta)$ are governed by the behavior of function $\tilde{v}(p,\omega)$ at small ω.[9,19] In this region, the second line of Eq. (18) applies to $\chi(\omega)$, whereas $K_0(\omega\sqrt{2p/\sigma_a})$ can be described by the approximate formula[18]

$$K_0\left(\omega\sqrt{\frac{2p}{n_0\sigma_a}}\right)\approx-\ln\left(\omega\sqrt{\frac{2p}{n_0\sigma_{\rm a}}}\right).\tag{A2}$$

The analysis shows that, in calculating function $\tilde{v}(p,\omega)$ in the region of small ω, it is convenient to express the logarithms in Eqs. (18) and (A2) in the form of the following limits of power series:

$$\chi(\omega)\approx1-\frac{(\omega\gamma_0)^2}{2(\alpha-4)}[1-(\omega\gamma_0)^{\alpha-4}],$$
$$K_0\left(\omega\sqrt{\frac{2p}{n_0\sigma_a}}\right)\approx\frac{1}{\beta-2}\left[1-\left(\omega\sqrt{\frac{2p}{n_0\sigma_a}}\right)^{\beta-2}\right]\tag{A3}$$

for $\alpha\to4$ and $\beta\to2$, respectively. Then the value of function $\tilde{v}(p,\omega)$ at small ω can be sought in the form of the following power expansion in ω (Ref. 19):

$$\tilde{v}(p,\omega)=\tilde{v}(p,\omega=0)(1+c_2\omega^2+c_\alpha\omega^\alpha+c_4\omega^4$$
$$+c_{\beta+2}\omega^{\beta+2}+c_{\alpha+\beta-2}\omega^{\alpha+\beta-2}+\cdots).\tag{A4}$$

After substituting Eqs. (A3) and (A4) in (A1) and equating coefficients of equal powers of ω, we can calculate the interesting parameters c_i. In particular, the coefficients c_α, $c_{\beta+2}$, and $c_{\alpha+\beta-2}$ of nonanalytic powers of ω, which contribute to the power expansion of $\tilde{v}$ in terms of reciprocal scattering angle, can be expressed as

$$c_\alpha=\frac{2(\gamma_0/2)^{\alpha-4}}{\sigma_a\alpha^2(\alpha-4)}\left[\frac{\sigma_{\rm tr}}{2\ln(2/\gamma_0)}\right.$$
$$\left.-\frac{\sigma_{\rm tr}\sigma_{\rm dep}}{\sigma_a\ln(2/\gamma_0)}\frac{\tilde{E}^{(V)}(p)}{\tilde{v}(p,\omega=0)(\beta-2)}\right],$$

$$c_{\beta+2}=-\frac{\sigma_{\rm tr}\sigma_{\rm dep}}{\sigma_a^2\ln(2/\gamma_0)}\frac{\tilde{E}^{(V)}(p)}{\tilde{v}(p,\omega=0)(\beta-2)(4-\alpha)(\beta+2)^2},\tag{A5}$$

$$c_{\alpha+\beta-2}=\frac{\sigma_{\rm tr}\sigma_{\rm dep}}{\sigma_a^2\ln(2/\gamma_0)}$$
$$\times\frac{\tilde{E}^{(V)}(p)}{\tilde{v}(p,\omega=0)(\beta-2)(4-\alpha)(\alpha+\beta-2)^2}$$
$$\times\left(\frac{\gamma_0}{2}\right)^{\alpha-4}\left(\frac{2p}{n_0\sigma_a}\right)^{\beta/2-1}.$$

Using the resulting equation (A5), performing the inverse Bessel transform in Eq. (A4), and then taking the limit as $\alpha\to4$ and $\beta\to2$, we obtain the following expression for $\tilde{v}(p,\theta)$:

$$\tilde{v}(p,\theta)\approx\frac{2\sigma_{\rm tr}}{\pi\sigma_a\ln(2/\gamma_0)}\left[\tilde{v}(p,\omega=0)\left(\frac{1}{\theta^6}+\cdots\right)\right]$$
$$-\frac{\sigma_{\rm dep}}{\sigma_a}\tilde{E}^{(V)}(p)\left[\frac{1}{\theta^6}\ln\left(\frac{\theta^2}{\gamma_0}\sqrt{\frac{n_0\sigma_a}{2p}}\right)+\cdots\right].\tag{A6}$$

This result is an expansion in powers of $1/\theta$ and $\sigma_{\rm dep}/\sigma_a$.

In the approximation of the lowest order, one can neglect in Eq. (A6) terms containing the additional small parameter $\sigma_{\rm dep}/\sigma_a$. In this approximation, the equality $\tilde{E}^{(V)}(p)=\tilde{v}(p,\omega=0)$ also holds (the contribution from the first term on the right-hand side of Eq. (31) to the integral with respect to directions is of order $\sigma_{\rm dep}/\sigma_a$). As a result, we have for $\tilde{v}(z,\theta)$ the expression

$$\tilde{v}(z,\theta)\approx\frac{\tilde{E}^{(V)}(z)}{\tilde{E}(z)}\tilde{I}_{\rm scal}(z,\theta),\tag{A7}$$

where $\tilde{I}_{\rm scal}(z,\theta)$ is determined by expansion (25).

To calculate $\tilde{V}(z,\theta)$, note that the first term on the right-hand side of Eq. (31) in the small-angle approximation for $\theta>\sqrt{l_a/z}$ can be transformed to

$$-\frac{n_0\sigma_{\mathrm{dep}}}{4\pi}\int_0^z dz'\exp\left(-\frac{n_0\sigma_a(z-z')\theta^2}{2}\right)\tilde{E}^{(V)}(z')$$
$$\approx-\frac{n_0\sigma_{\mathrm{dep}}\tilde{E}^{(V)}(z)}{2\pi\theta^2}\left[1-\frac{1}{\theta^2}\frac{2}{n_0\sigma_a\tilde{E}^{(V)}(z)}\frac{d\tilde{E}^{(V)}(z)}{dz}\right]$$
$$+\cdots \tag{A8}$$

The second term on the right-hand side of Eq. (A8) can be transformed using the equality[19]

$$\langle\theta^2\rangle_z=-\frac{2}{n_0\sigma_a\tilde{E}^{(V)}(z)}\frac{d\tilde{E}^{(V)}(z)}{dz}. \tag{A9}$$

Substituting Eqs. (A7) and (A8) into Eq. (31), we obtain Eq. (33).

Appendix B

Stokes's fourth parameter in the asymptotic regime ($z>l_d$) can be calculated in the following manner.

Neglecting backscattering (this approach is justified under conditions of high absorption, $l_a<l_{\mathrm{tr}}<l_{\mathrm{dep}}$), a solution of Eq. (6) [or (8)] that satisfies boundary condition (3) can be expanded in terms of eigenfunctions $\psi_n^{(V)}$ of the spectral problem corresponding to Eq. (6):

$$\tilde{V}(z,\mu)=\sum_n^{\infty}\psi_n^{(V)}(\mu=1)\psi_n^{(V)}(\mu)\exp(-\varepsilon_n^{(V)}z), \tag{B1}$$

where $\varepsilon_n^{(V)}$ are the corresponding eigenvalues. At large z, the sum contains only the one term with the smallest damping factor:

$$\tilde{V}(z,\mu)=\psi_0^{(V)}(\mu=1)\psi_0^{(V)}(\mu)\exp(-\varepsilon_0^{(V)}z). \tag{B2}$$

Under conditions of rare depolarizing collisions, when length l_{dep} is the maximum characteristic length, the values of $\varepsilon_0^{(V)}$ and $\psi_0^{(V)}(\mu)$ can be calculated by the perturbation theory, assuming that the intensity is known in this spectral problem and the last term of Eq. (8) can be treated as a perturbation.

The formulas for $\varepsilon_0^{(V)}$ and $\psi_0^{(V)}(\mu)$ in the first order of the perturbation theory should have the form[29]

$$\varepsilon_0^{(V)}\approx\varepsilon_0+\int d\mathbf{n}\int d\mathbf{n}'\frac{d\sigma_{\mathrm{dep}}(\mathbf{n}\cdot\mathbf{n}')}{d\mathbf{n}}\psi_0(\mu)\psi_0(\mu'), \tag{B3}$$

$$\psi_0^{(V)}(\mu)\approx\psi_0(\mu)+\sum_{n=1}^{\infty}\frac{\psi_n(\mu)}{\varepsilon_0-\varepsilon_n}\int d\mathbf{n}'$$
$$\times\int d\mathbf{n}''\frac{d\sigma_{\mathrm{dep}}(\mathbf{n}'\cdot\mathbf{n}'')}{d\mathbf{n}'}\psi_0(\mu')\psi_n(\mu''), \tag{B4}$$

where ε_n and $\psi_n(\mu)$ are the eigenvalues and eigenfunctions of the scalar transport equation for the intensity (i.e., Eq. (8) without the last term). The smallest eigenvalue determines l_d ($\varepsilon_0=l_d^{-1}$).

Equation (B4) can be written in a more compact form using the Green's function of the scalar transport equation:

$$\psi_0^{(V)}(\mu)\approx\psi_0(\mu)+\int_0^{\infty}dz\int d\mathbf{n}'\int d\mathbf{n}''$$
$$\times\exp(\varepsilon_0 z)[\tilde{G}(z|\mathbf{n},\mathbf{n}')-\psi_0(\mu)\psi_0(\mu')$$
$$\times\exp(-\varepsilon_0 z)]\frac{d\sigma_{\mathrm{dep}}(\mathbf{n}'\cdot\mathbf{n}'')}{d\mathbf{n}'}\psi_0(\mu''). \tag{B5}$$

Calculations by Eqs. (B3) and (B5) can be performed using our earlier results[9] for ε_0, ψ_0, and G obtained in the self-consistent small-angle diffusion approximation, assuming that the angular dependence of the cross section $d\sigma_{\mathrm{dep}}/d\mathbf{n}$ is isotropic. In this approximation[9]

$$\psi_0(\theta)=\sqrt{\frac{2}{\pi\langle\theta^2\rangle_\infty}}\exp\left(-\frac{\theta^2}{\langle\theta^2\rangle_\infty}\right), \tag{B6}$$

$$\int d\boldsymbol{\theta}'G(z|\boldsymbol{\theta},\boldsymbol{\theta}')\approx\frac{1}{\cosh(\varepsilon_0 z)}$$
$$\times\exp\left(-\frac{\theta^2}{\langle\theta^2\rangle_\infty}\tanh(\varepsilon_0 z)\right), \tag{B7}$$

and ε_0 is determined by Eqs. (28) and (70). In particular, in the interval $l_a<z<l_d$, it follows from Eq. (B7) that

$$\int d\boldsymbol{\theta}'G(z|\boldsymbol{\theta},\boldsymbol{\theta}')\approx\exp\left(-\frac{\theta^2\sigma_a z}{2}\right),\quad l_a<z<l_d. \tag{B8}$$

Equation (B8) provides an explicit illustration of how the angular distribution narrows in an absorbing medium. At $z>l_a$, only photons propagating along lines directed at small angles with respect to the z-axis contribute to the intensity due to an isotropic source (B8), and most of the angular distribution is concentrated in the region $\theta<\sqrt{l_a/z}$.

After substituting Eqs. (B6) and (B7) in (B5), we have Eqs. (38) and (39).

The angular dependence of $\tilde{V}(z,\mu)$ in the asymptotic regime in the region of relatively large scattering angles ($\theta>\sqrt{\langle\theta^2\rangle_\infty}$) is easier to calculate without using Eq. (B5) [or (B4)], but directly from the expansion of Eq. (33) in reciprocal powers of θ.

[*)] E-mail: gorod@theor.mephi.msk.su

[1)] A formula similar to Eq. (11) is widely used in describing elastic scattering of fast electrons by atoms.[16]

[2)] The error due to the inaccuracy of approximate equation (69) is of order $\gamma_{\mathrm{cr}}/\theta\ll1$.

---

[1] F. C. MacKintosh, J. X. Zhu, D. J. Pine, and D. A. Weitz, Phys. Rev. B **40**, 9342 (1989).

[2] R. Jullien and R. Botet, J. de Phys. **38**, 208 (1989).

[3] C. E. Mandt and L. Tsang, J. Opt. Soc. Am. A **9**, 2246 (1992).

[4] D. Bicout, C. Brosseau, A. S. Martinez, and J. M. Schmitt, Phys. Rev. E **49**, 1767 (1994).

[5] *Diffusing Photons in Turbid Media (special issue),* Appl. Opt. **36**, 9 (1997).

[6] V. L. Kuz'min and V. P. Romanov, Usp. Fiz. Nauk **166**, 247 (1996).

[7] S. G. Demos and R. R. Alfano, Opt. Lett. **21**, 161 (1996).

[8] V. V. Tuchin, Usp. Fiz. Nauk **167**, 517 (1997).

[9] E. E. Gorodnichev and D. B. Rogozkin, Zh. Éksp. Teor. Fiz. **107**, 209 (1995) [JETP **80**, 112 (1995)].

[10] L. S. Dolin, Dokl. Akad. Nauk SSSR **260**, 1344 (1981) [Sov. Phys. Dokl. **26**, 976 (1981)].

[11] V. S. Remizovich, D. B. Rogozkin, and M. I. Ryazanov, Izv. Vyssh. Uchebn. Zaved., Ser. Radiofizika **24**, 891 (1982).
[12] H. C. van de Hulst, *Multiple Light Scattering : Tables, Formulas, and Applications*, Academic Press, New York (1980).
[13] R. Newton, *Scattering Theory of Waves and Particles*, McGraw-Hill, New York (1966).
[14] A. Ishimaru, *Wave Propagation and Scattering Media*, Academic Press, New York (1978).
[15] V. S. Remizovich, Zh. Éksp. Teor. Fiz. **87**, 506 (1984) [Sov. Phys. JETP **60**, 290 (1984)].
[16] H. A. Bethe, Phys. Rev. **89**, 1256 (1953).
[17] S. V. Maleev and B. P. Toperverg, Zh. Éksp. Teor. Fiz. **78**, 315 (1980) [Sov. Phys. JETP **51**, 158 (1980)].
[18] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, Dover, New York (1974).
[19] E. E. Gorodnichev, A. I. Kuzovlev, and D. B. Rogozkin, Laser Phys. **6**, 1188 (1996).
[20] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Pergamon Press (1980).
[21] Yu. M. Gnedin, A. Z. Dolginov, and N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **57**, 988 (1969) [Sov. Phys. JETP **30**, 540 (1970)].
[22] K. Case and P. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, Massachusetts (1967).
[23] D. Eliyahu, M. Rosenbluch, and I. Freund, J. Opt. Soc. Am. A **10**, 477 (1993).
[24] Y. Kuga and A. Ishimaru, Appl. Opt. **28**, 2165 (1989).
[25] K. M. Yoo, L. Feng, and R. R. Alfano, Phys. Rev. Lett. **64**, 2647 (1990).
[26] M. P. van Albada, M. B. van der Mark, and A. Lagendjik, Phys. Rev. Lett. **58**, 361 (1987).
[27] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed., Pergamon, New York (1984).
[28] E. E. Gorodnichev, A. I. Kuzovlev, and D. B. Rogozkin, JETP Lett. **68**, 22 (1998).
[29] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Nonrelativistic Theory)*, 3rd ed., Pergamon, New York (1977).

Translation provided by the Russian Editorial office

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Kinetics of atoms in an elliptically polarized standing wave

O. N. Prudnikov,*) A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin

Novosibirsk State University, 630090 Novosibirsk, Russia

(Submitted 12 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 791–804 (March 1999)

We examine the kinetics of atoms with their ground and excited states being degenerate in the projection of angular momentum. The atoms are located in a standing wave with uniform elliptical polarization. Using the $j_g=1/2\rightarrow j_e=1/2$ transition as an example, we show that the friction and diffusion of atoms strongly depend on the ellipticity of the field. For instance, in the low saturation limit the frictional force contains, in addition to the ordinary Doppler friction term, a term that can be interpreted as Sisyphean friction. Under certain conditions, the contributions reflecting the degeneracy of the ground state are dominant, with the result that the values of the friction and diffusion coefficients (and hence the rate of kinetic processes) may differ from the values predicted by the two-level atomic model by several orders of magnitude. © *1999 American Institute of Physics.* [S1063-7761(99)00203-6]

1. INTRODUCTION

The kinetics of atoms in resonant light fields is being intensively studied in atomic physics. The achievements in the field of laser cooling and atom trapping have been marked by the 1997 Nobel Prize in physics (Steven Chu, Claude Cohen-Tannoudji, and William C. Phillips). The experimental and theoretical progress in this area of research in the period from 1970 to 1989 is thoroughly described in the monographs in Refs. 1 and 2. Special issues of journals have been devoted to the achievements after that period.[3,4] The current state of research is discussed in the Nobel Lectures of the laureates.[5] Here we will only mention the important papers directly related to the topic of our paper.

The first proposals to use resonant light pressure to trap and cool atoms in optical potential wells[6] stimulated the development of the theory of atomic motion in the field of resonant light in conditions of steady-state interaction. Historically, the first approach was the semiclassical one.[7,8] Within it the translational motion of an atom is described in classical terms (i.e., the atom is assumed to have a definite position and definite velocity), while the dynamics of the internal degrees of freedom is described by an optical Bloch equation for the density matrix with allowance for radiative damping. The force on the atom is calculated as the mathematical expectation of the force operator, and the diffusion coefficient is found from the two-time correlator of the force operator. For a two-level atom in the field of a standing wave, Gordon and Ashkin[8] derived analytic expressions for the frictional force and the coefficients of spontaneous and induced diffusion at arbitrary intensity and detuning.

The second approach to the kinetics of atoms in laser fields is based on a full quantum description of the internal and translational degrees of freedom of an atom in terms of the density matrix in the Wigner representation. By using the semiclassical expansion in powers of photon momentum the initial transport equation can be reduced (at the kinetic stage of the evolution) to a single equation for the Wigner distribution function in phase space.[9] The equation is of the Fokker–Planck type and contains not only the average force but also diffusion. The equivalence of the two approaches to calculating the frictional force and diffusion in the general case of arbitrary angular momenta of the ground and excited states was demonstrated by Dalibard and Cohen-Tannoudji[10] and Agarwal and Mølmer.[11] Furthermore, Dalibard and Cohen-Tannoudji[10] showed that the coefficient of friction (as well as the diffusion coefficient) can be expressed in terms of the two-time correlator of the force operator.

The third approach (also equivalent to the first two approaches) is based on the relationships between the moments of the Wigner matrix density and the force and diffusion.[12]

Despite the existence of a general formalism[10,13,14] that enables one (at least in principle) to study the kinetics of atoms with arbitrary level structure in a field of arbitrary configuration, the fundamental problem of the motion of an atom with degenerate ground and excited states in the uniformly polarized field of a standing wave has not been discussed before. This probably explains why researchers focused on effects of sub-Doppler cooling,[3,4] which emerge in fields with a polarization gradient. At the same time, the analytic form of the steady-state atomic density matrix in the zeroth approximation in the recoil parameter[15–17] points very definitely to the existence of polarization features in the kinetics of atoms in a field with spatially uniform polarization. The study of such features is the goal of the present paper.

We examine the motion of atoms with resonant optical transitions of the type $j_g=j\rightarrow j_e=j+1$ and $j_g=j'\rightarrow j_e=j'$ (here j' is a half-integer), where j_g and j_e are the total angular momenta of the ground and excited atomic levels, in the field of a monochromatic wave. We employ the semiclassical approach to derive general expressions for the gradient force, the force of friction, and diffusion, which provide a suitable basis for describing the kinetics of slow atoms. We then analyze these expressions in detail for atoms with the optical transition $j_g=1/2\rightarrow j_e\rightarrow 1/2$ in the field of a uniformly polarized standing wave of arbitrary elliptical polarization and intensity. We find that the frictional force and diffusion strongly depend on the ellipticity of the light and contain entirely new contributions, totally ignored in the two-level model. In particular, the frictional force in the weak-field limit is found to contain, in addition to the ordinary term due to the Doppler effect, a term that, depending on the ellipticity and detuning, may be dominant and emerges because of the spatial modulation of the probability of transitions between different adiabatic potentials, similar to the well-known Sisyphean effect.[2–4,15] For the $j_g=1/2 \rightarrow j_e=1/2$ transition this additional contribution disappears when the field is linearly polarized. However, for transitions with larger values of the orbital angular momentum j_g this term is nonvanishing even if the field is linearly polarized, which is demonstrated by the example of atoms with $j_g=1 \rightarrow j_e=2$ and $j_g=3/2 \rightarrow j_e=3/2$ transitions.

2. SEMICLASSICAL DESCRIPTION OF ATOMIC KINETICS

We assume that the atoms, with total angular momenta j_g in the ground state and j_e in the excited state, are in one-dimensional motion along the z axis in the field of a resonant plane wave:

$$\mathbf{E}(z,t)=E(z)\mathbf{e}(z)\exp(-i\omega t)+\text{c.c.},$$

$$\mathbf{e}(z)=\sum_{q=0,\pm 1} e^{q}(z)\mathbf{e}_q, \tag{1}$$

where $E(z)$ is the local field amplitude, and $\mathbf{e}(z)$ is the local unit polarization vector, with e^q the components of this vector in the cyclic basis $\{\mathbf{e}_0=\mathbf{e}_z,\ \mathbf{e}_{\pm 1}=\pm(\mathbf{e}_x+\pm i\mathbf{e}_y)/\sqrt{2}\}$.

The total Hamiltonian of an atom in a rotating basis (i.e., rotating in the energy pseudospin space) is $\hat{H}=\hat{H}_0+\hat{V}$, where H_0 is the free-atom Hamiltonian:

$$\hat{H}_0=\frac{\hat{p}^2}{2M}-\hbar\Delta\hat{\Pi}_e. \tag{2}$$

Here $\Delta=\omega-\omega_0$ is the detuning from the atomic transition frequency o_0, and the projection operator

$$\hat{\Pi}_e=\sum_{\mu_e}|j_e,\mu_e\rangle\langle j_e,\mu_e| \tag{3}$$

is constructed out of $|j_e,\mu_e\rangle$, the wave functions of the Zeeman sublevels of the excited states. In the dipole and resonance approximations, the operator representing the interaction with the field (1) can be written

$$\hat{V}(z)=\hbar\Omega(z)\sum_{q=0,\pm 1}\hat{T}_q e^{q}(z)+\text{H.c.}, \tag{4}$$

where $\Omega=dE/\hbar$ is the Rabi frequency, with d the reduced matrix element. The operator $\hat{T}_q$ can be expressed in terms of $3jm$-symbols:

$$\hat{T}_q=\sum_{\mu_e,\mu_g}|j_e,\mu_e\rangle(-1)^{j_e-\mu_e}\begin{pmatrix} j_e & 1 & j_g \\ -\mu_e & q & \mu_g\end{pmatrix}\langle j_g,\mu_g|. \tag{5}$$

In addition to having a coherent component, the dynamics of the atomic ensemble is determined by relaxation processes. To describe such processes we restrict attention to radiative relaxation, which means that we completely ignore all effects of interatomic interaction and assume that the atomic number density is low. In this approximation the transport equation for the one-particle density matrix of the atoms has the form

$$\begin{aligned}\frac{\partial}{\partial t}\hat{\rho}=&-\frac{i}{\hbar}[\hat{H}_0,\hat{\rho}]-\frac{i}{\hbar}[\hat{V}(\hat{z}),\hat{\rho}]-\frac{\gamma}{2}(\hat{\Pi}_e\hat{\rho}+\hat{\rho}\hat{\Pi}_e)\\ &+\gamma(2j_e+1)\sum_q\int_{-1}^{1}\hat{T}_q^{\dagger}\exp(-iks\hat{z})\hat{\rho}\\ &\times\exp(iks\hat{z})\hat{T}_qK_q(s)\,ds,\end{aligned} \tag{6}$$

where the quantization axis is the z axis, γ is the radiative relaxation rate, and $\hat{z}$ is the operator of an atom's center-of-mass coordinate. The last two terms on the right-hand side of Eq. (6) describe radiative relaxation. The first of these two terms gives the decay of the density matrix of the excited state, $\hat{\rho}^{ee}$, and of the off-diagonal elements $\hat{\rho}^{eg}$ and $\hat{\rho}^{ge}$. The second describes the arrival of atoms from the excited state to the ground state with allowance for recoil in spontaneous photon emission. The type of the functions $K_{\pm 1}=(1+s^2)3/8$ and $K_0=(1-s^2)3/4$ is determined by the probability of emission of a photon with polarization $q=\pm 1,0$ in the given direction $\mathbf{n}$ in relation to the quantization axis (the z axis): $\mathbf{n}\cdot\mathbf{e}_z=s$. In relation to the translational degrees of freedom, Eq. (6) has an invariant form. Assuming that the variance of the atomic momentum is much greater than the recoil momentum, $\hbar k/\delta p\ll 1$, below we describe the translational motion of the atoms semiclassically in the Wigner representation. Then the transport equation (6) can be expressed by a series expansion in powers of the small parameter $\hbar k/\delta p\ll 1$:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{p}{M}\frac{\partial}{\partial z}\right)\hat{\rho}(z,p)=&\hat{L}^{(0)}\{\hat{\rho}(z,p)\}+\hbar k\frac{\partial}{\partial p}\hat{L}^{(1)}\{\hat{\rho}(z,p)\}\\ &+(\hbar k)^2\frac{\partial^2}{\partial p^2}\hat{L}^{(2)}\{\hat{\rho}(z,p)\}+\cdots,\end{aligned} \tag{7}$$

where (z,p) are the coordinates of a point in phase space. In the zeroth approximation in the recoil effects, the evolution of the density matrix is determined by the operator

$$\begin{aligned}\hat{L}^{(0)}\{\hat{\rho}\}=&-[(\gamma/2-i\Delta)\hat{\Pi}_e\hat{\rho}+(\gamma/2+i\Delta)\hat{\rho}\hat{\Pi}_e]\\ &-\frac{i}{\hbar}[\hat{V}(z),\hat{\rho}]+\gamma(2j_e+1)\sum_q\hat{T}_q^{\dagger}\hat{\rho}\hat{T}_q.\end{aligned} \tag{8}$$

The first-order terms

$$\hat{L}^{(1)}\{\hat{\rho}\}=-\frac{1}{2\hbar k}(\hat{f}(z)\hat{\rho}+\hat{\rho}\hat{f}(z)) \tag{9}$$

can be expressed in terms of the force operator

$$\hat{f}(z)=-\frac{\partial\hat{V}(z)}{\partial z}. \tag{10}$$

The second-order terms include both induced terms, proportional to the second derivative of $\hat{V}(z)$ with respect to the coordinate, and a term reflecting the recoil effect in spontaneous emission:

$$\hat{L}^{(2)}\{\hat{\rho}\} = \frac{i}{8\hbar k^2}\left[\left(\frac{\partial^2}{\partial z^2}\hat{V}(z)\right), \hat{\rho}\right] + \gamma(2j_e+1)\sum_q C_q \hat{T}_q^{\dagger}\hat{\rho}\hat{T}_q, \tag{11}$$

where $C_0 = 1/10$, and $C_{\pm 1} = 1/5$.

The presence of the small parameter $\hbar k/\delta p$ makes it possible to separate the rapid processes related to ordering over the internal degrees of freedom from the slow processes related to translational motion. The kinetic stage of evolution in this problem corresponds to times $\tau \gg \tau_0 = \max\{\gamma^{-1}, (\gamma S)^{-1}\}$, where $S = |\Omega(z)|^2 (\gamma^2/4 + \Delta^2)^{-1}$ is the saturation parameter. At this stage a quasi-steady-state distribution over the internal degrees of freedom sets in and the dynamics of the atomic ensemble is determined by the slow processes of variation of the distribution function over the translational degrees of freedom. In the kinetic stage, the density matrix is a linear functional of the distribution function, $\mathscr{F}(z,p) = \mathrm{Tr}\{\hat{\rho}(z,p)\}$ (Refs. 1 and 2). For $\mathscr{F}(z,p)$ we can obtain a closed equation, which to within the second-order term in the recoil parameter $\hbar k/\delta p$ is the Fokker–Planck equation:

$$\left(\frac{\partial}{\partial t} + \frac{p}{M}\frac{\partial}{\partial z}\right)\mathscr{F} = -\frac{\partial}{\partial p} f(z,p)\mathscr{F} + \frac{\partial^2}{\partial p^2} D(z,p)\mathscr{F}, \tag{12}$$

where $f(z,p)$ is the force on the atom, and $D(z,p)$ is the diffusion coefficient at (z,p).

Note that Eqs. (8)–(11) are of a general nature and can be used to analyze the kinetics of atoms in an arbitrary plane monochromatic wave.

3. KINETIC COEFFICIENTS IN THE SLOW-ATOM APPROXIMATION

Finding the kinetic coefficients in analytic form is a very complicated problem involving tedious calculations. In this paper we restrict attention to slow atoms, which become displaced over distances much smaller than the wavelength in characteristic times of ordering over the internal degrees of freedom, i.e., $v\tau_0 \ll \lambda$, which with allowance for the definition of τ_0 yields the inequality

$$\frac{kp}{M} \ll \gamma,\ \gamma S. \tag{13}$$

When we are dealing with transitions of the type $j_g = j \to j_e = j+1$ and $j_g = j' \to j_e = j'$ (with j' a half-integer) and condition (13) is met, to take dissipation processes into account correctly we need only restrict attention to the linear approximation in velocity in the expression for the force:

$$f(z,p) \approx f_g(z) + \xi(z)\frac{p}{M}, \tag{14}$$

where $f_g(z)$ is the gradient force and $\xi(z)$ is the coefficient of friction, while for the diffusion coefficient we keep only the zeroth-order term: $D(z,p) \approx D(z)$. Then, finding the trace of Eq. (7) and allowing for the corresponding terms in the series expansion in the recoil parameter $\hbar k/\delta p$, we obtain expressions for $f_g(z)$, $\xi(z)$, and $D(z)$ that can written in invariant form.

Thus, the gradient force is determined by the expectation value of the force operator (1):

$$f_g(z) = \mathrm{Tr}\{\hat{f}(z)\hat{\sigma}(z)\}, \tag{15}$$

where the density matrix $\hat{\sigma}(z)$ is the solution of the equations

$$\hat{L}^{(0)}\{\hat{\sigma}\} = 0, \quad \mathrm{Tr}\{\hat{\sigma}\} = 1 \tag{16}$$

and describes the steady-state distribution of atoms over magnetic sublevels in the zeroth approximation in the recoil parameter and atomic velocity. Here and in what follows the symbol $\mathrm{Tr}\{\cdots\}$ stands for the trace over internal variables. Note that analytic expressions for $\hat{\sigma}(z)$ for all dipole transitions were found in Refs. 16–18.

The coefficient of friction is proportional to the spatial gradient of $\hat{\sigma}(z)$:

$$\xi(z) = -\hbar k \,\mathrm{Tr}\left\{\hat{\varphi}\frac{\partial\hat{\sigma}}{\partial z}\right\}. \tag{17}$$

The diffusion coefficient can be written as a sum of two terms:

$$D(z) = D_{\mathrm{sp}}(z) + D_{\mathrm{ind}}(z). \tag{18}$$

The first term on the right-hand side of Eq. (18),

$$\begin{aligned} D_{\mathrm{sp}}(z) &= (\hbar k)^2 \mathrm{Tr}\{\hat{L}^{(2)}\{\hat{\sigma}\}\} \\ &= (\hbar k)^2 \gamma(2j_e+1)\mathrm{Tr}\left\{\sum_q C_q \hat{T}_q^{\dagger}\hat{\sigma}\hat{T}_q\right\}, \end{aligned} \tag{19}$$

is due to the recoil effect in spontaneous emission. The second term,

$$\begin{aligned} D_{\mathrm{ind}}(z) &= -(\hbar k)^2 \left[\mathrm{Tr}\{\hat{\varphi}\hat{L}^{(1)}\{\hat{\sigma}\}\} - \mathrm{Tr}\{\hat{\varphi}\hat{\sigma}\}\mathrm{Tr}\{\hat{L}^{(1)}\{\hat{\sigma}\}\}\right. \\ &= \frac{\hbar k}{2}\mathrm{Tr}\{\hat{\varphi}(\hat{\sigma}\cdot\delta\hat{f} + \delta\hat{f}\cdot\hat{\sigma})\}, \end{aligned} \tag{20}$$

is determined by the fluctuation operator for the force (10):

$$\delta\hat{f}(z) = \hat{f}(z) - f_g(z). \tag{21}$$

In Eqs. (17) and (20) we introduced an auxiliary matrix $\hat{\varphi}$, which makes it possible to write the expression for the friction and diffusion coefficients in a unified manner. The matrix $\hat{\varphi}$ is a solution of the inhomogeneous linear equation

$$\begin{aligned} &\left[\left(\frac{\gamma}{2} + i\Delta\right)\hat{\Pi}_e\hat{\varphi} + \left(\frac{\gamma}{2} - i\Delta\right)\hat{\varphi}\hat{\Pi}_e\right] - \frac{i}{\hbar}[\hat{V}(z), \hat{\varphi}] \\ &\quad - \gamma(2j_e+1)\sum_q \hat{T}_q\hat{\varphi}\hat{T}_q^{\dagger} = \frac{\delta\hat{f}}{\hbar k}, \end{aligned} \tag{22}$$

where the source (the right-hand side) is the fluctuation of the force operator (Eq. (21)). Note that the system of equations (22) is linearly dependent. This can easily be verified if we multiply the right- and left-hand sides by the matrix $\hat{\sigma}$ [Eq. (16)] and calculate the traces, which vanish simultaneously since the left-hand side reduces to $-\mathrm{Tr}\{\hat{L}^{(0)}\{\hat{\sigma}\}\hat{\varphi}$ and vanishes by virtue of (16), and the right-hand side is of

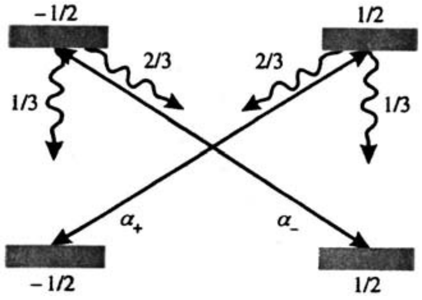

FIG. 1. Spontaneous (wavy lines) and light-induced (solid straight lines) transitions for $j_g=1/2\to j_e=1/2$. Relative amplitudes of the light-induced transitions are proportional to the circular components α_+ and α_- of the field [see Eq. (24)], and the numbers labeling the wavy and straight lines indicate the relative spontaneous-decay probabilities.

order $\mathrm{Tr}\{\hat{\sigma}\cdot\delta\hat{f}\}$ and vanishes by virtue of (15), (16), and (21). This means that the solution of the system of equations (22) is not unique, i.e., the matrix $\hat{\varphi}$ can be determined to within an arbitrary solution of a homogeneous equation, a solution which in the case in question is $c\hat{I}$, where $\hat{I}$ is the identity matrix. However, Eqs. (16), (17), and (20) show that this arbitrariness in defining $\hat{\varphi}$ has no effect on the kinetic coefficients. Thus, the induced diffusion coefficient D_{ind} is quadratic in $\delta\hat{f}$, which corresponds to the standard definition of diffusion in terms of the two-time correlator of the force operator (see, e.g., Ref. 10). What is important here is that the induced diffusion D_{ind} and the coefficient of friction ξ can be written in terms of the same matrix $\hat{\varphi}$. This can be considered a fluctuation–dissipation relationship linking momentum fluctuations (diffusion) to momentum dissipation (friction). Hence the above relationships are important from the general physical standpoint.

4. THE $j_g=1/2\to j_e=1/2$ TRANSITION IN AN ELLIPTICALLY POLARIZED STANDING WAVE

In this section we discuss the kinetics of atoms in the uniformly polarized field of a standing wave for the $j_g=1/2\to j_e=1/2$ transition (Fig. 1). In contrast to a two-level atom, this simple model of real atomic transitions makes it possible to establish and analyze a number of important polarization features of the kinetics of atoms for which Zeeman splitting of both the ground state and the excited state is responsible. In the case being discussed, the resonant field (1) has the form

$$\mathbf{E}(z,t)=E_0\cos(kz)\,\mathbf{e}\,\exp(-i\omega t)+\mathrm{c.c.},\tag{23}$$

where the unit polarization vector $\mathbf{e}$ is spatially uniform. In particular, if the quantization axis z is parallel to the wave vector $\mathbf{k}$, and the x axis is parallel to the semimajor axis of the polarization ellipse, we can write

$$\mathbf{e}=-\alpha_+\mathbf{e}_{+1}+\alpha_-\mathbf{e}_{-1},\quad \alpha_\pm=\cos(\varepsilon\mp\pi/4),\tag{24}$$

where ε is the ellipticity ($|\tan\varepsilon|$ is the ratio of the semiminor axis of the polarization ellipse to the semimajor axis). Recall that $\varepsilon=0$ corresponds to linear polarization and $\varepsilon=\pm\pi/4$ to circular polarization. Then the nonvanishing elements of the matrix $\hat{\sigma}$ have the form

$$\begin{aligned}
&\sigma^{ee}_{\mp1/2,\mp1/2}=\beta S_0\cos^2 kz,\\
&\sigma^{gg}_{\mp1/2,\mp1/2}=\beta(3\alpha_\pm^{-2}+S_0\cos^2 kz),\\
&\sigma^{eg}_{\mp1/2,\pm1/2}=-\frac{i\alpha_\mp^{-1}\beta\Omega_0\cos kz}{\sqrt{3}\,(\gamma/2-i\Delta)},\\
&\hat{\sigma}^{ge}=(\hat{\sigma}^{eg})^\dagger,
\end{aligned}\tag{25}$$

where

$$\Omega_0=\frac{dE_0}{\hbar},\quad S_0=\frac{|\Omega_0|^2}{\gamma^2/4+\Delta^2}\tag{26}$$

are the Rabi frequency and the saturation parameter, and

$$\beta=\frac{\alpha_+^2\alpha_-^2}{3+4S_0\cos^2(kz)\,\alpha_+^2\alpha_-^2}$$

is the normalization constant. The nonvanishing elements of the matrix $\hat{\varphi}$ have the form

$$\begin{aligned}
&\varphi^{ee}_{\mp1/2,\mp1/2}=\pm3\,\frac{\Delta\beta}{\gamma}\,\frac{\alpha_-^2-\alpha_+^2}{\alpha_-^2\alpha_+^2}\tan kz,\\
&\varphi^{gg}_{\mp1/2,\mp1/2}=\frac{\Delta\beta}{\gamma}\left(8S_0\cos^2 kz\mp9\,\frac{\alpha_-^2-\alpha_+^2}{\alpha_-^2\alpha_+^2}\right)\tan kz,\\
&\varphi^{eg}_{\mp1/2,\pm1/2}=\frac{\alpha_\mp\Omega_0\cos kz}{(\gamma/2+i\Delta)\sqrt{3}}\left[i\,\frac{\Delta\beta}{\gamma}\right.\\
&\qquad\left.\times\left(8S_0\cos^2 kz\pm6\,\frac{\alpha_-^2-\alpha_+^2}{\alpha_-^2\alpha_+^2}\right)+1\right]\tan kz,\\
&\hat{\varphi}^{ge}=(\hat{\varphi}^{eg})^\dagger.
\end{aligned}\tag{27}$$

Substituting (25)–(27) into (8)–(11), we obtain expressions for the gradient force, the coefficient of friction, and the diffusion coefficient for arbitrary ellipticity and field intensity.

4.1. Gradient force

For the given optical transition, Eqs. (15) and (25) readily yield

$$f_g=\frac{\hbar k\Delta}{2}\,\frac{S_\varepsilon\sin 2kz}{1+S_\varepsilon\cos^2 kz}.\tag{28}$$

The corresponding potential is

$$U=\frac{\hbar\Delta}{2}\ln(1+S_\varepsilon\cos^2 kz).\tag{29}$$

Note that Eqs. (28) and (29) coincide with the well-known result in the two-level model,[2] the only difference being that in our case the effective saturation parameter depends on the ellipticity of the field:

$$S_\varepsilon=\frac{S_0\cos^2 2\varepsilon}{3}.\tag{30}$$

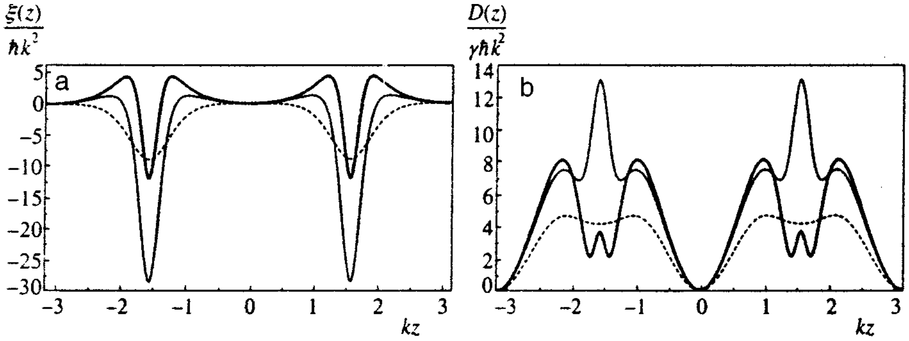

FIG. 2. Spatial dependence of the kinetic coefficients for various values of ellipticity: $\varepsilon=0$, solid heavy curves; $\varepsilon=\pi/8$, solid thin curves; and $\varepsilon=\pi/6$, dotted curves. The value of the saturation parameter $S_0=15$, and the detuning $\Delta=\gamma$. a) Coefficient of friction; b) diffusion coefficient.

4.2. Frictional force

Substituting the expressions for the matrices $\hat{\sigma}$ and $\hat{\varphi}$ into (17), we obtain a formula for the coefficient of friction:

$$\xi(z)=\frac{\gamma\hbar k^2\Delta S_\varepsilon\sin^2kz}{(\Delta^2+\gamma^2/4)(1+S_\varepsilon\cos^2kz)^3}\times\Bigg[1-S_\varepsilon\cos^2kz-\left(\Delta^2+\frac{\gamma^2}{4}\right)\frac{2S_\varepsilon^2\cos^4kz}{\gamma^2}+\left(\Delta^2+\frac{\gamma^2}{4}\right)\frac{6\sin^2 2\varepsilon}{\gamma^2}\Bigg]. \tag{31}$$

We see that the coefficient of friction (31) strongly depends on the ellipticity of the field (see Fig. 2a). In fact, the first three terms in square brackets coincide with the coefficient of friction in the two-level model.[2] Two physical mechanisms are known to correspond to this case. The first is the Doppler effect, which leads to a difference in the forces exerted by light on a moving atom from counterpropagating waves. This mechanism leads to friction at negative detunings and corresponds to the first term, which for low saturation, $S_\varepsilon \ll 1$, is the main mechanism in the two-level model. The second is lag effects, which lead to what is known as a Sisyphean force. This contribution is described by the third term; its sign is opposite that of Doppler friction, and it provides the dominant contribution in strong fields, $S_\varepsilon \gg 1$. The last term in square brackets emerges because we take into account the polarization features of atom–light interaction. Since this is a new term, there is considerable interest in elucidating the corresponding physical mechanism.

It turns out that the term we are interested in, as well as the third term, can be interpreted by using the concept of probabilities (or rates) of transitions between dressed states of an atom. Note that we call eigenstates of the Hamiltonian of an atom in a resonant field, without allowance for translational motion, dressed (adiabatic) states. For a two-level atom there can be only two dressed states (the normal state and the anomalous state; see Ref. 2). In our case of the $j_g=1/2\rightarrow j_e=1/2$ transition, there are four adiabatic states:

$$|n\pm\rangle=\cos(\theta_\pm)\,|j_g,\mp 1/2\rangle+\sin(\theta_\pm)\,|j_e,\pm 1/2\rangle,$$
$$|a\pm\rangle=-\sin(\theta_\pm)\,|j_g,\mp 1/2\rangle+\cos(\theta_\pm)\,|j_e,\pm 1/2\rangle. \tag{32}$$

Each of these states is a superposition of the wave functions of the Zeeman sublevels of the ground and excited states with spatially nonuniform coefficients. The angles $\theta_\pm$ determine the mixing of bare states of the atom:

$$\theta_\pm=\frac{1}{2}\arccos\sqrt{\frac{1}{1+4|\Omega_0|^2\cos^2(kz)\,\alpha_\pm^2/3\Delta^2}}. \tag{33}$$

States labeled with a plus (minus) are produced when the Zeeman sublevels interacting with the right (left) circular component of the field mix. When the field is zero (e.g., at a node), the mixing angles vanish and the dressed states, labeled n (a), become the wave functions of the Zeeman sublevels of the ground (excited) state. Hence, following Ref. 2, we call the $|n\pm\rangle$ states normal and the $|a\pm\rangle$ states anomalous. Eigenvalues corresponding to the states (32) are called adiabatic potentials. To within an unimportant constant, the potentials of normal states can be written

$$U_{n\pm}=\frac{\hbar\Delta}{2}\sqrt{1+\frac{4|\Omega_0|^2\cos^2(kz)\,\alpha_\pm^2}{3\Delta^2}}, \tag{34}$$

and the potentials of anomalous states have the opposite sign: $U_{a\pm}=-U_{n\pm}$. In the secular approximation $\Delta^2+\Omega^2 \gg \gamma^2$, the off-diagonal elements of the density matrix in the dressed-state basis are negligible, so that we can restrict attention to populations that satisfy the appropriate rate equations.[15] Transitions between the adiabatic states (32) occur because of spontaneous relaxation (since Landau–Zener transitions in the limit of low velocities considered here are unimportant). Lag effects (hysteresis), which lead to Sisyphean friction, result from the fact that the rates of direct and reverse transitions have different spatial localization. For instance, near a node in the field the rate of the transition from the normal state $|n-\rangle$ to the anomalous state $|a-\rangle$ is close to zero, while the reverse transition has a rate $\sim(2/3)\gamma$. At a crest of a strong field, where $\theta_-\sim\pi/4$, the rates of the direct and reverse transition become almost equal ($\sim\gamma/3$). As a result, some of the moving atoms transfer near a crest of the field from the normal state to the anomalous state, and return to the initial state near a node, which leads, as is well known,[2,15] to a retardation (acceleration) buildup for positive (negative) detunings due to the different spatial dependence of the normal and anomalous potentials.

In contrast to the model of a two-level atom, in our case we have additional degrees of freedom associated, for instance, with transitions between the "+" and "−" states. Allowing for such transitions is the reason for the last term on the right-hand side of Eq. (31). Since we are interested only in a qualitative interpretation, we examine the low satu-

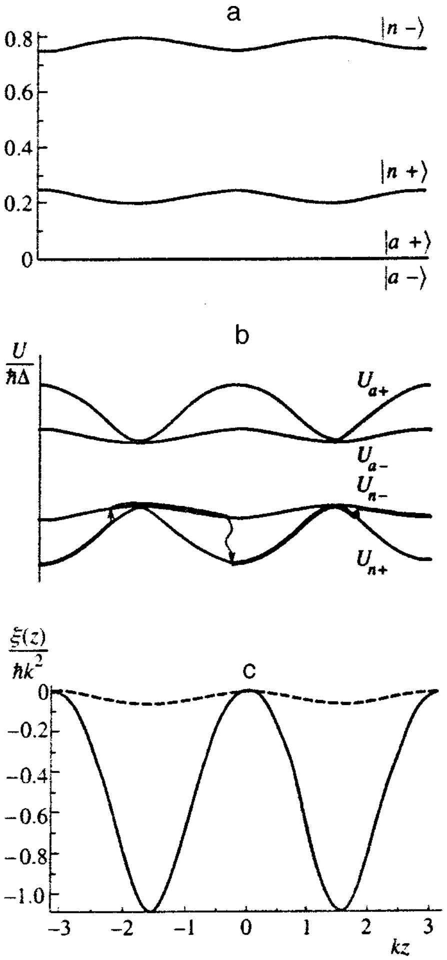

FIG. 3. Sisyphean mechanism of friction for $S_0<1$. a) Populations of dressed states. b) Adiabatic potentials. The heavy curve with wavy arrows depicts the most probable path of the atom. c) Coefficient of friction: the solid curve represents the additional contribution [the last term on the right-hand side of Eq. (31)], and the dashed curve represents the Doppler contribution to friction. All diagrams correspond to $\varepsilon=\pi/10$, $S_0=0.8$ ($S_\varepsilon=0.17$), and $\Delta=-3\gamma$.

ration limit ($S_0<1$). In this case the populations of the anomalous states are low everywhere and the entire effect is due to the transitions between normal states (in a weak field these are the Zeeman sublevels of the ground state). For instance, for $\alpha_-^2<\alpha_+^2$, most atoms at a node are in the $|n-\rangle$ state, while at a crest the populations of the "+" and "−" states equalize somewhat (see Fig. 3a). In the limit in question the adiabatic potentials $U_{n\pm}$ have the same spatial dependence ($\propto\cos^2 kz$) but different amplitudes (Fig. 3b). By analogy with the aforesaid, some of the moving atoms transfer near a crest of the field from a state with a lower potential into a state with a higher potential, and return to their former state near a node, which obviously leads to a systematic retardation (acceleration) for negative (positive) values of Δ. Thus, in its physical mechanism, the last term on the right-hand side of Eq. (31) represents Sisyphean friction but, in contrast to ordinary Sisyphean friction, has the opposite sign and depends differently on the field intensity, detuning, and ellipticity. In the weak-field approximation $S_\varepsilon\ll 1$, this contribution competes with the Doppler contribution (see Fig. 3c), and the ratio of the two is $6(\Delta/\gamma)^2\sin^2 2\varepsilon$, which can amount to several orders of magnitude if the detuning is large.

Summarizing, we can say that the difference of the kinetic of atoms with degenerate energy levels from the results of the two-level model stems primarily from the fact there are numerous adiabatic (normal and anomalous) states and possible transitions between them. For instance, there are transitions among various normal states, while for a two-level atom there can only be transitions between normal and anomalous states. Here, as Eq. (31) clearly shows, the difference between our model and the two-level model is important because of the different dependence on the intensity and frequency of the field. Thus, there is a range of parameters, $(\gamma/\Delta)^2\ll S_\varepsilon^2<3\sin^2 2\varepsilon$, for which even the direction of a kinetic process (cooling or heating) is opposite that predicted by the two-level model.

4.3. Diffusion coefficient

The diffusion coefficient defined in Eqs. (18)–(20) contains a contribution D_{sp} from spontaneous transitions and a contribution D_{ind} from induced transitions. Using Eqs. (18) and (25), we find the spontaneous diffusion coefficient,

$$D_{\mathrm{sp}}(z)=\frac{\gamma\hbar^2k^2}{12}\,\frac{S_\varepsilon\cos^2 kz}{1+S_\varepsilon\cos^2 kz}, \tag{35}$$

which coincides, to within a constant factor, in form with the expression for the spontaneous diffusion coefficient obtained in the two-level atomic model (see Ref. 2). It is proportional to the population of atoms in the excited stated and in our case of the $j_g=1/2\rightarrow j_e=1/2$ transition strongly depends on the ellipticity of the field, in accordance with (30). We can split the induced diffusion coefficient (as we did the coefficient of friction) into two terms:

$$D_{\mathrm{ind}}(z)=D_1(z)+D_2(z). \tag{36}$$

The first term is the well-known result[2] for a two-level atom with a new saturation parameter S_ε:

$$D_1(z)=\frac{\hbar^2k^2\gamma S_\varepsilon\sin^2 kz}{4}\left[1+4\right.$$
$$\left.\times\frac{\Delta^2}{\gamma^2}\,\frac{S_\varepsilon\cos^2(kz)\,(S_\varepsilon^2\cos^4 kz-\gamma^2(\gamma^2/4+\Delta^2)^{-1})}{(1+S_\varepsilon\cos^2 kz)^3}\right]. \tag{37}$$

The second term,

$$D_2=\frac{\hbar^2k^2S_\varepsilon\sin^2 2\varepsilon\sin^2 kz}{\gamma(1+S_\varepsilon\cos^2 kz)^3}\left[3\Delta^2+\frac{S_0\cos^2 kz}{3}\left(\frac{\gamma^2}{4}+\Delta^2\right)\right.$$
$$\left.\times(1+S_\varepsilon\cos^2 kz)^2\right], \tag{38}$$

is entirely new. It describes the contribution of diffusion processes generated by the Zeeman degeneracy of the energy levels. The diffusion coefficient (as well as the coefficient of friction) strongly depends on the ellipticity of the field (Fig. 2b).

As noted earlier, there is a direct relationship between friction and diffusion. Hence, by analogy with the above analysis of the frictional force, the additional contribution, Eq. (38), can be considered the result of force fluctuation that appears in the transitions between different adiabatic states. In particular, for small saturation parameters, $S_0 \ll 1$, this term is the consequence of transitions between the normal states $|n+\rangle$ and $|n-\rangle$ of (32) (see Fig. 3). The ratio of this contribution to the ordinary contribution (37) is $12(\Delta/\gamma)^2 \sin^2(2\varepsilon)$, which shows that for polarizations that are not linear and large detunings this term can be the dominant one and exceed the values obtained in the two-level model of the atom by several orders of magnitude.

5. LIMITS OF THE TWO-LEVEL MODEL

Recall that the need to allow for the Zeeman degeneracy of energy levels in the description of the motion of atoms in fields with nonuniform polarization was convincingly demonstrated in Refs. 3 and 4. Our results show that the Zeeman structure must be taken into account even when the polarization is uniform. The results of Secs. 4.2 and 4.3 suggest that the frictional force and diffusion contain additional contributions that do not exist in the two-level model. When the polarization is linear ($\varepsilon=0$), for the $j_g=1/2 \rightarrow j_e=1/2$ transition considered here these contributions vanish, since by virtue of the symmetry of the induced transitions the problem effectively reduces to the two-level one. However, it is obvious that in the general case there is no such symmetry and the additional terms in the friction and diffusion emerge even when the field polarization is linear. For instance, for the $j_g=3/2 \rightarrow j_e=3/2$ and $j_g=1 \rightarrow j_e=2$ transition, in the case of linear polarization the additional contributions to the friction are nonvanishing and at low saturations ($S_0 \ll 1$) have the form

$$\begin{aligned} \xi_{3/2 \rightarrow 3/2} &= \frac{24}{125} \frac{\hbar k^2 \Delta}{\gamma} S_0 \sin^2 kz, \\ \xi_{1 \rightarrow 2} &= \frac{576}{24\,565} \frac{\hbar k^2 \Delta}{\gamma} S_0 \sin^2 kz. \end{aligned} \tag{39}$$

As the additional term in (31) for the $j_g=1/2 \rightarrow j_e=1/2$ transition in a weak field, these contributions emerge because of transitions between the Zeeman sublevels of the ground state $|j_g, \mu\rangle$, whose adiabatic potentials

$$U_\mu = \hbar \Delta S_0 \begin{pmatrix} j_e & 1 & j_g \\ -\mu & 0 & \mu \end{pmatrix}^2 \cos^2 kz \tag{40}$$

have the same spatial dependence but different amplitudes, by virtue of the difference of the probabilities of the induced transitions between the respective magnetic sublevels.

Thus, we can say that in essentially all cases, the kinetics of slow atoms in fields with uniform polarization differs substantially from the kinetics predicted by the two-level model. The problem reduces to the two-level model only in the following exceptional cases: the $j_g=0 \rightarrow j_e=1$ transition for light of arbitrary elliptical polarization, the $j_g=1/2 \rightarrow j_e=1/2$ and $j_g=1/2 \rightarrow j_e=3/2$ transitions for linearly polarized light, and all $j_g=j \rightarrow j_e=j+1$ transitions for circularly polarized light.

6. CONCLUSION

In this paper we have used the semiclassical approach to analyze the kinetics of atoms in a uniformly polarized standing wave with allowance for the Zeeman degeneracy of the energy levels. For the atomic transition $j_g=1/2 \rightarrow j_e=1/2$ we have obtained analytic expressions for the gradient force, the friction and diffusion coefficient, which describe the kinetics of slow atoms, for arbitrary ellipticity, intensity, and detuning. We have shown that allowance for the Zeeman structure of the levels and optical pumping effects leads, first, to a dependence of the effective saturation of the transition (and hence the forces on an atom) on the ellipticity $S_\varepsilon = S_0 \cos(2\varepsilon)/3$ of the light. We have also found that the friction and diffusion coefficients contain additional terms, which are absent in the theory of a two-level atom and have not been discussed before. In the limit of low saturations and large detunings these terms are dominant and lead to much larger friction and diffusion in comparison to the results predicted by the two-level model. We have also given a quantitative interpretation in terms of dressed states,. We have shown that the additional contribution to friction is due to the transitions between the adiabatic potentials corresponding to different magnetic sublevels of the ground state, i.e., is of the Sisyphean origin.

Note that the analytic method developed in this paper [see Eqs. (15)–(22)] for finding the kinetic coefficients of the Fokker–Planck equations for the transitions $j_g=j \rightarrow j_e=j+1$ and $j_g=j' \rightarrow j_e=j'$ (j' is a half-integer) can be used in the general case of a field of arbitrary nonuniform polarization.

This work was supported by a grant from the Russian Fund for Fundamental Research (Grant No. 98-02-17794).

*)E-mail: llf@admin.nsu.ru

---

[1] V. G. Minogin and V. S. Letokhov, *Laser Light Pressure on Atoms*, Gordon and Breach, New York (1987).

[2] A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *Mechanical Effect of Light on Atoms*, World Scientific, Singapore (1990).

[3] Special Issue *Laser Cooling and Trapping of Atoms*, J. Opt. Soc. Am. B **6**, 11 (1989).

[4] Special Issue *Laser Cooling and Trapping*, Laser Phys. **4**, 5 (1994).

[5] S. Chu, Rev. Mod. Phys. **70**, 685 (1998); C. Cohen-Tannoudji, Rev. Mod. Phys. **70**, 707 (1998); W. Phillips, Rev. Mod. Phys. **70**, 721 (1998).

[6] V. S. Letokhov, JETP Lett. **7**, 272 (1968); A. Ashkin, Phys. Rev. Lett. **24**, 156 (1970); T. W. Hänsch and A. L. Schawlow, Opt. Commun. **13**, 68 (1975); V. S. Letokhov, V. G. Minogin, and B. D. Pavlik, Opt. Commun. **19**, 72 (1976).

[7] R. J. Cook, Phys. Rev. A **20**, 224 (1979).

[8] J. P. Gordon and A. Ashkin, Phys. Rev. A **21**, 1606 (1980).

[9] V. G. Minogin, Zh. Éksp. Teor. Fiz. **79**, 2044 (1980) [Sov. Phys. JETP **52**, 1032 (1980)]; V. S. Letokhov and V. G. Minogin, Phys. Rep. **73**, 1 (1981); R. J. Cook, Phys. Rev. A **22**, 1078 (1980).

[10] J. Dalibard and C. Cohen-Tannoudji, J. Phys. B **18**, 1661 (1985).

[11] G. S. Agarwal and K. Mølmer, Phys. Rev. A **47**, 5158 (1993).

[12] K. Berg-Sörensen, Y. Castin, E. Bonderup, and K. Mølmer, J. Phys. B **25**, 4195 (1992).

[13] J. Javanainen, Phys. Rev. A **44**, 5878 (1991).

[14] G. Nienhuis, P. van der Straten, and S.-Q. Shang, Phys. Rev. A **44**, 462 (1991).

[15] J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **2**, 1707 (1985).

[16] V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].

[17] A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **108**, 415 (1995) [JETP **81**, 224 (1995)].

[18] A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **110**, 1727 (1996) [JETP **83**, 949 (1996)]; A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **114**, 125 (1998) [JETP **87**, 70 (1998)].

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Absorption and scattering of high-power laser radiation in low-density porous media

A. É. Bugrov, S. Yu. Gus'kov, and V. B. Rozanov

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

I. N. Burdonskiĭ, V. V. Gavrilov,*) A. Yu. Gol'tsov,†) E. V. Zhuzhukalo, N. G. Koval'skiĭ,
V. N. Kondrashov, M. I. Pergament,‡) V. M. Petryakov, and S. D. Tsoĭ

Troitsk Institute of Innovative and Thermonuclear Studies (TRINITI), 142092 Troitsk, Moscow oblast', Russia

(Submitted 2 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 805–818 (March 1999)

We have experimentally investigated the interaction of high-power neodymium laser pulses in the intensity range $10^{13}-10^{14}$ W/cm^2 with flat low-density (0.5–10 mg/cm^3) agar-agar targets under conditions of interest for problems of inertial nuclear fusion. Optical and x-ray methods with high temporal and spatial resolution were used to examine the dependence of absorption and scattering of the incident beam on the initial mean density and thickness of the irradiated samples. We show that when a porous target is irradiated, a bulk absorption layer of high-temperature plasma is produced inside the target whose dimensions are determined by the initial density of the material. The time dependence and spectral composition of the harmonics $2\omega_0$ and $3\omega_0/2$ observed in the plasma-scattered radiation are measured. A theoretical model is developed that describes the interaction of high-power laser pulses with a porous medium. Predictions of the model, based on the hypothesis of two stages of homogenization of the target material—a fast stage (0.1–0.3 ns) and a slow stage (1–3 ns), are in good agreement with the experimental data.
[S1063-7761(99)00303-0]

1. INTRODUCTION

The interaction of high-intensity laser radiation (10^{12}–10^{14} W/cm^2) with porous media whose mean densities vary over a wide range—from 10^{-1} mg/cm^3 to 10^2 mg/cm^3—is an extraordinarily active area of study, due primarily to the capability it offers of investigating fundamental physical phenomena taking place in a nonequilibrium laser plasma. When porous material is heated by a laser pulse, a plasma is produced that is spatially homogeneous in density, temperature, and pressure. The production and interaction of plasma fluxes propagating from the heated elements of the porous material leads to a plasma nonequilibrium state: the temperature of the electron component of the ejected plasma flux exceeds the ion temperature whereas during hydrothermal dissipation incident to a collision of fluxes a plasma is produced whose ion temperature can significantly exceed the electron temperature.[1]

According to the ideas developed in Refs. 1–4, absorption of laser radiation inside a porous material occurs along a length determined by the geometrical transparency, which decreases with increasing dimensions of the regions filled by plasma with density exceeding some critical value. It can be expected that the production of an absorbing region is by no means over during the first collisions of plasma fluxes inside the pores. At laser intensities of $(10^{13}-10^{14})$ W/cm^2 the complete homogenization time of a porous material with density 10^{-1} g/cm^3–10^{-3} g/cm^3 during decay of the macro-oscillations of the plasma density can exceed by one to two orders of magnitude the characteristic time of collisions between the plasma fluxes from neighboring elements of the medium (which is of the same order of magnitude as the time it takes an ion, on average, to traverse the distance between neighboring elements of the structure—the mean transit time), and can reach several nanoseconds.[3] Thus, the nonequilibrium state of the laser plasma produced in a porous medium can be characterized by long-wavelength density oscillations, the excitation and collisions of shock waves, and hydrothermal dissipation processes taking place during under these conditions. These phenomena, without a doubt, must have a substantial effect on laser absorption processes and energy transport in the porous material, and also on the form of the equation of state of the nascent plasma.

Practical interest in the interaction of laser radiation with porous media and, in general, bulk-structured media of reduced density is motivated by their use as multipurpose elements in laser nuclear fusion (LNF) targets. Above all, this has to do with the ability to equalize heating inhomogeneities and ensure target stability under compression. It has been suggested that porous media consisting of light elements be used, for example, in ''laser greenhouse'' targets to implement distributed bulk absorption of the laser beam energy, thermal-conduction mediated equalization of heating inhomogeneities, and spatiotemporal profiling of the ablation pressure.[6] The properties of porous media with distributed heavy-element impurities have been actively investigated from the standpoint of controlling the spectral composition of x radiation from the nascent plasma.[7]

Novel capabilities for producing nonequilibrium, nonstationary plasma emerge from the irradiation of regularly structured material. As an example, we may cite the suggestion voiced in Ref. 8 of creating a relatively long-lived plasma with ion temperature several times the electron temperature, and achieving on this basis an intense source of thermonuclear neutrons with a yield in the range $10^{11}-10^{12}$ neutrons per joule of laser energy when a laser pulse acts on a medium consisting of either a system of plane-parallel films or thin shells containing thermonuclear fuel.

Physical processes played out in low-density bulk-structured targets when they are irradiated by intense laser pulses have recently attracted the attention of numerous researchers working in the LNF field. Thus, the authors of Ref. 4, working at the TRINITI *Mishen'* (*Target*) facility, obtained an overall picture of laser energy absorption processes taking place in a porous medium, the transformation and transport of energy in targets of finite thickness, and the acceleration of thin metallic foils situated on the back surface of such targets, and suggested a theoretical model for describing energy absorption and transport processes. The present work continues studies initiated in Ref. 4, and is dedicated to a detailed study of absorption and scattering of high-power laser radiation in low-density porous targets. On the basis of measurements of the scattered radiation and the radiation transmitted by the target, we have obtained data on the efficiency of absorption of laser energy in a porous medium at both sub- and supercritical mean density. We have experimentally established the density dependence of the longitudinal diameter of the laser absorption region over a wide range, from 5×10^{-4} g/cm^3 to 10^{-2} g/cm^3. The results are discussed from the standpoint of the properties of the nonequilibrium laser plasma produced in the porous medium. In particular, we have developed a theory of two-stage homogenization of a porous medium subject to a high-power laser pulse. Of special interest are studies, completed in the present work, on plasma emissions observed in experiments with porous targets at harmonics of the laser frequency. These studies are of special interest by virtue of the fact that the development of different kinds of anomalous processes (Mandel'shtam–Brillouin scattering, Raman scattering, parametric decay instabilities) is known to lead to consequences that are undesirable for LNF (a decrease in the absorption coefficient, generation of fast superthermal particles, reduction of the conversion coefficient of laser radiation into x radiation). Porous low-density media can only be applied to LNF target designs on the basis of detailed studies of laser interactions with the materials in question.

2. EXPERIMENTAL CONDITIONS AND DIAGNOSTICS

The interaction of intense laser radiation with low-density structured targets was studied experimentally at the *Mishen'* facility[9] under the following conditions of irradiation: wavelength $\lambda=1.054\,\mu$m, laser pulse energy up to 150 J for pulse duration ~2.5 ns with a 0.3-ns leading edge; energy contrast no worse than 10^6, radiation focused on the target by an 1:10 lens. For a focal-spot diameter of $\sim250\,\mu$m the mean light flux density at the surface of the irradiated target reached 10^{14} W/cm^2. Flat agar-agar $[C_{12}H_{18}O_9)_n]$ targets $100-1000\,\mu$m in thickness, with density in the range 0.5–10 mg/cm^3, were mounted in the interaction vacuum chamber. (For agar samples of mean density 3 mg/cm^3 the electron density corresponds to the critical value n_{cr} for $\lambda=1.054\,\mu$m.) Multilayer targets were also used in which aluminum layers of thickness $1-6\,\mu$m were deposited onto the irradiated surface, back surface, or both surfaces of the low-density material. The agar used in the experiments is a random structure of solid filaments of diameter $1-2\,\mu$m with an interfilament spacing of $10-50\,\mu$m. Figure 1 shows electron micrographs of the structure of agar samples with a density of 1 mg/cm^3 and 10 mg/cm^3.

The experiments mainly used optical and x-ray diagnostics. X-ray diagnostics allowed time-integrated calorimetric measurements in the wavelength range 0.5–1.5 nm, the construction of plasma images by camera-obscura with assorted filters, recording of the x-ray emission of the plasma by vacuum diodes with temporal resolution ~0.5 ns, measurements of x-ray spectra of the plasma in the spectral range 0.5–2 nm. The optical methods included multiframe shadowgraphs of the nascent plasma (exposure time of an individual frame ~0.3 s, wavelength of the probing radiation $\lambda=0.527\,\mu$m, spatial resolution $\sim30\,\mu$m); electro-optical recording with an Agat-SF camera of the time dependence of the luminescence of the back surface of the irradiated target in the wavelength range 400–700 nm with spatial resolution $\sim30\,\mu$m and temporal resolution 50 ps; measurements of the reflected, scattered, and transmitted laser radiation; both time-integrated and time-resolved spectral measurements of the radiation scattered by the plasma into the aperture of the focusing lens in the vicinity of the frequencies $2\omega_0$ and $3\omega_0/2$. The diagnostic setup is shown in Fig. 2.

3. EXPERIMENTAL RESULTS

Earlier, in Ref. 4 it was shown that absorption of laser radiation during interaction with a porous medium is a bulk phenomenon. As a result of energy absorption and transport inside the porous target, a layer of high-temperature plasma is produced. Figure 3 displays typical obscurograms (and results of their processing), obtained with a beryllium filter of thickness 50 μm observed perpendicular to the direction of the laser beam in the irradiation experiments with agar targets of thickness 500 μm with different mean densities. For comparison, this same figure shows the x-ray intensity distribution of the plasma for irradiation of a thick Lavsan (Mylar) film. It can be seen from the figure that upon irradiation of a porous target, plasma formation inside the target is highly extended in the direction of propagation of the laser beam.

Figure 4 depicts the dependence of the longitudinal diameter of the plasma layer on the initial density of the agar target. For a target with initial density 1 mg/cm^3 this length amounts to $300-400\,\mu$m, and for a target with initial density 10 mg/cm^3 it decreases to $100-150\,\mu$m. In the direction perpendicular to the laser beam, the diameter of the hot plasma region inside the porous target essentially coincides

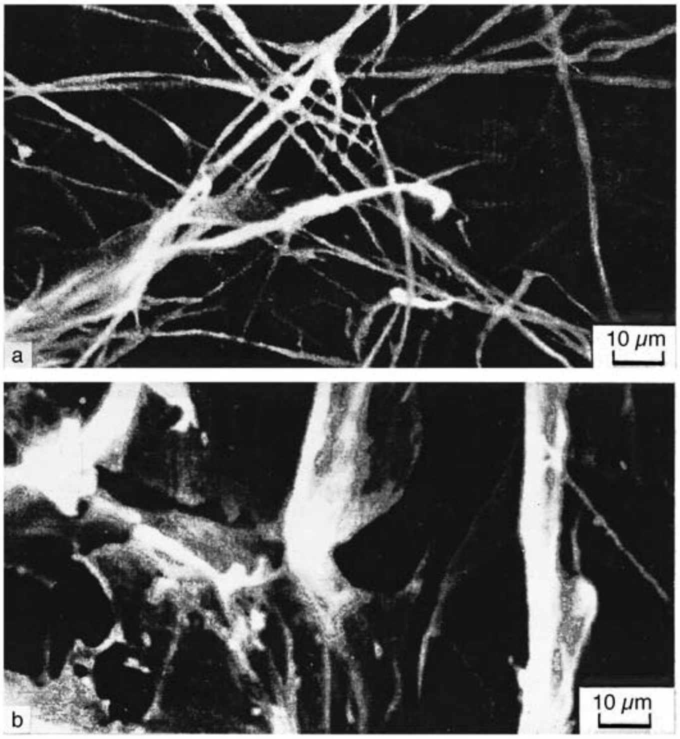

FIG. 1. Electron micrographs of the structure of agar samples with mean density 1 mg/cm^3 (a) and 10 mg/cm^3 (b).

with the diameter of the focal spot and depends weakly on the density of the agar target. Thus, for irradiation of an agar target with density 1 mg/cm^3 this length is approximately 1.2 times that of a target with density 10 mg/cm^3. The duration of the x-ray pulse recorded by the vacuum diodes is 3–4 ns, i.e., it essentially coincides with the duration of the laser pulse. The electron temperature of the plasma was determined by the method of filters and for porous targets of all types was 0.8–1 keV, decreasing somewhat with depth into the target. For a target with density 1 mg/cm^3, $T_e \approx 0.6 - 0.7$ keV at a distance $\sim 300\,\mu$m from the irradiated surface. It should be borne in mind that the displayed data are the result of processing time-integrated measurements whereas processes in the nascent absorbing layer can have a nonstationary character associated with evaporation along the extent of the laser pulse of solid elements of the structure, with interaction of plasma fluxes from separate elements, and with variations in the dimensions of the radiation-absorbing plasma layer and a displacement toward greater depths in the target of the zone of maximum energy liberation.

Important information about the interaction of laser radiation with the extended plasma produced in a porous medium is provided by measurements of the scattered and transmitted radiation. Results of measurements of the energy of the laser radiation scattered by the plasma into the aperture of the focusing lens are plotted in Fig. 5. Data obtained for irradiation of thin Lavsan films are also plotted for comparison. As can be seen from Fig. 5, the energy of the radiation

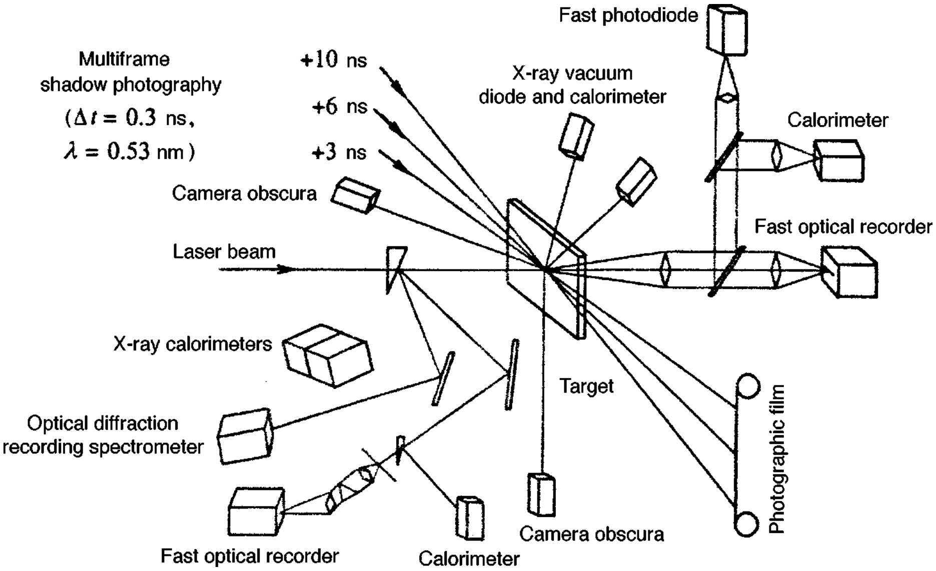

FIG. 2. Setup of the diagnostic system in the interaction vacuum chamber.

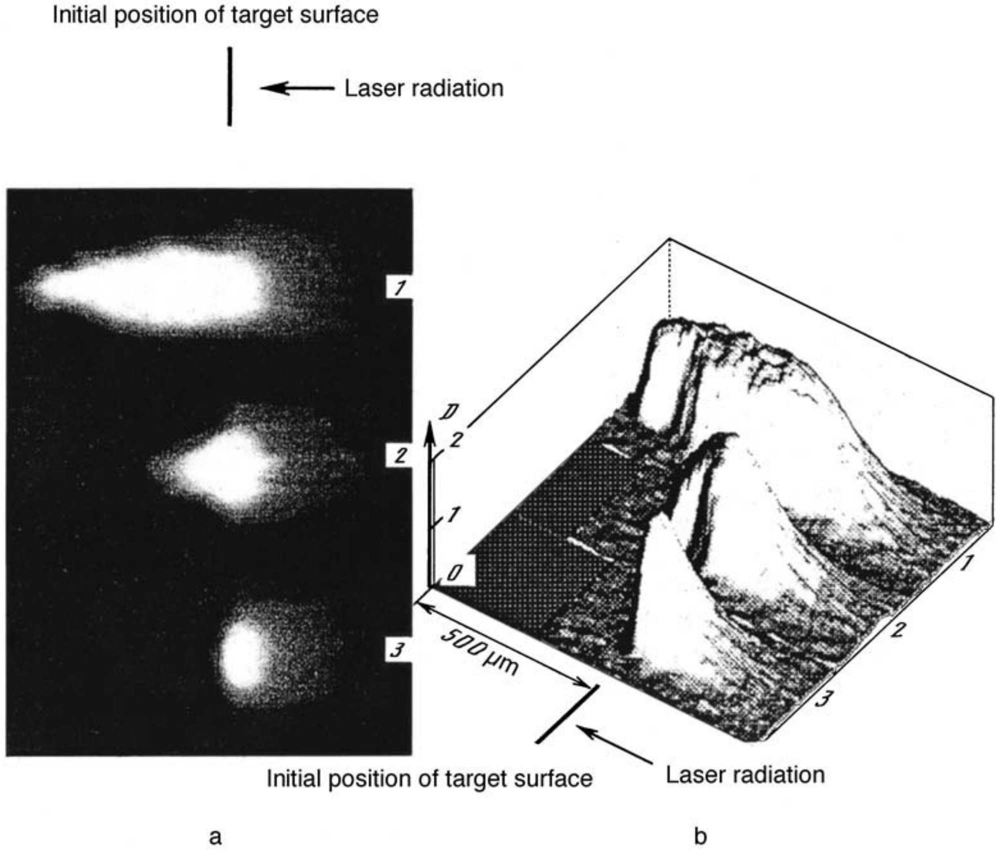

FIG. 3. Typical obscurograms (a) for irradiation of agar targets with mean density 1 mg/cm^3 (*1*) and 4 mg/cm^3 (*2*) and Lavsan targets with density 1.4 g/cm^3, and results of processing of these obscurograms (b).

scattered from low-density targets is similar to the corresponding values for the solid targets over the entire range of intensities of the focused laser beam. Figure 6 shows a characteristic directionality diagram for the plasma-scattered laser radiation, constructed from calorimetric data.

Measurements of the target-transmitted laser radiation (collection angle corresponded approximately to the angle of approach of the laser beam to the target $\approx 6°$)—both time-integrated and time-resolved—show that for all porous targets with thickness $>500\,\mu$m (i.e., exceeding the characteristic dimension L^* of the high-temperature plasma region) and densities from 0.5 mg/cm^3 to 10 mg/cm^3 the transmittance during the laser pulse did not exceed 0.1%. In the case of a low-density porous target (density ~ 0.5 mg/cm^3), as one goes to target thicknesses of $200-300\,\mu$m (less than L^*) an abrupt (by as much as $\sim 2\%$) growth in the transmittance is observed. This supports the conclusion of more efficient absorption of the heating laser radiation in the nascent high-temperature plasma layer. As the time-resolved measurements of the transmitted light intensity show, in these cases the transmittance grows up to the end of the heating

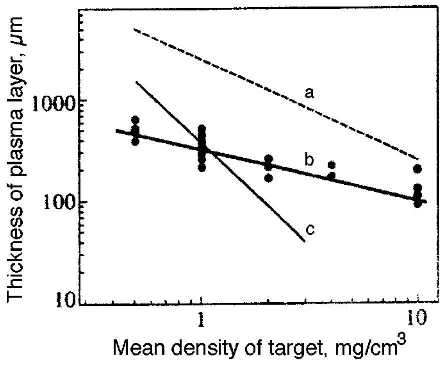

FIG. 4. Calculated dependence on the agar mean density of the characteristic length of geometrical transparency of the medium in the initial stage of irradiation (a) and after the first stage of homogenization (b), and of the characteristic length of the inverse bremsstrahlung radiation (c). The points plot experimental data on the longitudinal diameter of the bulk-absorbing plasma layer.

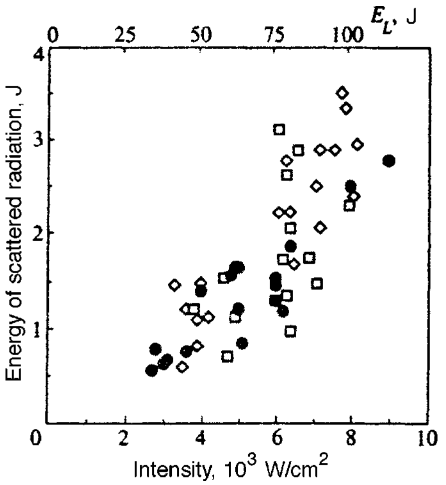

FIG. 5. Dependence of the energy of the radiation scattered by the plasma into the aperture of the focusing lens, on the power density of the heating radiation in experiments on irradiation of an agar target with mean density 1 mg/cm^3 (◇) and 10 mg/cm^3 (□) and a Lavsan film (●).

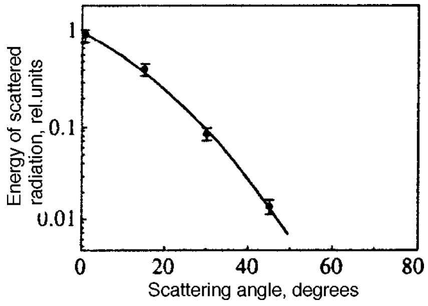

FIG. 6. Directionality diagram of the plasma-scattered laser radiation. Zero angle corresponds to scattering in the direction opposite the laser beam. Agar target with mean density 1 mg/cm^3, light flux density 5×10^{13} W/cm^2. The measurements were performed simultaneously by the four calorimeters.

pulse (see Fig. 7). This fact can be explained on the basis of ideas about ''clearing'' of the absorbing layer prior to termination of the laser pulse[4] and also by assuming that the zone of maximum energy liberation is shifted deeper into the absorbing layer at later stages of target irradiation.

To obtain additional information about the dynamics of processes in an extended plasma in porous targets, we performed time-integrated and time-resolved measurements of the radiation scattered into the aperture of the focusing lens in the vicinity of the frequencies $2\omega_0$ and $3\omega_0/2$. Figure 8 plots processed time scans of the luminescence at the harmonics $2\omega_0$ and $3\omega_0/2$ recorded from irradiated agar samples with mean densities of 0.5 mg/cm^3, 4 mg/cm^3, and 10 mg/cm^3. For comparison, this same figure plots time scans obtained with a Lavsan target. Figures 9 and 10 show density plots of time-integrated spectra of the harmonics $2\omega_0$ and $3\omega_0/2$, respectively. The spectra of both harmonics have a two-component structure with pronounced asymmetry in the case of the harmonic $2\omega_0$. Note that in each experiment the frequency-doubled component of the incident laser radiation also went through the entrance slit of the recording spectrometer. Considering all of the data obtained in the spectral measurements, we first note that in the irradiation of low-density targets harmonic generation takes place even when the mean electron density turns out to be subcritical for $\lambda=1.054\,\mu$m, even though the targets are fully ionized. Also noteworthy is the weak dependence of the spectral structure and time dependence of the harmonic $3\omega_0/2$ on the conditions of the experiment, specifically the time dependence of the laser radiation intensity, the intensity of the heating beam focused on the target surface, and the initial density of the target (see Figs. 8 and 10). The spectral structure of the $2\omega_0$ harmonic radiation (Fig. 9) also shows only a weak dependence on the target density and at a density of 10 mg/cm^3 it is essentially identical to the corresponding spectrum for the solid Lavsan targets. When the flux density of the light incident on the target is reduced to 10^{13} W/cm^2, the intensity of the ''red wing'' of the spectral distribution falls abruptly while the intensity of the primary maximum decreases only negligibly. Also note the ''redshift'' of the primary maximum in the spectrum of the second harmonic, which for targets with density 0.5 mg/cm^3 is $\sim$0.5 nm.

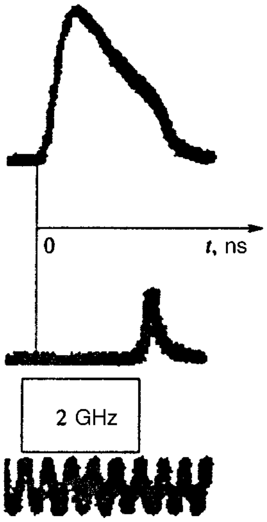

FIG. 7. Laser pulses: a — incident on target, b — after passing through an agar target with mean density 1 mg/cm^3 and thickness 250 μm.

4. THEORETICAL MODEL AND DISCUSSION OF EXPERIMENTAL RESULTS

The basic distinguishing property of porous materials is their inhomogeneity. In such bulk-structured media absorption of laser radiation, energy transport mechanisms, and hydrodynamic processes can be of a very specific nature. Following the ideas developed in Refs. 3 and 4, we may, nevertheless, expect that the structural distinctiveness of any given medium (fibrous, foamy, finely dispersed media, etc.) will show up only upon formation of a bulk-absorbing layer of plasma, i.e., in the initial stage of homogenization of the target material.

Let us consider homogenization of porous material irradiated by a focused laser beam. Homogenization of porous material acted upon by a powerful laser pulse takes place in two stages. The first, or fast, stage of partial homogenization is the result of evaporation of solid elements of the porous material and subsequent collisions between the nascent plasma fluxes. This stage culminates in the formation of an inhomogeneous plasma in which the dimensions of the dense regions of material are significantly greater, and the density in them is substantially less, than in the original elements of the porous material. Final equalization of the density takes place in the second—slower—stage of homogenization. The basic processes in this stage are collisions and dissipation of shock waves excited in the plasma filling the space between the regions of material with increased density. The duration of the first stage is of the same order of magnitude as the time it takes the evaporated material to traverse the distance between neighboring elements:

$$t_1=\frac{l_0}{v_{\text{eff}}}\approx\left[\frac{3\pi^{7/2}}{2[3(\gamma-1)]^{1/2}}\right]^{2/3}\left(\frac{\rho_s}{\rho_a}\right)^{2/3}\frac{\rho_s^{1/3}b_0^{5/3}}{\dot{E}_L^{1/3}}.$$

Here b_0 is the radius of the solid element, $l_0=b_0(\rho_s/\rho_a)^{1/2}$ is the mean distance between elements, ρ_s and ρ_a are respectively the density of the solid material and the mean density of the porous material, $\dot{E}_L$ is the power of the laser beam, and v_{eff} is given by

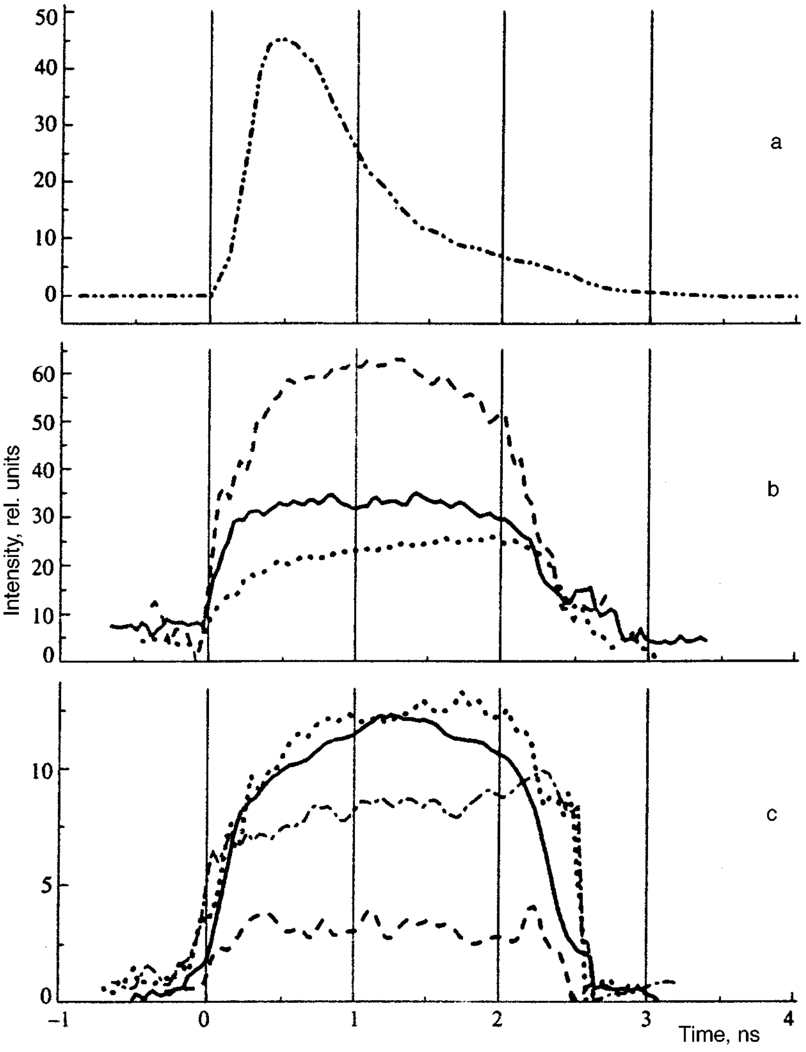

FIG. 8. Laser pulse (a) and intensities of luminescence of the $3\omega_0/2$ harmonic (b) and the $2\omega_0$ harmonic (c) for irradiation of Lavsan targets (dotted curves) and agar targets with mean density $0.5\,\mathrm{mg/cm^3}$ (dashed curve), $4\,\mathrm{mg/cm^3}$ (dash–dot curves), and $10\,\mathrm{mg/cm^3}$ (solid curve).

$$v_{\mathrm{eff}}=\left[\frac{(\gamma-1)\dot{E}_L}{3\pi^2\pi L_0^2\rho_a}\right]^{1/3}\left(\frac{\rho_s}{\rho_a}\right)^{1/6}.$$

This velocity is the effective ejection velocity of the material under conditions of a shrinking absorption region and, consequently, increasing plasma temperature.

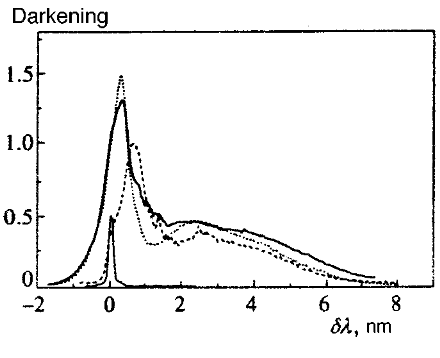

FIG. 9. Density plots of time-integrated spectra of the harmonic $2\omega_0$, recorded in irradiation experiments with Lavsan targets (dotted curve) and agar targets with mean density $0.5\,\mathrm{mg/cm^3}$ (dashed curve) and $10\,\mathrm{mg/cm^3}$ (solid curve). The thin line indicates the spectral position of the frequency-doubled laser radiation.

For the laser pulse power $\dot{E}_L\approx 1.6\times 10^{10}\,\mathrm{W}$ the duration of the first stage $t_1\approx 230\,\mathrm{ps}$ for $\rho_a=10^{-3}\,\mathrm{g/cm^3}$ and $t_1\approx 70\,\mathrm{ps}$ for $\rho_a=10^{-2}\,\mathrm{g/cm^3}$. During this initial interaction phase, laser radiation penetrates the porous material to a depth corresponding to the initial transparency depth[1]:

$$L_0\approx\frac{\pi^2}{2}\frac{\rho_s}{\rho_a}b_0. \tag{1}$$

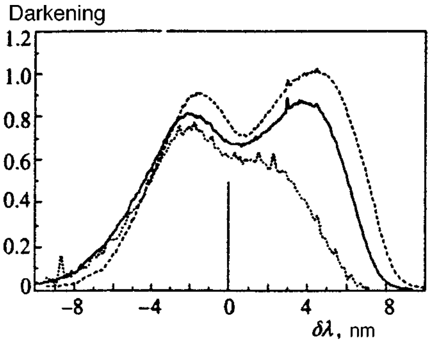

FIG. 10. Density plots of time-integrated spectra of the harmonic $3\omega_0/2$, recorded in irradiation experiments with Lavsan targets (dotted curve) and agar targets with mean density $0.5\,\mathrm{mg/cm^3}$ (dashed curve) and $10\,\mathrm{mg/cm^3}$ (solid curve).

For the densities 10^{-3} g/cm^3 and 10^{-2} g/cm^3 the quantity L_0 is equal to 4500 μm and 450 μm, respectively. Targets thinner than the indicated values are partially transparent to laser radiation during the initial stage of homogenization, as is observed experimentally.[4]

A calculation based on the picture of isothermal ejection of dense cylindrical elements and adiabatic compression of colliding plasma fluxes with conversion of hydrodynamic energy into heat in a plasma layer of thickness equal to the ion–ion collision length gives the following expression for the ratio of the diameter of the nascent dense plasma region to the initial thickness of the solid element[3]:

$$\frac{b_1}{b_0}=\left(\frac{2b_0}{l_0+\lambda_i}\right)^{1/(\gamma-1)}\left[\left(\frac{l_0-\lambda_i}{b_0}\right)\left(\frac{a}{2}\right)^{1/\gamma}+\frac{\lambda_i}{b_0}\right]^{\gamma/(\gamma-1)}. \qquad (2)$$

Here λ_i is the ion–ion collision length and a is the fraction of energy contained in the thermal component for isothermal ejection of the particle.

According to the self-similar solution,[10] the value of a for particles of differing flight geometry is

$$a=\frac{2}{2+3(\gamma-1)(\beta+1)},$$

where $\beta=0,1,2$ for planar, cylindrical, and spherical geometry, respectively. Under conditions in which a particle is ejected to distances significantly greater than its dimensions $(l_0 \gg b_0)$, we must choose a value for a intermediate between the cases of planar and cylindrical ejection: $a\approx 0.42$.

Formula (2) is valid for $\lambda_i \leqslant l_0$ when the hydrodynamic description of the colliding plasma fluxes is applicable, and as λ_i tends to l_0 from below it gives the valid limit $b_1 \to l_0$ corresponding to complete homogenization of the plasma fluxes under conditions in which $\lambda_i = l_0$. Since $\lambda_i \ll l_0, b_0$ and $b_0 \ll l_0$ in our case, Eq. (1) yields the approximate result

$$b_1 \approx a^{1/(\gamma-1)} l_0 \approx l_0/3.$$

Thus, as a result of the first stage of homogenization a plasma is produced in which the dimensions of the dense regions are equal to roughly half the original distance between the solid elements, and the density in them is several times the mean density of the porous material. In this case, enlargement of the dense regions is accompanied by a decrease in the geometrical transparency length:

$$L_1 \simeq \frac{\pi^2}{2} b_0 \left(\frac{\rho_s}{\rho_a}\right)\frac{b_0}{b_1} \approx \frac{\pi^2}{2a^{1/(\gamma-1)}} b_0 \left(\frac{\rho_s}{\rho_a}\right)^{1/2}. \qquad (3)$$

Now, for material with density $\rho_a = 10^{-3}$ g/cm^3 the geometrical transparency length is $\sim 300\,\mu$m. Figure 4 plots the calculated dependence of the geometrical transparency on the density of the medium: a — in the initial stage of irradiation (using (1) with $b_0 = 1\,\mu$m), b — after the first stage of homogenization (using (3) with $b_0 = 1\,\mu$m). This same figure plots the density dependence of the characteristic inverse bremsstrahlung absorption length (c) given by[1]

$$L_b \approx \frac{5.2\times 10^{-4} T_e^{3/2}}{\rho_a^2}\,\mu\text{m}$$

(where T_e is measured in keV and ρ_a in g/cm^3) for $T_e = 800$ eV and $\rho_a = 1$ mg/cm^3. The experimental data are in good agreement over the entire density range with the theoretical dependence corresponding to the plasma state after the first stage of homogenization.

In the second stage of homogenization the diameter of the dense region increases by roughly the ion–ion collision length during each traversal by a shock wave of the distance between neighboring elements:

$$\frac{db}{dt} \approx \frac{\lambda_i}{\tau}.$$

Here $\lambda_i = v/\nu_{ii}$, $\nu_{ii} = \nu_* \rho_a / T^{3/2}$ is the ion–ion collision frequency (it is assumed that $T = T_e = T_i$), $\nu_* \approx 1.4 \times 10^{14}\,(\text{keV})^{3/2}/(\text{g/cm}^3)\cdot\text{s}$, and $\tau \approx l_0/v$ is the shock transit time at velocity v between dense regions. Solving this equation yields an estimate for the duration of the second stage,

$$t_2 \approx \frac{(l_0-b_1)^2}{\lambda_i v} \approx 10^{-9}\frac{l_0^2 \rho_a}{T^{5/2}},$$

where t_2, l_0, ρ_a, and T, are measured respectively in ns, μm, g/cm^3, and keV. Estimates using this formula give the values 5 ns and 3 ns for the duration of the slow stage for $T = 1$ keV for material with density 1 mg/cm^3 and 10 mg/cm^3, respectively.

The brief duration of the first stage corresponds to the time interval at commencement of irradiation, during which partial transparency of the target is observed. In the second (slow) stage, whose duration under the conditions of our experiment exceeds the duration of the laser pulse, the model describes the dynamics of the plasma in the nascent extended layer inside the target. In this stage the influence of the specific structural properties of the target material is manifested to a lesser extent. The agreement between the measured values of the time-integrated length of the absorption region (corresponding to the entire duration of the laser pulse) and the theoretical dependence of the geometrical transparency length after the first (fast) stage of partial homogenization [Eq. (3)] is indirect proof of the presence of a second (slow) stage of complete homogenization.

Note that the existence of an extended stage of complete homogenization of porous material leads to the existence, over the course of the entire laser pulse, of an inhomogeneous plasma with regions having density both above and below critical. Such a state of the plasma allows the laser radiation to penetrate the porous material to the bottom the bulk-absorbing layer.

The above model of homogenization, which is in good agreement with the experimental data, leads to the conclusion that under the conditions of our experiments, during the greater extent of the laser pulse, the radiation interacts with the extended plasma layer in which spatial density modulations exist.

In light of the above discussion, let us analyze the experimental data obtained from measurements of the harmonics, generated in the porous-target plasma, of the heating radiation. The time dependence of the luminescence intensity of the $3\omega_0/2$ harmonic and its spectral structure for porous targets of different densities differs only slightly from the dependence observed for a Lavsan target (see Figs. 8 and 10). For each of the investigated porous targets with densities in the range 0.5–10 mg/cm^3 the luminescence intensity of the $3\omega_0/2$ harmonic grows rapidly in the leading edge of the laser pulse and remains essentially unchanged up to the end of the pulse. Generation of the $3\omega_0/2$ harmonic is a consequence of Raman scattering of the laser radiation by plasma waves with frequency $\omega_0/2$, excited as a result of the development of a resonant parametric instability of two-plasmon decay in a plasma whose electron density is one-fourth the critical value. Thus, the experimentally observed dependences indicate that by the earlier stages of irradiation, regions with plasma density $\sim n_{cr}/4$ are produced inside the target and exist for quite a long time. This is in good agreement with the foregoing theoretical model of homogenization. As for the harmonic $2\omega_0$, for which the denser regions of the plasma with electron density $\sim n_{cr}$, in agreement with the theoretical model a plasma with density $n > n_{cr}$ exists in the form of macroscopic density oscillations with a spatial scale of $3-10\,\mu$m for an extended time $\sim 2-3$ ns. A decrease in the intensity of the $2\omega_0$ harmonic before the end of the laser pulse might be expected either in the case of strong enough decay of such oscillations or by using sufficiently thin low-density targets. Under the conditions of our experiments for target thicknesses exceeding 500 μm and laser pulse duration 2.5 ns, no decrease in the intensity of the harmonic before the end of the laser pulse was observed. Extended duration of plasma regions with density $n > n_{cr}$ can also be facilitated by the effect of motion of the plasma produced by one-sided laser irradiation of the fibers of the porous material into the inner region of the porous medium (along the direction of propagation of the laser beam). This effect can also explain the redshift of the spectrum at $2\omega_0$ recorded in experiments on porous targets with subcritical mean density (see Fig. 9). The magnitude of the shift (~ 0.5 nm) corresponds to a directed velocity of 3×10^7 cm/s.

5. CONCLUSION

We have studied absorption and scattering of intense laser radiation ($\lambda = 1.054\,\mu$m, $\tau = 2.5$ ns, $I = 5\times 10^{13}$ W/cm^2), both theoretically and experimentally, by irradiating flat, low-density targets ($0.5-10$ mg/cm^3).

We have shown that when a porous target is irradiated, a bulk-absorbing layer of high-temperature plasma is produced in it. We have also derived the dependence of the dimensions of this layer on the initial mean density of the target.

We have established that the levels of absorption and scattering in porous targets and in thin films of a solid material with analogous elemental composition differ only slightly, one from another.

In the porous-target plasmas, generation of the $2\omega_0$ and $3\omega_0/2$ harmonics (ω_0 is the frequency of the heating radiation) was observed over the full range of mean density (from $3n_{cr}$ down to $0.2n_{cr}$), and the time dependence of emission at these frequencies was investigated as a function of the target density and conditions of irradiation.

A theoretical model has been developed which, taking account of the specifics of the structure of the porous material, describes the formation in the thickness of the irradiated target of a bulk-absorbing layer of plasma. The model is based on the concept of two stages of homogenization of the target material: a fast stage (0.1–0.3 ns) and a slow stage (1–3 ns). The structural details of each porous material are most important in the fast stage, i.e., the early stage of irradiation.

The predictions of the theoretical model are found to be in good quantitative agreement with x-ray and calorimetric measurements. Spectroscopic and time-resolved measurements of the intensity of the $2\omega_0$ and $3\omega_0/2$ harmonics are consistent with the theoretical picture.

We are indebted to S. F. Medovshchikov and V. A. Barsuk for preparation and testing of the porous targets, to B. F. Vasil'ev, V. V. Kryzhko, B. N. Mironov, V. G. Nikolaevskiĭ, and A. S. Skryabin for assistance with the experiments, and O. L. Dedova and M. Yu. Sadkov for their part in the processing of the experimental data.

This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grants No. 97-02-19157, 98-02-16660, and 98-02-16662).

*)E-mail: vvgavril@fly.triniti.troitsk.ru
†)E-mail: goltsov@fly.triniti.troitsk.ru
‡)E-mail: pergam@fly.triniti.troitsk.ru

---

[1] S. Yu. Gus'kov and V. B. Rozanov, Kvantovaya Élektron. **24**, 715 (1997).
[2] S. Yu. Gus'kov and V. B. Rozanov, in *Advances in Laser Interaction with Matter and Inertial Fusion*, G. Velarde *et al.* (eds.), World Scientific, Singapore (1997), p. 142.
[3] S. Yu. Gus'kov, Preprint No. 49, P. N. Lebedev Physical Institute, (1998).
[4] A. É. Bugrov, I. N. Burdonskiĭ, V. V. Gavrilov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 903 (1997) [JETP **84**, 497 (1997)].
[5] A. Caruzo, S. Yu. Gus'kov, N. N. Demchenko *et al.*, J. Russ. Laser Res. **18**, 464 (1997).
[6] S. Yu. Gus'kov, N. V. Zmitrenko, and V. B. Rozanov, Zh. Éksp. Teor. Fiz. **108**, 548 (1995) [JETP **81**, 296 (1995)].
[7] J. Lindl, Phys. Plasmas **2**, 3933 (1995).
[8] S. Yu. Gus'kov, V. B. Rozanov, and N. V. Zmitrenko, JETP Lett. **66**, 555 (1997).
[9] V. A. Bolotin, I. N. Burdonsky, V. V. Gavrilov *et al.*, Rev. Sci. Instrum. **61**, 3259 (1991).
[10] V. S. Imshennik, Dokl. Akad. Nauk kSSSR **131**, 1287 (1960) [Sov. Phys. Dokl. **5**, 253 (1960)].

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Effective interaction potential and ordered structures of dust particles in a gas-discharge plasma

O. M. Belotserkovskiĭ

Institute of Planning Automation, Russian Academy of Sciences, 123056 Moscow, Russia

I. E. Zakharov

Silicon Graphics Computer Systems, SGI Europe, 1196 Gland, Switzerland

A. P. Nefedov, V. S. Filinov, and V. E. Fortov

Joint Institute of High Temperatures, Russian Academy of Sciences Research Center for the Thermal Physics of Pulsed Phenomena, 127412 Moscow, Russia

O. A. Sinkevich

Moscow Energy Institute, 111250 Moscow, Russia

(Submitted 20 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 819–836 (March 1999)

An effective potential is proposed for the interaction between dust particles in a gas-discharge plasma which takes account of the following physical factors: the spatial dependence of the particle charges on the floating potential of the plasma, anisotropy of the interaction, resulting from focusing of the negatively charged particles of the drift ion current, and aspects of screening of the dust particles by plasma electrons and ions which interact strongly with them and recombine faster in their vicinity and on their surface. Monte Carlo calculations explain the formation of threadlike structures of dispersed particles, and also ‘‘transverse crystallization’’ of these ‘‘threads’’ in a stratified gas-discharge plasma. © *1999 American Institute of Physics.* [S1063-7761(99)00403-5]

1. INTRODUCTION

Growing interest in the properties of dusty plasma stems to a significant extent from the recent discovery of the formation of ordered structures of charged macroparticles in various types of laboratory plasma: in the near-cathode region of a high-frequency gas discharge,[1–11] in thermal plasma,[12–14] and in the standing strata of a glow discharge.[15–18] Dust crystals possess a number of unique properties: they are optically transparent, ranging in size from 1 to 100 μm, and the interstitial distance in the crystal is $100-1000\,\mu$m, which allows one to examine their properties in the visible with the unaided eye. Characteristic relaxation times are fractions of a second, which neatly distinguishes plasma crystals from colloidal crystals, in which these times reach several days. The parameters of the quasicrystalline structures can be varied by simply varying the gas pressure, current, or discharge power. Thanks to this, dust crystals are an effective instrument for studying the properties of strongly nonideal plasma, the fundamental properties of crystals, and the properties of gas discharges. Interesting experimental results have been obtained on polymorphic phase transitions between various crystal structures, on the melting of dust crystals, and on the propagation of sound waves in plasma crystals. Theoretical studies are being actively pursued on dust-particle charging processes , particle interactions with one another and with external fields, and collective effects in a highly nonideal dusty plasma. These questions are discussed in recently published reviews (Refs. 19–21).

Dust particles in a trap with an hf discharge plasma form a flat crystal consisting of several horizontal layers, each hexagonally structured. Thus, an important property of this structure is the fact that in the vertical direction, along the discharge axis, the dust particles are stacked one directly above another. A similar property in the arrangement of dust particles is observed in the plasma of a stationary glow-discharge,[15–17] where ordered structures are formed that have a hexagonal-like lattice in the horizontal plane and a similar vertical ordering along the discharge axis. Moreover, in a gas-discharge plasma, even isolated, randomly spaced threadlike structures are formed.

To understand the physical mechanisms leading to the emergence of ordered structures, an analysis of the interaction forces between dust particles and estimates of their charges are needed. The following expression is customarily used in the literature[19–21] to estimate the charge:

$$Q(\mathbf{r})=Z(\mathbf{r})e=C\phi_t(\mathbf{r}),\quad C\approx 4\pi\epsilon_0 R,$$

where C is a quantity of the order of the capacitance of the particle, R is its radius, and $\phi_t(\mathbf{r})$ is a floating potential governed by the difference in the electron and ion fluxes incident upon the dust particle, which depends substantially on the high-energy asymptotic behavior of the electron distribution function. It is well known that in an hf gas discharge and in a glow-discharge the asymptotic limits of the electron distribution function depend substantially on the spatial inhomogeneity of the electric fields accelerating the electrons, and on atomic ionization processes induced by those electrons.[19–21] This circumstance is the reason why the charge on a dust particle should depend on its location in the stratum, i.e., on its coordinates: $Q=Q(\mathbf{r})$. According to the above estimates, both in an hf discharge plasma and in a glow-discharge plasma, particle charges $Z(\mathbf{r})$ range from 10^4 to 10^6 (Refs. 19–21). Below we take $Z\approx 10^5$ as a typical value. Consequently, one of the main components of the interaction of dust particles in a plasma is, in one way or another, the screened Coulomb repulsion.

Besides the electrostatic forces associated with the negative charge of dust particles, a number of other physical mechanisms influencing their arrangement are under investigation, as an examination of the literature will show. Most notable here are the anisotropic forces associated with ion fluxes drifting toward the electrodes of the gas-discharge device.[22–26] The huge negative charge of the dust particles interacts with plasma ions and, as a result, in the vicinity of a dust particle regions are formed with elevated or reduced ion density, which polarizes the background plasma. The polarization of a dusty plasma is of great interest, as the resulting forces can induce capture of other dust particles if their mean kinetic energy is not too large. Laboratory experiments conducted by a number of authors have confirmed the existence of anisotropic interaction forces between dust particles.[1,11,27–31] The anisotropy of the interaction is considered at present to be the physical reason for the vertical ordering in the crystal-like and linear structures formed by dust particles. Numerical calculations based on the quasiparticle method[24,25] within the framework of a simple two-dimensional model, in which the plasma is treated as a liquid and the dust particles are treated as delocalized objects characterized by narrow Gaussian distributions, have demonstrated the possibility of there being many ordered equilibrium configurations, with energies that depend on both the charge-to-mass ratio of the dust particles and on the number of levitating particles in the supporting potential field of the discharge. As the number of levitating particles in the trap increases, vertical ordering becomes more likely.

Limitations of the model[24,25] have, to a significant degree, been overcome by a more realistic three-dimensional model,[26] which considers a two-layer crystal of dust particles in a trap with hf discharge plasma. The possibility of instabilities leading to horizontal oscillations of the dust particles when the neutral gas pressure is reduced has been investigated; such oscillations then result in anomalous heating and melting of the crystal structures. The given model considers ion fluxes, and collisions and charge exchange between ions with atoms of the neutral gas are taken into account. As a result of the strong attraction, not only can ions be captured by the potential wells of dust particles, but uncaptured ions can be focused, leading to the emergence of regions of enhanced positive charge density. Under these conditions, mechanical forces result from changes in the momenta and trajectories of ions flying past and impacting dust particles, and from electrical forces associated with the existence of stationary regions with enhanced ion density. These regions are situated near dust particles along the axis of the cathode plasma discharge.[24–26] Taking mechanical forces into account is important for very small particles when these forces exceed the weight of the particles and interaction forces with neutral gas atoms. Electric potential forces are important for both small and large dust particles, and their existence leads to attraction between the particles, which is the physical reason for their vertical ordering. Reference 26 also shows that a region with enhanced ion density can be replaced by an equivalent positive charge at a distance on the order of the Debye radius from the dust particle, where the charge and the distance are functions of the neutral gas pressure and other characteristics of the gas discharge. The introduction of an equivalent charge has made it possible to formulate an analytic model explaining the emergence of oscillations of the dust particles as the pressure is reduced, resulting in anomalous heating and melting of the crystal structure.

The main content of the present paper consists in the formulation of a model in which the effective interaction potential of dust particles with one another takes into account the main physical phenomena discovered in Refs. 22–26, and a Monte Carlo calculation of the ordered structures arising in a dusty plasma. The model possesses certain excellent features. It introduces the position dependence of the charge of the dispersed particles in the glow-dischargestratum, and considers potential interaction forces between the dispersed particles. It is shown that energy is exchanged between the power supply feeding the discharge and the dispersed particles due to a change in their charge, and also due to an interaction with the electric field of the stratum, which counteracts the weight of the dispersed particles and levitates them. Inhomogeneities in the spatial distribution of charge in the background plasma of the stratum are taken into account; these result from interaction of the huge negative charge of a dispersed particle with the ion flux of the discharge. The action of the negative charge of the dispersed particles on the ions flying from the anode to the cathode is the physical reason for their focusing, the emergence of zones with enhanced ion density, and the emergence of a spatial dipole moment in the interaction of the dispersed particles. Emergent anisotropy of the interaction between dispersed particles leads to a certain amount of mutual attraction in the direction of the discharge axis in the cylindrical tube and repulsion in the perpendicular direction.

Monte Carlo calculations explain the experimentally observed formation of threadlike structures of dispersed particles in a stratum, and ‘‘transverse crystallization’’ of these ‘‘threads’’ under the conditions of the experiment.[15–17]

2. ANALYSIS OF PHYSICAL CONDITIONS OF A GAS DISCHARGE, AND CONSTRUCTION OF A GENERIC DUST-PARTICLE INTERACTION POTENTIAL

In general, to find the forces acting on a particle and the effective interaction potential, it is necessary to solve the Poisson equation with allowance for the external source and local charges: electrons, ions, and solid particles whose charge itself depends on the total local potential. Finding the self-consistent electric potential and knowing the charge of the particle, it is possible to determine the force acting on the particle by the remaining particles. However, such a self-consistent task is hard to fulfill. Therefore it is convenient to separate out from the total interaction the force due to the interaction between the dust particles, considering the plasma as a background against which this interaction unfolds. Separating out the interaction forces and construction of an effective potential must be done in such a way as to satisfy the equations of electrostatics following from Maxwell's equations. Here it is necessary to recall that the question of the form of the effective potential has still not been solved.

To solve the given problem, we consider the physical conditions necessary for the emergence of strata in a gas-discharge plasma, and we also discuss the characteristics of a potential trap holding charged dust particles, which is created by the electrostatic fields of the stratum and the walls of the discharge tube. The strata in a low-pressure discharge have been well examined experimentally.[32,33] In a positive discharge column under the conditions of interest, the loss of energy by the electrons in elastic collisions is negligibly small and the electron distribution function is formed under the action of the electric field and the inelastic collisions. This can lead to the emergence of strata, i.e., spatial periodicity of the plasma parameters with characteristic scale on the order of a few centimeters. The electron density, the electron energy distribution, and the electric field are highly inhomogeneous along the length of a stratum. The electric field is relatively large at the head of the stratum (maximum around 10–15 W/cm)—the region occupying 25–30% of the length of the stratum, and small (around 1 W/cm) outside this region. The maximum in the electron density is shifted relative to the maximum of the field strength toward the anode.[33] The electron energy distribution is strongly bimodal,[33] and at the head of the stratum a second maximum dominates, whose center lies near the excitation potential ε_1 of the neutral gas atoms. Due to the high floating potential of the walls of the discharge tube, the potential relief of the stratum has a strongly two-dimensional character: the center-to-wall potential difference at the head of the stratum reaches 20–30 V. Thus, at the head of each stratum there is an electrostatic trap, which in the case of vertical orientation is capable of keeping particles with high enough charge and low enough mass from falling to the bottom, while the strong radial field prevents them from reaching the walls of the discharge tube.

For the purpose of theoretical estimates, according to the viewpoint adopted in the literature,[19–21] it may be assumed that the charge of the dust particles is proportional to the floating potential of the plasma, which in turn is determined by the balance of electron and ion fluxes incident upon the dust particle. The electron fluxes depend on the asymptotic behavior of the electron distribution function at high energies. The specific form of the function $Q(\mathbf{r})$ is determined by the design of the gas-discharge setup, the arrangement of the electrodes, choice of neutral gas, conditions of the gas discharge, etc. The variation of the floating potential, which follows from the observed variation of the potential of the walls along the length of the stratum,[33] leads primarily to a height dependence of the charge on the particles. As a result, the dust particles do not comprise a closed subsystem, and they can exchange charge and energy with the gas-discharge plasma. Below, in constructing the model, we assume that the charge of a dust particle is a prescribed function of its spatial coordinates, $Q(\mathbf{r})$, which requires some modification of traditional electrostatics.

As is well known,[34] it follows from Maxwell's equations that the electromagnetic static and time-independent fields are described by two pairs of independent equations:

$$\operatorname{div}\mathbf{E}=\rho/\varepsilon_0, \quad \operatorname{curl}\mathbf{E}=0 \tag{1}$$

and

$$\operatorname{curl}\mathbf{B}=\mathbf{j}/c^2\varepsilon_0, \quad \operatorname{div}\mathbf{B}=0. \tag{2}$$

The first pair of equations describes the electric fields, and the second, the magnetic fields. Despite the fact that motion of the particles can lead to local variation of the electric and magnetic fields, if the conditions

$$|\triangle\mathbf{E}|/\triangle t \ll j/\varepsilon_0, \quad |\triangle\mathbf{B}|/\triangle t \ll |\triangle\mathbf{E}|/|\triangle r|$$

are satisfied, the mutual influence of the electric and magnetic fields can be neglected. It is just this case that is realized under the conditions of our experiment. In what follows, we will be interested in the electric fields that emerge in a dusty glow-discharge plasma, and the equations of electrostatics (1) will be used as a consistency check on the model.

As was already noted, the spatial variation of the charge $Q(\mathbf{r})$ in the plasma can be described, assuming that

$$Q(\mathbf{r})=C\phi_t(\mathbf{r}), \tag{3}$$

where $C\simeq 4\pi\varepsilon_0 R_p$ is a coefficient that has the order of magnitude of the capacitance of the dust particle (R_p is the radius of the particle) and $\phi_t(\mathbf{r})$ is the position-dependent floating potential of the plasma. Despite the fact that formula (3) is outwardly similar to the familiar electrostatic relation, it is important to note an important difference in content. Whereas in electrostatics the analogous relation establishes a linear relationship between an electric charge and the electric potential created by it, for a plasma with condensed particles this formula establishes a relationship between the electric charge of a particle and the electric potential produced, in the final analysis, by external sources, e.g., electrode systems or other charges in the plasma. One can imagine a capacitor filled with particles whose charge is proportional to the electric potential at a given point. It is easy to show that the capacitance of such a capacitor, defined as the ratio of the total electric charge between the plates to the potential difference of the plates will differ from the capacitance of a capacitor filled with particles whose electric charge does not depend on the electric potential at a given point.

The work required to move dust particles with position-dependent charge is given by the expression

$$W_{12}=-\frac{Q(0)}{4\pi\varepsilon_0}\left(\frac{\tilde{Q}(\mathbf{r}_2)}{|\mathbf{r}_2|}-\frac{\tilde{Q}(\mathbf{r}_1)}{|\mathbf{r}_1|}\right), \tag{4}$$

whose validity is postulated in the present work in line with the viewpoint adopted in the literature[19–21] that the electrical forces describing the interaction between the dust particles are potential forces. Let us consider the physical meaning of expression (4). The quantity W_{12} is equal to the work performed by the charge $\tilde{Q}(\mathbf{r})$ as it is moved from the point $\mathbf{r}_1$ to the point $\mathbf{r}_2$, where this charge has the value $\tilde{Q}(0)$ at the origin. Let the points $\mathbf{r}_1$ and $\mathbf{r}_2=\mathbf{r}_1+\overline{\delta s}$ lie infinitesimally close to each other and be separated by the distance $|\overline{\delta s}|$. Then to within linear terms in $|\overline{\delta s}|$ the work W_{12} can be represented in the form

$$\begin{aligned}\delta W_{12}=(F,\overline{\delta s})&=-\frac{Q(0)}{4\pi\varepsilon_0}\left[\frac{(Q(\mathbf{r}_2)-Q(\mathbf{r}_1))}{|\mathbf{r}_2|}\right.\\&\qquad\left.-\tilde{Q}(\mathbf{r}_1)\left(\frac{1}{|\mathbf{r}_2|}-\frac{1}{|\mathbf{r}_1|}\right)\right]\\&\simeq-\frac{Q(0)\tilde{Q}(\mathbf{r}_1)}{4\pi\varepsilon_0}\left[\left(\overline{\nabla}\left(\frac{1}{|\mathbf{r}_1|}\right),\overline{\delta s}\right)\right.\\&\qquad\left.+\frac{\overline{\nabla}\ln(|\tilde{Q}(\mathbf{r})|,\overline{\delta s})}{|\mathbf{r}_1|}\right],\end{aligned} \tag{5}$$

where $\overline{\nabla}$ is the gradient operator.

As a result, the force due to the charge $Q(0)$ acting on the charge $\tilde{Q}(\mathbf{r}_1)$ in the plasma has the form

$$\begin{aligned}F(\mathbf{r}_1)&=\frac{Q(0)\tilde{Q}(\mathbf{r}_1)}{4\pi\varepsilon_0}\left[\frac{\mathbf{r}_1}{|\mathbf{r}_1|^3}-\frac{\nabla\ln|\tilde{Q}(\mathbf{r}_1)|}{|\mathbf{r}_1|}\right]\\&=\frac{Q(0)\tilde{Q}(\mathbf{r}_1)}{4\pi\varepsilon_0}\left[\frac{\mathbf{r}_1}{|\mathbf{r}_1|^3}-\nabla[1+\ln|\phi_t(\mathbf{r}_1)/\phi_t(0)|]|\mathbf{r}_1|\right].\end{aligned} \tag{6}$$

Accordingly, we define the electric field strength as the ratio of the force acting on a charge to the magnitude of that charge

$$\begin{aligned}E(\mathbf{r}_1)&=\frac{Q(0)}{4\pi\varepsilon_0}\left[\frac{\mathbf{r}_1}{|\mathbf{r}_1|^3}-\frac{\nabla\ln|\tilde{Q}(\mathbf{r}_1)|}{|\mathbf{r}_1|}\right]\\&=\frac{Q(0)}{4\pi\varepsilon_0}\left[\frac{\mathbf{r}_1}{|\mathbf{r}_1|^3}-\frac{\nabla[1+\ln|\phi_t(\mathbf{r}_1)/\phi_t(0)|]}{|\mathbf{r}_1|}\right].\end{aligned} \tag{7}$$

The first term in expression (6) and the first term in expression (7) correspond to the usual Coulomb force calculated for constant charges. The second term in each of those expressions corresponds to the component of the force due to the dependence of the charges of the dust particles on the size and type of dust particle, and are completely determined by the gradient of the logarithmic derivative of the floating potential of the particle in the plasma. For like charges this component of the force acts in the direction of regions with the smallest charge (in magnitude), and for unlike charges it acts in the opposite direction. As a result, the dust particles in a stationary gas-discharge plasma can move into regions with minimal energy of repulsion. Thus, thanks to time-independent charge and energy exchange with the plasma, the open subsystem of dust particles tends to seek configurations with minimal energy of repulsion corresponding to the equilibrium state at a certain effective temperature of the dust particles. The latter can result from the influx of energy from the power supply via the aforementioned processes, and the efflux of energy from the dust particles as a result of collisions with the neutral gas atoms.

Bearing in mind that the charge $Q(\mathbf{r}_0)$ is located at the point $\mathbf{r}_0$, and also that

$$\mathbf{E}(\mathbf{r}_1,\mathbf{r}_0)=-\nabla\varphi(\mathbf{r}_1,\mathbf{r}_0),$$

we find from Eq. (7) that the electric potential of the field created by a charge that depends on particle position is

$$\varphi(\mathbf{r},\mathbf{r}_0)=\frac{1}{4\pi\varepsilon_0}\,\frac{Q(\mathbf{r}_0)[1+\ln|\phi_t(\mathbf{r})/\phi_t(\mathbf{r}_0)|]}{|\mathbf{r}-\mathbf{r}_0|}+\text{const.} \tag{8}$$

If the magnitudes of the charges do not depend on the spatial variables, then expression (8) automatically goes over to the familiar expression of electrostatics. Recall that the work and the difference in potential energies in this case are determined by relation (4), and that the transformation to relation (4) from expression (8) through the electric field strength and the interaction potential must be performed in reverse order.

We now check whether Maxwell's equations (1) still hold in the proposed model. Obviously, the second equation $\operatorname{curl}\mathbf{E}=0$ is satisfied. It is not hard to show that the divergence of the electric field $\mathbf{E}$ satisfies the equation

$$\begin{aligned}\operatorname{div}\mathbf{E}&=\lim_{V\to 0}\frac{1}{V}\oint_S\mathbf{E}\,d\boldsymbol{\sigma}=\lim_{V\to 0}\frac{Q(0)}{4\pi\varepsilon_0}\,\frac{1}{V}\left[4\pi\int_V\delta(\mathbf{r})d\mathbf{r}_1\right.\\&\quad\left.+\oint_S\frac{(\nabla\ln|Q(\mathbf{r})|,\mathbf{n})}{|\mathbf{r}_1|}|\mathbf{r}_1|^2\sin\theta d\theta d\varphi\right],\end{aligned}$$

where $\delta(\mathbf{r}_1)$ is the Dirac delta function, $|\mathbf{r}_1|^2\sin\theta\,d\theta\,d\varphi$ is the surface element (of the surface S), and $\mathbf{n}$ is the normal to the surface S. To reduce this to differential form, we let the volume V containing the origin tend to zero, assuming for the estimates that the volume is a sphere of radius κ. Then for small κ we have

$$\operatorname{div}\mathbf{E}=\frac{Q(0)}{4\pi\varepsilon_0}[4\pi+O\{\kappa\}]\frac{1}{\Delta v},$$

where $O\{\kappa\}$ is a small quantity of order κ and Δv is an infinitesimal volume. In the limit $\kappa\to 0$ we obtain

$$\operatorname{div}\mathbf{E}=\frac{\rho}{\varepsilon_0},$$

where $\rho=Q(0)/\triangle v$ is the charge density. In the above treatment the origin can be chosen at any point in space; consequently, the first of the equations of electrostatics (1) is fulfilled.

Thus, the treatment based on the postulate (4) shows that in a model taking into account the dependence of the charges of the dust particles on the spatial variables, the generic pairwise potential energies of the interaction of the particles in the plasma can be described by a function of the form

$$V(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{4\pi\varepsilon_0}\frac{Q(\mathbf{r}_1)Q(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}. \tag{9}$$

The potential energy of the interaction of two identically oriented dipoles with moment arm d, in which the negative charge is greater than the positive charge, can be written in the form

$$V_d(\mathbf{r}_1,\mathbf{r}_2)=\int_V\int_V\frac{\rho_d(\mathbf{r}_1,\tilde{\mathbf{r}}_1)\rho_d(\mathbf{r}_2,\tilde{\mathbf{r}}_2)d\tilde{\mathbf{r}}_1 d\tilde{\mathbf{r}}_2}{|\tilde{\mathbf{r}}_2-\tilde{\mathbf{r}}_1|},$$

where

$$\rho(\mathbf{r}_1,\tilde{\mathbf{r}}_1)=\delta(\tilde{\mathbf{r}}_1-\mathbf{r}_1)Q_-(\tilde{\mathbf{r}}_1)+\delta(\tilde{\mathbf{r}}_1-(\mathbf{r}_1+d\mathbf{e}_+))Q_+(\tilde{\mathbf{r}}_1),$$

$$\rho(\mathbf{r}_2,\tilde{\mathbf{r}}_2)=\delta(\tilde{\mathbf{r}}_2-\mathbf{r}_2)Q_-(\tilde{\mathbf{r}}_2)+\delta(\tilde{\mathbf{r}}_2-(\mathbf{r}_2+d\mathbf{e}_+))Q_+(\tilde{\mathbf{r}}_2)$$

are the spatial charge densities, $\delta(\mathbf{r})$ is the Dirac delta function, Q_- is the negative charge, Q_+ is the positive charge, and $\mathbf{e}_+$ is the unit radius vector extending from the negative charge of the dipole to the positive charge.

If the moment arm d of the dipole is less than the distance between dipoles, then invoking the smallness of the ratio $d/|\mathbf{r}_2-\mathbf{r}_1|$, we can perform a multipole expansion. As a result, the potential $V_d(\mathbf{r}_1,\mathbf{r}_2)$ can be written in the form

$$V_d(\mathbf{r}_1,\mathbf{r}_2)=\frac{Q_s^1Q_s^2}{|\mathbf{r}_{12}|}+\frac{P^1Q_s^2-P^2Q_s^1}{|\mathbf{r}_{12}|^2}+\frac{\sigma^1Q_s^2+\sigma^2Q_s^1-3P^1P^2+Q_+^1Q_+^2d^2}{|\mathbf{r}_{12}|^3}+\dots, \tag{10}$$

where $\mathbf{r}_{12}=\mathbf{r}_2-\mathbf{r}_1$ is the radius vector extending from the negative charge of the first dipole to the negative charge of the second dipole,

$$Q_s^l=Q_s(\mathbf{r}_l)=\int_V\rho(\mathbf{r}_l,\tilde{\mathbf{r}}_1)d\tilde{\mathbf{r}}_1=Q_-(\mathbf{r}_l)+Q_+(\mathbf{r}_l+d\mathbf{e}_+)$$

is the excess negative charge of the dipole,

$$P^l=P(\mathbf{r}_l)=\int_V|\tilde{\mathbf{r}}_1|\rho(\mathbf{r}_l,\tilde{\mathbf{r}}_1)\cos\theta\, d\tilde{\mathbf{r}}_1=dQ_+(\mathbf{r}_l+d\mathbf{e}_+)\cos\theta_d$$

is the dipole moment (θ is the angle between the vectors $\mathbf{r}_{12}$ and $\mathbf{e}_+$), and

$$\sigma^l=\sigma(\mathbf{r}_l)=\int_V|\tilde{\mathbf{r}}_1|^2((3\cos^2\theta-1)/2)\rho(\mathbf{r}_l,\tilde{\mathbf{r}}_1)d\tilde{\mathbf{r}}_1=d^2Q_+(\mathbf{r}_l+d\mathbf{e}_+)(3\cos^2\theta_d-1)/2$$

is the quadrupole moment.

3. EFFECTIVE INTERACTION POTENTIAL BETWEEN DUST PARTICLES

At present several physical mechanisms are being discussed in the literature that influence both the balance of gravitational and electrical forces of the levitating dust particles and the interaction between them. From our point of view, along with the partial screening of the charge of the dust particles by the electrons and ions of the plasma that interact strongly with them, the emergence of regions with enhanced free ion density as a consequence of the focusing action of the large negative charge of the dust particles on the ion current of the plasma discharge is very important.

As has already been mentioned, the existence of zones with enhanced ion density along the path of the ion current behind the dust particles was proved in Ref. 26, where the Poisson equation was also solved and charge exchange processes and collisions of ions with atoms were taken into account by a Monte Carlo calculation. In this same reference, characteristics of the positive point charge equivalent to the ion clouds were calculated and the distance from it to the dust particle was found. In Ref. 26 it was also shown that the magnitude of the equivalent positive charge can reach one-third of the charge on the dust particle, and its distance d from the particle is at most of the order of the Debye radius r_D. Usually, for dust particles in a gas-discharge plasma the mean distance between the dust particles $\langle r\rangle$ is greater than the Debye radius, and consequently according to Ref. 26, it is also greater than the distance d.

Plasma screening of the dust particles and the effect on them of the electron and ion currents were taken into account, for example, in Ref. 9, where it was shown that the pairwise interaction potential $\tilde{\varphi}(\mathbf{r}_1,\mathbf{r}_2)$ at distances less than the mean distance between the particles is screened by the Debye exponential factor, and at distances of the order of a few Debye radii it converges to its asymptotic limit, which is proportional to the inverse square of the distance between the particles.

Thus, introducing the effective positive point charge Q_+ to account for regions with enhanced ion density and returning to the question of the effective pairwise interaction potential between dust particles $V(\mathbf{r}_{ij})$ located a distance $\mathbf{r}_{ij}$ from each other, we can adopt the following approximation, based on the results of numerical calculations:[9,24–26]

$$\begin{aligned}U(\mathbf{r}_i,\mathbf{r}_j)=&\frac{1}{4\pi\varepsilon_0}\frac{Q_s(\mathbf{r}_i)Q_s(\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|}\exp\left(-\frac{|\mathbf{r}_i-\mathbf{r}_j|}{r_D}\right)\\&+\frac{\tilde{A}}{|\mathbf{r}_i-\mathbf{r}_j|^2}(1-D(|\mathbf{r}_i-\mathbf{r}_j|))\\&+\frac{P^iQ_s^j-P^jQ_s^i}{|\mathbf{r}_i-\mathbf{r}_j|^2}(1-\tilde{D}(|\mathbf{r}_i-\mathbf{r}_j|))\\&+\frac{\sigma^iQ_s^j+\sigma^jQ_s^i-3P^iP^j+Q_+^1Q_+^2d^2}{|\mathbf{r}_i-\mathbf{r}_j|^3}\\&\times(1-\tilde{D}(|\mathbf{r}_i-\mathbf{r}_j|)),\end{aligned} \tag{11}$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the radius vectors of the ith and jth particle ($|\mathbf{r}_{ij}|=|\mathbf{r}_i-\mathbf{r}_j|$), $\widetilde{D}(|\mathbf{r}_{ij}|)$ and $D(|\mathbf{r}_{ij}|)$ are matching functions, which range from 1 to 2 under our conditions at distances of 1 and 2 Debye radii, respectively, and $\widetilde{A}$ is the matching constant.

In the calculations it is convenient to choose as the unit of length to be the Debye radius, which in the given model can be taken to be independent of the spatial variables. We then have

$$\beta U(\mathbf{r}_i,\mathbf{r}_j)=\frac{\beta e^2}{4\pi\varepsilon_0}\frac{Z_s(\mathbf{r}_i)Z_s(\mathbf{r}_j)}{r_D}\left\{\frac{\exp(-|\mathbf{r}_{ij}|)}{|\mathbf{r}_{ij}|}+\frac{A}{|\mathbf{r}_{ij}|^2}(1-D(|\mathbf{r}_{ij}|))+\frac{d\cos(\theta)[Z_+(\mathbf{r}_i)/Z_s(\mathbf{r}_i)-Z_+(\mathbf{r}_j)/Z_s(\mathbf{r}_j)]}{|\mathbf{r}_{ij}|^2}(1-\widetilde{D}(|\mathbf{r}_{ij}|))\right.$$
$$\left.+\frac{d^2[Z_+(\mathbf{r}_i)Z_+(\mathbf{r}_j)/Z_s(\mathbf{r}_i)Z_s(\mathbf{r}_j)][(Z_-(\mathbf{r}_i)/Z_+(\mathbf{r}_i)+Z_-(\mathbf{r}_j)/Z_+(\mathbf{r}_j))(3\cos^2(\theta)-1)/2]}{|\mathbf{r}_{ij}|^3}(1-\widetilde{D}(|\mathbf{r}_{ij}|))\right\}. \qquad (12)$$

Here $\beta=1/kT_p$, kT_p is the energy temperature of the dust particles, $Z(\mathbf{r}_i)$ is the charge on a dust particle expressed in units of the electron charge, $Q(\mathbf{r}_i)=Z(\mathbf{r}_i)e$, and distances are reduced to dimensionless form by scaling to the Debye radius r_D, which for simple estimates can be chosen in the form $r_D^2=kT_g/4\pi e^2(n_e+n_i)$ (T_g is the gas temperature).

We simplify the given potential by taking d, $\mathbf{e}_+$, and the ratios Z_+^i/Z_s^i and Z_-^i/Z_+^i to be equal for all the dust particles ($i=1,\dots,N$). Then the second term of the order of d in the effective potential (12), which describes an interaction of charge–dipole type, vanishes. We finally obtain

$$\beta U(\mathbf{r}_i,\mathbf{r}_j)=\Gamma(\mathbf{r}_i,\mathbf{r}_j)\left\{\frac{\exp(-|\mathbf{r}_{ij}|)}{|\mathbf{r}_{ij}|}+\frac{\widetilde{A}}{|\mathbf{r}_{ij}|^2}(1-D(|\mathbf{r}_{ij}|))\right.$$
$$\left.+\Xi\frac{[|\mathbf{r}_{ij}|^2-3(\mathbf{r}_{ij}^z)^2]}{|\mathbf{r}_{ij}|^5}(1-\widetilde{D}(|\mathbf{r}_{ij}|))\right\}. \qquad (13)$$

Here $\Xi=d^2|\chi|/(1+\chi^2)$ ($\chi=\langle Z_-(\mathbf{r}_i)/Z_+(\mathbf{r}_i)\rangle$ is the mean ratio of charges in the pair),

$$\Gamma(\mathbf{r}_i,\mathbf{r}_j)=\frac{\beta e^2}{4\pi\varepsilon_0}\frac{Z_s(\mathbf{r}_i)Z_s(\overline{r_j})}{r_D}\approx\Gamma=\frac{Z_p^2e^2}{4\pi\varepsilon_0kT_pr_D}$$

is the interaction parameter, which can also be represented in the form

$$\Gamma=\frac{\gamma_p\langle r\rangle}{r_D},$$

where $\gamma_p=Z_p^2e^2/4\pi\varepsilon_0kT_p\langle r\rangle$ is the nonideality parameter, Z_p is the mean particle charge, and $\langle r\rangle=(4\pi n_p/3)^{-1/3}$ is the mean distance between macroparticles. It follows from the results obtained in Ref. 26 that the quantity χ under gas-discharge conditions satisfies the inequality $\chi<-3$.

Note again that the potential (13) allows for a number of physical factors that influence the interaction of dust particles in the plasma. First, it allows for the spatial dependence of particle charge on the floating potential. Second, the first term in brackets, describing the spherically symmetric part of the interaction, takes account of the screening of dust particles by plasma electrons and ions, which interact strongly with them. The third term describes the anisotropic part of the interaction of the dust particles resulting from focusing of the ion current by highly charged dust particles. It follows from the form of the effective potential that if the dust particles are arranged vertically (stacked) one above another (along the discharge (z) axis), i.e., $r\approx r^z$), then the third term in the potential becomes negative, which corresponds to the emergence of attraction between the particles. If the particles are arranged in a horizontal plane ($r\gg r^z$), then this term is positive, which implies that the particles repel. The matching function $\widetilde{D}$ also allows for destruction of ion clouds when dust particles approach to within distances on the order of the Debye radius.

Equation (13) for the interaction potential of two dust particles in the plasma can be verified experimentally. Measurements of the forces between two dust particles must be performed under the same conditions as the dust crystal preparation experiments. In such an experiment it will be possible to determine the parameters entering into Eq. (13): χ, the ratio of the effective screening charge of the ions (if it exists) to the particle charge, and d, the dipole moment. Such experiments themselves on the interaction of two dust particles, apart from their stated purpose to measure the parameters used to model the dust crystal, are fundamentally important in their own right, since they enable one to choose a faithful model from among existing candidates for the forces that act on a dust particle in a plasma.

To conclude this section we note that the mechanical forces associated with variation of the momenta and trajectories of the ions flying into the field of a dust particle can be taken into account with the help of an inhomogeneous correction to the gravitational component of the forces on the dust particle. This correction should be proportional to the spatial distribution of the ion current in the discharge and can be prescribed *a priori* or found by a self-consistent approach.

4. NUMERICAL RESULTS

Numerical modeling of plasma with a dispersed phase was performed by the standard Monte Carlo method,[35] which considers a finite number of particles N distributed over a cell of extent L. In the present calculations, considering the speed of presently available computers and a reasonable calculation time (one point per day), we decided to limit the number of particles to $N=3000$. The cell size L, which it is

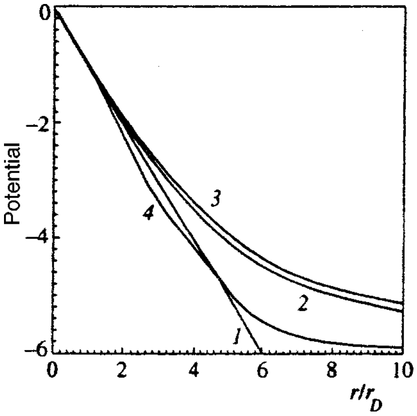

FIG. 1. Logarithm of the ratio of the interparticle interaction potential to the Coulomb potential: *1* — Debye potential, *2* — Debye potential going over to the asymptotic dependence r^{-2}. *3, 4* — upper and lower boundaries of the potential (13) for $\Xi = 0.1$ and $|\chi| \approx 12$.

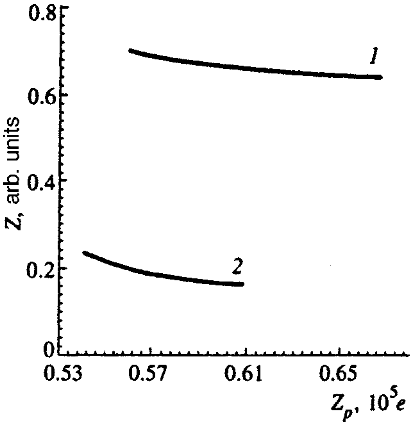

FIG. 2. Probability distribution of the charge on the dust particles, Z_p, in units of $10^5 e$ in the upper and lower strata for $\Xi = 0.05$ and $|\chi| \approx 22$.

also convenient to express in units of the Debye radius, was chosen such that two strata could be situated in it. This length turns out to be $L = 700 r_D$.

The parameters of the electric fields of the stratum were chosen to be close to the experimentally measured values.[33] The walls of the cylindrical tube, of radius $R_t = 15$ mm, exerted a force on the dust particles, whose potential can be expressed in the form

$$\phi_w = \phi_t (r/R_t)^{3/2}, \quad r^2 = x^2 + y^2,$$

$$\phi_t = 10 + 20/(1 + ((z - z_0)/d_w)^2)$$

(potential in volts at the walls of the tube), $z_0 = 0.8$ mm, and $d_w = 4$ mm. The interaction potential of the dust particles was given by Eq. (13) with Debye radius $r_D = 0.1$ mm. Along the z axis the force of gravity and the supporting electric field with potential $\phi_z = 14/(1 + (z/d_z)^2)$, where $d_z = 1$ mm, acted on the dust particles. In the course of the calculations the quantities Γ and L were fixed while the parameter Ξ was varied from 0.05 to 0.4, which corresponds for $d \approx 1$ to the range of values $|\chi| = |Q_- / Q_+|$ from 22 to 4. The physical meaning of an increase in the absolute value of the parameter χ depends on how the neutral gas pressure in the discharge plasma is varied. For example, increasing neutral gas pressure reduces the mean free path of the ions and hinders their focusing beyond the dust particle. According to the calculations of Ref. 26, the variation of the absolute value of the parameter χ over the given range corresponds to gas-discharge conditions and variation of the pressure over the range 20–200 Pa.

The results of Monte Carlo calculations presented below were obtained with a model in which the interaction of dust particles is described by the potential (13). Figure 1 compares the potential (13) with the Coulomb and Debye potentials, and with the Debye potential trailing off at large distances to its asymptotic limit r^{-2}. For convenience, the ratio of the given potentials to the Coulomb potential is plotted on a logarithmic scale in the figure. Thus, the ordinate of the Coulomb potential in Fig. 1 is identically equal to zero, the Debye potential is represented by the linear dependence *1*, the Debye potential with asymptotic limit r^{-2} corresponds to curve *2*, and curves *3* and *4* are the potential (13) for particles located respectively in a horizontal plane and vertically one above the other. All of the represented repulsion potentials are much softer than the Coulomb potential.

Let us consider the results of the Monte Carlo calculations. For the given axially symmetric problem, Fig. 2 depicts the probability distribution of the dust particles over charge for two strata located one above the other ($\Xi = 0.05$ and $|\chi| \approx 22$). The height of the particle in the discharge tube is plotted along the vertical axis, and the magnitude of the

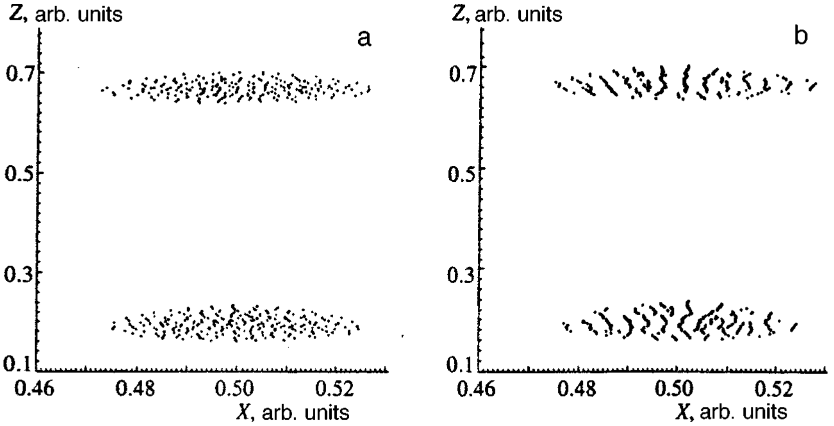

FIG. 3. Thin vertical cross sections (of finite thickness) in the upper and lower strata for $\Xi = 0.05$ and $|\chi| \approx 22$ (a) and $\Xi = 0.2$ and $|\chi| \approx 7$ (b).

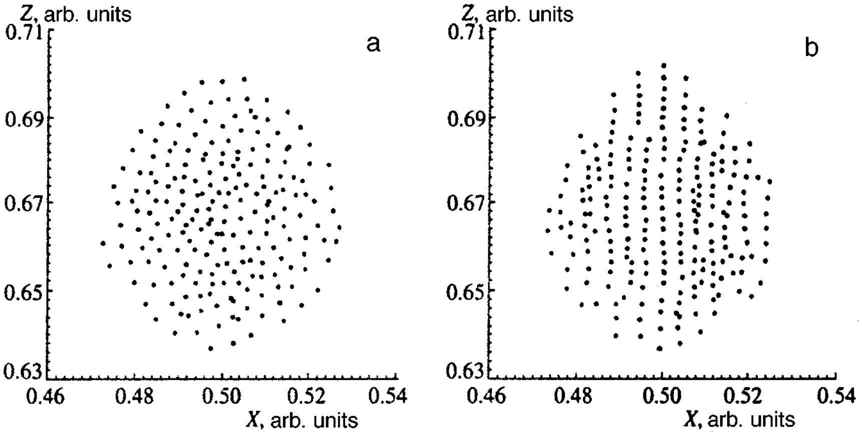

FIG. 4. Thin vertical cross sections (of finite thickness) of dust structures in the strata with $\Xi=0.05$ and $|\chi|\approx 22$ (a) and $\Xi=0.2$ and $|\chi|\approx 7$ (b).

charge Z_p is plotted along the horizontal axis in units of $10^5 e$. The particle distribution in the upper stratum is more extended. However, the spread in charge does not exceed 17% since the dust particles are mainly located in that part of the stratum where the magnitude of the floating potential is less and varies quite smoothly. Varying the pressure for $\Xi=0.2$ and $|\chi|\approx 7$ has hardly any effect on the distribution. These data confirm our earlier conclusion about the preferred location of particles in that part of the stratum where the potential energy of repulsion between particles is lower. It appears that the formation of very long, isolated threadlike structures is also possible. This can be energetically more favorable, not only thanks to the additional attraction in the vertical direction, but also thanks to the presence of the non-Coulomb correction to the Coulomb forces discussed above, which is directed opposite the gradient of the dependence $Q(\mathbf{r})$ (in the given case, from the cathode to the anode). Questions of the possible existence of such structures and conditions for their emergence require further study.

Figure 3 shows vertical axial cross sections of finite thickness, of two dust clouds filling out strata located one above the other. The vertical distance between strata is roughly $50\cdot 7r_D\approx 3.5\,\text{cm}$, which approximately corresponds to the conditions of our experiment. In the given figure the structure of the dust clouds in the strata is distorted since the scale of the vertical axis is 10 times larger than the horizontal. A properly scaled picture of the vertical structure of the dust clouds is shown in Fig. 4 (upper stratum), where the scales of the vertical and horizontal axes are equal. No appreciable difference in the structure of the dust particles in the upper and lower strata emerges in the given model. Generally speaking, it must be borne in mind that varying the gas pressure alters the characteristics of the strata, since the electric field in a gas discharge depends on the product of the pressure and the radius of the discharge tube.

Figures 4 and 5 show vertical and horizontal axial thin cross sections, of finite thickness, of typical structures of dust particles in a stratum for $\Xi=0.05$ ($|\chi|\approx 22$) and $\Xi=0.2$ ($|\chi|\approx 7$), respectively. Distances are given in units of the Monte Carlo cell L. Figure 4 plots distance along the diameter of the discharge tube along the horizontal axis, with the point 0.5 corresponding to the charge axis; each division (0.01) corresponds to roughly $7r_D$ or $700\,\mu\text{m}$. Distance along the z axis is plotted along the vertical axis. From an analysis of the figures we may conclude that for high neutral gas pressures the dust particles for the most part are arranged randomly, while at lower pressures the dust particles form a hexagonal crystal lattice of threadlike structures of arbitrary length in the transverse direction. The threadlike structures can be vertically displaced relative to one another along the discharge axis. The given configuration of threadlike structures is reminiscent of structures resulting from the emergence of an ordered phase in liquid crystals.

The two-point correlation functions corresponding to the

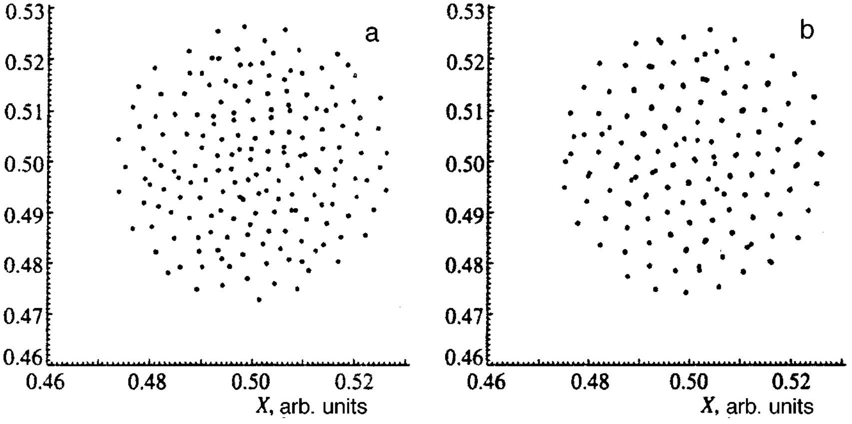

FIG. 5. Thin horizontal cross sections (of finite thickness) of dust structures in the strata with $\Xi=0.05$ and $|\chi|\approx 22$ (a) and $\Xi=0.2$ and $|\chi|\approx 7$ (b).

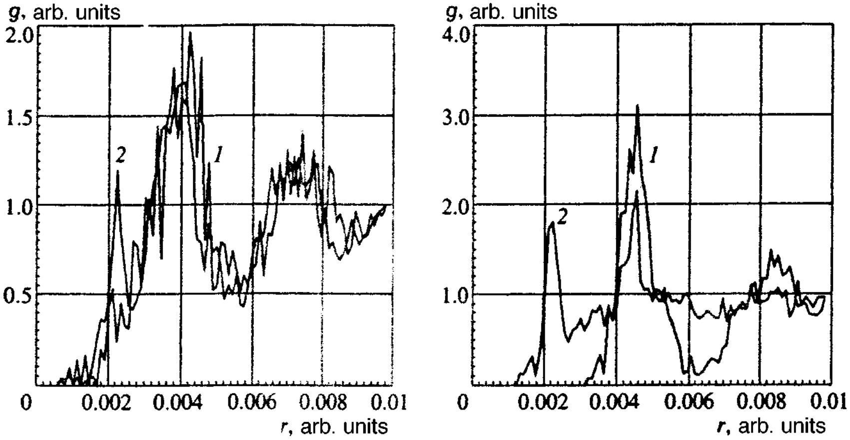

FIG. 6. Two-point correlation functions of the horizontal (curves *1*) and vertical (curves *2*) cross sections of the dust structures depicted in Figs. 5 and 4, respectively.

given cross sections for $\Xi=0.05$ and $\Xi=0.2$ are plotted in Figs. 6a and 6b. The correlation functions were calculated for the interior particles of the depicted cross sections. For $\Xi=0.05$ the two-point correlation functions of the vertical and horizontal cross sections are essentially identical. The only significant difference, associated with attraction of the dust particles in the vertical direction, leads to an additional peak at small distances. Decreasing the pressure ($\Xi=0.2$) leads to substantial differences in the pairwise correlation functions of the vertical and horizontal cross sections [Fig. 6(b)]. The two-point correlation functions of the vertical cross section have pronounced peaks corresponding to the emergence of ordered structures. The peaks increase at small distances as a consequence of attraction between particles in the vertical direction. The correlation function of the horizontal cross section has a form characteristic of crystal-like structures.

As noted above, the physical reasons for the emergence of these structures when the neutral gas pressure is reduced are an increase in the mean free path of the ions and, as a consequence of the focusing action of the negatively charged dust particles on the ion current, the formation of clouds with enhanced ion density between them. As a result, in the direction of the discharge axis, against the background of the spherically symmetric part of the effective potential, an additional attraction emerges between the dust particles, while in the horizontal plane an additional positive repulsion emerges. This phenomenon is noticeable in Figs. 4, 5, and 6. In the displayed cross sections it is possible to discern an increase in the horizontal mean distance between the dust particles as the pressure is reduced. The particles in rectilinear threadlike structures become essentially completely ordered in the vertical direction.

For a gas discharge, typical configurations of dust particles in a stratum are shown in Figs. 7 and 8 ($p=0.5$ and 0.2 Torr, $I=0.4$ and 0.7 mA, dimensions of the frame 6×7 mm).[17] Comparison of Figs. 4, 5, and 6 and Figs. 7 and 8 reveals good agreement between the numerical and experimental results. In the horizontal cross section, the emergence of hexagonal ordered structures is observed. In the vertical cross sections the emergence of threadlike structures can be distinctly made out. The mean distance between particles in both the horizontal and vertical cross sections is roughly 280 μm.

Note that in Ref. 17 the formation of extended threadlike structures several centimeters in length was also observed upon the coalescence of several strata, which, as noted above, is possibly related to the non-Coulomb force directed opposite the gradient of the dependence $Q(\mathbf{r})$ (upwards, toward the anode). This force should contribute to the balance between the gravitational force and the supporting electrical force that causes the dust particles to levitate.

The resultant threadlike structures[17] are distinctly different from the formations observed in erosion discharges,

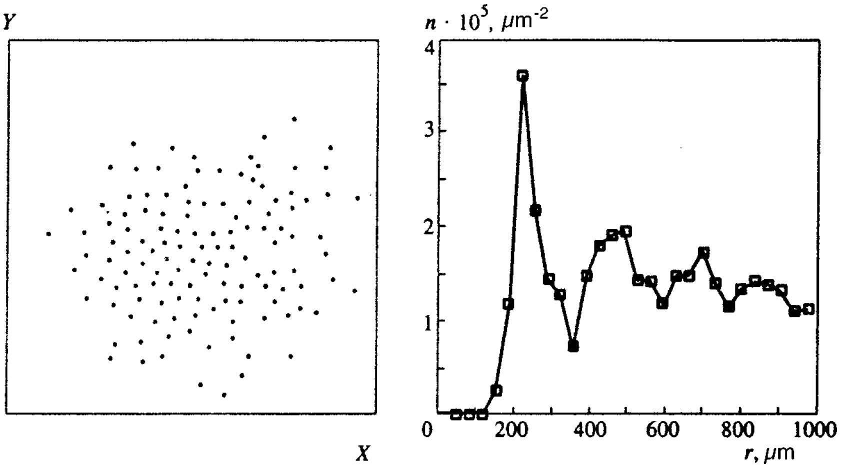

FIG. 7. Thin horizontal cross section (of finite thickness) of the dust structure in the stratum and the corresponding two-point correlation function (Ref. 17).

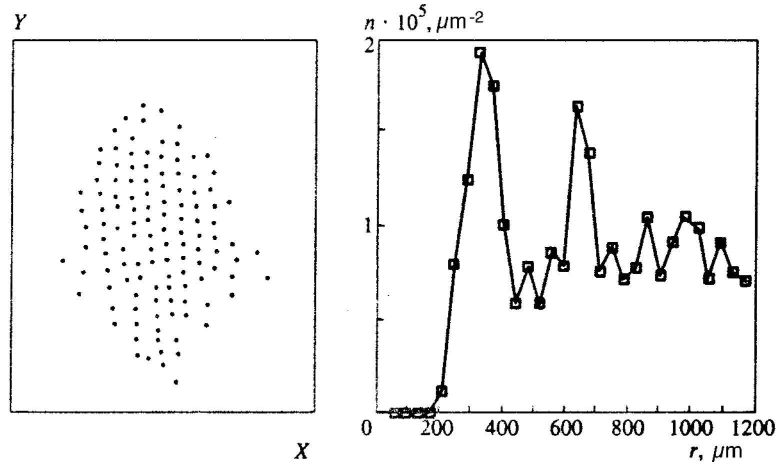

FIG. 8. Thin vertical cross section (of finite thickness) of the dust structure in the stratum and the corresponding two-point correlation function (Ref. 17).

where rapid coagulation of submicron aerosols leads to the emergence of threadlike three-dimensional structures.[35]

5. CONCLUSION

A model has been formulated in which the pairwise interaction potential of dust particles takes into account the spatial dependence of the particle charges on the floating potential, and screening of the dust particles by plasma electrons and ions that interact strongly with them (and themselves undergo vigorous recombination). The pairwise interaction potential at distances less than the mean interparticle distance is screened by the Debye exponential factor, and at distances on the order of a few Debye radii the interaction potential trails off to its asymptotic limit, proportional to the inverse square of the interparticle distance. We have shown that the spatial dependence of the particle charges leads to the emergence of a non-Coulomb component of the interaction forces between particles, which is directed opposite the gradient of the logarithmic derivative of the floating potential of the plasma. As a result, the open subsystem of dust particles, thanks to steady-state charge and energy exchange with the plasma, tends to seek configurations with minimal energy of repulsion, corresponding to the equilibrium state. The model also takes into account focusing of the ion beam by large-radius dust particles, leading to polarization of the background plasma and the emergence of anisotropy in particle interactions.

We have performed a Monte Carlo study of ordered structures that emerge in a dusty plasma. These calculations explain the formation of threadlike structures of dispersed particles in a stratum, and the ''transverse crystallization'' of these ''threads'' under the conditions of our experiment.[15–17] We obtained the probability distribution of charges on the dust particles levitating in a stratum. The results of our calculations agree with the experimental data.[15–17]

We are deeply grateful to Yu. E. Lozovik, A. G. Khrapak, V. I. Molotkov, V. M. Torchinskiĭ, and V. V. Zhakovskiĭ for helpful discussions and valuable remarks. This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grants Nos. 97-02-16572, 97-02-17565, 97-1-00931, and 96-15-96462) and INTAS (Grant No. 95-1335).

[1] J. H. Chu and I. Lin, Phys. Rev. Lett. **72**, 4009 (1994).
[2] H. Thomas, G. E. Morfill, V. Demmel *et al.*, Phys. Rev. Lett. **73**, 652 (1994).
[3] A. Melzer, T. Trottenberg, and A. Piel, Phys. Lett. A **191**, 301 (1994).
[4] J. B. Pieper and J. Goree, Phys. Rev. Lett. **77**, 3137 (1996).
[5] A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).
[6] M. D. Kilgore, J. E. Daugherty, R. K. Porteous, and D. B. Graves, J. Appl. Phys. **73**, 7195 (1993).
[7] J. Perrin, P. Molinas Mata, and P. Belenguer, J. Phys. D **27**, 2499 (1994).
[8] J. P. Boeuf, Phys. Rev. A **46**, 7910 (1992).
[9] J. E. Daugherty, M. D. Kilgore, R. K. Porteous, and D. B. Graves, J. Appl. Phys. **72**, 3934 (1992).
[10] S. J. Choi and M. J. Kushner, J. Appl. Phys. **75**, 3351 (1994).
[11] Y. Hayashi and K. Tachibana, Jpn. J. Appl. Phys., Part 2 **33**, L804 (1994).
[12] V. E. Fortov, A. P. Nefedov, O. F. Petrov, A. A. Samarian, A. V. Chernyschev, and A. M. Lipaev, JETP Lett. **63**, 187 (1996).
[13] V. E. Fortov, A. P. Nefedov, O. F. Petrov, A. A. Samarian, and A. V. Chernyschev, Phys. Lett. A **219**, 89 (1996).
[14] V. E. Fortov, V. S. Filinov, A. P. Nefedov, O. F. Petrov, A. A. Samaryan, A. M. Lipaev, Zh. Éksp. Teor. Fiz. **111**, 889 (1997) [JETP **84**, 489 (1997)].
[15] V. E. Fortov, A. P. Nefedov, V. M. Torchinskiĭ, V. I. Molotkov, A. G. Khrapak, O. F. Petrov, K. F. Volykhin, JETP Lett. **64**, 92 (1996).
[16] V. E. Fortov, A. P. Nefedov, V. M. Torchinsky, V. I. Molotkov, O. F. Petrov, A. A. Samarian, A. M. Lipaev, and A. G. Khrapak, Phys. Lett. A **229**, 317 (1997).
[17] A. M. Lipaev, V. I. Molotkov, A. P. Nefedov, O. F. Petrov, V. M. Torchinsky, V. E. Fortov, A. G. Khrapak, and S. A. Khrapak, Zh. Éksp. Teor. **112**, 2030 (1997) [JETP **85**, 1110 (1997).
[18] V. V. Zhakovskiĭ, V. I. Molotkov, A. P. Nefedov, V. M. Torchinskiĭ, A. G. Khrapak, and V. E. Fortov, JETP Lett. **66**, 419 (1997).
[19] H. M. Thomas and G. E. Morfill, Nature (London) **379**, 806 (1996).
[20] V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997).
[21] A. P. Nefedov, O. F. Petrov, and V. E. Fortov, Usp. Fiz. Nauk **167**, 1215 (1997).
[22] S. V. Vladimirov and M. Nambu, Phys. Rev. E **52**, 2172 (1995).
[23] M. Nambu, S. V. Vladimirov, and P. K. Shukla, Phys. Lett. A **203**, 40 (1995).
[24] F. Melandso and J. Goree, Phys. Rev. E **52**, 5312 (1995).
[25] F. Melandso and J. Goree, J. Vac. Sci. Technol. A **14**, 511 (1996).
[26] V. A. Schweigert, I. V. Schweigert, A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **54**, 4155 (1996).
[27] G. Praburam and J. Goree, Astrophys. J. **441**, 830 (1997).
[28] H. M. Thomas, G. E. Morfill, V. Demmel, J. Goree, B. Feuerbacher, and D. Mohlmann, Phys. Rev. Lett. **72**, 4009 (1994).
[29] J. B. Pieper, J. Goreeand, and R. A. Quinn, J. Vac. Sci. Technol. A **14**, 519 (1996).
[30] G. E. Morfill and H. M. Thomas, J. Vac. Sci. Technol. A **14**, 490 (1996).
[31] Y. Hayashi and K. Tachibana, J. Vac. Sci. Technol. A **14**, 506 (1996).
[32] Yu. P. Raĭzer, *Physics of Gas Discharge* [in Russian], Nauka, Moscow (1987).

[33] Yu. B. Golubovskiĭ and S. U. Nisimov, Zh. Tekh. Fiz. **66**, 20 (1996) [Tech. Phys. **41**, 645 (1996)].
[34] V. M. Zamalin, G. É. Norman, V. S. Filinov, *The Monte Carlo Method in Statistical Thermodynamics* [in Russian], Nauka, Moscow (1977).
[35] V. Ya. Aleksandrov, I. N. Borodin, E. V. Kichenko *et al.*, Zh. Tekh. Fiz. **52**, 818 (1982) [Sov. Phys. Tech. Phys. **27**, 527 (1982)].

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Dynamics of formation of ordered structures in a thermal plasma with macroparticles

A. P. Nefedov, O. F. Petrov,*) Ya. K. Khodataev, and S. A. Khrapak*)

Research Center for the Thermal Physics of Pulsed Phenomena, Russian Academy of Sciences, 127412 Moscow, Russia

(Submitted 6 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 837–845 (March 1999)

The molecular dynamics method is used to model the formation of ordered structures of charged macroparticles in a thermal plasma at atmospheric pressure. The results of the numerical calculations are compared with the experimental data. It is shown that the ordered structure of macroparticles detected experimentally is far from steady state, since the existence time of the plasma in the experiment is less than the characteristic time of formation of the structure. © *1999 American Institute of Physics.* [S1063-7761(99)00503-X]

1. INTRODUCTION

One of the main reasons for the significant interest manifested recently in the physics of a dusty plasma (a plasma with macroparticles) is the recently discovered formation of ordered structures of macroparticles in such a plasma.[1–9] It is customary to explain this ordering by the existence of a strong electrical interaction between the charged macroparticles. Most experiments on the formation of ordered structures of macroparticles have been performed in a high-frequency (hf) plasma discharge.[1–4] Ordered structures have also been obtained in the laminar jet of an atmospheric-pressure thermal plasma at a temperature around 1700 K,[5,6] in a stratified constant-current glow-discharge,[8,9] and in a double electric layer.[7]

The present work examines the dynamics of formation of ordered structures of macroparticles in experiments with thermal plasma. The experimental setup used in these experiments was described in detail in Ref. 6. It includes a plasma generator (two-flame propane–air torch of Mekker type) creating a laminar plasma jet with uniform distribution of its parameters (temperature, electron and ion densities) in the region of the inner flame, a macroparticle feed system, and an extensive system of diagnostics allowing one to measure the electron n_e and ion n_i densities, plasma temperature T_g, and macroparticle radius R_d and density n_d. In addition, a photon correlation method was devised to allow detailed study of the dynamical characteristics of the system of macroparticles. To measure the spatial arrangement of the macroparticles, we used a time-of-flight counter, which allowed us to obtain the two-point correlation function.

The conditions of the described experiment differ significantly from those of experiments in a gas discharge, which gives reason to hope for new insights into the self-organization of macroparticles in a plasma. First, thanks to thermal-electron emission the macroparticles acquire a positive charge. The second important feature characterizing the system of thermal plasma plus macroparticles is the relative simplicity of achieving uniform conditions in the plasma and the application of a wide array of diagnostic tools to determine the parameters of the plasma and macroparticles. Finally, in contrast to gas-discharge experiments the system in question is nonstationary by design. The point here is that the two-point correlation function is measured 35 mm above the torch nozzle. Taking into account the velocity of the plasma jet (≈ 5 m/s), we obtain for the plasma existence time $t_{ft} \approx 7$ ms. Thus, it would be wrong to compare the experimental data with calculations for stationary conditions (e.g., with Monte Carlo calculations). It is necessary to model the very process of formation of the ordered structure. The present work investigates the temporal evolution of the two-point correlation function. The calculations are based on the molecular dynamics method. The calculated results are compared with the results of experiment. Questions of the charging and dynamical behavior of the system of macroparticles are also discussed.

2. MODEL

In order to ensure that the numerical model is correct, it is necessary to analyze the main physical processes influencing the dynamics of the formation of the ordered structures of macroparticles. Under the conditions of the experiment the randomly arranged neutral particles incident upon the plasma region are heated to the temperature of the ambient gas, acquire an electric charge, and begin to interact. A simple estimate of the thermalization time gives (see also Ref. 6)

$$\tau_{th} = \left[\pi R_d^2 n_g \sqrt{\frac{8T_g}{\pi m_g}} \, \frac{2m_g}{m_d} \right]^{-1}, \qquad (1)$$

where m_d is the macroparticle mass, and n_g and m_g are the density and mass of the neutral component. For the conditions of the experiment (see Table I) the thermalization time is of the order of $\tau_{th} \approx 5 \times 10^{-6}$ s. This is significantly less than the existence time of the plasma. Consequently, it may be assumed that heating of the macroparticles is essentially instantaneous.

Under the conditions of our experiment the macroparticle charge is determined by the absorption of electrons and ions of the plasma, and by emission of electrons from the

TABLE I. Parameters used in the simulation.

$T_g = 1700$ K	$P = 1$ atm	$Z_d = 500$
$n_e = 7\times10^{10}$ cm^{-3}	$n_i = 4\times10^{10}$ cm^{-3}	$n_d = 5\times10^{7}$ cm^{-3}
$l = 17\ \mu$m	$\lambda_d = 11\ \mu$m	$R_d = 0.4\ \mu$m
$m_d = 1.6\times10^{-12}$ g	$\nu_{fr} = 9.6\times10^{4}$ s^{-1}	$\Gamma_c = 150$
$\Gamma_d = 30$	$N = 200$	$\tau_s = 0.3\ \mu$s

surface of the macroparticle via thermal-electron emission. The charging dynamics are governed by the equation

$$\frac{dZ_d}{dt} = I_e^- + I_i^+ - I_e^+ , \tag{2}$$

where I_e^- is the emitted electron flux, and I_i^+ and I_e^+ are the fluxes of the absorbed ions and electrons, respectively. We utilize expressions for the fluxes in the case of thermal-emissive charging, given in Ref. 10:

$$I_e^+ = \pi R_d^2 n_e \sqrt{\frac{8T_g}{\pi m_e}} \left[1 + \frac{Z_d e^2}{R_d T_g}\right],$$

$$I_i^+ = \pi R_d^2 n_i \sqrt{\frac{8T_g}{\pi m_i}} \exp\left(-\frac{Z_d e^2}{R_d T_g}\right), \tag{3}$$

$$I_e^- = 2\pi R_d^2 \left(\frac{m_e T_g}{2\pi\hbar^2}\right)^{3/2} \sqrt{\frac{8T_g}{\pi m_e}} \left[1 + \frac{Z_d e^2}{R_d T_g}\right] \times \exp\left(-\frac{W_e}{T_g}\right) \exp\left(-\frac{Z_d e^2}{R_d T_g}\right), \tag{4}$$

where $n_{e(i)}$ and $m_{e(i)}$ are the density and mass of the electrons (ions), Z_d is the charge of a macroparticle, and W_e is the work function of the electron. It is easy to see from Eqs. (3) that

$$\frac{I_e^+}{I_i^-} \gg \sqrt{\frac{m_i}{m_e}} \gg 1,$$

so ion absorption can be neglected. The equilibrium charge of the macroparticles in this case is given by $I_e^- = I_e^+$, which for the conditions of our experiment ($T_g = 1700$ K, $R_d = 0.4\ \mu$m, $n_e = 7\times10^{10}$ cm^{-3}, $W_e = 2.1$ eV) gives $Z_d \approx 550$. Equation (2) can in principle be solved numerically to determine the time characterizing the charging of the macroparticles. To estimate the characteristic charging time, we use the simple expression $\tau_{ch} = Z_d / I_e^+$, which gives $\tau_{ch} \approx 4 \times 10^{-9}$ s. This time is more than three orders of magnitude less than the existence time of the plasma. By virtue of this fact, we may assume that the charging is essentially instantaneous. Moreover, this estimate shows that the charge on the macroparticles ''tracks'' their temperature, since $\tau_{ch} \ll \tau_{th}$.

Another problem of correct modeling of a system of macroparticles in a plasma consists in the recently discovered experimental fact of a significant departure of the macroparticle temperature corresponding to their random motion from the temperature of the neutral component.[11–13] It has been shown that under certain conditions the kinetic temperature of the macroparticles in a gas-discharge plasma can exceed the neutral gas temperature by a factor of a thousand, despite the high efficiency of energy dissipation via friction with the neutral component in a weakly ionized plasma. Processes leading up to such a nonequilibrium condition are presently unknown. Therefore it has become necessary to examine the dynamical characteristics of the system of macroparticles with regard to the conditions of our experiment with thermal plasma. Toward this end, a photon correlation method was devised on the basis of the ''Malvern'' serial correlator.[14] The essence of the method consists in examining the autocorrelation function of the laser radiation scattered by the system of macroparticles. Under certain conditions,[14,15] an analysis of the autocorrelation function of the scattered radiation yields information about the magnitude of the diffusion coefficient of the macroparticles, and consequently about their kinetic temperature.

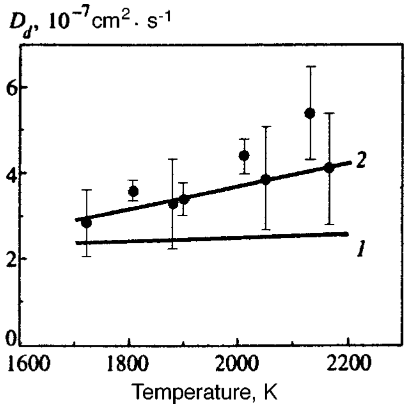

FIG. 1. Temperature dependence of the diffusion coefficient of the charged macroparticles D_d. Curves are the results of theoretical calculation (*1* — hydrodynamic regime, *2* — free-molecule regime).

The scheme of the photon correlation method is described in detail in Ref. 16. Here we will dwell only briefly on the main results of the measurements. Measurements were performed over a wide range of plasma parameters (ion, electron, and macroparticle densities) and revealed the weak influence of these parameters on the dynamical characteristics of the system of macroparticles. Figure 1 plots the dependence of the macroparticle diffusion coefficient on the temperature. For comparison, the two theoretically calculated curves are also plotted in the same figure. Curve *1* corresponds to the hydrodynamic regime $\lambda_{fp} \ll R_d$ (λ_{fp} is the mean free path of the air molecules) and was calculated according to the formula

$$D_d = \frac{T_g}{6\pi\eta R_d},$$

where η is the viscosity of the medium. Curve *2* corresponds to the inverse free-molecule regime ($\lambda_{fp} \gg R_d$) and was calculated according to the formula (see, e.g., Ref. 16)

$$D_d = \frac{3T_g^{3/2}}{8PR_d^2(2\pi m_g)^{1/2}},$$

where P is the gas pressure. Under the conditions of our experiment $\lambda_{fp} \approx 0.8$–$1.1\ \mu$m, i.e., $\lambda_{fp} \sim R_d$ and, consequently, a regime intermediate between the hydrodynamic regime and the free-molecule regime is realized. Nevertheless, Fig. 1 shows that curve *2* (the free-molecule regime) provides a better description of the experimental results. Thus, the experiment shows that under the conditions of the experiment, an ensemble of charged macroparticles behaves like an ordinary collection of Brownian particles with diffusion coefficient calculated in the free-molecule approximation.

3. MODELING RESULTS

The system was modeled using the molecular dynamics method programmed into the KARAT computer code.[17] The two-dimensional equation of motion was solved for each macroparticle with allowance for the interaction between macroparticles, the frictional force exerted by the neutral component, and a random force arising from collisions with molecules of the ambient gas (the Brownian force):

$$m_d \frac{d^2\mathbf{r}_k}{dt^2} = \sum_{j\neq k} \Phi(r)\bigg|_{r=|\mathbf{r}_k-\mathbf{r}_j|} \frac{\mathbf{r}_k-\mathbf{r}_j}{|\mathbf{r}_k-\mathbf{r}_j|} - m_d \nu_{fr}\frac{d\mathbf{r}_k}{dt} + \mathbf{F}_{br}. \tag{5}$$

Here ν_{fr} is the friction frequency, calculated in accordance with Sec. 2 with the formula that is valid for the free-molecule regime:

$$\nu_{fr} = \frac{8PR_d^2}{3m_d}\sqrt{\frac{2\pi m_g}{T_g}}$$

(see also Ref. 16), $\mathbf{F}_{br}$ is the random Brownian force, and $\Phi(r)$ can be written in the form

$$\Phi(r) = -Z_d e\frac{\partial \phi_d}{\partial r} = \frac{Z_d^2 e^2}{r^2}\left[1+\frac{r}{\lambda}\right]\exp\left(-\frac{r}{\lambda}\right) \tag{6}$$

under the assumption that the interaction potential between the macroparticles has the Debye form; here λ is the screening length. At the initial moment the charged macroparticles (charging is essentially instantaneous) are randomly disposed in the calculation region, consisting of a square with side length L_0. To model an infinite system, we used periodic boundary conditions, which make it possible to avoid edge effects and determine the macroparticle density. The two-dimensional macroparticle density was chosen such that the mean interparticle distance l coincided with its value obtained in a real three-dimensional experiment $l=(4\pi n_d/3)^{-1/3}$. The interaction between macroparticles, described by the screened Coulomb potential (6), was cut off at very small distances in order to avoid too small a time step in the initial stage of the calculations; thus we set $\Phi(r) = \Phi(l_0)$ for $r<l_0$ ($l_0=0.3l$). In the calculations the time step was set equal to $\tau_s = 0.03/\nu_{fr}$. For these assignments, 250,000 time steps were calculated per run. Therefore, we were forced to use two-dimensional geometry with a relatively small number of particles ($N=200$). A larger number of particles would have required too much calculation time. Nevertheless, since we were calculating a system with a not-too-large nonideality parameter Γ_c, such that the correlation radius did not exceed several l, the given number of particles was completely sufficient to obtain reliable results.

The parameter values used in the simulation are listed in Table I. They correspond to the conditions of an actual experiment[5,6] and also to the parameters used in Monte Carlo calculations of the given system.[18] The Coulomb interaction parameter Γ_c and the interaction parameter with screening taken into account, Γ_d, were calculated as follows:

$$\Gamma_c = \frac{Z_d^2 e^2}{lT_g}, \quad \Gamma_d = \frac{Z_d^2 e^2}{lT_g}\exp\left(-\frac{l}{\lambda}\right). \tag{7}$$

Note that because of the significant concentration and relatively large charge of the macroparticles, the ion and electron densities turn out to be substantially different. The problem of determining the screening radius for such a system requires a special analysis, which lies beyond the scope of the present paper. In our calculations, it was set equal to the electron Debye radius $\lambda_d = \sqrt{T_g/4\pi e^2 n_e}$, so that screening by ions could be neglected.

We go on now to the main results of the numerical simulation. They show that after relaxation the system arrives at a state which we can nominally call liquid-like. This result is in agreement with the values of the nonideality parameters Γ_c and Γ_d. Figures 2a–2d show the temporal evolution of the two-point correlation function $R(r)$, which we calculated from the positions of the particles at a certain moment of time. Figure 2e was obtained by averaging $R(r)$ over time, which is possible since in the time interval $40\,\text{ms} < t < 70\,\text{ms}$ the system reaches the equilibrium state and the two-point correlation function becomes essentially constant. In the construction of the two-point correlation function the step in r was chosen equal to $0.1l$. For comparison, Fig. 2 also plots the correlation function obtained directly in the experiment (dashed curve).

Analysis of Fig. 2 Initially, closely spaced particles disperse, as a result of which a region is formed where the two-point correlation function vanishes [Figs. 2(a) and 2(b)]. This process concludes rapidly, since the repulsive force between particles increases abruptly with decreasing distance between them. Next, the first maximum increases with the passage of time, and simultaneously higher-order peaks begin to form [Fig. 2d]. The steady-state correlation function [Fig. 2e] is characterized by several pronounced maxima, which is characteristic of systems with close-range order. It was in this sense that we called the final state of the system liquid-like.

The determination of the formation time of the ordered structure remains, to a certain degree, arbitrary. Obviously, it depends on which spatial scale of the correlations is of interest. The greater the distance at which the two-point correlation function approaches its final form, the longer the time needed for this to happen. Thus, for example, in our case we can say that the first three peaks form in $t_f \approx 35\,\text{ms}$. One can also introduce the formation time of the first maximum of the two-point correlation function, t_1. In essence, this is the time needed for the emergence of any close-range order in the system. Numerical calculations shows that under our conditions $t_1 \approx 5$ ms.

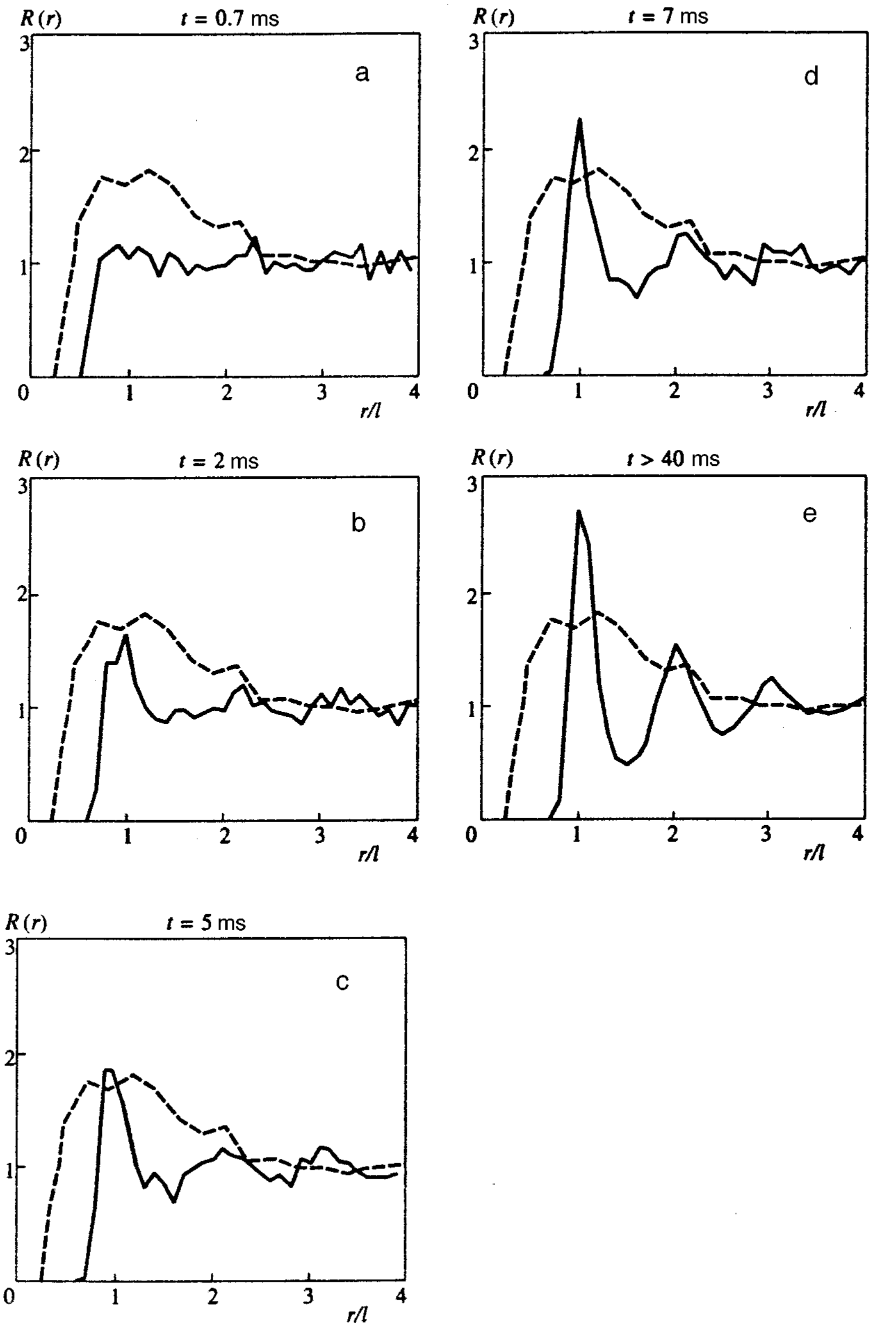

FIG. 2. Temporal evolution of the two-point correlation function. Solid curve — simulation, dashed curve — experiment. Formation time $t=0.7$ (a), 2 (b), 5 (c), 7 (d), >40 ms (e).

As noted in the Introduction, one peculiarity of the present experiment is the finite existence time of the plasma, being $t_{ft}\approx 7$ ms. According to the numerical simulation this implies that the structure being diagnosed in the experiment is still being formed, and consequently the two-point correlation function that has been measured is not the two-point correlation function corresponding to the steady state. Nevertheless, the existence time of the plasma turns out to be completely sufficient for the emergence in the system of close-range order.

Figures 2a–2e enable us to compare the shape of the experimental two-point correlation function with that of the calculated function. We make two remarks here. First, the experimental correlation function has only one peak, in complete agreement with the modeling results. Indeed, Fig. 2(d) shows that for $t=7$ ms the first maximum is already close to its final shape while the higher-order peaks are only beginning to form. Thus, the existence of just one peak is explained by the fact that during the existence time of the plasma the ordering process does not have time to play out completely. Second, the first maximum of the experimental correlation function is significantly broader (by almost fivefold) than the calculated function. One of the possible reasons for this broadening, associated with peculiarities of the use of a laser time-of-flight counter, is discussed in Ref. 18. Another reason for this broadening could, in principle, be stochastic fluctuations of the macroparticle charges.[19] However, as numerical estimates show, the frequency of these fluctuations $\sim\tau_{ch}^{-1}$ is large while their amplitude $\delta Z_d/Z_d \sim 10^{-2}$ is so small that they probably cannot explain the observed effect. A definitive assessment of the reasons for this broadening of the autocorrelation peak requires further study. The numerical calculation presented here only enables us to claim that it is in no way associated with the formation of ordered structures due to electrostatic interaction between macroparticles.

4. CONCLUSION

Modeling the ordering of charged macroparticles in an atmospheric-pressure thermal plasma in application to the conditions of our experiments[5,6] has required us to consider a number of processes. We have shown that macroparticles are charged very rapidly (in comparison with other characteristic times) when they are injected into the plasma region. The photon correlation method has been used to examine the dynamic behavior of the macroparticles. The numerical modeling results presented here show that the formation of ordered structures can be explained by the electrostatic interaction between macroparticles. The finite existence time of the plasma explains the presence of only one peak in the experimental correlation function, since the structure is observed in the process of formation.

This work was carried out with partial financial support from the Russian Fund for Fundamental Research (Project No. 98-02-16825) and INTAS–RFBR (Project No. 95-1335).

*⁾E-mail: ipdustpl@redline.ru

---

[1] J. H. Chu and Lin I, Phys. Rev. Lett. **72**, 4009 (1994).
[2] H. Thomas, G. E. Morfill, V. Demmel *et al.*, Phys. Rev. Lett. **73**, 652 (1994).
[3] Y. Hyashi and K. Tachibana, Jpn. J. Appl. Phys., Part 2 **33**, L804 (1994).
[4] A. Melzer, T. Trottenberg, and A. Piel, Phys. Lett. A **191**, 301 (1994).
[5] V. E. Fortov, A. P. Nefedov, O. F. Petrov *et al.*, JETP Lett. **63**, 187 (1996).
[6] V. E. Fortov, A. P. Nefedov, O. F. Petrov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 467 (1997) [JETP **84**, 256 (1997)].
[7] A. Barkan and R. L. Merlino, Adv. Plasma Phys. **2**, 3261 (1995).
[8] V. E. Fortov, A. P. Nefedov, V. M. Torchinskiĭ *et al.*, JETP Lett. **64**, 69 (1996)].
[9] V. E. Fortov, A. P. Nefedov, V. M. Torchinsky *et al.*, Phys. Lett. A **229**, 317 (1997).
[10] M. Sodha and S. Guha, Advances in Plasma Phys. **4**, 219 (1971).
[11] H. Thomas and G. E. Morfill, Nature (London) **379**, 806 (1996).
[12] A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).
[13] J. B. Pieper and J. Goree, Phys. Rev. Lett. **77**, 3137 (1996).
[14] P. N. Pusey and R. J. A. Tough, in *Dynamic Light Scattering: Application of Photon Correlation Spectroscopy*, R. Pecora (Ed.), Plenum Press, New York (1985).
[15] K. Shatzel, Adv. Colloid Interface Sci. **46**, 309 (1993).
[16] A. P. Nefedov, O. F. Petrov, S. A. Khrapak *et al.*, Teplofiz. Vys. Temp. **36**, 141 (1998).
[17] V. Tarakanov, *User's Manual for the Code KARAT*, BRA, Inc. (1992).
[18] V. E. Fortov, V. S. Filinov, A. P. Nefedov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 889 (1997) [JETP **84**, 449 (1997)].
[19] S. Cui and J. Goree, IEEE Trans. Plasma Sci. **22**, 151 (1994).

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Electron kinetics in collisionless shock waves

A. M. Bykov*) and Yu. A. Uvarov†)

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted 7 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 846–864 (March 1999)

We study the kinetic model of the formation of the energy spectrum of nonthermal electrons near the front of a quasilongitudinal, supercritical, collisionless shock wave. Nonresonant interactions of the electrons and the fluctuations generated by kinetic instabilities of the ions in the transition region inside the shock front play the main role in the heating and preacceleration of electrons. We calculate the electron energy spectrum in the vicinity of the shock wave and show that the heating and preacceleration of electrons occur on a scale of the order of several hundred ion inertial lengths in the vicinity of the viscous discontinuity. Although the electron distribution function is significantly nonequilibrium near the shock front, its low-energy part can be approximated by a Maxwellian distribution. The effective electron temperature T_2^{eff} behind the front, obtained in this manner, increases with the Mach number of the shock wave slower than it would if it followed the Hugoniot adiabat. We determine the condition under which the electron heating is ineffective but the electrons are effectively accelerated to high energies. The high-energy asymptotic behavior of the distribution function is that of a power law, with the exponent determined by the total compression ratio of the plasma, as in the case of acceleration by the first-order Fermi mechanism. The model is used to describe the case (important for applications) of acceleration of electrons by shock waves with large total Mach numbers, with the structure of these waves modified by the nonlinear interaction of nonthermal ions and consisting of an extended prefront with a smooth variation of the macroscopic parameters and a viscous discontinuity in speed with a moderate value of the Mach number. © *1999 American Institute of Physics.* [S1063-7761(99)00603-4]

1. INTRODUCTION

The physical processes of relaxation of highly nonequilibrium perturbations of plasma, responsible for the formation and structure of collisionless shock waves, are of fundamental importance in modeling various objects with a high energy release in cosmic plasma.[1–5] Collisionless shock waves in rarefied plasma serve as a universal source of nonthermal charged particles and the observed radiation.[1,6] In the case of strong shock waves (with Alfvén Mach numbers M exceeding several units), the dissipation due to the anomalous resistance of electrons proves to be insufficient and the structure of the shock front is determined by kinetic instabilities of the ions. Such collisionless shock waves are called supercritical.[7] Here, to avoid any misunderstanding, we must bear in mind that the concept of a supercritical, collision shock wave in radiative gas dynamics is defined differently.[8] In this paper we deal only with collisionless, supercritical shock waves. Modeling collisionless shock waves in laboratory conditions is extremely difficult, but there is a vast body of observational data on the structure of shock waves in the interplanetary medium.[9,10] Computer simulations of the structure of collisionless shock waves that use hybrid codes, which interpret protons as particles and electrons as a liquid, have made it possible to describe the main features of supercritical, quasilongitudinal shock waves (with the angle between the normal to the front and the local magnetic field smaller than $\pi/4$) (see Refs. 11 and 12). Note that the structure of the front and other manifestations differ substantially for quasilongitudinal and quasitransverse waves. The theory of quasitransverse, collisionless shock waves and the methods used in observing such waves are described, e.g., in Refs. 13–16. Here we examine only quasilongitudinal shock waves.

The front of a quasilongitudinal, supercritical shock wave is an extremely extended transition region occupied by magnetic-field fluctuations with amplitudes $\delta B/B \sim 1$ and characteristic frequencies below the ion gyrofrequency[10] (Fig. 1). The generation of fluctuations is due to instabilities in the interpenetrating multicurrent ion movements.[11] The width Δ of the front (the transition region) of a quasilongitudinal shock wave reaches several dozen ion inertial lengths $l_i = c/\omega_{pi}$.

Computer simulation that uses hybrid codes has made it possible to arrive at a very important result, i.e., the fact that a group of reflected nonthermal ions detaches itself in the process of relaxation of fluctuations within the front of a supercritical, quasilongitudinal shock wave.[11] The reflected ions with a gyroradius exceeding the width of the shock front are then accelerated very efficiently, via the Fermi mechanism, by converging plasma fluxes carrying MHD fluctuations.[1,17,18,4] Electrons with gyroradiuses larger than the front width are also efficiently accelerated by the first-order Fermi mechanism in the vicinity of a quasilongitudinal shock wave.[19] However, a nonrelativistic electron must have an energy that is m_p/m_e times higher than that of the respec-

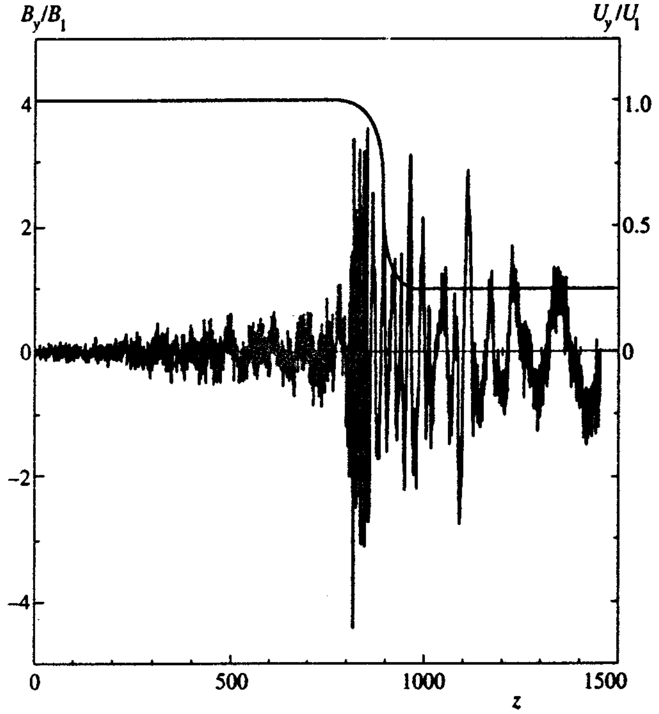

FIG. 1. Profile of a fluctuating magnetic field in the vicinity of a viscous discontinuity (according to the data of Bennett and Ellison[12]) with the characteristic flow speed profile superimposed on it (z is measured in units of c/v_{pi}).

tive proton to be injected into the Fermi acceleration mechanism. Thus, the problem of electron injection actually reduces to forming a nonthermal electron distribution in energy up to energies of order $(m_p/m_e)T_1$, where T_1 is the temperature of the plasma in the undisturbed region.[20] Here and in what follows we use labels 1 and 3 to indicate quantities that refer, respectively, to the incoming (toward the front) and outgoing (away from the front) flows of plasma. The label 2 indicates quantities that refer to the transition region.

There is no way in which the hybrid-code description of electrons as a liquid can provide information about nonthermal electrons. However, since magnetic-field fluctuations in the transition region of a supercritical shock wave are determined by kinetic instabilities of the ions, which carry a significant portion of the energy dissipated in the shock wave, we can study the effect of magnetic-field fluctuations on the kinetics of the electrons, interpreting the latter as test particles. In such an approach the electrons do not dynamically affect the fluctuations. However, the effect of the electron liquid on the dispersion properties of the waves is taken into account in computer simulation based of hybrid codes. Our model relies on this fact. It allows calculating the electron energy spectrum near a fast quasilongitudinal shock wave with a local Alfvén Mach number of the viscous discontinuity, M, smaller than $M<M_* \equiv (\beta m_p/m_e)^{1/2}$, where $\beta = 4\pi P/B^2$, with P the plasma pressure. Moreover, the model can be used for computer simulations of electron kinetics in the vicinity of extended, collisionless shock waves with large Mach numbers ($M>M_*$) propagating in a turbulent plasma, in particular, of shock waves of supernova shells with Mach numbers ~ 100 and larger (see Sec. 5). In this case, due to nonlinear effects of the interaction of the accelerated particles with the incoming flow of ions there is effective smoothing of the ion speed profile by the pressure of the fact accelerated particles, which penetrate the region of ''undisturbed'' flow before the shock front.[1,4] The characteristic scale of the region of smooth deceleration of the incoming flow, known as the prefront, is $l_{pf} \sim (v_i/u)\lambda_i$, where λ_i and v_i are the transport mean free path and the speed of the accelerated nonthermal ions, which carry a significant portion of the flux energy thermalized in the shock wave, and u is the speed of the shock front. A detailed computer simulation of the large-scale structure of such a shock wave that uses standard hybrid codes with a spatial resolution of order l_i is difficult since $l_{pf} \gg l_i$. Such a shock wave can be successfully modeled by the Monte Carlo method.[4] It has also been shown that the front of a strong collisionless shock wave consists of an extended prefront and a viscous discontinuity in speed, corresponding to the local Mach number M, which is smaller than the total Mach number of the shock wave.[1,4,21] Here the compression of matter at the discontinuity (on the scale of several hundred l_i) can be much lower than the total compression of the medium in the shock wave with allowance for compression in the prefront.

A remark is in order here. Long before the first results describing the structure of a collisionless shock wave modified by the pressure of nonthermal particles were obtained, physical models of collision shock waves were being used in many-liquid radiative gas dynamics.[8,22,23] These models predicted the structure of strong shock waves with a prefront. Furthermore, the conditions needed for the realization of discontinuous solutions for shock waves with an isoelectron-thermal discontinuity and of continuous flows as functions of the strength of the wave and the radiation pressure before the front were determined.[23]

We allow for the interaction of electrons with the prefront (on scales of order l_{pf}) in the adiabatic approximation, but study the problem of nonadiabatic heating and acceleration of electrons at the density discontinuity with a Mach number smaller than M_* by employing the model developed below. If the local Alfvén Mach number of the incoming flow in a strong shock wave exceeds M_*, the thermal electron distribution becomes highly anisotropic and mode generation effects of the whistler type become important. Levinson[24,25] did a detailed study of electron acceleration in strong quasilongitudinal shock waves with $M \geqslant M_*$. Earlier, Cargill and Papadopoulos[26] modeled nonadiabatic electron heating in a quasitransverse shock wave with $M \geqslant M_*$ by the hybrid-code method. The calculations showed the heating to be highly efficient. As noted earlier, nonthermal electron distributions could not be studied directly by the hybrid code method.

Note that we have limited our discussion to nonrelativistic shock waves, in which flow dynamics is determined by the ion component. The modeling of relativistic shock waves in an electron–positron plasma is discussed, e.g., in Refs. 27 and 28. In Sec. 2 we discuss the kinetic model for describing the electron spectrum in the vicinity of a collisionless shock wave. In Sec. 3 we examine the effective electron temperature. Applications of the theory to direct observations of electrons in interplanetary shock waves are studied in Sec. 4. Finally, in Sec. 5 we give a brief description of the applications of our model to galactic sources of hard electromagnetic radiation.

2. DESCRIPTION OF MODEL

We consider a collisionless shock wave propagating in a plasma with $\beta\sim 1$. If the Alfvén Mach number of the shock wave meets the condition $M<M_*$, the velocity of thermal electrons exceeds that of the shock front. Here the angular distribution of the electrons is almost isotropic.

We assume that the front of a supercritical shock wave is the transition region of finite width Δ (between the incoming and outgoing plasma flows) within which are strong magnetic-field fluctuations and a vortex electric field. In this work we limit ourselves to a shock wave with a flat front, i.e., front curvature is ignored. The characteristic correlation length l_c of fluctuations is comparable to the gyroradii of the ions in the plasma and is much longer than the gyroradii of the electrons whose energy is close to thermal. What is important is that such fluctuations can be directly observed in quasilongitudinal shock waves and are clearly manifest themselves in computer simulations.[29] Since the distribution of fluctuations in the transition region is probably close to isotropic,[10] on scales larger than l_c the electron motion can be interpreted as spatial diffusion. The transport mean free path of the magnetized electrons, Λ_2, in this case is approximately equal to the correlation length of strong magnetic-field fluctuations, $\Lambda_2\sim l_c$, for electrons with gyroradii smaller than l_c and is almost energy independent for $E\leqslant T_1(m_p/m_e)$ if the spectrum of magnetic-field fluctuations is close to a single-scale one or decreases with increasing wave number κ according to a $\kappa^{-\nu}$-law with $\nu\geqslant 2$.

The longitudinal microscopic diffusion coefficient for electrons whose gyroradii are smaller than l_c is $k_{zz}\approx v\Lambda/3$, where v is the electron velocity, and the z axis is directed along the normal to the shock front. The magnetic-field fluctuations are assumed uniformly distributed over the front width and are characterized by the mean-square quantity $\alpha=\overline{(\delta B/B)^2}$. In the highly turbulent plasma of the transition region of the shock front, the transport of particles with moderately epithermal energies can be due to transport by stochastic vortices with scales l lying in the interval $l_c\geqslant l\geqslant\Delta$. We use the methods of describing of highly turbulent plasmas discussed in Ref. 30. Here the macroscopic coefficient of turbulent electron diffusion (averaged over vortex movements in the scale interval mentioned earlier) in the low-energy limit is independent of particle velocity.

The fluctuation vortex electric fields induced by chaotic ion flows in the transition region determine the statistical acceleration of electrons. For a strong, quasilongitudinal, supercritical shock wave the effect of electron acceleration by a potential electric field in the transition region can be ignored, since the potential discontinuity at the front is smaller than $0.05m_pu_1^2$ for $M\sim 5$ and decreases with increasing M (see the review in Ref. 10). A potential field of such magnitude does not allow the electrons to acquire an energy of order $(m_p/m_e)T_1$.

To describe the heating and acceleration of a magnetized electron in the shock wave transition region by a vortex electric field, we use the drift transport equation. More precisely, we employ the method developed by Chandrasekhar *et al.*[31] for obtaining a transport equation averaged over rapid rotations of the electron, which is then averaged over the ensemble of short-correlated fluctuations of magnetic and hence vortex electric fields. Here, following the approach developed by Toptygin,[32] we arrive at a transport equation for the isotropic part of the electron distribution function. In the reference frame in which the front is at rest, the isotropic part of the quasi-steady-state electron momentum distribution function $N(z,p)$ normalized to the phase volume element satisfies the following equation in the transition region:

$$k_{zz}(p)\frac{\partial^2N(z,p)}{\partial z^2}-u(z)\frac{\partial N(z,p)}{\partial z}+\frac{p}{3}\frac{\partial N}{\partial p}\frac{\partial u(z)}{\partial z}$$

$$+\frac{1}{p^2}\frac{\partial}{\partial p}p^2D(p)\frac{\partial N}{\partial p}=0, \tag{1}$$

where

$$D(p)\approx p^2\overline{G}\left(\frac{\delta B}{B_0}\right)^2\left(\frac{v_a}{v}\right)^2\frac{v}{l_c}. \tag{2}$$

Here $u(z)$ is the mean velocity of the ion component of the flux averaged over fluctuations of scales of order l_c. Within this model, allowance for the effect of slow cyclic front reformation, studied by Scholer,[33] can be described by a slow time dependence of the quantities averaged over the rapid fluctuations in the transition region. Here, however, we limit ourselves to the statistically steady-state problem. The factor $\overline{G}$ depends on the index of the distribution of the power-law spectrum and the polarization of the magnetic-field fluctuations.[32] For power-law spectral distributions of fluctuations with an index $\nu\geqslant 2$ and for a Gaussian distribution of unpolarized magnetic-field fluctuations, $\overline{G}\sim 0.1$. Here the mean-square magnetic-field fluctuation $\alpha\approx\overline{G}(\delta B/B_0)^2$. The distributions of magnetic-field fluctuations and of the ion-component velocity depend on the Mach number of the shock wave.[10] A detailed procedure of finding these distributions based, say, on hybrid simulations, is also needed for a precise description of the low-energy asymptotic behavior of $D(p)$. The results discussed in Ref. 30 show that in the region of moderate epithermal momenta, $D(p)$ exhibits an asymptotic behavior of the type $D(p)\sim p^2G_1u_1\Delta^{-1}$, which differs from (2) (the latter holds in the high-momentum region). The numerical factor G_1 and the related momentum limit p_a, at which the expression (2) is matched with the low-energy asymptotic formula for $D(p)$, depend on the details of the distribution of vortices over scales.[30] When calculating the models, we parametrized this quantity and studied the dependence of the resulting electron distributions on the parameter p_a. As we will show shortly, the choice of the value of the parameter for matching the two regimes of the D vs. p behavior has a strong influence on the effective electron temperature behind the shock front (see Sec. 3) but has very little influence on the fluxes of highly epithermal electrons accelerated by the shock wave. Note that the electron acceleration, which in our model is described by the coefficient $D(p)$, is an analog of the well-known betatron effect.

In some cases, outside the transition region we can ignore the statistical acceleration effect, which is described in the equation by a second-order differential operator (diffusion in momentum due to inelastic electron scattering by magnetic-field fluctuations). This is the case in interplanetary shock waves, for example, where the boundary of free particle escape (see condition (10) below) is at a distance of the order of several Δ from the transition region of the shock front. Here the derivative $\partial u/\partial z$ is usually also small, since the smoothed flow speed changes little outside the transition region. Then in the incoming and outgoing flows the isotropic part of the electron distribution function satisfies the equation

$$k_i\frac{\partial^2 N_i}{\partial z^2}-u_i\frac{\partial N_i}{\partial z}=0, \quad i=1,3, \tag{3}$$

where $i=1$ corresponds to the incoming flow, $i=3$ corresponds to the outgoing flow, and $k_i=k_{zz_i}$. This equation has a simple analytic solution, which makes it possible to simplify and speed up calculations of the model. Numerically, it is more convenient to solve the boundary value problem with an equation obtained from (1) for the function $p^2N(z,p)$, since for this function the boundary conditions $p=0$ and $p=\infty$ have the simple form $p^2N(z,p)=0$. The calculation procedure and the diagrams were set up for this function, but here it is more convenient to discuss the function $N(z,p)$.

At the boundary $z=-\infty$ the $N_1(p)=N_{-\infty}(p)$ electron distribution function was taken in form of the thermal Maxwellian distribution with a temperature T_1 and a power-law correction, the latter describing the background nonthermal electron flux in the incoming flow far from the shock wave (this additional term contains contributions of previous shock waves).

At the boundary between the incoming flow and the transition region we impose the matching conditions for the distribution functions and the fluxes. We also place the origin at this boundary $(z=0)$. Then at $z=0$ we have the boundary conditions

$$N_1(p)=N_2(p), \quad J_1(p)=J_2(p), \tag{4}$$

where the flux J is given by the formula

$$J_i=-k_i\frac{\partial N_i}{\partial z}-u_i\frac{p}{3}\frac{\partial N_i}{\partial p}. \tag{5}$$

The first term describes the diffusion flux and the second, the convection flux (related to the motion of medium). In the incoming-flow region, Eq. (2) has an analytic solution satisfying the condition at $-\infty$:

$$N_1=C_1(p)\exp\frac{u_1 z}{k_1}+N_{-\infty}, \tag{6}$$

where $C_1(p)$ is an arbitrary function found from the boundary conditions. Using this solution and the condition at the boundary $z=0$, we find that the problem reduces to solving the equation to the right of the boundary $z=0$ with the following conditions at that boundary:

$$N_{-\infty}+\frac{k_2}{u_1}\frac{\partial N_2}{\partial z}=N_2, \tag{7}$$

where u_1 is the speed of the incoming flow, and the subscript ''2'' corresponds to the transition region. Reasoning along similar lines, we can impose the conditions at the viscous discontinuity in speed inside the transition region at $z=z_t$,

$$N_{2-}(p)=N_{2+}(p), \quad J_{2-}(p)=J_{2+}(p), \tag{8}$$

and at the boundary between the outgoing flow and the transition region at $z=\Delta$,

$$N_2(p)=N_3(p), \quad J_2(p)=J_3(p). \tag{9}$$

This one-dimensional model with diffusion propagation of the particles in the spatially unrestricted outgoing-flow region does not allow for the finiteness of the shock front and the related restrictions on the acceleration efficiency. In particular, one must bear in mind that most electrons that have moved away from the front in the outgoing-flow region to distances of the order of the transverse size of the shock wave leave the system, due to diffusion and the drag by large-scale flow, and do not participate in further acceleration. We allow for this effect by imposing an additional boundary condition in the outgoing-flow region at a distance z_{fe} of the order of the transverse dimensions of the front. The condition is that of free escape of the electrons that have passed this boundary $(z=z_{fe})$ and are freely moving in the right half-plane:

$$N_{3-}(p)=N_{3+}(p), \quad \mathbf{J}_{3-}(p)=\mathbf{J}_{3+}(p),$$

$$\mathbf{J}_{3+}(p)=\int_{p_z>0}\mathbf{v}F\,d\Omega_{\mathbf{p}}, \tag{10}$$

with F the electron distribution function, which in the diffusion approximation[32] has the form

$$F(\mathbf{r},\mathbf{p},t)=\frac{1}{4\pi}\left(N(\mathbf{r},p,t)+\frac{3}{v^2}\mathbf{v}\cdot\mathbf{J}(\mathbf{r},p,t)\right). \tag{11}$$

Equation (2) in the outgoing-flow region has an analytic solution that is similar to that in the incoming-flow region:

$$N_3=C_2(p)\exp\frac{u_3(z-\Delta)}{k_3}+C_1(p), \tag{12}$$

where $C_1(p)$ and $C_2(p)$ are arbitrary functions determined by the boundary conditions. Using the explicit form of the solution and the conditions at the boundaries $z=\Delta$ and $z=z_{fe}$, we find that the problem reduces to solving Eq. (1) in the transition region with a condition at the boundary $z=\Delta$,

$$\begin{aligned}-k_2\frac{\partial N_2}{\partial z}f-u_3\frac{p}{3}\frac{\partial}{\partial p}\left(N_2-\frac{k_2}{u_3}\frac{\partial N_2}{\partial z}+\frac{k_2}{u_3}\frac{\partial N_2}{\partial z}\right)\\ =\frac{v}{2}\left(N_2-\frac{k_2}{u_3}\frac{\partial N_2}{\partial z}+\frac{k_2}{u_3}\frac{\partial N_2}{\partial z}f\right),\end{aligned} \tag{13}$$

where $f=\exp\{[u_3(z_{fe}-\Delta)]/k_3\}$, and a condition at the boundary $z=0$,

$$N_{-\infty}+\frac{k_2}{u_1}\frac{\partial N_2}{\partial z}=N_2. \tag{14}$$

It is now convenient to go over to the dimensionless variables $\tilde{z}=z/\Delta$ and $\tilde{p}=p/p_T$ and the dimensionless speed $u(z)=u(z)/u_1$, where $p_T=\sqrt{2m_ekT}$ is the thermal momentum. Then the problem can be characterized in a natural way by two dimensionless quantities, the parameter α and the parameter $\Gamma=u_1\Delta/v\Lambda$ calculated for electrons with $\tilde{p}=1$. The parameter Γ depends on the strength of the shock wave and is linked to the Alfvén Mach number through the relationship $\Gamma\propto M(m_e/\beta m_p)^{1/2}(\Delta/l_c)$. Generally speaking, the width Δ of the transition region depends on the Mach number of the shock wave.[10] Recall that the applicability of the adopted model is limited by the condition imposed on the local Mach number of the viscous discontinuity in speed: $M(m_e/\beta m_p)^{1/2}\leqslant 1$. The parameter α, introduced above, characterizes the amplitude and spectral distribution of magnetic-field fluctuations in the shock wave.

After we have found the distribution function, we can determine the electron flux registered by the detector:

$$J(E)=vp^2N(p(E))\frac{dp}{dE}. \tag{15}$$

For typical magnetic-field fluctuation spectra, the electron mean free path $\Lambda(p)$ increases with momentum.[32] The exact shape of this dependence is determined by the spectrum of magnetic inhomogeneities and the large-scale motion of the medium.

In our calculations of the electron distribution function we used the following dependence:

$$\Lambda_2(p)=\begin{cases} l_c, & \tilde{p}\leqslant\tilde{p}_*, \\ l_c(\tilde{p}/\tilde{p}_*)^{\zeta}, & \tilde{p}\geqslant\tilde{p}_*. \end{cases}$$

According to the results of Sec. 2, we have $\tilde{p}_*=(m_p/m_e)^{1/2}$. The exponent ζ depends on the spectrum of magnetic fluctuations on scales larger than the proton gyroradius. Under typical conditions, $0\leqslant\zeta\leqslant 1$. What is important is that if $\Gamma\gg 1$, epithermal electron transport in the transition region and in region 3 is determined by turbulent vortex fluctuations of the plasma macroscopic speed. Then, in the low-energy region, both the transport electron mean free path and the spatial diffusion coefficient are energy-independent and are determined by the properties of the turbulence.[30] The value of the turbulent diffusion coefficient is estimated at $k_2\approx u_2\Delta_*$ for electron with momenta $p\leqslant p_a$. Here Δ_* is the mixing length, which is much smaller than Δ. Measurements from observations or numerical modeling of the polarization and the correlation function of fluctuations in the macroscopic speed would make it possible to use, instead of estimates, the detailed theory of calculations of transport coefficients in highly turbulent plasma described in Ref. 30. We know of no measurements of these fluctuation parameters. Hence in this paper we study two cases: that of electron transport within the transformation region of a shock wave due to scattering by magnetic-field fluctuations ($p_a\sim p_T$), and that of low-energy electron transport via turbulent diffusion ($p_a\gg p_T$).

Figures 2 and 3 depict the results of calculating the electron distribution function at the transition-region boundaries: the two solid curves in each figure correspond to the spectra at the points $\tilde{z}=0$ and $\tilde{z}=1$. The modeling was done for a bare distribution $N_{-\infty}(p)$ equal to the Maxwellian distribution of electrons with a temperature T_1 and of unit concentration. The dotted curves in Figs. 2 and 3 represent the approximations of the low-energy part of the electron distribution by Maxwellian functions; also indicated are the effective temperatures measured in units of T_1 (for more details see Sec. 3). The calculations were done for two different values of the shock parameters. Figure 2 depicts the results of calculations for a boundary condition at point $\tilde{z}=1$ corresponding to the diffusion propagation of electrons in the outgoing flow. The results of calculations of the electron distribution function for a shock wave with electron transport in the transition region due to scattering by magnetic-field fluctuations ($p_a\sim p_T$) are shown in Figs. 2a and 2b. The model with turbulent low-energy electron transport ($p_a\gg p_T$) in the transition region of the shock wave leads to the distributions depicted in Figs. 2c and 2d. Note the substantial difference between the effective temperatures under relatively small variations in the high-energy asymptotic behavior for these two cases.

The electron spectra for finite-sized systems in which the particles freely leave the plane $\tilde{z}=1$ are depicted in Fig. 3. Figures 3a and 3b correspond to the case where $p_a\sim p_T$, and Figs. 3c and 3d to the case where $p_a\gg p_T$. The differences between systems with different conditions of particle escape from the acceleration region become especially evident when one analyzes high-energy asymptotic behavior. For systems with diffusion particle propagation in the outgoing-flow region there forms an asymptotic power-law distribution of high-energy electrons with an exponent determined by the shock-wave compression ratio (Fig. 2). This distribution corresponds to standard first-order Fermi acceleration of test particles.[1] For systems with free escape, the asymptotic distribution of high-energy particles drops off exponentially (see Fig. 3).

3. EFFECTIVE ELECTRON TEMPERATURE

A specific feature of all shock waves, both collisional and collisionless, is the possibility of strong nonadiabatic heating of the medium after the passage of the shock front.[8,34] In standard single-liquid collision shock waves, the heating of the medium behind the front is determined via the Hugoniot shock adiabat, which for strong shock waves yields the $T_2^H/T_1\propto M^2$ law. In a multicomponent system, the quantity T_2^H calculated via the shock adiabat is no more the temperature of the separate component even for a collision shock wave and determines an average quasiequilibrium temperature. For a collisionless shock wave, a substantial portion of the energy of the incoming flow is transferred to nonthermal particles and vibrational modes. The assumption that there is local thermodynamic equilibrium behind the front, which is used to obtain the Hugoniot shock adiabat, may be violated very significantly.

The spectrum of the electrons behind the front of a collisionless shock wave is highly nonthermal, but in some

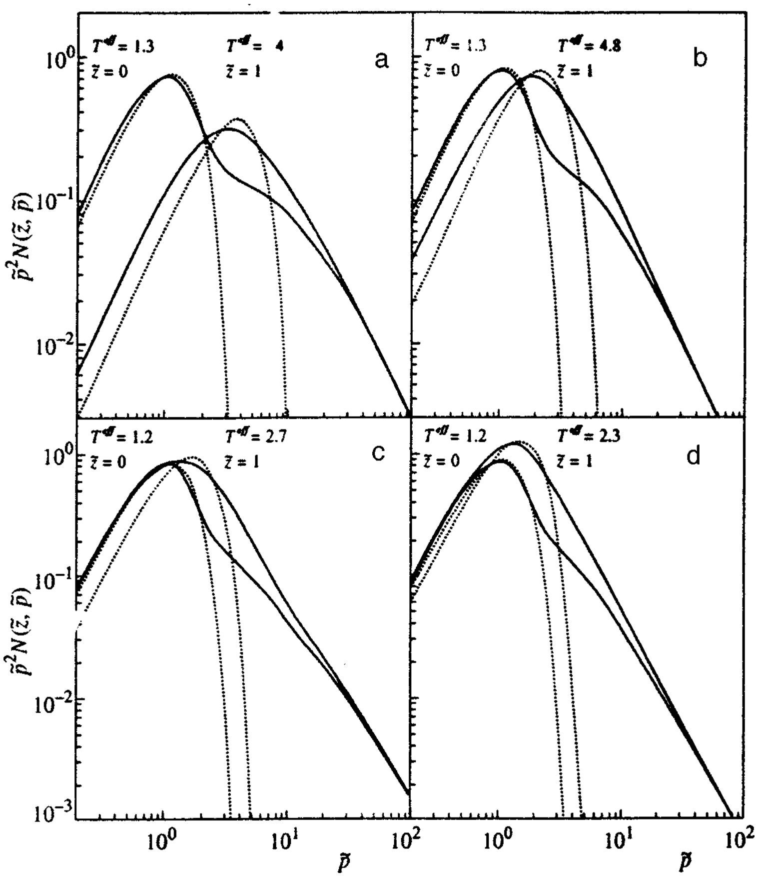

FIG. 2. Electron distribution function in the vicinity of a shock wave with $\Gamma = 5$ for a diffusion boundary condition. The dotted curves represent the approximations of the peak in the distribution by Maxwellian functions with effective temperatures: (a) and (b), calculations for $p_a \sim p_T$; (c) and (d), calculations for $p_a \gg p_T$. Here $\alpha = 0.4$ ((a) and (c)), and $\alpha = 0.1$ ((b) and (d)).

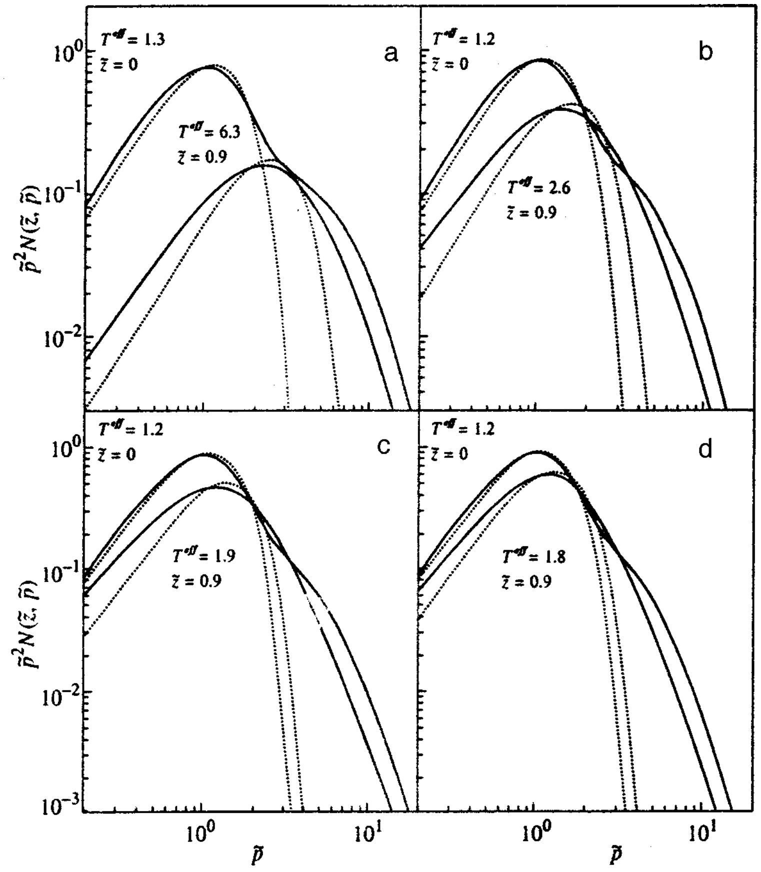

FIG. 3. The same as in Fig. 2 but for the case in which the electrons freely leave the surface $\tilde{z} = 1$.

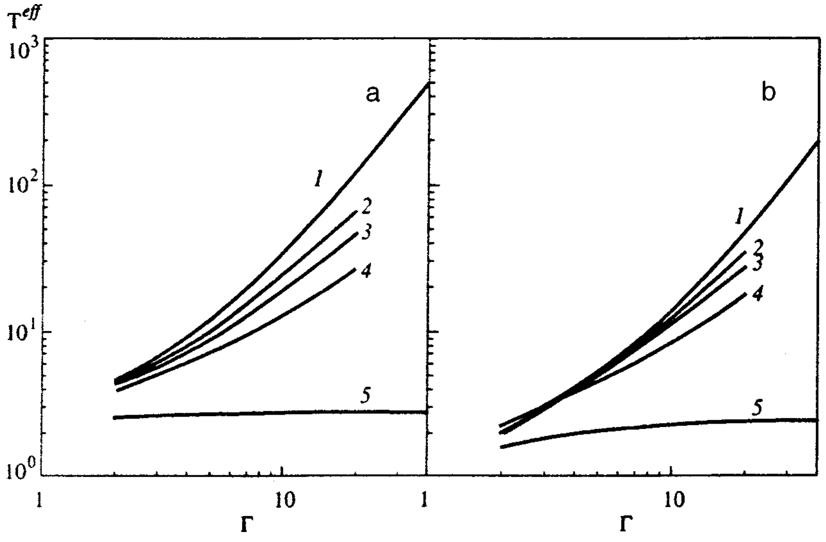

FIG. 4. Dependence of the effective temperature of accelerated electrons on parameter Γ at $\alpha = 0.5$. Curves *1, 2, 3,* and *4* have been calculated for $p_a \sim p_T$, and curves *5* have been calculated for $p_a \gg p_T$. For curves *1* and *5*, the length of the electron path was assumed to be momentum-independent. For curves *2, 3,* and *4* the exponent in the momentum dependence of the length of the electron path is 0.3, 0.5, and 1, respectively.

cases in the low-energy region near the peak in the distribution it can be approximated adequately by the Maxwellian distribution with an effective temperature $T_2^{\rm eff}$, which distribution then smoothly transforms into a power-law distribution for the accelerated particles (see Fig. 2). The problem of the ratio of the effective temperature of the particles behind the front of the collisionless shock wave to the temperature T_2 obtained via the Hugoniot adiabat is of interest both from the viewpoint of understanding the transport energy-relaxation processes in a shock wave and from the viewpoint of numerous applications. In collisionless shock waves, the temperature T_2^H is not an observable, but the ratio $T_2^{\rm eff}/T_2^H$ may serve as a measure of effectiveness of electron heating. The effective temperature can be observed. In shock waves in the interplanetary medium, this temperature is determined by direct measurements of spectral electron fluxes.[10] Observations of soft x-rays emitted by collisionless shock waves make it possible to estimate the effective temperature $T_2^{\rm eff}$ of the electrons behind the fronts of such waves in supernova remnants (see the review in Ref. 34) and in the vicinity of high-speed clouds, where a large nonthermal excess of x-rays is also observed.[35] The above model makes it possible to explain the observed excess of x-rayphotons as the bremsstrahlung from the high-energy part of the nonthermal electron distribution.

We have numerically modeled the effective electron temperature as a function of the Mach number. Calculations in the modeling process involved calculating $T_2^{\rm eff}$ as a function of Γ for various values of the parameter p_a. Recall that the parameter p_a characterizes the role of turbulent electron transport by vortices inside the transformation region. If $p_a \sim p_T$, the role of turbulent transport is minor, and transport occurs only due to scattering by magnetic-field fluctuations.

Remaining within the model of collisionless heating and accelerating of electrons in the shock wave transition region by magnetic-field fluctuations with a coherence length $l_c \sim l_i$, we can approximate the effective heating by the $T^{\rm eff}/T_1 \propto \Gamma^a$ law (curves *1* in Fig. 4). Here the quadratic law holds approximately for both a diffusion boundary condition at the boundary of the extended outgoing-flow region (curve *1* in Fig. 4a) and in condition of free escape (curve *1* in Fig. 4b). This result coincides with the dependence that holds for strong shock waves in a medium with collisions, which ensure local equilibrium in the outgoing-flow region. Nonadiabatic heating in a collisionless shock wave with $\Gamma \sim 10$ is predominant up to mean-square field amplitudes $\alpha \leqslant 0.01$.

In collisionless shock waves, for models with different electron scattering laws in the transition region of the front the function $T^{\rm eff}(\Gamma)$ differs dramatically. In our model, the scattering of electrons with an energy lower than $(m_p/m_e)T_1$ is nonresonant. To formulate observational tests for verifying the model, we also modeled the $T^{\rm eff}(\Gamma)$ functions that emerge if the conditions for nonresonant scattering are violated.

When the transition region contains small-scale ($< l_i$) resonant fluctuations, the transport electron mean free path depends on electron momentum. This changes the function $T^{\rm eff}(\Gamma)$. Curves *2, 3,* and *4* in Fig. 4, corresponding to $\zeta = 0.3$, 0.5, and 1, respectively, illustrate this fact. For instance, for an electron mean free path proportional to the electron gyroradius ($\zeta = 1$), which corresponds to Bohm diffusion, we have $T^{\rm eff}/T_1 \propto M^a$ with $a \approx 1$. The curves *4* in Fig. 4 represent the function $\Gamma^{\rm eff}(\Gamma)$ in this model.

When $p_a \gg p_T$, turbulent electron transport by developed vortex turbulence in the shock wave transition region is predominant. In this case, effective electron heating can be suppressed even for strong shock waves, which is illustrated by curves *5* in Fig. 4.

The difference in the models discussed makes it possible to formulate observational tests. Indeed, fluxes of observed soft x-rays from the shock waves from supernovas are sensitive to the value of $T^{\rm eff}$. Hence there is the possibility of statistically analyzing the residues with the measured (via optical lines) speeds of shock waves and the effective temperatures (determined through x-ray measurements) so as to verify the dependence of the effective electron temperature on the front speed and hence to find $T^{\rm eff}(\Gamma)$. Another possible test is related to the fact that efficient electron heating in a shock wave with developed vortex turbulence is extremely weak in the transition region (greater by a factor of approximately 2.5 even for $\Gamma \geqslant 10$), while T^{eff} is at least ten times higher if electron transport due to scattering by magnetic-field fluctuations is the predominant mechanism (Fig. 4). Since the high-energy electron fluxes accelerated by the shock wave are of the same order for these two cases, we can select the correct electron-transport mechanism by analyzing the ratio of the fluxes of radiation from the shock wave in the x-rayand γ-ray windows.

4. ACCELERATION OF ELECTRONS BY SHOCK WAVES IN THE INTERPLANETARY MEDIUM

Modeling the effects of generation of nonthermal electrons by collisionless shock waves in a laboratory is extremely difficult. Information about the spectra of electrons accelerated by shock waves outside the heliosphere can be extracted only indirectly, i.e., either by analyzing the observed electromagnetic radiation from objects (radio waves and gamma rays) or by studying the features of spectra of cosmic electrons and positrons observed near the Earth's orbit. At present, measuring the electron spectra in the vicinity of shock waves in the interplanetary medium with space probes is the only direct method of investigating such phenomena. Such measurements have been carried out for several decades.[13,9] The results of observing 37 shock waves by the ISEE-3 in the interplanetary medium have been summed up in the review in Ref. 9. The detectors on board ISEE-3 and several other satellites have established the existence of several typical classes of events related to the passage of shock waves. The nature of these events differs for protons and electrons and strongly differs for longitudinal and transverse shock waves. We modeled the spectra of nonthermal electrons in the vicinity of quasilongitudinal shock waves on the basis of the theory discussed above, with allowance for the features of collisionless shock waves in interplanetary plasma.

Usually, shock waves in the interplanetary medium are not strong enough to form a prefront but are strong enough to ensure a compression ratio of about four at the viscous discontinuity. The region of the incoming flux, the region of the outgoing flow, and the transition region (the discontinuity) are clearly visible in the speed profile of the ion component. In the transition region, the electron distribution function obeys Eq. (1). For shock waves in the interplanetary medium, at distances of about 1 AU from the Sun (i.e., at the Earth's orbit, where many satellites measurements are carried out), the characteristic front width Δ is probably 10^9 cm and the characteristic transverse dimensions do not exceed 10^{13} cm. The typical value of the electron mean free path in the interplanetary medium (far from the shock front), Λ_∞, is roughly 10^{12} cm at energies of about 100 keV, but the mean free path is much shorter in the vicinity of the shock wave transition region because of the sharp increase in the magnetic-field fluctuation amplitude. Since the electron mean free path is comparable to the size of the shock wave, we must allow for the fact that the problem is not one-dimensional. This is effectively taken into account by our model by introducing a boundary for the free escape of the accelerated particles.

In the region of the incoming flow before the shock front, the electrons freely leave the acceleration region in view of the large value of Λ_∞. This corresponds to the condition for free escape at $\tilde{z}=0$. The electron spectrum and the time dependence of the fluxes measured by the ISEE-3 satellite can be obtained via calculations if one assumes that the electron mean free path near the shock wave behind the front, Λ_3, is roughly 10^9 cm, which agrees with estimates made on the basis of the model. Within the model, this leads to the condition for free escape at the boundary $\tilde{z}=300$. Electrons that have reached this boundary leave the acceleration region.

According to the observational data provided by the ISEE-3 satellite, the electron spectrum in the interplanetary medium far from shock waves has a highly nonthermal nature and can be described by the sum of the Maxwellian thermal spectrum and a nonthermal power-law background. According to observations,[9] the nonthermal background undergoes strong temporal and spatial variations. Our model requires specifying a bare electron distribution function in the incoming flow at $-\infty$. Hence, to describe the acceleration in the entire energy range, we must know the electron temperature and concentration before the shock wave and the parameters of the nonthermal background. The background flux of nonthermal electrons in the ''undisturbed'' interplanetary medium in the course of observing the shock wave was also measured by the same detector on board the ISEE-3 satellite.[9] The calculations fixed the initial distribution of the electrons in the incoming flow ($z \to -\infty$) in the form of a superposition of the thermal Maxwellian distribution characterized by two parameters, the temperature T and the concentration N, and a nonthermal background flux in the undisturbed medium far from the shock front.

The results of calculations of the electron flux in our model with the parameters $\alpha=0.2$ and $\Gamma \approx 5$ and a typical (for the interplanetary medium) incoming flow with $N_e = 3\ \mathrm{cm}^{-3}$ and $T_1 = 10^5$ K are depicted in Fig. 5 along with the results from ISEE-3 observations of a quasiperiodic shock wave in the interplanetary medium.[9] The speed of the shock front amounted to $422\ \mathrm{km\,s}^{-1}$, and the angle of inclination the field to the front was 22°. The parameters α and Γ were determined via magnetic field observations[9] and the results of calculations of the shock front structure based on hybrid codes,[12] which yielded $\alpha \approx 0.2$ and $\Gamma \approx 5$. The most significant uncertainty in determining α is related to the assumption about the nature of the polarization and about the magnetic-field fluctuation spectrum in the transition region, which affects $\overline{G}$. Our model makes it possible to quantitatively describe the typical observable features of electron events in supercritical shock waves: a sharp (severalfold) increase in the intensity of the electron fluxes in the vicinity of the transition region and the subsequent slow decrease of the intensity of the flux of nonthermal electrons over times of about one hour (see Fig. 5). Moreover, the observed electron flux exhibits intense short bursts about ten minutes long (''spikes''). Two such spikes are clearly visible in Fig. 5. The duration of the spikes corresponds to the characteristic time it takes the electrons to be accelerated to energies of

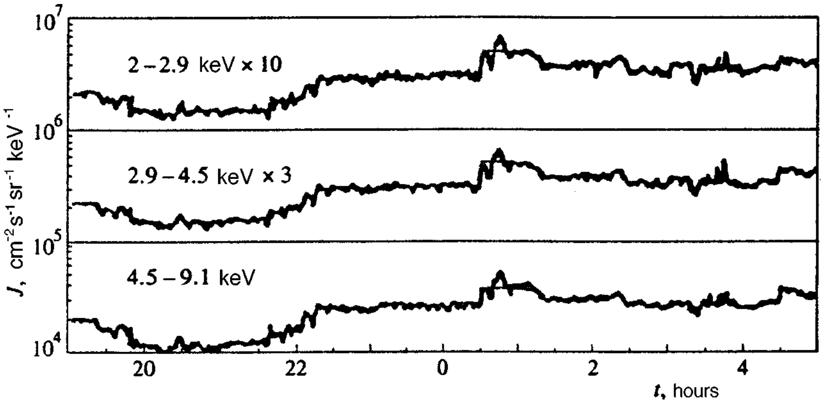

FIG. 5. Calculation of the time evolution of the electron flux (thin lines) in the vicinity of an interplanetary shock wave (the speed of the shock front is 422 km s^{-1}, and the field's inclination angle is 22°) for three energy intervals. Electron flux measurements obtained by the ISEE-3 satellite[9] are depicted by heavy lines. The observation time in hours labels the horizontal axis.

order 1 keV at the shock front. Hence, in our model, the origin of the spikes can be explained by the passage of narrow fluctuations of the electron density (with a duration of less than 1 min), registered in the flow before the shock front, through the shock front (see Fig. 5). Figure 5 clearly shows that the model provides a satisfactory description of the observed evolution of the nonthermal electron flux in the vicinity of the shock wave.

The study of electron spectra in the solar wind was one of the goals of the ACE space laboratory launched by NASA in 1997. In particular, the SWEPAM monitor analyzes the energy spectra of low-energy electrons (from 1 eV to 1240 eV). Two spectrometers of the EPAM device with a geometric factor of 0.48 cm^2 sr each are intended for observing fluxes of electrons with an energy higher than 30 keV. Along with observations of electron fluxes, observations of magnetic-field fluctuations (by the MAG magnetometer) and other parameters of the solar wind can be conducted. Observations of nonthermal electrons done by the ACE laboratory will make it possible to thoroughly compare the data and the results obtained in our model at low (up to 1.25 keV) and at high energies.

5. ACCELERATION OF ELECTRONS BY EXTENDED SHOCK WAVES IN A TURBULENT MEDIUM

The theory of generation of nonthermal electrons by shock waves described in Secs. 1–3 can also be used in modeling distant galactic and extragalactic objects. Direct measurements of electron distributions outside the solar system are impossible, so that the model must describe the generation of electromagnetic radiation, both thermal and nonthermal, by electrons accelerated by shock waves. Here we briefly touch on two classes of objects.

The high-speed clouds falling onto the galactic plane actively interact with the matter of the disk. The velocities of these extended objects occupying an appreciable fraction of the celestial sphere above the Galaxy are determined from radio observations of neutral hydrogen HI in the 21-cm line. The typical values of these velocities are $\sim 150\,\mathrm{km\,s}^{-1}$. Since the velocities of the clouds exceed the phase velocities of magnetoacoustic and Alfvén perturbations in the galactic medium, the interaction of the clouds and the matter of the halo and disk of the Galaxy must be accompanied by shock wave formation. Observations of the hard electromagnetic radiation indicated that there are large surpluses of such radiation, which probably formed in the regions where the high-speed clouds interact with the Galaxy.[35] One can assume that the soft x-rays and the hard γ-radiation have the same nature and origin. Then the sources of such radiation are the nonthermal particles, mostly electrons, accelerated by shock waves formed as a result of interaction of the cloud and the matter of the halo and disk of the Galaxy. Taking the values $\sim 1\,\mu\mathrm{G}$ and $\sim 3\times 10^{-3}\,\mathrm{cm}^{-3}$ as estimates of the magnetic field strength and the plasma density in the halo of the Galaxy, we find that the magnetoacoustic Mach number of the shock wave is roughly 3. This is quite sufficient for the wave to be supercritical, although the wave is not strong enough to form a prefront and is even not strong enough to ensure the maximum possible compression ratio equal to 4. Using this model yields results that are fully consistent with the observed spectra (for more details see Ref. 35).

Supernova remnants are among the brightest nonthermal sources, emitting radiation over a broad range, from radio waves to hard γ rays.[36–40] A sizable fraction of the enormous amount of energy released in the explosion of a star ($\sim 10^{51}$ erg) goes into the shell in the form of kinetic energy; the shell expands at supersonic velocity into the ambient medium. This expansion leads to formation of a strong collisionless shock wave with a Mach number reaching several hundreds (see, e.g., Refs. 1 and 3). An important role in the formation of the observed spectrum of nonthermal electromagnetic radiation emitted by supernova remnants is played by the emission from electrons accelerated by shock waves at the shell boundary. At present there is clear observational evidence of electron acceleration in the supernova remnant SN1006 up to energies of order 10^{14} eV (Ref. 36). Let us discuss the possibility of applying our model of generation of nonthermal electrons to shock waves from supernovas.

In extended shock waves with large Mach numbers that propagate in a turbulent medium there is effective transformation of the kinetic energy of the incoming flow into nonthermal particles and waves. Here, due to the interaction between nonthermal ions penetrating the incoming flow to a distance of order l_{pf} and slowing it down, the speed profile is the region of adiabatic decrease of the speed, known as the prefront. The prefront is followed by the viscous discontinuity in speed. The characteristic prefront size is usually several orders of magnitude greater than the characteristic size

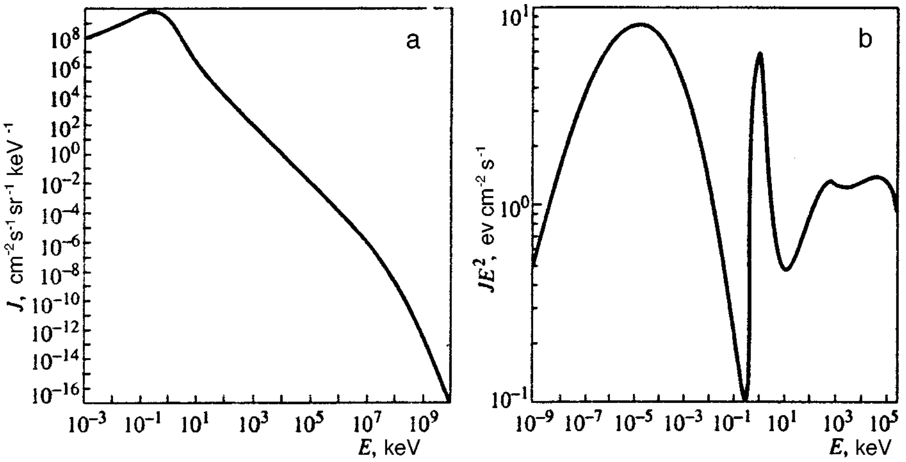

FIG. 6. (a) Calculated electron flux in the vicinity of the shock wave from a supernova shell, and (b) the expected spectrum of the electromagnetic radiation for the electron distribution in Fig. 6a.

of the transition region and the viscous discontinuity. The viscous discontinuity in speed has a much smaller amplitude than the discontinuity in a shock wave with the same compression ratio but without a prefront.

In this problem one must distinguish between the total Mach number of a shock-wave with a prefront and the local Mach number of the viscous discontinuity in speed, which is much smaller than the total Mach number. A description of the large-scale ($\sim l_{pf}$) structure of the front of a strong shock wave can be found in Ref. 19, where it is shown that the local Mach number M at a viscous discontinuity does not exceed M_*. For this reason we used the model in calculating the absolute electron fluxes in the vicinity of the strong shock wave of a supernova remnant propagating in a medium with typical parameters, a temperature $T = 3\times10^5$ K and a concentration $n = 1$ cm^{-3}. The electron transport mean free path in the external medium ahead of the shock front depends on the spectrum of magnetic inhomogeneities in the undisturbed interstellar medium: $\Lambda_1(p)\sim\sqrt{p}$ (see Ref. 3). The magnetic field strength inside a remnant is 10^{-4} G. The total compression ratio in the shock wave amounted to 50, while the compression ratio at the viscous discontinuity amounted to roughly 2. The results of the calculations are illustrated by Fig. 6. Figure 6a depicts the calculated electron spectrum for a strong shock wave modified by the accelerated particles. Here the nonthermal electron spectrum exhibits three characteristic sections, whose origin is due to the structure of the strong shock wave modified by the accelerated particles. The low-energy part of the spectrum corresponds to electron heating in a narrow transition region of the shock wave in the vicinity of a viscous discontinuity. Note that here we modeled the case in which $p_a\sim p_T$. As the energy of the accelerated electrons increases, so does their mean free path in the prefront, and with the increasing mean free path the electrons effectively ''feel'' the increasing compression ratio in the shock wave, with the result that the electrons are accelerated by the Fermi mechanism more efficiently (see also Refs. 19 and 24). This leads to a situation in which, as the energy increases, the spectrum first becomes less steep and then, at high energies, again becomes steeper due to energy losses by relativistic electrons to synchrotron radiation and escape of particles from the acceleration region.

Such a shape of the electron spectrum makes it possible to match the spectra of radio waves and x-rays observed in some supernova remnants. The results of calculating the electrons of the electromagnetic radiation emitted by electrons that are accelerated by shock waves are depicted in Fig. 6b. The emission spectrum is calculated with allowance for the synchrotron radiation emitted by electrons in the magnetic fields of a supernova remnant and the bremsstrahlung from high-energy electrons in the surrounding plasma. A contribution to radiation with an energy higher than 100 MeV (not shown in Fig. 6) will also be provided by the Compton scattering of low-energy (radio and optical) photons by relativistic electrons (the inverse Compton effect). A detailed quantitative comparison of the calculated spectra and the observed radiation from supernova remnants lies outside the scope of the present work, whose main goal was to develop a theory of generation of electrons by collisionless shock waves.

The authors would like to express their deep gratitude to Lee Bennett and Donald C. Ellison for making available the results of hybrid simulations of shock wave structures. The present work was partially supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17711) and INTAS (Grant No. 96-0390).

*)E-mail: byk@astro.ioffe.rssi.ru
†)E-mail: uv@astro.ioffe.rssi.ru

---

[1] R. Blandford and E. Eichler, Phys. Rep. **154**, 2 (1987).

[2] A. A. Galeev, in *Space Research* [in Russian] (Itogi Nauki i Tekhniki), Vol. 27, All-Union Institute of Scientific and Technical Information, Moscow (1988), p. 3.

[3] S. I. Vaĭnshteĭn, A. M. Bykov, and I. N. Toptygin, *Turbulence, Current Layers, and Shock Waves in Cosmic Plasma* [in Russian], Nauka, Moscow (1989).

[4] F. C. Jones and D. C. Ellison, Space Sci. Rev. **58**, 259 (1991).

[5] V. S. Ptuskin, Nuovo Cimento C **19**, 755 (1996).

[6] A. A. Galeev, Zh. Éksp. Teor. Fiz. **86**, 1655 (1984) [Sov. Phys. JETP **59**, 965 (1984)].

[7] C. F. Kennel, J. P. Edmiston, and T. Hada, in *Collisionless Shocks in the Heliosphere*, B. T. Tsurutani and R. G. Stone (eds.), AGU, Washington (1985), p. 1.

[8] Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Academic Press, New York (1967), Chap. 7.
[9] B. T. Tsurutani and R. P. Lin, J. Geophys. Res. **90**, 1 (1985).
[10] J. R. Kan, M. E. Mandt, and L. H. Lyu, Space Sci. Rev. **57**, 201 (1991).
[11] L. B. Quest, J. Geophys. Res. **93**, 9649 (1988).
[12] L. Bennett and D. C. Ellison, J. Geophys. Res. A **100**, 3439 (1995).
[13] D. L. Vaĭsberg, A. A. Galeev, G. N. Zastenker, S. I. Klimov, M. N. Nozdrachev, R. Z. Sagdeev, A. Yu. Sokolov, and V. D. Shapiro, Zh. Éksp. Teor. Fiz. **85**, 1232 (1983) [Sov. Phys. JETP **58**, 716 (1983)].
[14] A. A. Galeev, M. A. Malkov, and H. J. Völk, J. Plasma Phys. **54**, 59 (1995).
[15] P. L. Biermann and J. P. Cassinelli, Astron. Astrophys. **277**, 691 (1993).
[16] K. G. McClements, R. O. Dendy, R. Bingham *et al.*, Mon. Not. R. Astron. Soc. **291**, 241 (1997).
[17] L. Drury, Rep. Prog. Phys. **46**, 973 (1983).
[18] E. G. Berezhko and G. F. Krymskiĭ, Usp. Fiz. Nauk **154**, 49 (1988) [Sov. Phys. Usp. **31**, 27 (1988)].
[19] D. C. Ellison and S. P. Reynolds, Astrophys. J. **382**, 242 (1991).
[20] A. M. Bykov and Yu. A. Uvarov, JETP Lett. **57**, 644 (1993).
[21] I. N. Toptygin, Zh. Éksp. Teor. Fiz. **112**, 1584 (1997) [JETP **85**, 862 (1997)].
[22] V. D. Shafranov, Zh. Éksp. Teor. Fiz. **32**, 1453 (1957) [Sov. Phys. JETP **5**, 1183 (1957)].
[23] V. S. Imshennik, Zh. Éksp. Teor. Fiz. **42**, 236 (1962) [*sic*].
[24] A. Levinson, Astrophys. J. **401**, 73 (1992); **426**, 327 (1994).
[25] A. Levinson, Mon. Not. R. Astron. Soc. **278**, 1018 (1996).
[26] P. J. Cargill and K. Papadopoulos, Astrophys. J. **329**, L29 (1988).
[27] P. Schneider and J. G. Kirk, Astron. Astrophys. **217**, 344 (1989).
[28] Y. A. Gallant and J. Arons, Astrophys. J. **435**, 230 (1994).
[29] D. Krauss-Varban, Adv. Space Res. **15**, 271 (1995).
[30] A. M. Bykov and I. N. Toptygin, Usp. Fiz. Nauk **163**(11), 19 (1993) [Phys. Usp. **36**, 1020 (1993)].
[31] S. Chandrasekhar, A. N. Kaufman, and K. M. Watson, Ann. Phys. **5**, 1 (1958).
[32] I. N. Toptygin, *Cosmic Rays in Interplanetary Magnetic Fields* [in Russian], Nauka, Moscow (1983).
[33] M. Scholer, J. Geophys. Res. **98**, 47 (1993).
[34] B. T. Draine and C. F. McKee, Annu. Rev. Astron. Astrophys. **31**, 373 (1993).
[35] J. J. Blom, H. Bloemen, A. M. Bykov *et al.*, Astron. Astrophys. **321**, 288 (1997).
[36] K. Koyama, R. Petre, E. V. Gotthelf *et al.*, Nature (London) **378**, 255 (1995).
[37] R. A. Chevalier, Astrophys. J. **258**, 790 (1982).
[38] A. I. Asvarov, O. H. Guseinov, F. K. Kasimov, and V. A. Dogiel, Astron. Astrophys. **229**, 196 (1990).
[39] S. P. Reynolds, Astrophys. J. **493**, 375 (1998).
[40] T. K. Gaisser, R. J. Protheroe, and T. Stanev, Astrophys. J. **492**, 219 (1998).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Collective effects in spin-polarized Boltzmann gases

T. L. Andreeva and P. L. Rubin*)

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117942 Moscow, Russia

(Submitted 15 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 865–876 (March 1999)

As a result of an investigation of the real collision integral for paramagnetic atoms, we have obtained a criterion for propagation of spin waves in a spin-polarized Boltzmann gas. The main condition for propagation of weakly damped spin waves is a high anisotropy of the atomic scattering amplitude with predominance of the forward scattering. This condition is different from those suggested in earlier publications. Our results indicate that the range of paramagnetic gases where weakly damped spin waves can propagate at the room temperature is considerably wider than it was assumed previously. One example is vapors of alkali metals (Na, Cs, and Rb), where the degree of electron spin polarization can be very high. © *1999 American Institute of Physics.* [S1063-7761(99)00703-9]

1. INTRODUCTION

In gases described by the Boltzmann statistics in the absence of applied electric and magnetic fields, the only propagating collective mode is an acoustic wave. A possibility of collective spin oscillations in a paramagnetic gas was first suggested by Silin.[1] In 1977 Aronov[2] analyzed spin-wave oscillations in semiconductors. Intense interest in propagation of spin waves in gases, however, was generated by the publication by Bashkin in the 1980s.[3] That paper predicted the existence of weakly damped spin waves in spin-polarized Boltzmann gases and determined conditions for their propagation. The criterion for the existence of a new collective mode in a spin-polarized gas was formulated in terms of its ''quantum'' property: the average de Broglie wavelength Λ should be considerably larger than the range of interaction between atoms, r_0.

Although the number of publications dedicated to the dynamics and kinetics of spin-polarized gases has been quite considerable (see Ref. 4 and references therein), the criterion for the existence of spin waves remained unchanged, irrespective of the fact that it was based on purely qualitative reasons.[5] An important point is that this criterion led to a very severe limitation on the gas temperature, so that most gases should condense at temperatures higher than those prescribed by this criterion. For this reason, the only remaining candidates were the spin-polarized hydrogen, H↑, and ^{3}He↑. The existence of spin waves in these two gases was confirmed by experiments, although the detected waves were not weakly damped.[6,7] Note that it is the nuclear spin that is considered in this specific case.

This paper suggests a criterion for propagation of spin waves in a spin-polarized Boltzmann gas based on a detailed investigation of the collision integral. It turns out that the main parameter that controls propagation of spin waves in Boltzmann paramagnetic gases is the real part of the forward exchange scattering amplitude. Moreover, in conventional gases, where scattering of atoms is semiclassical in nature and highly anisotropic, the situation is more favorable than in ''quantum'' gases, where scattering is almost isotropic (s-wave scattering).[8] This criterion differs from those given in previous publications.[3–6] As a result, the range of paramagnetic gases where weakly damped spin waves can propagate at temperatures about the room temperature is extended considerably. One can take as examples vapors of alkali metals (Na, Cs, and Rb), where the degree of spin polarization can be high.[9]

2. COLLISION INTEGRAL IN BOLTZMANN SPIN-POLARIZED GAS

This paper considers a collision integral calculated previously.[10,11] The spin sublevels in question are degenerate in the absence of magnetic field. In such cases, the collision integral[11] is identical to the Waldmann–Snider integral.[10] Nonetheless, since our previous publication[11] contains misprints, here we give first a general expression for the collision integral St for arbitrary, not necessarily degenerate, systems:

$$\mathrm{St}_{\alpha\alpha'}(p)=-i[I_{\alpha\alpha'}(p)-I^{*}_{\alpha'\alpha}(p)];$$

$$I_{\alpha\alpha'}(p)=(2\pi)^3\hbar^2\left\{\int dp_1\,T_{\alpha\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\exp\left(i\frac{\mathscr{E}_{\alpha'\lambda\theta\sigma}t}{\hbar}\right)f_{\sigma\lambda}(p_1)f_{\theta\alpha'}(p)+\int dp'\,dp_1'\,dp_1\,\delta(p+p_1-p'-p_1')\right.$$

$$\left.\times\exp\left(i\frac{\mathscr{E}_{\tau_1\tau_2\tau_3\tau_4}t}{\hbar}\right)\frac{T_{\alpha\lambda\tau_3\tau_4}((p-p_1)/2,(p'-p_1')/2)\,T^{*}_{\alpha'\lambda\tau_1\tau_2}((p-p_1)/2,(p'-p_1')/2)}{(p-p_1)^2/4m-(p'-p_1')^2/4m+\mathscr{E}_{\tau_1\tau_2\alpha'\lambda}-i0}f_{\tau_3\tau_1}(p')f_{\tau_4\tau_2}(p_1')\right\}.$$

Here p is the momentum, m is the mass, the subscripts are quantum numbers of internal particle states, $\hat{f}(p)$ is the Wigner matrix, $\hat{T}$ is the scattering T-matrix, and, finally,

$$\mathscr{E}_{\alpha\beta\gamma\delta}=E_\alpha+E_\beta-E_\gamma-E_\delta$$

(E_α and E_β are the energies of internal particle states).

This paper considers the case of full degeneracy of the internal (spin) states of particles: $\mathscr{E}_{\alpha\beta\gamma\delta}=0$. In this case, the collision integral has the form (compare with Ref. 10)

$$\begin{aligned}\mathrm{St}_{\alpha\alpha'}(p)={}&-i(2\pi)^3\hbar^2\int dp_1\Bigg[T_{\alpha\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\\&\times f_{\sigma\lambda}(p_1)f_{\theta\alpha'}(p)-T^*_{\alpha'\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\\&\times f_{\lambda\sigma}(p_1)f_{\alpha\theta}(p)\Bigg]+(2\pi)^4\hbar^2\\&\times\int dp'dp_1'dp_1\,\delta(p+p_1-p'-p_1')\\&\times\delta\left(\frac{(p-p_1)^2}{4m}-\frac{(p'-p_1')^2}{4m}\right)T_{\alpha\lambda\tau_3\tau_4}\\&\times\left(\frac{p-p_1}{2},\frac{p'-p_1'}{2}\right)T^*_{\alpha'\lambda\tau_1\tau_2}\\&\times\left(\frac{p-p_1}{2},\frac{p'-p_1'}{2}\right)f_{\tau_3\tau_1}(p')f_{\tau_4\tau_2}(p_1').\end{aligned}\tag{1}$$

Here the Hermitian property of the Wigner matrix is taken into account.

In studying small perturbations, we will use a linearized collision integral, in which the Wigner matrix has the form

$$f_{\alpha\alpha'}(p)=\frac{1}{2}f^{(0)}(p)[C_{\alpha\alpha'}+\varphi_{\alpha\alpha'}(p)].\tag{2}$$

Here $f^{(0)}(p)$ is the equilibrium Maxwell distribution function over particle momenta, $C_{\alpha\alpha'}$ is a constant $(\alpha\alpha')$-tenso r, and $\varphi_{\alpha\alpha'}(p)$ is a small perturbation of the Wigner function. In the absence of additional conservation laws, the equilibrium function is diagonal in internal quantum numbers, i.e., $C_{\alpha\alpha'}=\delta_{\alpha\alpha'}$. But this is not the case if such conservation laws hold, if only to a reasonably good approximation (in this specific case, we are dealing with the total spin of the system).

The linearized collision operator (1) takes the form

$$\begin{aligned}J(\varphi_{\alpha\alpha'})={}&-i(2\pi)^3\hbar^2\int dp_1f^{(0)}(p)f^{(0)}(p_1)\\&\times\Bigg\{\Bigg[C_{\theta\alpha'}T_{\alpha\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\varphi_{\sigma\lambda}(p_1)\\&-C_{\alpha\theta}T^*_{\alpha'\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\varphi_{\lambda\sigma}(p_1)\Bigg]\\&+\Bigg[C_{\sigma\lambda}T_{\alpha\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\varphi_{\theta\alpha'}(p)\\&-C_{\lambda\sigma}T^*_{\alpha'\lambda\theta\sigma}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\varphi_{\alpha\theta}(p)\Bigg]\Bigg\}\\&+(2\pi)^4\hbar^2\int dp'dp_1'dp_1\,\delta(p+p_1-p'\\&-p_1')\,\delta\left(\frac{(p-p_1)^2}{4m}-\frac{(p'-p_1')^2}{4m}\right)\\&\times f^{(0)}(p)f^{(0)}(p_1)T_{\alpha\lambda\tau_3\tau_4}\left(\frac{p-p_1}{2},\frac{p'-p_1'}{2}\right)\\&\times T^*_{\alpha'\lambda\tau_1\tau_2}\left(\frac{p-p_1}{2},\frac{p'-p_1'}{2}\right)[C_{\tau_4\tau_2}\varphi_{\tau_3\tau_1}(p')\\&+C_{\tau_3\tau_1}\varphi_{\tau_4\tau_2}(p_1')].\end{aligned}\tag{3}$$

In the latter equation we use the relation[12] $f^{(0)}(p)f^{(0)}(p_1)=f^{(0)}(p')f^{(0)}(p_1')$ for equilibrium distribution functions. Summation over repeated subscripts is assumed in this case.

In what follows, we consider a polarized gas with spin 1/2. Presently such gases are well known and have been actively investigated both theoretically and experimentally.[6,7,13] In the case of identical particles with spin 1/2, the T-matrix has the form

$$\begin{aligned}T_{\alpha\beta\mu\nu}(P,P')&=\hat{A}[t(P,P')\,\delta_{\alpha\mu}\delta_{\beta\nu}+\theta(P,P')\,\sigma_{\alpha\mu}\sigma_{\beta\nu}]\\&=t(P,P')\,\delta_{\alpha\mu}\delta_{\beta\nu}-t(-P,P')\,\delta_{\alpha\nu}\delta_{\beta\mu}\\&\quad+\theta(P,P')\,\sigma_{\alpha\mu}\sigma_{\beta\nu}\\&\quad-\theta(-P,P')\,\sigma_{\alpha\nu}\sigma_{\beta\mu},\end{aligned}\tag{4}$$

where P and P' are the relative momenta of colliding particles (associated with the reduced mass $m/2$), $\hat{A}$ denotes antisymmetrization with respect to particle exchange (incident and scattered), and the product of Pauli matrices is the scalar product of $\sigma^{(j)}_{\alpha\mu}$ and $\sigma^{(j)}_{\beta\nu}$. The factors t and θ should satisfy the following symmetry properties related to the identity of particles and time-reversal invariance of their interaction:

$$\begin{aligned}&t(P,P')=t(-P,-P')=t(P',P),\\&\theta(P,P')=\theta(-P,-P')=\theta(P',P).\end{aligned}\tag{5}$$

This notation means that we consider collisions in which the total spin is conserved. This approximation is fairly good, especially in the case of the nuclear spin.

Tensor C for a polarized gas has the form

$$C=1+M\sigma \text{ or } C_{\alpha\alpha'}=\delta_{\alpha\alpha'}+M\sigma_{\alpha\alpha'},\tag{6}$$

M characterizes the constant (induced) spin polarization of the gas. It is convenient to expand the small perturbation of the Wigner matrix into scalar φ and vector μ parts:

$$\phi_{\alpha\alpha'}(p)=\frac{1}{2}f^{(0)}(p)[\varphi(p)\,\delta_{\alpha\alpha'}+\mu(p)\,\sigma_{\alpha\alpha'}].\tag{7}$$

By substituting Eqs. (7) and (12) in Eq. (3), we obtain the following expression for the linearized collision operator describing the polarized gas:

$$\begin{aligned}
&J(\varphi\delta_{\alpha\alpha'}+\mu^l\sigma^l_{\alpha\alpha'})\\
&\quad=32\pi^4\hbar^2\int dp'dp_1dp_1'W(p,p_1|p',p_1')\\
&\quad\times\{\delta_{\alpha\alpha'}A_0[\varphi(p')+\varphi(p_1')-\varphi(p)-\varphi(p_1)]\\
&\quad+\sigma_{\alpha\alpha'}[A_1\,\mu(p_1')+A_2\,\mu(p')-A_3\,\mu(p_1)-A_0\,\mu(p)]\\
&\quad+\delta_{\alpha\alpha'}A_3M[\mu(p')+\mu(p_1')-\mu(p)-\mu(p_1)]\\
&\quad+\sigma^3_{\alpha\alpha'}|M|[A_1\,\varphi(p')+A_2\,\varphi(p_1')-A_0\,\varphi(p_1)-A_3\,\varphi(p)]\\
&\quad-A_4[[\mu(p_1'),M,\sigma_{\alpha\alpha'}]-[\mu(p'),M,\sigma_{\alpha\alpha'}]]\}\\
&\quad+16\pi^3\hbar^2\int dp_1f^{(0)}(p)f^{(0)}(p_1)\mathrm{Re}\left[t_1\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\right.\\
&\quad\left.-t_2\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\right][[\mu(p_1),M,\sigma_{\alpha\alpha'}]\\
&\quad-[\mu(p),M,\sigma_{\alpha\alpha'}]]. \qquad (8)
\end{aligned}$$

Here

$$\begin{aligned}
W(p,p_1|p',p_1')&=f^{(0)}(p_1)\,\delta(p+p_1-p'-p_1')\\
&\quad\times\delta((p-p_1)^2/4m-(p'-p_1')^2/4m);
\end{aligned}$$

the coefficients A_n are

$$\begin{aligned}
&A_0=(|t_1|^2+|t_2|^2+|t_1-t_2|^2)/2=|t_1|^2+|t_2|^2-\mathrm{Re}\,(t_1t_2^*),\\
&A_1=|t_1|^2-\mathrm{Re}\,(t_1t_2^*),\quad A_2=\mathrm{Re}\,(t_1t_2^*),\\
&A_3=(|t_1|^2-|t_2|^2-|t_1-t_2|^2)/2=\mathrm{Re}\,(t_1t_2^*)-|t_2|^2,\\
&A_4=\mathrm{Im}\,(t_1t_2^*),
\end{aligned}\qquad (9)$$

where parameters t_1 and t_2 are related to the T-matrix elements as follows [see Eq. (4)]:

$$\begin{aligned}
t_1(P,P')&=t(P,P')-t(-P,P')\\
&\quad+\theta(P,P')-\theta(-P,P'),\\
t_2(P,P')&=t(P,P')-\theta(P,P')-2\,\theta(-P,P'),
\end{aligned}$$

$P=(p-p_1)/2$, and $P'=(p'-p_1')/2$. Bracketed expressions of the form $[a,b,c]$ denote the scalar triple product of three vectors.

In deriving Eq. (8), we have used the optical theorem.[14] Note that the last integral on the right of Eq. (8) is linear in the forward scattering amplitude, whereas the rest of the terms are quadratic in the scattering amplitude.

3. KINETIC EQUATION FOR THE TRANSVERSE MAGNETIZATION COMPONENT IN SPIN-POLARIZED GAS

Let us first consider the transverse (with respect to the z-axis) component of magnetization in a spin-polarized gas. It is more convenient to use the following combinations of μ_x and μ_y:

$$\mu_\pm=\mu_x\pm i\mu_y\,.$$

The kinetic equation for $\mu_\pm$ can be derived from the full equation (8):

$$\begin{aligned}
&-i\omega\mu_\pm(p)+ikv\mu_\pm(p)\\
&\quad=\frac{1}{2}\mathrm{Tr}\{(\sigma^x\pm i\sigma^y)J(p)\}=J_\pm(\mu_\pm),
\end{aligned}\qquad (10)$$

where $v=p/m$, $J_\pm$ are the corresponding components of the collision integral. This equation is expressed in terms of time and space Fourier components of the Wigner matrix: $\mu=\mu(\omega,k)$, ω and k are the frequency and wave vector, respectively.

The collision integral $J_\pm(\mu_\pm)$ now has the form

$$\begin{aligned}
J_\pm(\mu_\pm)&=32\pi^4\hbar^2\int dp'dp_1dp_1'W(p,p_1|p',p_1')\\
&\quad\times\{[A_1\mu_\pm(p_1')+A_2\mu_\pm(p')-A_3\mu_\pm(p_1)\\
&\quad-A_0\mu_\pm(p)]\pm2i|M|A_4[\mu_\pm(p_1')-\mu_\pm(p')]\}\\
&\quad\mp16i|M|\pi^3\hbar^2\int dp_1f^{(0)}(p_1)\\
&\quad\times\mathrm{Re}\left[t_1\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)-t_2\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)\right]\\
&\quad\times[\mu_\pm(p)-\mu_\pm(p_1)].
\end{aligned}\qquad (11)$$

Operators $J_\pm$ can be expressed in the form of a sum, i.e.,

$$J_\pm=Q_R\pm i|M|Q_I\mp i|M|L_I\,. \qquad (12)$$

Here Q_R and Q_I are integral operators quadratic in the T-matrix (scattering amplitude), whose structure is similar to that of the corresponding parameters in the familiar Boltzmann equation. There is good reason to suppose that the eigenvalues of Q_R and Q_I in the case of electron spin are of the same order of magnitude: $\nu_s\simeq n\overline{v}\sigma$, where σ is the gas-kinetic collision cross section and $\overline{v}$ is the average thermal velocity of atoms. For example, the exchange cross section of the cesium atom is of the same order as the gas-kinetic cross section, hence all factors A are of the same order of magnitude.

The structure of operator L_I is notably different:

$$\begin{aligned}
L_I(\mu_\pm)&=16\pi^3\hbar^2\int dp_1f^{(0)}(p_1)\\
&\quad\times\mathrm{Re}\left\{T_{\mathrm{ex}}\left(\frac{p-p_1}{2},\frac{p-p_1}{2}\right)[\mu_\pm(p)-\mu_\pm(p_1)]\right\}.
\end{aligned}\qquad (13)$$

Here T_{ex} is the T-matrix of spin-exchange scattering $(\uparrow\downarrow\to\downarrow\uparrow)$:

$$\begin{aligned}
T_{\mathrm{ex}}(P,P')&=t_1(P,P')-t_2(P,P')\\
&=2\,\theta(P,P')+\theta(-P,P')-t(-P,P'),
\end{aligned}$$

P and P' are relative momenta of colliding particles associated with the reduced mass $m/2$.

It follows from the symmetry properties of the T-matrix (5) that all three operators, Q_R, Q_I, and L_I, are Hermitian in the Hilbert space of functions of momentum p with the conventional scalar product for the kinetic gas theory:[16]

$$\langle\mu_1|\mu_2\rangle=\frac{1}{n}\int f^{(0)}(p)\mu_1^*(p)\mu_2(p)\,dp,$$

where $n=\int f^{(0)}(p)\,dp$ is the particle density in space. Thus, operators J_+ and J_- are Hermitian conjugate.

One can easily check that the function $\mu\equiv1$ makes the collision integral (11) vanish, thus being its eigenfunction with eigenvalue 0. Physically, this is a consequence of total spin conservation, which derives from the specific form of the *T*-matrix (4). Similarly, and for the same reason, function $\mu_z\equiv1$ makes the scalar component of the total collision integral (8) vanish, as will be discussed in detail below. Owing to the fact that the operators J_+ and J_- are mutually conjugate, the function $\mu\equiv1$ is at the same time the left and right eigenfunction of both these operators. Thus, the eigen projector corresponding to the zero eigenvalue is identical for both operators, orthogonal, and takes the form (in Dirac's notation)

$$P=|1\rangle\langle1|. \tag{14}$$

Here $|1\rangle$ is the function identically equal to unity.

Function $|1\rangle$ should be the only eigenfunction with a vanishing eigenvalue of operators J_+ and J_-, which signifies the absence of alternative conservation laws in the μ-subspace. The latter statement does not conflict with the applicability of other conservation laws. It will be shown below that functions p (p_x, p_y, p_z) and $p^2/4m$, along with the constant, are eigenfunctions of the diagonal components of the collision integral (8) with vanishing eigenvalues, which is a consequence of the conservation of momentum, energy, and number of particles.

4. DYNAMICS OF TRANSVERSE SPIN POLARIZATION

Equation (10) is a mathematical expression of the problem of eigenvalues of operators $J_\pm-ik\hat{v}$, where $\hat{v}$ denotes multiplication by the velocity treated as an operator. At small k ($|kv|\ll\nu$, where ν is of order of the kinetic collision frequency) the eigenvalue problem (11) can be solved with the help of perturbation theory in the operator $ik\hat{v}$, i.e., in the hydrodynamic approximation.[16] In the present case, as noted above, the existence of only one eigenfunction $\mu=1$ of operator J with eigenvalue 0 is assumed. It follows from Eqs. (10) and (11) that the correction of the first order in $ik\hat{v}$ is zero.

The second-order correction in this case can be expressed as

$$-i\omega_\pm=k^2\langle1|\hat{v}J_\pm^{-1}\hat{v}|1\rangle. \tag{15}$$

Here $J_\pm^{-1}$ denotes the operator inverse of $J_\pm$ in the subspace $(1-P)L$ (in the original space L operator $J_\pm$ does not have an inverse because one of its eigenvalues is zero).

First consider an unpolarized gas with $M=0$. In this case $J_+=J_-=Q_R$. The eigenvalues of Q_R and Q_R^{-1} are real and the frequency is purely imaginary. Thus, we have in this case simple spin diffusion. As usual,[16] in this situation one can only estimate the corresponding diffusion coefficient:

$$D_s=\langle1|\hat{v}Q_R^{-1}\hat{v}|1\rangle. \tag{16}$$

To order of magnitude, $D_s\sim\overline{v^2}/3\nu_s$. Here $\overline{v}$ is the average thermal velocity of atoms and $1/\nu_s$ is a typical eigenvalue of operator Q_R^{-1}. In this case, as one can easily see from Eqs. (8) and (13), the diffusion coefficient D_s is, generally speaking, different from the usual diffusion coefficient, since the corresponding collision frequencies are different.

In a polarized gas the situation is significantly different. Let $|M|\sim1$ (the condition that the density operator be positive definite leads to $0\leqslant|M|\leqslant1$, and in the case of optical spin polarization in a gas a value of $|M|$ close to unity is achieved in a natural way). Now the operators $J_\pm$ essentially reduce to the third term on the right of Eq. (12), whereas the first two can be treated as small corrections.

In order to prove this, let us estimate the real part of the exchange forward scattering amplitude $\mathrm{Re}[T_{\mathrm{ex}}(0)]$. It follows from the general formula for the *T*-matrix[14] that the main contribution to the forward scattering operator for fast-moving atoms is due to the Born component, since in this component oscillations due to the factor exp(ikx) describing the incident wave completely cancel ($k=p/\hbar$ is the atom wave vector). At room temperature (and in fact even at lower temperatures) the condition $|k|a_0\gg1$ is easily satisfied for the electron spin exchange (a_0 is the effective interaction range). The latter inequality is the hallmark of fast atomic motion. Thus, the characteristic eigenvalue ν_{ex} of operator L_I is proportional to the Born forward scattering (a_0 is the effective interaction range):

$$A(0)=-\frac{m}{4\pi\hbar^2}\int U\,d^3x\simeq-\frac{m}{4\pi\hbar^2}|U|a_0^3,$$

where U is the exchange interaction potential.

Now we can easily estimate the ratio ν_{ex}/ν_s:

$$\frac{\nu_{\mathrm{ex}}}{\nu_s}\sim|M|\frac{|U|a_0}{\hbar\overline{v}}. \tag{17}$$

Note that the right-hand side of this relation is the so-called ''Born parameter''[8] multiplied by $|M|$, which is usually large in classical (nonquantum) gases. The Born parameter for cesium atoms will be estimated below for illustration.

Thus, in the first approximation with respect to the (reciprocal) Born parameter, Eq. (10) can be expressed as

$$(kv-\omega)\mu_\pm(p)=\mp16|M|\pi^3\hbar^2\int dp_1f^{(0)}(p_1)\times\mathrm{Re}\,T_{\mathrm{ex}}(0)[\mu_\pm(p)-\mu_\pm(p_1)]. \tag{18}$$

This equation can again be treated using the perturbation theory, as was done previously. In this case, however, we will use another estimation method. Suppose for simplicity that $T_{\mathrm{ex}}=$const, i.e., it is constant with energy. Then

$$(kv-\omega)\mu_\pm(p)=\mp\nu_{\mathrm{ex}}\left[\mu_\pm(p)-\frac{1}{n}\times\int dp_1f^{(0)}(p_1)\mu_\pm(p_1)\right]$$

(this time $\nu_{\mathrm{ex}}=16|M|\pi^3\hbar^2n\mathrm{Re}T_{\mathrm{ex}}$).

In the hydrodynamic approximation ($|k\overline{v}/\nu_{\text{ex}}|\ll 1$, $|\omega/\nu_{\text{ex}}|\ll 1$), we have the dispersion relation

$$\omega=\pm\frac{k^2\overline{v^2}}{3\nu_{\text{ex}}}. \tag{19}$$

This equation describes a magnetization wave of constant amplitude. The diffusion damping of this wave characterized by parameter Γ is the effect of the next order in the reciprocal Born parameter ν_s/ν_{ex}. The damping parameter Γ is estimated by the expression

$$\Gamma\sim\frac{k^2\overline{v^2}}{\nu_{\text{ex}}}\frac{\nu_s}{\nu_{\text{ex}}}\ll\omega. \tag{20}$$

Thus, weakly damped spin waves whose frequency and damping parameter are estimated by Eqs. (19) and (20) can propagate in a spin-polarized paramagnetic gas. Note that the diffusion damping factor Γ of the spin wave in a polarized gas given by Eq. (20) is a factor of $(\nu_s/\nu_{\text{ex}})^2$ smaller than in an unpolarized gas [Eq. (16)].

By way of example, consider the parameters of spin waves in a vapor of spin-polarized cesium atoms, whose parameters have been measured experimentally. First, the electron spin magnetization in cesium vapor is conserved to high accuracy:[15] the spin nonconserving collision cross section is about 1% of the spin-conserving cross section. The effective interaction range is $a_0\sim 10^{-7}$ cm.[15,17] Let us estimate the amplitude $|U|$ by assuming that it is twice the binding energy of the Cs_2 molecule: $|U|\sim 1$ eV.[18] Then the Born parameter at the thermal velocity $\overline{v}=2\times 10^4$ cm/s of cesium atoms is of order 10^3. Such a large value of the Born parameter is closely related to the clear-cut anisotropy of the scattering amplitude angular dependence for fast-moving particles.[8] Note that the temperature dependence of this parameter is largely controlled by the dependence $\overline{v}\propto\sqrt{T}$. Therefore the Born parameter is close to unity only at temperatures below 10^{-4}–10^{-3} K.

The ratio between the spin wave frequency and its damping parameter, according to Eqs. (19) and (20), is

$$\omega/\Gamma\sim 10^3|M|.$$

Hence it follows, in particular, that the resonance line should narrow as $|M|$ increases. Note that this estimate, generally speaking, may need revision with a view to taking into account actual spin nonconservation.

Thus, propagation of spin waves in a polarized Boltzmann gas is determined by two factors: spin conservation or, to be exact, its slow decay due to collisions, and the high anisotropy of the scattering amplitude at room temperature. It is known that in this case atoms mostly undergo small-angle scattering in the angular range $(|k|a_0)^{-1}$, where $|k|=m\overline{v}/\hbar$. For cesium, the corresponding angular range at the room temperature is $\simeq 10^{-3}$. Note that, in accordance with the previously adopted criterion $\Lambda/r_0\gg 1$ (see above), lightly-damped spin waves should not propagate in cesium vapor at room temperature, since $\Lambda/r_0\sim 10^{-3}$.

In addition to magnetic resonance techniques,[19] it is also possible to use Rayleigh light scattering near the resonant D lines in the atomic spectra of alkali metals in detecting spin waves. In the presence of spin waves, the spectrum of fluctuations in the electron magnetization should have the form of a well-resolved doublet with narrow spectral components.[20]

5. DYNAMICS OF LONGITUDINAL SPIN POLARIZATION

In contrast to the transverse spin polarization $\mu_\pm$, the longitudinal polarization μ_z in a polarized gas cannot be treated independently and is related to the scalar component of the Wigner function φ (gas density). The collision integral in this case is determined by the first four terms in the general collision integral (8). The parameters invariant with respect to collisions in this case include the longitudinal spin component in addition to the usual five parameters, namely, the number of particles, three momentum components, and energy. These invariants correspond to six eigenfunctions of the collision integral with vanishing eigenvalues. The corresponding functional space consists of paired-off coefficients of the unit matrix and the Pauli z-matrix. A calculation of correct zeroth-order functions (diagonalization of the operator $ik\hat{v}$) is described in the Appendix.

Among the six modes listed above, the only propagating mode (first-order perturbation theory in the operator $ik\hat{v}$) is the acoustic mode propagating at the sound velocity $u_s=\sqrt{5T/3m}$ independent of gas polarization. Note that the polarization-dependent correction to the sound velocity turns up only in third-order perturbation theory in $ik\hat{v}$. Here we have in mind a classical gas with no virial corrections.[5] The eigenfunctions of transverse (shear) modes are independent of the gas polarization degree, whereas the others, namely, those of the acoustic, thermal conductivity, and spin-wave modes, depend on the gas polarization degree.

At this point, our aim is not to calculate accurate second-order corrections, which determine damping of these modes, since this would require specific data concerning interaction between particles. We merely note some features of second-order corrections in $ik\hat{v}$. In contrast to the case of transverse modes, here all these corrections γ are of the same order of magnitude. At small M we have $\gamma\sim\gamma_0+\gamma_1 M^2$. Here γ_0 is the damping in unpolarized gas, and both parameters, γ_0 and γ_1, are of the order of the gas-kinetic collision frequency. An important point is that this situation is significantly different from the case of a transverse spin-wave mode, for which the damping in polarized gas is considerably weaker than in an unpolarized gas [see Eqs. (17)–(20)]. This circumstance makes propagation of a weakly damped spin wave possible.

This work was supported by the Russian Fund for Fundamental Research (Grant 96-02-17312a).

APPENDIX

Here results of diagonalization of operator $ik\hat{v}$ in the basis of zero eigenfunctions of the collision operator (8) are given. This analysis ignores off-diagonal (transverse) elements of the Wigner matrix, since we can independently analyze the dynamics of the transverse and longitudinal spin components.

The collision operator is Hermitian in the scalar product:

$$\langle [a(p),b(p)] | [c(p),d(p)] \rangle = (2\pi mT)^{-3/2} \int (a(p)c(p)+b(p)d(p)) \times \exp\left(-\frac{p^2}{2mT}\right) d^3p.$$

Here the initial linear space is composed of pairs $[a(p),b(p)]$ of diagonal elements of the Wigner matrix. These are coefficients of the unit matrix and the Pauli z-matrix. The weighting factor is the normalized equilibrium momentum distribution function.

The initial, zeroth-order subspace consists of six orthonormal pairs of functions corresponding to the following conservation laws:

$$\left.\begin{array}{l} f_1=[1,0]\ \text{(number of particles)},\\ f_2=[p_x/\sqrt{mT},0],\\ f_3=[p_y/\sqrt{mT},0],\\ f_4=[p_z,Mp_z]/\sqrt{(1+M^2)mT} \end{array}\right\}\ \text{(momentum)},$$

$$f_5=[p^2,Mp^2]/mT\sqrt{3(1+M^2)}\ \text{(energy)},$$

$$f_6=[0,1]\ (z-\text{component of the spin}).$$

Diagonalization of the perturbation matrix $ik\hat{v}=ik\hat{p}/m$ in this basis leads to the following correct functions in zeroth order:

$$f_s^{(+)}=\frac{1}{\sqrt{2(1+M^2)}}\left[\frac{p_z}{\sqrt{mT}} + \frac{1}{\sqrt{15}}\frac{p_x^2+p_y^2+p_z^2}{mT}, \frac{M}{\sqrt{15}}\frac{p_x^2+p_y^2+p_z^2+p_z\sqrt{15mT}}{mT}\right],$$

$$f_s^{(-)}=\frac{1}{\sqrt{2(1+M^2)}}\left[-\frac{p_z}{\sqrt{mT}} + \frac{1}{\sqrt{15}}\frac{p_x^2+p_y^2+p_z^2}{mT}, \frac{M}{\sqrt{15}}\frac{p_x^2+p_y^2+p_z^2-p_z\sqrt{15mT}}{mT}\right],$$

$$f_v^{(x)}=[p_x/\sqrt{mT},0],$$

$$f_v^{(y)}=[p_y/\sqrt{mT},0],$$

$$f_t=\frac{1}{mT\sqrt{2(2M^2+5)(1+M^2)}}[p_z+p_x^2+p_y^2+p_z^2-5mT -2M^2mT, M(p_z+p_x^2+p_y^2+p_z^2-3mT)],$$

$$f_m=[-M/\sqrt{(1+M^2)},1].$$

The first two modes, $f_s^{(+)}$ and $f_s^{(-)}$, describe sound propagation in a monoatomic gas. Modes $f_x^{(-)}$ and $f_y^{(-)}$ describe viscous damping of transverse velocity and are independent of the gas polarization. Mode f_t describes heat transfer in polarized gas and, generally speaking, depends on the gas polarization degree. Finally, f_m describes the transverse magnetic polarization.

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

---

[1] V. P. Silin, Zh. Éksp. Teor. Fiz. **33**, 945 (1957) [Sov. Phys. JETP **6**, 726 (1958)].

[2] A. G. Aronov, Zh. Éksp. Teor. Fiz. **73**, 577 (1977) [Sov. Phys. JETP **46**, 301 (1977)].

[3] E. P. Bashkin, JETP Lett. **33**, 8 (1981).

[4] A. E. Meyerovich, S. Stepaniants, and F. Laloë, Phys. Rev. B **52**, 6808 (1995).

[5] E. P. Bashkin, Usp. Fiz. Nauk **148**, 433 (1986) [Sov. Phys. Usp. **29**, 238 (1986)].

[6] B. R. Johnson, J. S. Den'ker, N. Bigelow *et al.*, Phys. Rev. Lett. **52**, 1508; **53**, 302 (1984).

[7] N. Bigelow, P. J. Nacher, and M. Leduc, J. de Phys. II **2**, 2159 (1992).

[8] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Nonrelativistic Theory)*, Pergamon Press New York (1977).

[9] R. J. Knize, Phys. Rev. A **40**, 6219 (1989).

[10] R. F. Snider, J. Chem. Phys. **32**, 1051 (1960).

[11] T. L. Andreeva and P. L. Rubin, Zh. Éksp. Teor. Fiz. **111**, 831 (1997) [JETP **84**, 457 (1977)].

[12] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics*, Pergamon Press, New York (1981).

[13] A. E. Meyerovich, J. H. Naish, J. R. Owers-Braley, and S. Stepaniants, Fiz. Nizk. Temp **23**, 553 (1997) [Low Temp. Phys. **23**, 411 (1997)].

[14] M. L. Goldberger and K. M. Watson, *Collision Theory*, Wiley, New York (1964).

[15] N. D. Bhaskar, J. Pietras, J. Camparo *et al.* Phys. Rev. Lett. **44**, 930 (1980).

[16] R. Balescu, *Equilibrium and Nonequlibrium Statistical Mechanics*, Wiley, New York (1975); O. Resibois and M. de Leener, *Classical Kinetic Theory of Fluids*, Wiley, New York (1977).

[17] A. A. Radtsig and B. M. Smirnov, *Handbook of Atomic and Molecular Physics* [in Russian], Atomizdat, Moscow (1980).

[18] W. Happer, Rev. Mod. Phys. **44**, 169 (1972).

[19] L.-J. Wey, N. Kalenchofsky, and D. Candela, Phys. Rev. Lett. **71**, 879 (1993).

[20] T. L. Andreeva, P. L. Rubin, and E. A. Yukov, Zh. Éksp. Teor. Fiz. **107**, 1160 (1995) [JETP **80**, 645 (1995)].

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Structure of a microparticle crystal in a radio-frequency discharge plasma

V. A. Schweigert and V. M. Bedanov

Institute of Theoretical and Applied Mechanics, Siberian Branch, Russian Academy of Sciences, 630090 Novosibirsk, Russia

I. V. Schweigert*)

Institute of Semiconductor Physics, Siberian Branch, Russian Academy of Sciences, 630090 Novosibirsk, Russia

A. Melzer, A. Homann, and A. Piel

Institut für Experimentalphysik, Christian–Albrechts–Universität Kiel, 24098 Kiel, Germany

(Submitted 22 January 1998)
Zh. Éksp. Teor. Fiz. **115**, 877–893 (March 1999)

By employing the particle-in-cell method we study the distributions of the electric field strength and of the electron and ion concentrations in the microparticle crystal in the electrode sheath in a radio-frequency discharge in helium. The coordinates and charges of the microparticles are found from the balance condition for the forces acting on the particles and the balance of electron and ion fluxes to the particles. With periodic boundary conditions introduced, we investigate the three-dimensional problem for the unit cell of the microparticle crystal. We examine the dependence on gas pressure and discharge voltage of the main crystal parameters: the critical particle separation (at which a phase transition from a monolayer crystal to a double-layer crystal occurs), the particle potentials, and the distances between the layers in the double-layer crystal. We obtain the critical values of the friction coefficient for the particles in the gas, i.e., values below which the crystal becomes unstable against the development of particle oscillations in the transverse direction, and compare the experimental data on crystal structure and stability with the theoretical results. Finally, we set up an approximate model that makes it possible to calculate the main parameters of the microparticle crystal. © *1999 American Institute of Physics.* [S1063-7761(99)00803-3]

1. INTRODUCTION

A collection of charged microparticles in a plasma constitutes a unique macroscopic object for the physics of nonideal plasmas, an object that makes possible an experimental observation of the crystal–liquid–gas transition in Coulomb systems. The possibility of forming a microparticle crystal in a plasma was predicted by Ikezi.[1] Such crystals were detected in magnetron-discharge experiments,[2–4] in a radio-frequency discharge between plane-parallel electrodes,[5–8] and in standing striations of a stationary glow discharge.[9] A microparticle crystal in a radio-frequency discharge consists of several layers, each of which contains about ten thousand particles occupying the sheath of the lower electrode, where the force of gravity acting on the particles is balanced by the electric field. In the transverse direction (in the plane of the electrode), the crystal is usually limited by the edge of the electrode or by protrusions specially manufactured on the electrodes. Longitudinal confinement of mutually repulsive negatively charged particles is achieved by the positive space charge in the layer.

The results of experimental observations of the structure of a microparticle crystal cannot be described by the classical theory of Coulomb systems. According to the calculations done in Refs. 10–12, the structure of multilayer Coulomb crystals corresponds to different types of close packing (depending on the distance between the layers), where the particles in adjacent levels are shifted in relation to each other. The experiments reported in Refs. 4,5,7,8, and 13 reveal another type of packing, where the particles in different layers are one above another, while in a plane the particles form a regular hexagonal lattice. Note that sometimes a close-packed crystal structure is also observed (face-centered[3] and body-centered[3,13,14]). According to the results of the experiments conducted by Hayashi and Tachibana,[13] the transition from a bcc lattice to a loosely packed hexagonal lattice occurs as the particle radius increases to $R \sim 1\ \mu$m. On the basis of studies done with the collisionless model of ion motion, Melandsø and Goree[15] and Vladimirov *et al.*[16] suggested that the reason for the observed loose packing lies in the formation, due to the focusing of the ion paths by the particle field, of regions of enhanced ion number density behind particles. However, in typical experimental conditions, the ion mean free path is much shorter than the particle separation, with the result that the collisionless model breaks down. Depending on the mean free path of the ions, regions of enhanced or reduced ion concentration form behind a particle.[17,18] Monte Carlo calculations of the ion motion with allowance for scattering in the double-layer crystal of charged microparticles in the electrode sheath[17,18] have made it possible to find the forces acting on the particles and corroborated the assumption of Melandsø and Goree[15] and Vladimirov *et al.*[16] for the experimental conditions described in Refs. 7,8, and 19. However, mutual repulsion of the ions, which impedes the formation of ion clouds behind the particles, was not taken into account in Refs. 17 and 18. Thus, in Refs. 15–18 an explanation was given of the loose packing of the microparticle crystal, but no quantitative analysis was done of the crystal's structure, an analysis that would make it possible to determine the particle charges, the distance between the layers, and other lattice parameters.

We know of no self-consistent calculations of the microparticle charges in the electrode sheath, charges measured in the experiments described in Refs. 7,8, and 19. Theoretical studies of the charging of particles[20–22] and of the interaction between the particles[23] have been done only for quasineutral plasma conditions. In Ref. 24 the charging of a single particle by ions in the electrode sheath of an rf discharge was examined by a non-self-consistent approach. In their self-consistent calculations of the structure of a microparticle crystal in a layer, Melandsø and Goree[25] used the collisionless hydrodynamic description of ion motion and modeled the particles by infinitely long cylinders. A one-dimensional crystal in which the microparticles are replaced by infinite charged planes was examined by Otani and Bhattacharjee,[26] who used the kinetic approach for the ions. A more realistic case of spherical particles was studied by Melandsø.[27] However, the error introduced by the use of the hydrodynamic approach and the linear response approximation for calculating the particle interaction forces in Ref. 27 is unknown. Note that the researchers whose papers we have just cited use models in their statement of the problem and do not compare the results of crystal structure calculations with the known experimental data. Thus, a quantitative theory of the microparticle crystal in a plasma has yet to be developed. The present paper describes the results of a self-consistent modeling of the structure of a microparticle crystal in the electrode sheath for an rf discharge in helium in conditions coinciding with those of the experiments described in Refs. 7,8, and 19.

2. MODEL OF A MICROPARTICLE CRYSTAL

In the experiments discussed in Refs. 7,8, and 19, microparticles of radius $R=4.7\,\mu$m and density $\varrho=1.5\,\Gamma/\mathrm{cm}^{-3}$ levitate in helium. The characteristic microparticle separations amount to hundreds of micrometers. The discharge frequency f_{rf} is 13.56 MHz, the electrode separation is 6 cm, the gas pressure P is varied from 40 to 120 Pa, and the amplitude of the radio-frequency component of the voltage across the electrodes, U_{rf}, is varied from 40 to 80 V. Below we show that the presence of microparticles has a small effect on the discharge characteristics (the distributions of the electron temperature in the electrode sheath, of the discharge current density, and of the ion flux to an electrode), which are needed in setting up a self-consistent model of a microparticle crystal. Hence first a one-dimensional rf discharge in helium without microparticles was modeled by the particle-in-cell method,[28] with the set of electron scattering cross sections taken from Ref. 29 and with allowance for ionization in the collision of two metastable helium atoms.[30] As for ions, we allowed only for their scattering in the gas consisting of these ions, i.e., for resonant charge exchange, with a constant cross section of $3.5\times10^{-15}\,\mathrm{cm}^2$ (Ref. 31). Without going into the details of the analysis of an rf discharge, we examine only the effect of variations of gas pressure and discharge voltage on the characteristics of interest. For the pressure and discharge-voltage ranges involved ($P=40-120\,\mathrm{Pa}$ and $U_{\mathrm{rf}}=40-80\,\mathrm{V}$), the discharge current density $j=(0.5-1.3)\times10^{-3}\,\mathrm{A\,cm}^{-2}$ and the ion flux to the cathode $j_i=(0.5-2.2)\times10^{14}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ are weakly dependent on the gas pressure and increase with the discharge voltage. The average electron energy increases only slightly (by 15–20%) as we move from the electrode to the boundary of the electrode sheath and is somewhat higher (by $\sim 20\%$) than the average electron energy U_e at the discharge center, a value that is almost independent of the discharge voltage and decreases almost linearly with increasing pressure: $U_e=4.5-3$ eV for $P=40-120$ Pa, which agrees with the experimental data listed in Refs. 32 and 33.

According to the measurements carried out by Godyak *et al.*,[32] which are corroborated by the results of our calculations, the electron distribution in the volume of the rf discharge in helium is close to the Maxwellian. We are interested in the behavior of the electron distribution function in electrode sheaths and, in particular, the rate at which the microparticles are charged. The characteristic mean free path of the electrons is much longer than the microparticle radius, which means that the charging of the particles can be studied in the orbital motion approximation.[34] The electron distribution over the velocity directions is close to isotropic, and the electron flux density j_e at the microparticle surface whose potential is U is given by the formula[35]

$$j_e=\frac{n_e}{4}\int_U^\infty\left(1-\frac{U}{\epsilon}\right)v f_e(\epsilon)\,d\epsilon, \tag{1}$$

where the electron energy distribution function $f_e(\epsilon)$ is normalized as $\int_0^\infty f_e(\epsilon)\,d\epsilon=1$, and n_e, $v=\sqrt{2\epsilon/m_e}$, and m_e are the electron number density, velocity, and mass. With the Maxwellian electron distribution function $f_e(\epsilon)=2\sqrt{\epsilon/\pi T_e}\exp(-\epsilon/T_e)\,T_e^{-1}$ we have

$$j_e=\frac{n_e v_t}{4}\exp(-U/T_e),\quad v_t=\sqrt{\frac{8T_e}{\pi m_e}}, \tag{2}$$

where $T_e=2U_e/3$ is the electron temperature, with U_e the average electron energy. A comparison of the results of calculations of the electron flux density to the particle surface in the electrode sheath with allowance for Eqs. (1) and (2) is illustrated by Fig. 1 for different particle potentials. Since the average electron energy changes little in the electrode sheath, the rapid decrease of the charging rate in the direction to the electrode is due primarily to the decrease in the electron number density.

Crystals used in experiments are always finite, but each layer contains a very large number, $\sim 100\times100$, of unit cells. Hence we consider an ideal crystalline structure, for which all the characteristics of the system are strictly periodic, $f(\boldsymbol{\rho}+\mathbf{l})=f(\boldsymbol{\rho})$, in the electrode plane $\boldsymbol{\rho}=xy$, where

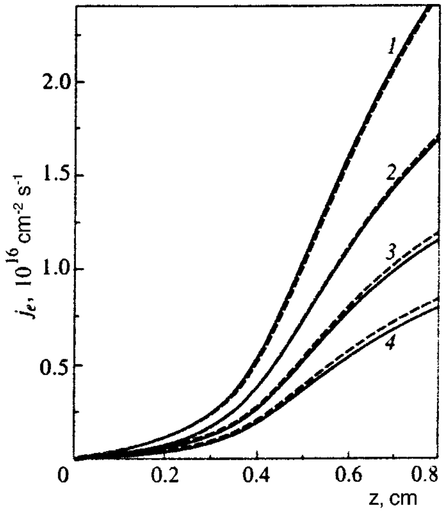

FIG. 1. Period average of the electron flux to a particle in the electrode sheath for various particle surface potentials *U*. The gas pressure $P = 80$ Pa, and the discharge voltage $U_{\rm rf} = 60$ V. The solid curves represent the results of calculations that use an electron distribution function obtained by the Monte Carlo method, and the dashed curves represent the results of calculations that use a Maxwellian electron distribution function. Curve *1*, $U = 2$ V; curve *2*, $U = 3$ V; curve *3*, $U = 4$ V; and curve *4*, $U = 5$ V.

$\mathbf{l} = l_1\mathbf{a}_1 + l_2\mathbf{a}_2$, with $\mathbf{a}_1$ and $\mathbf{a}_2$ the two-dimensional basis vectors of the lattice. Then the calculation can be restricted to one unit cell in the *xy* plane. The characteristic time in the course of which the rf discharge becomes time-periodic is determined by the ambipolar diffusion of the ions through the discharge gap and is much longer than the time it takes the electrode sheath to form (the latter time is determined by the time of flight of an ion through the sheath). Hence, to speed up calculations we consider only the electrode sheath that contains the microparticle crystal. The lower limit of the calculation interval corresponds to the lower electrode, and the upper is in the quasineutral plasma near the boundary of the electrode sheath. Since the electron flux to the particles in much smaller than the electron flux to an electrode, the presence of microparticles affects the electron distribution in the sheath only weakly, and this distribution may be assumed Maxwellian (see above). Then the electron number density n_e has a Boltzmann distribution:

$$n_e = n_0 \exp(e\phi/T_e), \tag{3}$$

where ϕ is the electric field potential, n_0 is the density of the quasineutral plasma at the upper limit $z = d$ of the calculation interval, with $\phi(z = d) = 0$, and the spatial distribution of the electron temperature, $T_e(z)$, is taken from the results of rf-discharge calculations.

To calculate the potential distribution, we used the Fourier transform

$$\phi = \phi_0(z) + \phi_1(z,\boldsymbol{\rho}) = \phi_0 + \sum_{\mathbf{g},|\mathbf{g}|\neq 0} \phi_{\mathbf{g}}(z)\exp(i\mathbf{g}\cdot\boldsymbol{\rho}), \tag{4}$$

where $\mathbf{g} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2$, with $\mathbf{b}_1$ and $\mathbf{b}_2$ the basis vectors of the reciprocal lattice. The part of the potential that depends on the transverse coordinates, ϕ_1, is much smaller than the electron temperature everywhere except a small region near the particle with a size of several particle radii. Then $n_e(z,\boldsymbol{\rho}) \approx n_e(z)(1 + e\phi_1/T_e)$, where $n_e(z) = n_0 \times \exp(e\phi_0/T_e)$. The Fourier transforms $\phi_{\mathbf{g}}$ can be found by solving Poisson's equation

$$\frac{\partial^2 \phi_0}{\partial z^2} = 4\pi e(n_{d,\mathbf{g}=0} + n_e - n_{i,\mathbf{g}=0}), \tag{5}$$

$$\frac{\partial^2 \phi_{\mathbf{g}}}{\partial z^2} - |\mathbf{g}|^2\phi_{\mathbf{g}} - \frac{4\pi e^2 n_e}{T_e}\phi_{\mathbf{g}} = 4\pi e(n_{d,\mathbf{g}} - n_{i,\mathbf{g}}), \tag{6}$$

where $n_{d,\mathbf{g}}$ and $n_{i,\mathbf{g}}$ are the Fourier transforms of the microparticle charge density and the ion number density (the latter was found by modeling the ion motion by the Monte Carlo method).

In the entire range of the discharge parameters we have $|\mathbf{g}|^2 \gg 4\pi e^2 n_e/T_e$, and the Debye screening of the particle charge by the electrons is insignificant. At gas pressures $P = 40-120$ Pa, the mean free path of ions in helium, $\lambda = 100-300\,\mu$m, is much shorter than the thickness of the electrode sheath, and the characteristic time of flight of ions through the sheath is much longer than the discharge period. On the whole, ions are produced in the quasineutral plasma, and the ion flux density j_i in the electrode sheath can be assumed constant. When we solved the transport equation for the ions, we assumed that the velocity distribution function for the ions is fixed at the upper limit of the calculation interval and that the electrode and microparticles are ideal absorbers. To solve Eq. (5) with an electron number density (3), one must impose a boundary condition for the potential at the electrode surface and determine the parameter n_0 in (3). The electron number density at the upper limit of the calculation interval, n_0, was fixed by the quasineutrality condition $n_0 = n_{i,z=d}$. The electric field strength at the electrode, $E_k = -\partial\phi_0/\partial z$, was found from the law of total current conservation:

$$\frac{1}{4\pi}\frac{\partial E_k}{\partial t} + e\left(j_i - \frac{1}{4}v_t n_0 \exp\frac{e\phi_0(z=0,t)}{T_e}\right) = j\sin\omega t, \tag{7}$$

where the amplitude j of the discharge current density was taken from calculations of an rf discharge without particles, and $\omega = 2\pi f_{\rm rf}$. The number of ions in the particle-in-cell method was chosen between 32 000 and 64 000. A decrease in the statistical calculation error depicted in some of the figures was achieved by averaging over hundreds of rf-discharge periods. In calculating the three-dimensional electric field potential we used the fast Fourier transform.[28] The model of the electrode sheath in an rf discharge described in the present paper is in good agreement with the results of self-consistent calculations of rf discharges by the particle-in-cell method.

The charge Z and the potential $U = eZ/R$ of the particle surface were found from the balance of the electron and ion fluxes to the particle, $4\pi R^2 j_e = J_i$, where the electron current density j_e is given by (2). The ion flux to the particle,

$$J_i = \int n_i(\mathbf{v}_i\cdot d\boldsymbol{\sigma}_i), \tag{8}$$

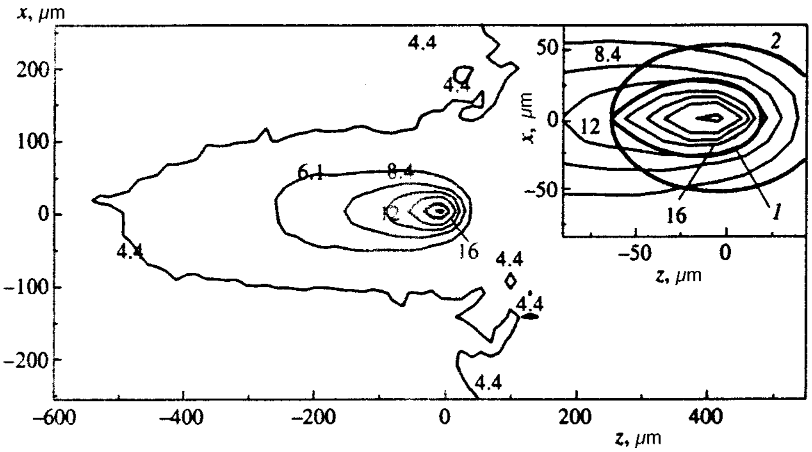

FIG. 2. Ion concentration distributions, measured in units of 10^8 cm^{-3}, in the *xz* crosssection for the electrode sheath with a monolayer microparticle crystal. The coordinates are measured from the point $(x=0,\ z=0)$ where a particle is situated. The gas pressure $P=80$ Pa, the discharge voltage $U_{\mathrm{rf}}=60$ V, and the particle separation $a=519\,\mu$m. The heavy curves in the inset indicate the following: curve *1*, the boundary of the potential well surrounding the particle (within these boundaries ions are in finite motion); and curve *2*, the boundary of the region from which zero-energy ions land on the particle.

and the force

$$\mathbf{F}=\int m_i\mathbf{v}_i n_i(\mathbf{v}_i\cdot d\boldsymbol{\sigma}_i) \tag{9}$$

acting on the particle as a result of ion momentum transfer were found by modeling the ion motion. Here integration is carried out over the particle surface, **n** is a unit vector normal to the surface, and $\mathbf{v}_i$ and m_i are the ion velocity and mass. The particle is also subjected to the force of gravity $\mathbf{F}_g=(0,0,-Mg)$ and the electric-field force $\mathbf{F}_k=eZ\,\partial\phi/\partial\mathbf{r}$, where $g=980\,\mathrm{cm\,s^{-2}}$ is the acceleration of free fall, and $M=6.73\times10^{-10}$ g is the particle mass.

3. RESULTS AND DISCUSSION

3.1. Charging of microparticles by ions

The role that the friction force (9) plays in the balance of forces is insignificant for micrometer-sized particles. The levitation of particles in a layer is determined primarily by the equilibrium between the force of gravity and the force exerted by the electric field, which makes it possible to estimate the strength E_p of the field acting on a particle at Mg/eZ. For a typical value $U\approx3$ V of the potential at the surface of a particle of radius $R\approx5\,\mu$m, the characteristic field strength E_p is approximately $50\,\mathrm{V\,cm^{-1}}$. The positive ionic space charge concentrated near the particle is small compared to the particle charge, and the distribution of the potential energy U_i of the ions is

$$U_i=-\frac{e^2Z}{r}-eE_pr\cos\theta, \tag{10}$$

where r is the distance to the particle, and θ is the angle between the radius vector and the z axis. For a potential energy given by (10) there exists a region of finite motion, $r<R_f(\theta)$, and a region of infinite motion, $r>R_f(\theta)$, where R_f is the boundary of the first region,

$$R_f=R_w1-\sqrt{1-\cos\theta}/\cos\theta,\quad R_w=\sqrt{eZ/E_p},$$

depicted in Fig. 2 by the curve *1*. In conditions of the experiment described in Ref. 19, the characteristic size of the potential well for ions, $R_w\approx50\,\mu$m, is much smaller than the mean free path of the ions, $\lambda=100-300\,\mu$m, which allows making analytical estimates of the rate at which the ions charge the particles.

The total ion flux to the particle, $J_i=J_{i,1}+J_{i,2}$ consists of the flux of high-energy ions with an ordered-motion characteristic energy $\epsilon_i=eE_p\lambda=(0.5-1.5)$ eV and the flux of ions involved in a charge exchange near the particle. Since $\epsilon_i\gg eE_pR_w, T_g$ and $\lambda\gg R_w$, in estimating the flux of high-energy ions, $J_{i,1}$, we can use the collisionless approximation, ignore the effect of field E_p near the particle, and employ the following expression for the distribution function of the ions impinging on the particle:[36]

$$F_i(\mathbf{v},v_z\geqslant0)=j_i\frac{m_i}{e\lambda E_p}\,\delta(v_x)\,\delta(v_y)\exp\left(-\frac{m_iv_z^2}{e\lambda E_p}\right),$$

$$F_i(\mathbf{v},v_z<0)=0, \tag{11}$$

which is valid for a strong uniform field. According to classical mechanics,[37] the particle's capture cross section for an ion with energy $\epsilon=m_iv_z^2/2$ is $\sigma_c=\pi R_p^2(1+U/\epsilon)$, and with allowance for (11) the high-energy ion flux can be written

$$J_{i,1}=\int\sigma_cv_zF_i(\mathbf{v})\,d\mathbf{v}=\pi R^2j_i\left(1+\frac{U}{\epsilon_i}\int\frac{e^{-x}\,dx}{x}\right)$$

$$\approx\pi R^2j_i\left(1+\frac{U}{\epsilon_i}\ln\Lambda\right), \tag{12}$$

where $\Lambda=\epsilon_iR_w^2/UR^2\gg1$. Here integration with respect to the impact parameter was truncated at the characteristic size of the potential well, and the terms small compared to $\ln\Lambda$ were dropped.

Near a particle, the kinetic energy of the ions after charge exchange is much lower than the potential energy $U_i(R_w)\sim eUR/R_w\approx0.3$ eV. Hence in analyzing the flux $J_{i,2}$ of the ions to the particle that have undergone resonant scattering near the particle we can use the cold-ion approximation. We must determine the volume V_w of the region surrounding the particle from which zero-energy ions pass to the particle surface. Note that this region is larger than the potential well since some of the ions whose motion is infinite also go over to the particle. Dimensional analysis suggests that $V_w=C_wR_w^3$, with the value of the coefficient C_w depending only on R_w/R. We found this coefficient numerically. To this end, for certain fixed initial coordinates of the ion, we solved the equations of motion of an ion in the potential (10) with a zero initial ion velocity. If at a certain moment in time the ion lands on the particle, the initial coordinates belong to the region involved, whose cross section is depicted in Fig. 2 (the outline of the cross-sectional area is curve *2*). For typical experimental conditions ($R_w/R\sim 10$), the value $C_w\approx 3.3$ is weakly dependent on R_w/R. A variation in R_w/R by a factor of two leads only to a 10% variation in C_w. Since $\lambda \gg R_w$ and the probability of ions being scattered near the particle is low,

$$J_{i,2}\approx j_i\frac{V_w}{\lambda}\approx 3.3 j_i\frac{R_w^3}{\lambda}. \tag{13}$$

Since $R_w \gg \lambda R^2$, we have $J_{i,2}\gg J_{i,1}$, and the rate at which the particle is charged is determined primarily by cold ions. Knowing the velocity $v_{is}\approx\sqrt{2U/m}$ of initially cold ions at the particle surface, we can estimate the number density of these ions at the particle surface:

$$n_{i\,s}\approx\frac{J_{i,2}}{4\pi R^2 v_{is}}\approx C_w n_i\frac{R_w^2}{4\pi R^2}\left(\frac{R}{\pi\lambda}\right)^{1/2}, \tag{14}$$

where $n_i=j_i/v_d$ is the average ion number density in the layer; here we have used the expression $v_d=\sqrt{2e\lambda E_p/\pi m_i}$ for the drift velocity of ions in a strong field.[31] For a pressure of $P=80\,\text{Pa}$, the estimate (14) yields $n_{is}/n_i\sim 3$, which is roughly equal to the results of numerical calculations (Fig. 2). According to the estimate (14) and the results of numerical calculations, the total ionic charge concentrated in the potential well around the particle is much smaller than the charge of the particle proper, which justifies the use of the expression (10) for the ion potential energy.

To estimate the particle charge we must determine the electron concentration distribution in the electrode sheath. With the rf discharge considered here, the following conditions are met: $j\gg j_i$, $U_k\gg T_e$, and $\omega\gg d_k/v_d$, where U_k and d_k are the voltage drop across the electrode sheath and the sheath thickness. Then the standard model of the electrode sheath[38] and the linear-profile approximation of the period-averaged field strength $\overline{E}=\overline{E}_e(1-z/d_k)$ (Ref. 35) make it possible to obtain analytic expressions for the electron and ion concentration distributions in strong fields, where $v_d \propto\sqrt{E}$:

$$n_i=\frac{n_{ik}}{\sqrt{1-z/d_k}},\quad n_e=n_i\frac{\varphi}{\pi},$$
$$\cos\varphi=1-\frac{2E_e(1-\sqrt{1-z/d_k})-E_m}{E_e-E_m}. \tag{15}$$

Here $E_e=E_m+4\pi j/\omega$ and $n_{ik}=j_i/v_d(E_e)$ are the field strength and ion concentration at the electrode; $d_k=E_e/4\pi e n_{ik}$; $E_m\approx(8\pi e n_{ik}T_e\ln(v_e/2\sqrt{2\pi}v_d))^{1/2}$ is the minimum field strength at the electrode, which can be found from the balance of electron and ion fluxes,[39] with $v_e=\sqrt{2T/2\pi m_e}$ the thermal velocity of the electrons; and φ/π is the fraction of the rf-field period in which the electron number density in the electrode sheath is equal to the ion number density. Equations (2), (12), (13), and (15) make it possible to find the positions of particles and their charges in a monolayer crystal after a solution of a system of ordinary equations has been found. The results of numerical calculations of particle charges are in good agreement with the approximate model discused earlier (Fig. 3a), in which the electron temperature in the electrode sheath is assumed constant. The particle charge is relatively weakly dependent on the discharge voltage and increases with decreasing gas pressure, i.e., it behaves in the same way as the average electron energy. The difference between the experimental data and the results of numerical calculations is about 30% and lies within experimental error.

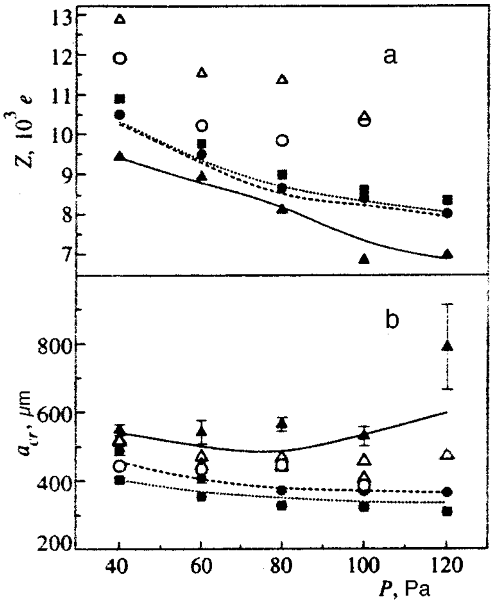

FIG. 3. Gas-pressure dependence of the particle charge (a) and the critical particle separation (b) at which the monolayer crystal transforms into a double-layer crystal, for various discharge voltages: $U_{\text{rf}}=40$ V(▲), 60 V(●), and 80 V(■). The curves represent the results obtained by using an approximate crystal model, the black symbols represent the results of self-consistent calculations, and the open symbols represent the experimental data obtained at $U_{\text{rf}}=40$ V(△) and ≈ 50 V(○).

3.2. Transition from a monolayer crystal to a double-layer crystal

As the particle number density increases, the formation of a double-layer crystal is observed in the experiments. Below we give the main results of Dubin's theoretical analysis[10] of the structure of a Wigner crystal consisting of particles with charge eZ that are immersed in a uniform compensating background with an ion number density ϱ and interact via the Coulomb law. When the particle number density is low, they form a monolayer crystal with a hexagonal lattice in which the particle separation is a. As this density increases, the monolayer Coulomb crystal becomes disadvantageous from the energy viewpoint and transforms into a triple-layer crystal with hexagonal close packing when

$$(\varrho/Z)^{1/3}a\leq\alpha,\quad \alpha\approx 1.02. \tag{16}$$

A microparticle crystal in a plasma is an open Hamiltonian system, so that there is no way in which the concept of potential energy can be introduced.[17,18] Hence we examined the stability of a monolayer crystal against stratification by initiating a small longitudinal displacement of particles from their equilibrium position. In a double-layer crystal with a rectangular unit cell, $\mathbf{a}_1=a(1,0)$ and $\mathbf{a}_2=a(0,\sqrt{3})$, containing two particles the distance δz between the particles was fixed, while the position z_p of the crystal's center of mass was found from the balance of all forces acting on the two particles. Note that at $\delta z=0$ the crystal forms a monolayer hexagonal lattice. By varying the particle separation in a layer we found the difference $\delta F_z=F_{z,1}-F_{z,2}$ of the forces acting on the higher and lower particles. Stratification of the crystal occurs at the critical value $a=a_{\text{cr}}$, when $\delta F_z=0$. For $a<a_{\text{cr}}$, the repulsion of the particles of the upper and lower layers is balanced by the attraction of the ionic space charge concentrated between the layers, which leads to the formation of a double-layer crystal. A microparticle crystal differs from a Coulomb crystal placed 'inside'' a uniform background primarily in the dependence of the charges on their coordinates. Since the electron number density decreases as we move closer to the electrode, the charge of the particles belonging to the lower layer is smaller than the charge of the particles belonging to the upper layer, which facilitates stratification of the crystal. Suppose that the lower layer is shifted in relation to the upper by the distance δz. The conditions for equilibrium of the two layers can be written $E_uZ_u=E_lZ_l=Mg$, where E_u, $E_l=E_u+4\pi(\varrho-\alpha^3Z_u/a^3)\delta z$ and Z_u, $Z_l=Z_u-\delta z\,\partial Z/\partial z$ are the field strengths and the charges of the particles in the upper and lower layers, respectively; $\varrho=n_i(z_p)-n_e(z_p)$ is the electron and ion space charge density; and $\alpha=1.02$ is a dimensionless coefficient (taken from Dubin's theory[10]) that determines the strength of the field generated by the upper layer of charged particles. Then the critical particle separation in the dust crystal is given by the formula

$$a_{\text{cr}}=\alpha\left(\frac{Z}{\varrho(1-\delta)}\right)^{1/3}, \tag{17}$$

$$\delta=\frac{E_p}{4\pi\varrho}\frac{\partial\ln U}{\partial z}=\frac{T_e}{3T_e+U}\left(\frac{1}{2}+\frac{\sqrt{1-z_p/d_k}\,(1+E_m/E_e)}{\varphi\sin\varphi}\right), \tag{18}$$

where the particle charge Z and the phase φ are taken at point z_p. Depending on the discharge parameters, we have $\delta=0.1-0.5$, so that the difference of the charges of the lower and upper layers provides a sizable contribution to a_{cr}. In Fig. 3b we compare the results of calculations with those provided by the analytic formula (17). As the gas pressure grows, the microparticle charge decreases and the ion number density increases. The two effects lead to a decrease in the critical particle separation at $U_{\text{rf}}=60$ and 80 V, which agrees with the experimental data. At $U_{\text{rf}}=40$ V the statistical calculation error is several times larger than at $U_{\text{rf}}=60$ V. For $U_{\text{rf}}=40$ V and $P\leqslant100$ Pa, the critical particle separation varies only slightly with pressure and amounts to ~540 μm within statistical calculation error. A further increase in the gas pressure leads to a rapid increase in a_{cr} due to the increase in the difference between the charges of the lower and upper particles. The experimental data on critical particle separation obtained for $U_{\text{rf}}=40$ V and $U_{\text{rf}}=50$ V lie between the calculated data obtained for $U_{\text{rf}}=40$ V and $U_{\text{rf}}=60$ V. Note that the experimenter observes not the phase transition of a monolayer crystal into a double-layer crystal but the formation of a second layer of particles at $a=a_{\text{cr}}$. When $a>a_{\text{cr}}$, the addition of new particles leads to a decrease in the particle separation, and at $a=a_{\text{cr}}$ there appears a second layer, which gets filled as the number of particles increases.

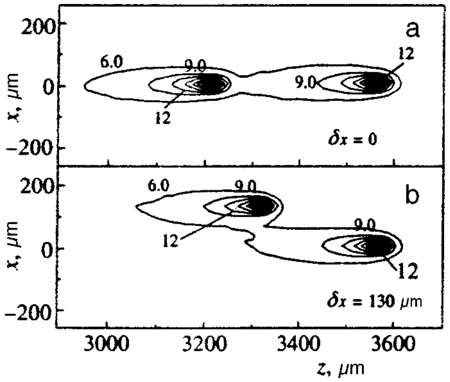

FIG. 4. Ion concentration distributions, measured in units of 10^8 cm^{-3}, in the *xy* sectionfor the electrode sheath with a double-layer microparticle crystal at zero, $\delta x=0$ (a), and finite, $\delta x=130$ μm (b), shifts of the lower layer in relation to the upper layer along the x axis. The regions corresponding to ''condensation'' of ion concentration coincide with the positions of particles. The gas pressure P is 80 Pa, the discharge voltage U_{rf} is 60 V, and the particle separation a in a layer is 519 μm.

3.3. Structure of a double-layer crystal

Immediately after stratification of the monolayer crystal the particle separation amounts to $\sqrt{2}\,a_{\text{cr}}$. A further increase in the particle number density leads first to a decrease in the particle separation, and for $a<a_{\text{cr,2}}$ there is a transition to a triple-layer structure. Probably $a_{\text{cr,2}}\sim a_{\text{cr}}$, but we did not examine this transition quantitatively. We calculated the double-crystal parameters for two limits in particle separation, $\sqrt{2}\,a_{\text{cr}}$ and a_{cr}. Qualitatively, the spatial distributions of the electric field potential, the electron number density, and the ion number density are the same for different gas pressures and discharge voltages. In the longitudinal plane they are almost axisymmetric. In view of the focusing of the ion paths by the particle field, an elongated ion cloud is formed downstream of the particles (Fig. 4a). The maximum ion concentration in the cloud exceeds severalfold the average ion number density at the points occupied by a particle. Excess ionic charges near a particle are small compared to the particle charges proper. For instance, for the case depicted in Fig. 4, the excess ionic charges in a region with a radius of 50 μm near the particles are $5\times10^2\,e$, while the particle charges are approximately $(8-9)\times10^3\,e$. Since a fraction of the ion flux lands on the particles, in the region between the electrode and the crystal the ion number density is somewhat

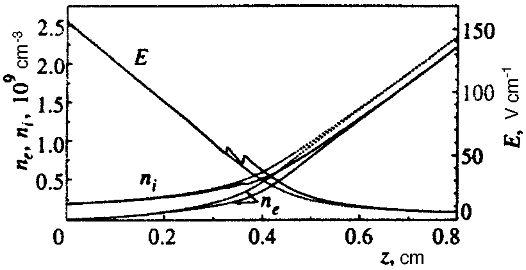

FIG. 5. Cross-section and discharge-period averages of the distributions of the ion concentration (n_i), electron concentration (n_e), and electric-field strength (E) in the electrode sheath at a gas pressure $P=80$ Pa and a discharge voltage $U_{\mathrm{rf}}=60$ V. The solid curves were obtained for a layer containing a two-layer crystal with a particle separation $a=519\,\mu$m in the layer, and the dotted curves were obtained for a layer without microparticles.

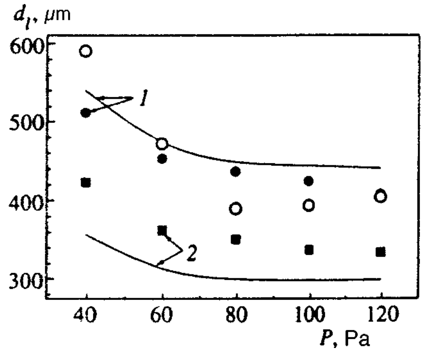

FIG. 6. Dependence of gas pressure on the distance between the layers in a double-layer microparticle crystal for a discharge voltage $U_{\mathrm{rf}}=60$ V and various particle separations in a layer: $a=a_{\mathrm{cr}}$ (*1*) and $a=\sqrt{2}\,a_{\mathrm{cr}}$ (*2*). The curves represent the results obtained by using an approximate crystal model, ● and ■ represent the calculated data, and ○ represent the experimental data at $U_{\mathrm{rf}}\approx 40$ V and $a=a_{\mathrm{cr}}$.

lower than in a layer without particles. The presence of particles in a layer only weakly affects all layer characteristics except the layer thickness (Fig. 5). Hence separate modeling of an rf discharge without particles and for the electrode sheath with particles is justified. In equilibrium, the mutual repulsion of two charged layer is balanced by the attraction to the space charge between the layers. If we take into account the particle charge, the crystal can be considered a quasineutral region. Hence the increase in the characteristic thickness of the electrode sheath is approximately equal to the distance between the outmost planes of the crystal, a fact corroborated by specific calculations of a triple-layer crystal.

Since the electron number density decreases as we move closer to the electrode, the lower particles have a smaller charge than the upper. A transition from a monolayer crystal to a double-layer crystal and a further variation of in particle separation in a layer have a small effect on the particle charge. As the pressure increases, the average electron energy decreases, and so does the potential at the particle surface, with the potential difference between the upper and lower particles remaining essentially unchanged. For a classical Coulomb double-layer crystal, the distance d_l between the layers can be estimated from the condition for the equilibrium between the repulsion of two uniformly charged planes and the attraction of the planes to the background charge ϱ in the space between the planes: $d_l=2Z/\sqrt{3}\,\varrho a^2$. For a microparticle crystal in a layer, the difference of particle charges, $\delta Z=Z_u-Z_l$, of the upper (Z_u) and lower (Z_l) layers plays an important role. Here the equilibrium condition can be approximately written

$$\int_{z_1}^{z_1+d_l}(n_i-n_e)\,dz=\frac{2Z}{\sqrt{3}\,a^2}+\frac{\delta Z}{Z}Mg. \tag{19}$$

Since $2Z/\sqrt{3}\,a^2 \ll Mg$, a slight difference in particle charges leads to an appreciable increase in the distance between the layers, which decreases as the separation of particles belonging to a single layer increases (Fig. 6). When we used the expression (19) to estimate the distance between the layers (the solid curves in Fig. 6), we took the ion and electron concentrations from (15). We also allowed for a reduction of electron concentration near a charged layer due to Debye screening. No calculations of the distance between the layers at $U_{\mathrm{rf}}=40$ V were carried out, since a large statistical error is involved in determining the transition from a monolayer crystal to a double-layer crystal (Fig. 3b). The experimental data on layer distances obtained with $U_{\mathrm{rf}}=40$ V agree fairly well with the results of calculations for $U_{\mathrm{rf}}=60$ V and $a=a_{\mathrm{cr}}$ (Fig. 6). As noted earlier, in experiments the particle separation in a layer must be equal to the critical value a_{cr}.

3.4. Stability of double-layer crystals

To analyze the stability of a loosely packed hexagonal lattice, we calculated the forces (Fig. 7) acting on the particles when the lower layer is shifted as a whole in relation to the upper layer along the x coordinate (the direction of the x axiscoincides with the direction to the nearest particle in the hexagonal lattice). In this series of calculations we examined a hexagonal unit cell containing two particles with fixed transverse coordinates, while the longitudinal coordinates of the particles were determined by the balance of forces acting on the particles. Under a transverse shift of the lower layer,

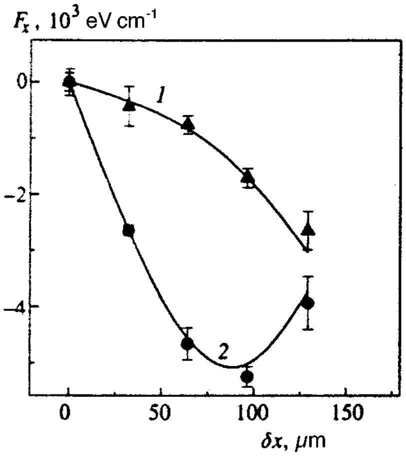

FIG. 7. Forces acting on the particles of the upper (*1*) and lower (*2*) layers when the lower layer is shifted in relation to the upper layer by δx at $U_{\mathrm{rf}}=60$ V, $P=80$ Pa, and $a=519\,\mu$m. The curves represent the results obtained by using the approximation (20), and ● and ▲ represent the results of self-consistent calculations.

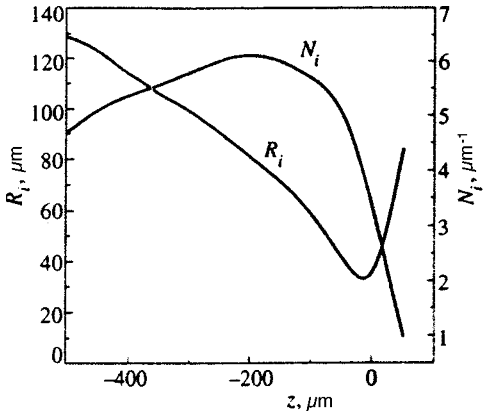

FIG. 8. Effective radius R_i of the ion cloud and the ion number density N_i in the cloud generated by a particle in a monolayer crystal at $U_{\rm rf}=60$ V, $P=80$ Pa, and $a=519$ μm. The coordinate $z=0$ corresponds to the position of the particle.

the distance between the layers decreases, which is especially evident at $a=\sqrt{2}\,a_{\rm cr}$ (Fig. 4). The forces $F_{x,u}$ acting on the particles of the upper layer in the direction x are negative, i.e., the Coulomb repulsion between the lower and upper particles tend to disrupt the equilibrium in which the particles of the lower layer are under the particles of the upper layer. However, the attraction of the lower particles to the ion clouds that are below the upper particles exceeds the Coulomb repulsion between the particles. Hence the forces $F_{x,l}$ acting on the particles in the lower level are also negative, i.e., the lower particles tend to occupy positions under the upper particles. Since $|F_l|>|F_u|$, as shown in Ref. 18, such crystal packing is stable against a shift of the lower layer as a whole in relation to the upper layer. As noted in Ref. 18 and suggested by Fig. 4, a shift of the lower particles has a relatively small influence on the ion concentration distribution in the cloud above the upper particles. This makes it possible to approximately determine the interparticle forces for a known ion concentration distribution in ion clouds, which can be found by analyzing the monolayer crystal (Fig. 2). The results of calculations reveal that with an accuracy of 1–3% the spatial distribution of ions behind a particle in a monolayer crystal can be approximated by the formula

$$n_i(z,\boldsymbol{\rho})=n_{i0}(z)+n_{i1}(z,\boldsymbol{\rho}),$$

$$n_{i1}(z,\boldsymbol{\rho})=\frac{N_i}{\pi R_i}\exp\left(-\frac{|\boldsymbol{\rho}|^2}{R_i^2}\right), \tag{20}$$

where N_i is the ion number density per unit length in the cloud, and R_i is the effective cloud radius (both depend on the distance to the particle; Fig. 8). The minimum in the ion cloud radius is attained near particles and is approximately equal to the characteristic radius R_w of the potential well. As the distance $|z|$ from the generating particle down the flux increases, the effective cloud radius increases approximately according to the square-root law

$$R_i=\sqrt{R_{\rm min}^2+4D_\perp|z|/v_d}$$

due to the transverse ion diffusion with a diffusion coefficient $D_\perp$. The ion number density N_i in the cloud reaches its maximum value at distances of 100–200 μm behind the particle, and then decreases due to the Coulomb repulsion of ions. Note that the total ionic charge in the cloud,

$$N_t=e\int N_i(z)\,dz\approx(2-3)\times10^3\,e,$$

remains several times smaller than the particle charge. Hence ion clouds below lower particles have almost no effect on the upper particles. Nevertheless, the attraction of the lower particles to ion clouds above the upper particles is stronger than the Coulomb repulsion from the upper particles, since on the average the ions are closer to the lower particles.

To analyze the motion of particles in the crystal we must know the interparticle forces, whose self-consistent calculation in the general case, where the crystal loses its symmetry, is almost impossible. Ignoring the momentum transferred from ions to particles in collisions and using formula (20) for the ion concentration distribution in the cloud, we can reduce the many-body problem to finding the two-body forces between the particles. To do this, in addition to accounting for Coulomb repulsion, we must allow for the attractive force

$$\mathbf{F}_l(\mathbf{r})=eZ\frac{\partial}{\partial\mathbf{r}}\sum_i\int\frac{n_{i1}(\mathbf{r}'-\mathbf{r}_i)}{|\mathbf{r}'-\mathbf{r}|}d\mathbf{r}' \tag{21}$$

between the lower particles and the ion clouds that can be considered being rigidly coupled to the upper particles that give rise to them.[18] Summation in (21) is done over the upper particles with coordinates $\mathbf{r}_i$. The results obtained in this manner for the shift of the lower layer in relation to the upper layer (a problem considered earlier) agrees, within statistical error, with a complete calculation of the forces (Fig. 7), which suggests that this approach remains meaningful in more complex situations.

As shown in Ref. 18, a decrease in the gas pressure makes a double-layer crystal unstable against the development of short-wavelength perturbations. In Ref. 18, non-self-consistent calculations of ion motion were carried out, while the ion concentration distribution in the cloud was replaced by a point charge positioned at a certain distance downstream of the upper particles. Using the expression (21) for the force $\mathbf{F}_l$ of interaction between the lower particles and the ion cloud together with the approximation (20) for the ion concentration distribution, we study the stability of the crystal more thoroughly. Following the line of reasoning developed in Ref. 18, we examine the motion of particles only in the transverse plane $\boldsymbol{\rho}=xy$, with allowance for the friction of the particles in the gas with a coefficient ν, the Coulomb repulsion between particles, and the attraction of the lower particles to the ion clouds above the upper particles. Then the equations of motion of the upper ($\boldsymbol{\rho}_{k,u}$) and lower ($\boldsymbol{\rho}_{k,l}$) particles can be written

$$\frac{d^2\boldsymbol{\rho}_{k,u}}{dt^2}=-\nu\frac{d\boldsymbol{\rho}_{k,u}}{dt}+\frac{e^2Z_u^2}{M}\sum_i\frac{\boldsymbol{\rho}_{k,u}-\boldsymbol{\rho}_{i,u}}{|\boldsymbol{\rho}_{k,u}-\boldsymbol{\rho}_{i,u}|^3}$$

$$+\frac{e^2Z_uZ_l}{M}\sum_i\frac{\boldsymbol{\rho}_{k,u}-\boldsymbol{\rho}_{i,l}}{|\boldsymbol{\rho}_{k,u}-\boldsymbol{\rho}_{i,l}+\mathbf{e}_zd_l|^3}, \tag{22}$$

TABLE I.

$U_{\rm rf}$, V	$a/a_{\rm cr}$	ν_*, s^{-1}	ω_*, s^{-1}	φ_*, deg
60	1	17	59	73
60	$\sqrt{2}$	18	49	43
80	$\sqrt{2}$	24	61	45

$$\frac{d^2\boldsymbol{\rho}_{k,l}}{dt^2}=-\nu\frac{d\boldsymbol{\rho}_{k,l}}{dt}+\frac{e^2Z_l^2}{M}\sum_i\frac{\boldsymbol{\rho}_{k,l}-\boldsymbol{\rho}_{i,l}}{|\boldsymbol{\rho}_{k,l}-\boldsymbol{\rho}_{i,l}|^3}+\frac{e^2Z_uZ_l}{M}\sum_i\frac{\boldsymbol{\rho}_{k,l}-\boldsymbol{\rho}_{i,u}}{|\boldsymbol{\rho}_{k,l}-\boldsymbol{\rho}_{i,u}+\mathbf{e}_z d_l|^3}+\frac{1}{M}\mathbf{F}_{k,l}, \tag{23}$$

where $\mathbf{e}_z$ is the unit vector along the z axis, with the lower index indicating the equilibrium position of the particle in a layer. The forces acting on a particle are nonpotential due to the interaction of the lower particles and the ion clouds, which in the final analysis leads to the development of particle oscillations as the coefficient of friction of particles in the gas decreases. To analyze crystal stability, we examine the perturbations

$$\delta\boldsymbol{\rho}_{k,u}=\xi_u\exp(i\mathbf{q}\cdot\boldsymbol{\rho}^0_{k,u}+\lambda_i t),$$
$$\delta\boldsymbol{\rho}_{k,l}=\xi_l\exp(i\mathbf{q}\cdot\boldsymbol{\rho}^0_{k,l}+\lambda_i t)$$

of the coordinates of the particles with respect to the equilibrium particle positions $\boldsymbol{\rho}^0_{k,u}$ and $\boldsymbol{\rho}^0_{k,l}$ in the crystal lattice, linearize Eqs. (22) and (23) in the small quantities $\delta\boldsymbol{\rho}_{k,u}$ and $\delta\boldsymbol{\rho}_{k,l}$, and find the eigenvalues λ_i for various wave vectors $\mathbf{q}$. The statement of the problem differs from that in Ref. 18 only by the fact that here we specify the particle charges and the force $\mathbf{F}_l$ self-consistently. Hence, without going into the cumbersome calculations, which can be found in Ref. 18, we write only the results for the main physical characteristics of instability: the critical friction coefficient ν_* for the particles in the gas (below this value instability begins to develop), the corresponding oscillation frequency ω_*, and phase shift in the oscillations of the lower and upper particles, $\sin\varphi_*=\mathrm{Im}(\xi_u/\xi_l)(|\xi_u/\xi_l|)^{-1}$ (see Table I). In accordance with the experimental data listed in Refs. 5 and 19, the critical value of the friction coefficient of the particles in the gas increases with the discharge power or, in other words, the critical gas pressure (the pressure at which the crystal becomes unstable against the development of particle oscillation) increases with the discharge power. As the particle separation in a layer decreases, the value if ν_* changes only slightly, and the oscillation frequency ω_* and phase shift φ_* increase. Experiments conducted at $U_{\rm rf}\approx 40$ V and $a=a_{\rm cr}$ yield $\nu_*\approx 28\,{\rm s}^{-1}$, $\omega_*=72\,{\rm s}^{-1}$, and $\varphi_*=40°$. The difference between the experimental and theoretical data may possibly be due to the defects existing in real crystals.

4. CONCLUSION

As a result of our calculations for various gas pressures and discharge voltages, we have obtained (in a self-consistent manner) the main parameters of a microparticle crystal in a plasma: the particle potentials, the critical particle separation in a layer at which, due to the Coulomb repulsion of particles, the monolayer crystal transforms into a double-layer crystal, and the distance between layers in a double-layer crystal. We have proposed a model for calculating interparticle forces. This model has been used to determine the main characteristics of instability of a double-layer crystal against the development of transverse particle oscillations: the critical friction coefficient for the particles in the gas, the oscillation frequency, and the phase shift in the oscillations of the lower and higher particles. The discrepancy between our results and the experimental data on the various lattice parameters amounts to 25–40%, which is of the order of the experimental error. We have also built an approximate model of a microparticle crystal in the electrode sheath, a model that makes it possible to find all the crystal characteristics to a high accuracy.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-19134-a), RFFI–NNIO(Grant No. 96-02-00241-G), and INTAS (Grant No. 96-0235).

[*)] E-mail: ischweig@isp.nsc.ru

---

[1] H. Ikezi, Phys. Fluids **29**, 1764 (1986).

[2] J. H. Chu, J.-B. Du, and Lin I, J. Phys. D **27**, 296 (1994).

[3] J. H. Chu and Lin I, Physica A **205**, 183 (1994).

[4] J. H. Chu and Lin I, Phys. Rev. Lett. **72**, 4009 (1994).

[5] H. Thomas, G. E. Morfill, V. Demmel, J. Goree, B. Feuerbacher, and D. Möhlmann, Phys. Rev. Lett. **73**, 652 (1994).

[6] Y. Hayashi and K. Tachibana, Jpn. J. Appl. Phys., Part 2 **33**, L804 (1994).

[7] A. Melzer, T. Trottenberg, and A. Piel, Phys. Lett. A **191**, 301 (1994).

[8] T. Trottenberg, A. Melzer, and A. Piel, Plasma Sources Sci. Technol. **4**, 450 (1995).

[9] V. E. Fortov, A. P. Nefedov, V. M. Torchinskiĭ, V. N. Molotkov, A. G. Khrapak, O. F. Petrov, and K. F. Volykhin, JETP Lett. **64**, 92 (1996).

[10] D. H. E. Dubin, Phys. Rev. Lett. **71**, 2753 (1993).

[11] V. A. Shveĭgert and M. S. Obrekht, Pis'ma Zh. Tekh. Fiz. **21**(10), 57 (1995) [Tech. Phys. Lett. **21**, 377 (1996)].

[12] G. Goldoni, V. A. Schweigert, and F. M. Peeters, Surf. Sci. **361/362**, 163 (1996).

[13] Y. Hayashi and K. Tachibana, J. Vac. Sci. Technol. A **14**, 506 (1996).

[14] J. Pieper, J. Goree, and R. Quinn, J. Vac. Sci. Technol. A **14**, 519 (1996); Phys. Rev. E **54**, 5636 (1996).

[15] F. Melandsø and J. Goree, Phys. Rev. E **52**, 5312 (1995).

[16] S. V. Vladimirov and M. Nambu, Phys. Rev. E **52**, 2172 (1995); S. V. Vladimirov and O. Ishihara, Phys. Plasmas **3**, 444 (1996); O. Ishihara and S. V. Vladimirov, Phys. Plasmas **4**, 1 (1997).

[17] A. Melzer, V. A. Schweigert, I. V. Schweigert, A. Homann, S. Peters, and A. Piel, Phys. Rev. E **54**, R46 (1996).

[18] V. A. Schweigert, I. V. Schweigert, A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **54**, 4155 (1996).

[19] A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).

[20] J. Goree, Phys. Rev. Lett. **69**, 277 (1992).

[21] S. J. Choi and M. J. Kushner, J. Appl. Phys. **75**, 3352 (1994).

[22] G. Lapenta, Phys. Rev. Lett. **75**, 4409 (1995).

[23] Ya. K. Khodataev, R. Bingham, V. P. Tarakanov, and V. N. Tsytovich, Fiz. Plazmy **22**, 1028 (1996) [Plasma Phys. Rep. **22**, 932 (1996)].

[24] V. A. Shveĭgert, Pis'ma Zh. Tekh. Fiz. **21**(12), 69 (1995) [Tech. Phys. Lett. **21**, 476 (1995)].

[25] F. Melandsø and J. Goree, J. Vac. Sci. Technol. A **14**, 511 (1996).

[26] N. Otani and A. Bhattacharjee, Phys. Rev. Lett. **78**, 1468 (1997).

[27] F. Melandsø, Phys. Rev. E **55**, 7495 (1997).

[28] R. Hockney and J. Eastwood, *Computer Simulation Using Particles*, Adam Hilger, Philadelphia (1988).

[29] R. Lagushenko and J. Maya, J. Appl. Phys. **59**, 3293 (1984).

[30] M. E. Riley, K. E. Greenberg, G. A. Hebner, and P. Drallos, J. Appl. Phys. **75**, 2789 (1994).
[31] E. W. McDaniel and E. A. Mason, *The Mobility and Diffusion of Ions in Gases*, Wiley, New York (1973).
[32] V. A. Godyak, R. B. Piejak, and B. M. Alexandrovich, Plasma Sources Sci. Technol. **1**, 36 (1992).
[33] V. A. Godyak, R. B. Piejak, and B. M. Alexandrovich, Phys. Rev. Lett. **68**, 40 (1992).
[34] B. V. Alekseev and V. A. Kotel'nikov, *The Probe Method in Plasma Diagnostics* [in Russian], Énergoatomizdat, Moscow (1988).
[35] Yu. P. Raizer, *Gas Discharge Physics*, Springer-Verlag, Berlin (1991).
[36] Yu. M. Kagan and V. I. Perel', Dokl. Akad. Nauk SSSR **108**, 222 (1956) [Sov. Phys. Dokl. **1**, 289 (1956)].
[37] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).
[38] M. A. Lieberman, IEEE Trans. Plasma Sci. **17**, 338 (1989).
[39] V. A. Schweigert, Preprint, Institute of Theoretical and Applied Mechanics, Siberian Branch of the USSR Academy of Sciences, Novosibirsk (1990).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Cherenkov interaction of vortices with a free surface

E. A. Kuznetsov*) and V. P. Ruban

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia

(Submitted 23 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 894–919 (March 1999)

The interaction of vortex filaments in an ideal incompressible fluid with the free surface of the latter is investigated in the canonical formalism. A Hamiltonian formulation of the equations of motion is given in terms of both canonical and noncanonical Poisson brackets. The relationship between these two approaches is analyzed. The Lagrangian of the system and the Poisson brackets are obtained in terms of vortex lines, making it possible to study the dynamics of thin vortex filaments with allowance for finite thickness of the filaments. For two-dimensional flows exact equations of motion describing the interaction of point vortices and surface waves are derived by transformation to conformal variables. Asymptotic steady-state solutions are found for a vortex moving at a velocity lower than the minimum phase velocity of surface waves. It is found that discrete coupled states of surface waves above a vortex are possible by virtue of the inhomogeneous Doppler effect. At velocities higher than the minimum phase velocity the buoyant rise of a vortex as a result of Cherenkov radiation is described in the semiclassical limit. The instability of a vortex filament against three-dimensional kink perturbations due to interaction with the ''image'' vortex is demonstrated. © *1999 American Institute of Physics.* [S1063-7761(99)00903-8]

1. INTRODUCTION

It is generally known (see, e.g., Ref. 1) that fluid flow in the presence of oscillations of the surface of water under the influence of the forces of gravity and capillary tension can be treated very accurately as potential flow: $\mathbf{v}=\nabla\Phi$. The small nonpotential component of the velocity field is attributable to fluid viscosity. The system becomes Hamiltonian in the zero viscosity limit. The actual profile of the surface $z=\eta(x,y,t)$ and the value of the potential on the surface $\phi(x,y,t)$ emerge as canonical variables, whereupon the equations of motion are written in the form

$$\frac{\partial\eta}{\partial t}=\frac{\delta H}{\delta\phi},\quad \frac{\partial\phi}{\partial t}=-\frac{\delta H}{\delta\eta},\tag{1}$$

where the Hamiltonian H coincides with the total energy of the fluid.

A Hamiltonian formulation (1) of the equations of motion for potential flows of a fluid with a free surface was published by Zakharov at the end of the 1960's.[2,3] Since then the Hamiltonian approach has been used successfully in the investigation of a great many phenomena: the modulation instability of surface waves,[2] the nonlinear stage of development of Kelvin–Helmholtz instability,[4] the formation of hexagonal relief on the surface of liquid dielectrics in the presence of an external electric field,[5] and the formation of singularities on the surface of ideal fluids.[6,7] Two important contributions to research on surface wave turbulence are Refs. 3 and 8, in which the first systematic theory of Kolmogorov spectra—flux-type power-law distributions—is formulated.

Nowadays the theory of surface wave turbulence is one of the most advanced of all (see, e.g., the survey in Ref. 9). All the same, the theory has lingering unsolved problems, the crux of which could well be the problem of the interaction of nonpotential flows, i.e., rotational (vortex) flows with a free surface.

Probably the first theoretical treatment of this problem was published by Keldysh and Lavrent'ev,[10] and a section of a later book[11] is also devoted to it. Novikov has subsequently investigated the problem.[12] In all these publications the authors have calculated the losses of a two-dimensional point vortex in steady-state motion as a result of the excitation of surface waves. In Novikov's work[12] this process is analyzed in the linear approximation with respect to the surface wave amplitude on the assumption that the vortex is situated at considerable depth from the free surface. In the zeroth approximation in this case the vortex combines with its mirror image to form a dipole pair, which moves at a constant velocity along the surface. The generated flow away from the vortex is inhomogeneous along the surface. In a coordinate frame where the vortex is at rest, the flow is homogeneous at large distances from the vortex. As we approach the vortex, the flow velocity on the surface changes sign at some point, so that the flow velocity above the vortex has the opposite sign and three times the magnitude of the velocity at infinity.

It is clear from this situation that the generated wave experiences an inhomogeneous Doppler effect by virtue of the Cherenkov process. It should be recalled that Cherenkov radiation occurs when a vortex propagates at a velocity greater than the minimum phase velocity of surface waves. It is shown in the cited paper that when the Cherenkov wavelength λ is much shorter than the distance from the vortex to the surface, i.e., the vortex depth h, the Cherenkov radiation can be described semiclassically. Surface waves are known to be localized within a surface layer having a thickness of the order of the wavelength, the fluid velocity decreasing exponentially at greater distances. For $\lambda \ll h$, therefore, the back influence of radiation on the vortex is exponentially small, so that perturbation theory can be used to good advantage.

In the present article we investigate the nonlinear interaction of vortices with a free surface. We are concerned primarily with the Cherenkov radiation of surface waves by a moving point vortex. Our work differs fundamentally from previous studies in that the inhomogeneous Doppler effect is taken into account. Its influence is significant not only for Cherenkov radiation, but also in the subcritical regime, where the vortex propagation velocity is smaller than the minimum phase velocity of the waves. The entrainment of surface waves is possible in this case, along with the formation of steady-state bound states involving the vortex and the waves propagating with it. It is possible for waves to be entrained by the vortex when the propagation velocities v are much lower than the minimum phase velocity $V_{\min}$. For example, when $\lambda \ll h$, the velocity v at which entrainment sets in attains a value of the order of $V_{\min}/3$.

We use the Hamiltonian approach to describe the interaction of vortices with a free surface. In the canonical[13,14] formulation the fluid velocity is represented in terms of Clebsch variables. Even though this kind of parametrization of the velocity describes a special type of flow, the canonical Poisson bracket (expressed in terms of Clebsch variables) admits conversion to a noncanonical bracket, which is expressed at once in terms of the velocity, the potential on the surface, and the surface profile itself; most important, it can be used to describe flows having an arbitrary topology. However, the noncanonical bracket is found to be degenerate, owing to the existence of a special symmetry, which forms an entire group—the group of relabeling transformations of Lagrangian markers (the details are given in a recent survey[14]). This symmetry generates all known vorticity conservation laws. The principal law is the freezing-in of vortex lines in the fluid. It corresponds to a local Lagrangian invariant—the Cauchy invariant. In a previous paper[15] we propose a technique for removing the degeneracy for an incompressible ideal fluid without a free boundary by introducing new variables called Lagrangian markers, which enumerate each vortex line. It will be shown in Sec. 2 that this approach can be extended to ideal hydrodynamics with a free surface. In particular, the transition to a vortex-line representation provides the means for writing a variational principle and a simple way to achieve a self-consistent Hamiltonian description of a system of thin vortex filaments interacting with a free surface.

In Sec. 3 we use the variational principle and conformal mapping into the half-plane (by analogy with Ref. 16) to derive the equations of motion of point vortices and a free surface for a two-dimensional geometry. In the next section we find steady-state solutions of these equations in the form of an asymptotic expansion in powers of the Froude parameter F. In Sec. 5 we discuss bound surface wave states. In Sec. 6, assuming that $h \gg \lambda$, we determine the law governing the buoyant rise of a point vortex as a result of Cherenkov radiation. In the last section we investigate the influence of three-dimensional perturbations on the dynamics of vortex filaments. We show that the interaction of the vortex with its own image renders a vortex filament unstable against kink perturbations, where the parts of the filament closest to the free surface accelerate toward it, while more distant parts move in the opposite direction, away from the surface. This instability is analogous to the Crow instability[17] of two antiparallel vortex filaments. Without a free surface the development of this instability leads to reconnection of the vortex lines and the formation of vortex rings. In this situation we can expect a vortex line to ‘‘reconnect’’ with its own mirror image, inevitably producing a large set of vortex half-rings that begin and end at the surface of the fluid. The ‘‘reconnection’’ process must be accompanied by the Cherenkov radiation of surface waves in this case.

2. BASIC EQUATIONS; HAMILTONIAN FORMALISM

We consider an incompressible ideal fluid of constant density ρ (we assume from now on that $\rho=1$) in the presence of a constant gravitational field $\mathbf{g}$ antiparallel to the z axis. Let the profile of the free surface be described by the equation $z=\eta(x,y,t)$.

The equations of motion of the fluid are the Euler equations

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\nabla)\mathbf{v}\right)=-\nabla p, \quad \operatorname{div}\mathbf{v}=0 \tag{2}$$

subject to the boundary conditions: 1) extinction of the velocity ($\mathbf{v}\to 0$) at a large depth $z\to-\infty$; 2) the dynamic condition at the free boundary

$$p|_{z=\eta}=\sigma \operatorname{div}\frac{\nabla \eta}{\sqrt{1+(\nabla \eta)^2}}, \tag{3}$$

and the kinematic condition

$$\frac{\partial \eta}{\partial t}=v_n\sqrt{1+(\nabla \eta)^2}=v_z-(\mathbf{v}_\perp \nabla)\eta. \tag{4}$$

Here p is the pressure, and σ is the coefficient of surface tension.

The system of equations (2)–(4) is classified as Hamiltonian. For pure potential flows $\mathbf{v}=\nabla\Phi$, Eq. (2) with the boundary conditions (3) and (4) can be written in the canonical form (1). For nonpotential flows that can be parametrized in terms of Clebsch variables, we have

$$\mathbf{v}=\hat{P}\lambda\nabla\mu+\nabla\Phi, \tag{5}$$

and the canonical equations (1) are augmented with two equations in the variables λ and μ (see, e.g., Refs. 14 and 18):

$$\frac{\partial\lambda}{\partial t}=\frac{\delta H}{\delta\mu}=-(\mathbf{v}\nabla)\lambda, \quad \frac{\partial\mu}{\partial t}=-\frac{\delta H}{\delta\lambda}=-(\mathbf{v}\nabla)\mu. \tag{6}$$

Here the Hamiltonian H is identical to the total energy of the fluid:

$$H=\int_{z\leqslant\eta}\frac{\mathbf{v}^2}{2}\,d\mathbf{r}+\int d\mathbf{r}_\perp\left\{\frac{g\eta^2}{2}+\sigma\left[\sqrt{1+(\nabla\eta)^2}-1\right]\right\}, \tag{7}$$

and the velocity $\mathbf{v}$ can be expressed in terms of λ and μ by means of Eq. (5). In Eq. (5) $P_{\alpha\beta}=\delta_{\alpha\beta}-\nabla_\alpha\Delta^{-1}\nabla_\beta$ is the transverse projector. The operator Δ^{-1} is interpreted everywhere in this section as the inverse operator for the Dirichlet problem $\Delta f=g$ with vanishing boundary conditions

$$f|_{z=\eta}=0 \quad \text{and} \quad f\rightarrow 0 \ \text{as}\ z\rightarrow-\infty.$$

For this choice the symbol ϕ in Eqs. (1) retains its definition as the value of Φ at the free surface, and the potential Φ itself is a harmonic function:

$$\Delta\Phi=0, \quad \Phi|_{z=\eta}=\phi. \tag{8}$$

It is important to note that here λ and μ are completely unconstrained at the free surface. The vortex lines, which are specified by the intersection of the surfaces $\lambda=\text{const}$ and $\mu=\text{const}$, can have arbitrary orientation at the surface $z=\eta(\mathbf{r}_\perp,t)$, i.e., they can either be tangent to the free surface or run into it.

The substitution (5) is known (see, e.g., Ref. 19) to describe a special type of flow; in particular, it does not describe knotted flows. Nonetheless, despite the special character of this substitution, it can be used to obtain a Hamiltonian description for flows having an arbitrary topology in terms of so-called noncanonical Poisson brackets.[20] This fact was first demonstrated for flows without a free boundary.[14] We now show that it is also valid for flows with a free surface. We convert the canonical Poisson bracket, specified in terms of bulk variables (λ,μ) and surface variables (ϕ,η),

$$\{F,G\}=\int\left(\frac{\delta F}{\delta\lambda}\frac{\delta G}{\delta\mu}-\frac{\delta G}{\delta\lambda}\frac{\delta F}{\delta\mu}\right)d\mathbf{r} + \int\left(\frac{\delta F}{\delta\eta}\frac{\delta G}{\delta\phi}-\frac{\delta G}{\delta\eta}\frac{\delta F}{\delta\phi}\right)d\mathbf{r}_\perp, \tag{9}$$

to a bracket expressed in terms of the velocity $\mathbf{v}$ and the surface profile. These calculations are based on a recalculation of the variational derivatives.

Invoking the definition (5) and the self-adjointness of the operator $\hat{P}$, we have

$$\left.\frac{\delta F}{\delta\lambda}\right|_{\mu\eta\psi}=\left(\nabla\mu,\frac{\tilde{\delta}F}{\delta\mathbf{v}}\right), \quad \frac{\delta F}{\delta\mu}_{\lambda\eta\psi}=\left(\nabla\lambda,\frac{\tilde{\delta}F}{\delta\mathbf{v}}\right). \tag{10}$$

Here

$$\frac{\tilde{\delta}F}{\delta\mathbf{v}}=\hat{P}\frac{\delta F}{\delta v},$$

so that $\operatorname{div}(\tilde{\delta}F/\delta\mathbf{v})=0$.

Next we calculate the variational derivative of ϕ:

$$\left.\frac{\delta F}{\delta\phi}\right|_{\lambda\mu\eta}=\int\frac{\delta F}{\delta\mathbf{v}}\nabla\frac{\delta\phi(\mathbf{r})}{\delta\phi(\mathbf{r}_\perp)}d\mathbf{r}. \tag{11}$$

Making use of the relation $\Delta\,\delta\Phi(\mathbf{r})/\delta\phi(\mathbf{r}_\perp)=0$, we replace the variational derivative $\delta F/\delta\mathbf{v}$ by $\tilde{\delta}F/\delta\mathbf{v}$ in the integral (11). Then, transforming the integral to a surface integral, we obtain

$$\left.\frac{\delta F}{\delta\phi}\right|_{\lambda\mu\eta}=\left(\mathbf{n}\frac{\tilde{\delta}F}{\delta\mathbf{v}}\right)_{z=\eta}. \tag{12}$$

In the meantime, the variational derivative with respect to η remains unchanged:

$$\left.\frac{\delta F}{\delta\eta}\right|_{\lambda\mu\phi}=\left.\frac{\delta F}{\delta\eta}\right|_{v}. \tag{13}$$

Next, substituting relations (10), (12), and (13) into the bracket (9), we arrive at an expression for the noncanonical Poisson bracket[21]:

$$\{F,G\}=\int\left(\operatorname{curl}\mathbf{v}\left[\frac{\tilde{\delta}F}{\delta\mathbf{v}}\times\frac{\tilde{\delta}G}{\delta\mathbf{v}}\right]\right)d\mathbf{r} + \int\left(\frac{\delta F}{\delta\eta}\left(\mathbf{n}\frac{\tilde{\delta}G}{\delta\mathbf{v}}\right)-\frac{\delta G}{\delta\eta}\left(\mathbf{n}\frac{\tilde{\delta}F}{\delta\mathbf{v}}\right)\right)d\mathbf{r}_\perp. \tag{14}$$

The first part of this bracket takes bulk variables into account, and without a free boundary it goes over to a bracket expression first published in terms of the curl of the velocity $\mathbf{\Omega}=\operatorname{curl}\mathbf{v}$ in Ref. 19:

$$\{F,G\}=\int\left(\mathbf{\Omega}\left[\operatorname{curl}\frac{\delta F}{\delta\mathbf{\Omega}}\operatorname{curl}\frac{\delta G}{\delta\mathbf{\Omega}}\right]\right)d\mathbf{r}. \tag{15}$$

When the vortex lines do not run into the free surface, the second part of the bracket can be expressed in terms of the variables ϕ and η:

$$\int\left(\frac{\delta F}{\delta\eta}\frac{\delta G}{\delta\phi}-\frac{\delta G}{\delta\eta}\frac{\delta F}{\delta\phi}\right)d\mathbf{r}_\perp \tag{16}$$

and coincides with the canonical Zakharov bracket.

Accordingly, the equations of motion (2) and (4) can be written by virtue of the bracket (14) in the form

$$\mathbf{v}_t=\{\mathbf{v},H\}, \quad \eta_t=\{\eta,H\}.$$

In the investigation of thin vortices it is more convenient to use a third Hamiltonian formulation, in terms of vortex lines. We have shown previously[15] that for ideal incompressible hydrodynamics without a free boundary the transformation to vortex lines as new variables removes the degeneracy of the Poisson brackets and makes it possible to write a variational principle. It should also be noted that the limit of infinitely thin vortex filaments of finite vorticity in the presence of a free boundary has been analyzed previously.[22] The formulation set forth below gives a Hamiltonian description of distributed vortices interacting with a free surface. Our proofs are simpler and, most importantly, they can be used to check the limiting transition to infinitely thin vortices.

We represent a vortex tube of finite thickness by a continuous distribution of vortex filaments, each of which we label with the coordinate ν. We assume that the coordinate ν belongs to a certain fixed two-dimensional domain. For example, ν can be interpreted as the initial coordinates of a vortex filament in a cross section of the vortex tube.

We specify the position of each vortex filament by the function

$$\mathbf{r}=\mathbf{R}(\nu,s,t), \tag{17}$$

where s is a parameter that varies along the filament. For a closed vortex line the function $\mathbf{R}(\nu,s,t)$ is periodic with a period that depends on ν:

$$\mathbf{R}(\nu,s,t)=\mathbf{R}(\nu,s+l(\nu),t).$$

We shall assume below for simplicity that the vortex lines do not reach the surface anywhere and are closed. The scalar velocity potential is therefore well defined at the boundary. (Note that its value $\phi(\mathbf{r}_\perp)$ contains, in addition to the wave potential, a contribution from vortices situated in the depth of the fluid.)

Let the velocity circulation around the vortex tube, which rests on an area element $d^2\nu$ in the vicinity of the point ν, be equal to $\rho(\nu)d^2\nu$. By the conservation of circulation, $\rho(\nu)$ is independent of the time.

In this case the curl of the velocity $\mathbf{\Omega}(\mathbf{r},t)$ can be written in the form

$$\mathbf{\Omega}(\mathbf{r};\{\mathbf{R}(\nu,s,t)\})=\int d^2\nu\rho(\nu)\int ds\,\frac{\partial\mathbf{R}(\nu,s,t)}{\partial s}\times\delta(\mathbf{r}-\mathbf{R}(\nu,s,t)). \tag{18}$$

We see at once that the choice of the parameter s in Eqs. (17) and (18) is nonunique, up to the replacement $s\to\tilde{s}(s,t)$. Irrespective of the choice of s, the vector $\partial\mathbf{R}(\nu,s,t)/\partial s$ is always tangent to the vortex line with the given ν.

We have shown previously[15] that the equation of motion for a vortex line $\mathbf{r}=\mathbf{R}(\nu,s,t)$,

$$[\mathbf{R}_s\times(\mathbf{R}_t-\mathbf{v}(\mathbf{R},t))]=0, \tag{19}$$

governs only the ''transverse'' dynamics relative to the vortex line. The equation of motion (19) follows directly from the Euler equation for the curl of the velocity $\mathbf{\Omega}=\operatorname{curl}\mathbf{v}$. As should be the case, the longitudinal ''variation'' does not affect the deformation of the curve, leaving Eq. (19) invariant under all smooth substitutions $s\to\tilde{s}(s,t)$.

The description of the vortex lines by Eqs. (17), (18), and (19) is a hybrid Lagrangian–Eulerian description. Here the parameter ν has a transparent Lagrangian origin, whereas the coordinate s remains Eulerian. For planar flows the coordinate s is naturally identified with the coordinate z perpendicular to the corresponding plane.

It can be checked by direct substitution that the equations of motion (4) and (19) for the variables $\mathbf{R}$ and η and also for ψ follow from the variational principle for the action $S=\int\mathscr{L}dt$:

$$\delta S=0,$$

where the Lagrangian $\mathscr{L}$ is given by

$$\mathscr{L}=\frac{1}{3}\int d^2\nu\rho(\nu)\int([\mathbf{R}_t(\nu,s)\times\mathbf{R}(\nu,s)]\mathbf{R}_s(\nu,s))ds+\int\phi\eta_t d\mathbf{r}_\perp-\mathscr{H}[\mathbf{R},\phi,\eta]. \tag{20}$$

The key factor in this check is embodied in the two relations

$$\rho(\nu)\left[\mathbf{R}_s\times\operatorname{curl}\frac{\delta F}{\delta\mathbf{\Omega}}\right]=\frac{\delta F}{\delta\mathbf{R}} \tag{21}$$

and

$$\mathbf{v}=\operatorname{curl}\frac{\delta\mathscr{H}}{\delta\mathbf{\Omega}},$$

the first of which is valid only for functionals that depend on $\mathbf{\Omega}$, ϕ, and η, i.e., gauge-invariant functionals that are independent of the choice of the parameters s and ν.

The equation of motion for $\mathbf{R}$ (19) deduced from (20) acquires the Hamiltonian form

$$\rho(\nu)[\mathbf{R}_s(\nu,s)\times\mathbf{R}_t(\nu,s)]=\left.\frac{\delta\mathscr{H}}{\delta\mathbf{R}(\nu,s)}\right|_{\phi,\eta}, \tag{22}$$

and the equations for ϕ and η retain their canonical form (1).

Making use of the property (21), by rewriting the bracket (14) we can obtain the Poisson bracket (between two gauge-invariant functionals) expressed in terms of vortex lines:

$$\{F,G\}=\int\frac{d^2\nu ds}{\rho(\nu)|\mathbf{R}_s(\nu,s)|^2}\left(\mathbf{R}_s(\nu,s)\left[\frac{\delta F}{\delta\mathbf{R}(\nu,s)}\times\frac{\delta G}{\delta\mathbf{R}(\nu,s)}\right]\right)+\int\left(\frac{\delta F}{\delta\eta}\frac{\delta G}{\delta\phi}-\frac{\delta G}{\delta\eta}\frac{\delta F}{\delta\phi}\right)d\mathbf{r}_\perp. \tag{23}$$

It is evident from this expression that the new bracket (23) does not contain variational derivatives with respect to ρ, although ρ could have been assumed to be invariant in the given calculations. This result indicates that ρ is a Casimir variable in relation to the original bracket, permitting us to introduce the variational principle (20). The variables ϕ and η now remain canonically conjugate.

The canonical formulation (20) also benefits us in the ease with which the limiting transition is made to a finite number of very thin vortex filaments $\mathbf{R}_n(s,t)$ having finite vorticities γ_n:

$$\mathscr{L}_\epsilon=\frac{2\pi}{3}\sum_n\gamma_n\int ds(\mathbf{R}_{ns}[\mathbf{R}_n\times\mathbf{R}_{nt}])+\int\phi\eta_t d\mathbf{r}_\perp-\mathscr{H}[\mathbf{R}_n,\phi,\eta]. \tag{24}$$

All that is required is to take accurate account of the fact that the ''self-energy'' gained by the filament in the immediate vicinity of its axis increases logarithmically as its thickness tends to zero. In three-dimensional space finite-thickness effects often play an important role, because the stretching of a certain part of a filament is accompanied by a decrease in its thickness and, as a result, an increase in the self-energy density per unit length. In principle, this property of the system can be incorporated into the discussion by retaining additional degrees of freedom other than $\mathbf{R}(s)$ in our crude description, specifically the filament cross section $\Sigma(s)$ and, canonically conjugate to it, the angle $\theta(s)$ of rotation of the cross section as a whole about the axis of the filament.

This problem never arises in the investigation of planar flows. The vortex has a constant area, which is associated only with the additive constant in the expression for the Hamiltonian. In this case the variational principle for the action in the Lagrangian (20) leads to the well-known conclusion (see Ref. 13) that the coordinates of each point vortex, $X_n(t)$ and $Y_n(t)$, are canonically conjugate variables, and the Lagrangian assumes the form

$$\mathscr{L}=\sum_n 2\pi\gamma_n X_n(t)\dot{Y}_n(t)+\int \phi\eta_t dx - \mathscr{H}[X_n, Y_n, \phi, \eta]. \tag{25}$$

3. DYNAMICAL EQUATIONS IN CONFORMAL VARIABLES

In this section we address planar flows, assuming point vortices. Our primary concern being the interaction of vortices with a free surface, we confine the discussion to a single vortex (any generalization to N vortices is more or less obvious). For two-dimensional flows it is helpful to follow Ref. 16 and map the fluid-filled domain $D:\{y\leqslant\eta(x,t)\}$ into the lower half-plane of the complex variable $w=u+iv$. We assume that the conformal mapping is specified by the analytic function

$$z(w)=x+iy=w+A(w), \tag{26}$$

which takes the boundary of the fluid $y=\eta(x,t)$ into the real axis $v=0$, with the function $A(w)$ tending to zero at infinity.

To describe the fluid flow, we introduce the complex velocity potential

$$\Pi=\Phi+i\Theta,$$

where Φ is the hydrodynamic potential used above, and Θ is the stream function. In the presence of vortices the potential Π is not unique: the potential acquires a nonzero increment when it skirts the vortex. Consequently, wherever necessary in the ensuing discussion, we assume that the potential Π is specified in the (lower) half-plane with a cut that extends from the point $w=W=U+iV$ where the vortex is located to $w=-i\infty$.

Rectification of the free surface by means of the conformal transformation $z=z(w)$ permits the vortex component and vortex image to be exactly separated out of the complex potential:

$$\Pi(w)=\Pi_0(w)+\Psi(w)=i\gamma\ln\left(\frac{w-W}{\overline{W}-w}\right)+\Psi(w). \tag{27}$$

This component [the first term in (27)] has the important property that it makes no contribution to the velocity component normal to the surface. The other term in Π can be interpreted as the surface wave potential.

We assume everywhere that $\gamma>0$, and accordingly that the vortex rotates clockwise.

For $v=0$ (i.e., at the free surface) the real potential is

$$\phi(u)=\Phi_0(u)+\psi(u)=2\gamma\arctan\left(\frac{u-U}{-V}\right)+\psi(u). \tag{28}$$

We introduce the projectors $P^{(\pm)}$ and the operator $|\hat{k}|$, which will be needed below and which have the Fourier representation

$$P^{(\pm)}=\frac{1}{2}(1\pm\mathrm{sgn}(k)), \tag{29}$$

$$|\hat{k}|=|k|. \tag{30}$$

When the projectors $P^{(\pm)}$ act on an arbitrary function, rendering its Fourier components with negative or positive wave numbers k equal to zero, from the given function they separate out functions analytically continuable into the upper or lower complex half-plane of the argument, respectively. These operators can be expressed in terms of the Hilbert transform

$$\hat{H}f(x)=\frac{1}{\pi}\,\mathrm{V.\ P.}\int_{-\infty}^{\infty}\frac{f(x')dx'}{x'-x}$$

by means of the equation

$$P^{(\pm)}=\frac{1}{2}(1\mp i\hat{H}). \tag{31}$$

Accordingly, the operator $|\hat{k}|$ can be written in terms of $\hat{H}$ in the form

$$|\hat{k}|=-\hat{H}\frac{\partial}{\partial x}.$$

We now write the kinematic condition on the surface, along with the Bernoulli equation in conformal variables. We use the relation

$$\eta_t=\frac{\partial(y,x)}{\partial(t,x)}=\frac{1}{x_u}(y_t x_u-y_u x_t)$$

to transform the kinematic condition to

$$y_t x_u-y_u x_t=|\hat{k}|\psi. \tag{32}$$

Following Ref. 16, we can solve this equation for the time derivatives:

$$\frac{z_t}{z_u}=P^{(-)}\left(\frac{2i|\hat{k}|\psi}{|z_u|^2}\right). \tag{33}$$

In no way does this equation reflect the presence of the vortex, owing to the representation of the velocity potential in the special form (27). The right-hand side of Eq. (32) ($|\hat{k}|\psi$) represents the normal velocity component, which has does not have any vortex contribution.

In the transformation of the dynamical boundary condition, i.e., the Bernoulli equation at the free boundary $v=0$, the term $(\nabla\Phi)^2$ can be written in the form

$$(\nabla\Phi)^2|_{y=\eta}=\frac{1}{|z_u|^2}(\Phi_u^2+\Phi_v^2)|_{v=0}=\frac{1}{|z_u|^2}(\phi_u^2+\psi_v^2),$$

and the derivative is

$$\left.\frac{\partial\Phi}{\partial t}\right|_{y=\eta}=\phi_t-\phi_u\hat{H}\left(\frac{\hat{H}\psi_u}{|z_u|^2}\right)-\frac{(\hat{H}\psi_u)^2}{|z_u|^2}.$$

Hence, simple transformations reduce the Bernoulli equation for ϕ in conformal variables to the form

$$\phi_t-\phi_u\hat{H}\left(\frac{\hat{H}\psi_u}{|z_u|^2}\right)-\frac{(\hat{H}\psi_u)^2-(\phi_u)^2}{2|z_u|^2}+gy$$

$$-\sigma\frac{1}{1+|\hat{k}|y}\frac{\partial}{\partial u}\left(\frac{y_u}{|z_u|}\right)=0. \tag{34}$$

In the absence of a vortex $\phi=\psi$, in which case the equation describes surface waves proper together with their nonlinearities, so that the entire wave–vortex interaction is embodied in the difference between the potentials ϕ and ψ.

We have yet to write the vortex equation of motion. We know that the velocity $\tilde{\mathbf{v}}$ can be obtained from the total flow velocity field by subtracting the centrosymmetric field of the vortex itself and passing to the limit $x\to X$, $y\to Y$:

$$X_t=\tilde{v}_x(X,Y,t),\quad Y_t=\tilde{v}_y(X,Y,t)$$

or

$$\overline{Z}_t=\tilde{v}_x-i\tilde{v}_y=\left.\frac{\partial\tilde{\Pi}}{\partial z}\right|_{z=Z}.$$

Upon transformation to conformal variables, the latter equation can be rewritten as

$$Z'(W)\frac{d\overline{Z}}{dt}=\tilde{\Pi}'|_{w=W}.$$

To complete this set of equations, we must now determine the potential $\tilde{\Pi}$. From Eq. (27),

$$\tilde{\Pi}=i\gamma\ln\left(\frac{w-W}{\overline{W}-w}\right)+\Psi(w)-i\gamma\ln(z(w)-Z(W)).$$

Differentiating this equation with respect to w and passing to the limit $w\to W$, we finally have

$$Z'(W)\frac{d\overline{Z}}{dt}=\frac{-i\gamma}{W-\overline{W}}-\frac{i\gamma}{2}\frac{Z''(W)}{Z'(W)}+\Psi'(W). \tag{35}$$

Equations (33), (34), and (35) in conjunction with the equations

$$x-u=-\hat{H}y, \tag{36}$$

$$Z(W)-W=\frac{i}{2\pi}\int\frac{(z(u)-u)du}{u-W} \tag{37}$$

form a closed system describing the interaction of a point vortex with the free surface in conformal variables.

We close this section with an expression for the Lagrangian in conformal variables:

$$\mathscr{L}_{conf}=i\pi\gamma Z\frac{d\overline{Z}}{dt}+\int(\Phi_0(u)+\psi)\left(\frac{z_t\overline{z}_u-\overline{z}_tz_u}{2i}\right)du$$

$$-\int\frac{\psi|\hat{k}|\psi}{2}du+\frac{g}{2}\int\frac{(z-\overline{z})^2}{4}\frac{z_u+\overline{z}_u}{2}du$$

$$-\sigma\int(\sqrt{z_u\overline{z}_u}-1)du-\pi\gamma^2\left(\ln\left(\frac{\overline{W}-W}{i}\right)\right.$$

$$\left.+\frac{1}{2}\ln(Z'(W)\overline{Z'}(\overline{W}))\right). \tag{38}$$

The terms in the second row of this equation are responsible for surface waves. In the limit of small wave amplitudes and no vortex, the Lagrangian is quadratic in the variables ψ and η:

$$\mathscr{L}_s=\int du\,\psi y_t-\frac{1}{2}\int du(\psi|\hat{k}|\psi+gy^2+\sigma y_u^2).$$

The corresponding linear equations of motion give the surface wave dispersion law

$$\omega=\sqrt{gk+\sigma k^3}; \tag{39}$$

a surface wave with the wave vector k is localized in a layer having a thickness of the order of the wavelength ($\sim k^{-1}$).

The logarithmic terms in the Lagrangian (38) correspond to the interaction of a vortex with its mirror image and take into account the change in size of the vortex under the conformal transformation: $|\Delta z|\approx|Z'(W)||\Delta w|$. It will be shown in the Appendix that the last term in (38) causes the term $(i\gamma/2)Z''(W)/Z'(W)$ to appear in Eq. (35), and likewise gives the Bernoulli equation (34) a term characterizing vortex–wave interaction.

4. STEADY-STATE MOTION

We consider the steady-state solutions of Eqs. (33), (34), and (35) describing the propagation of a vortex along the x axis at a constant velocity c, and the surface deformations associated with it. In this case

$$\dot{U}\equiv\dot{W}=c,$$

and

$$\frac{\partial}{\partial t}=-c\frac{\partial}{\partial u}.$$

As a result, the kinematic condition in the form (32) can be integrated simply:

$$x(u')-u'=c\psi, \tag{40}$$

which by virtue of the analyticity of A yields

$$A(w')=c\Psi. \tag{41}$$

Here $u'=u-ct$ and $w'=w-ct$ (from now on we drop the primes from u and w).

The Bernoulli equation (34) can also be simplified considerably in the steady state. In a coordinate frame with the vortex at rest the flow moves from right to left with the velocity c. At the boundary (by virtue of the steady state) the stream function is constant $(=0)$, and the velocity therefore has only one tangential component:

$$V(u)=-c+\frac{i\gamma(W-\overline{W})}{|u-W|^2}. \tag{42}$$

As a result, Eq. (34) can now be written in the form

$$\frac{V^2}{2|z_u|^2}+gy-\sigma\frac{1}{1+|\hat{k}|y}\frac{\partial}{\partial u}\left(\frac{y_u}{|z_u|}\right)=\frac{c^2}{2}. \tag{43}$$

In the steady state with Eq. (41) taken into account, the vortex equation of motion (35) gives the velocity c as a function of the derivatives of $z(w)$ at the point $w=W$:

$$c=\frac{-i\gamma}{W-\overline{W}}-\frac{i\gamma}{2}\frac{Z''(W)}{Z'(W)}. \tag{44}$$

Equations (44) and (43) form a closed system, in which c plays the role of an eigenvalue.

We now consider the asymptotic behavior of Eq. (43) for large u, assuming that the deviation $y(u)$ is small. Linearizing, we have

$$\hat{L}y(u)=\frac{2c\gamma h}{u^2}. \tag{45}$$

Here the operator $\hat{L}$, which is given by the expression

$$\hat{L}=|\hat{k}|(V_{\mathrm{ph}}^2-c^2), \quad V_{\mathrm{ph}}^2=g|k|^{-1}+\sigma|k|,$$

determines the asymptotic behavior of y. The operator $\hat{L}$ is positive definite if the square of the velocity c is less than the square of the minimum phase velocity $V_{\mathrm{ph}}=\omega(k)/k$ of linear waves:

$$c^2<\min V_{\mathrm{ph}}^2=2gk_0,$$

where $\omega(k)$ is given by Eq. (39), and $k_0=\sqrt{g/\sigma}$. In this case the operator $\hat{L}$ is invertible, and accordingly Eq. (43) can have localized solutions. If

$$c^2>\min V_{\mathrm{ph}}^2,$$

the operator $\hat{L}$ is not invertible, and the solution of the linear equation (45) therefore contains an oscillating asymptotic form with wave numbers k that are roots of the equation

$$g|k|^{-1}+\sigma|k|=c^2. \tag{46}$$

This equation is the Cherenkov radiation condition, which is satisfied by two values of $k_{1,2}$:

$$k_{1,2}=\frac{c^2}{2\sigma}[1\pm\sqrt{1-(V_{\min\mathrm{ph}}/c)^4}]. \tag{47}$$

Here k_1 lies in the capillary region of the spectrum ($k_1>k_0$) and k_2 lies in the gravity region ($k_2<k_0$). The group velocity at $k=k_1$ is greater than c, and for gravity waves with $k=k_2$, in contrast, it is less than c. Consequently, for $c^2>\min V_{\mathrm{ph}}^2$ an oscillating capillary wave front is produced in front of the vortex, and a gravity wave front is produced behind it (see Sec. 6 for more details on this subject).

Thus, the localized solutions differ from nonlocalized solutions primarily in the relation between the vortex velocity c and the minimum surface wave phase velocity. An important consideration is the fact that for $c\geqslant\min V_{\mathrm{ph}}$ the localized object—the bound states of the vortex and the waves—cannot be a steady-state formation; it will necessarily radiate surface waves on account of the Cherenkov effect, losing energy in the process. In general, however, a lack of Cherenkov radiation still does not fully guarantee steady-state motion of the vortex and the surface waves bound to it (see below in this regard).

We now turn to the solution of Eqs. (43) and (44). Above all, we note that these solutions are stationary points of the Hamiltonian $\mathscr{H}$ for a fixed x component of the momentum $\mathscr{P}$:

$$\delta(\mathscr{H}+c\mathscr{P})=\delta S_c=0, \tag{48}$$

where

$$\mathscr{P}=-2\pi\gamma Y+\int\phi\eta_x dx,$$

and the action for the steady-state solutions S_c has the form

$$\begin{aligned}S_c=\int\Bigg(&-\frac{ic^2}{4}\overline{A}A'+\sigma(\sqrt{(1+A')(1+\overline{A}')}-1)\\&+\frac{g}{2}\left(\frac{A-\overline{A}}{2i}\right)^2\left(1+\frac{A'+\overline{A}'}{2}\right)+\overline{\mu}P^{(+)}A\\&+\mu P^{(-)}\overline{A}\Bigg)du-\pi\gamma ci(W-\overline{W})+\pi\gamma^2\Bigg[\ln\left(\frac{W-\overline{W}}{-i}\right)\\&+\frac{1}{2}\ln(1+A'(W))(1+\overline{A}'(\overline{W}))\Bigg].\end{aligned} \tag{49}$$

Here we have made use of relations (40), which can be used to eliminate the potential Ψ, and we have introduced the relation $P^{(+)}A=P^{(-)}\overline{A}=0$, which guarantees the analyticity of A and $\overline{A}$ (μ and $\overline{\mu}$ are the corresponding Lagrange multipliers).

Now the variation of S_c with respect to W gives an equation for the vortex depth W, which coincides with (44). It follows at once from this equation that for a small deformation of the free surface (in which case the second term on the right-hand side of Eq. (44) can be disregarded) the vortex velocity is inversely proportional to the distance $2h$ from the image vortex:

$$c\approx c_0=\frac{\gamma}{2h}.$$

Examining the variation with respect to A, we obtain the steady-state Bernoulli equation in a form that differs from (43), but is more practical for the ensuing analysis[1]:

$$\begin{aligned}P^{(-)}\Bigg(&ic^2A'+2\sigma\frac{d}{du}\sqrt{\left(\frac{1+A'}{1+\overline{A}'}\right)}+g(\overline{A}A'-AA'-A)\\&+\frac{i\gamma^2}{1+\overline{A}'(\overline{W})}\frac{1}{(u-\overline{W})^2}\Bigg)=0.\end{aligned} \tag{50}$$

The solutions of Eqs. (44) and (50) depend on two dimensionless parameters: the Froude number $F=\gamma^2/gh^3$ and $T=\sigma/gh^2$, which represent the ratio of c_0^2 and the characteristic velocity of capillary waves (with wavelength $\sim h$) to the characteristic velocity of gravity waves.

Let a vortex be situated at the point $W=-ih$ in a coordinate frame moving at the velocity c. Introducing the dimensionless variables

$$\xi=u/h, \quad a(\xi,F,T)=A/h, \quad c=c_0\tilde{c},$$

we can rewrite Eqs. (44) and (50) in the form

$$\tilde{c}=1-i\frac{a''(-i)}{1+a'(-i)}, \tag{51}$$

$$a=P^{(-)}\left(iF\frac{\tilde{c}^2}{4}a'+2T\frac{d}{d\xi}\sqrt{\left(\frac{1+a'}{1+\overline{a'}}\right)}+\bar{a}a'-aa' \right.$$
$$\left. +\frac{iF}{1+\overline{a'}(-i)}\frac{1}{(\xi+i)^2}\right). \tag{52}$$

For small flow velocities the solution of the system (51), (52) can be represented by an asymptotic expansion in powers of the parameter F.

The expansion of the velocity c as a function of the depth is given by the expression

$$\tilde{c}=1+\frac{3}{8}F+\frac{21}{16}F^2+\left(\frac{2439}{256}-\frac{15}{8}T\right)F^3$$
$$+\left(\frac{52629}{512}-\frac{213}{8}T\right)F^4+\dots\,. \tag{53}$$

The surface profile can be determined parametrically from the relations

$$A(u)=h(a_1(\zeta,T)F+a_2(\zeta,T)F^2+a_3(\zeta,T)F^3+\dots), \tag{54}$$

where

$$a_1=\frac{i}{\zeta^2}, \tag{55}$$

$$a_2=-\frac{2}{\zeta^5}+\frac{1}{\zeta^3}+\frac{(3/4)i}{\zeta^2}-\frac{3/8}{\zeta}, \tag{56}$$

$$a_3=-\frac{14i}{\zeta^8}+\frac{10i}{\zeta^6}-\frac{11/2}{\zeta^5}+\frac{i(6T-9/2)}{\zeta^4}+\frac{17/4}{\zeta^3}$$
$$+\frac{(123/32)i}{\zeta^2}-\frac{39/16}{\zeta}. \tag{57}$$

(Here $\zeta=\xi-i$.)

The first-order terms in F can be determined directly from the equations. In this order the surface profile can be determined in explicit form. The following equation is obtained to within higher-order terms from Eq. (55):

$$y=Fh^3\frac{x^2-h^2}{(x^2+h^2)^2}.$$

At $x=0$ (strictly above the vortex) the surface has a minimum with a negative value of y (dip), and at $x^2>h^2$ the surface rises above its average level, consistent with the results of numerical integration.[12] The occurrence of a dip in the surface is attributable to the fact that the velocity V (42) above the vortex is greater in absolute value than the value of V at infinity, so that by virtue of Bernoulli's equation the pressure above the vortex is lower than the pressure at infinity.

We used the Maple V software package to compute the leading terms of the expansion (up to fourth order in F) in Eqs. (53) and (54). It is important to note that the leading coefficients of the expansion $a(\xi)$ decrease more slowly than $a_1(\xi)$ at large ξ. The existence of a finite coefficient of surface tension in this case is a fundamental prerequisite for the existence of a stationary solution. As we have seen above, only under this condition is it impossible to have Cherenkov radiation over a certain velocity range. However, the expansion (53), (54) can be obtained even at $T=0$, implying that in reality these series in F diverge, and their use must be restricted to a finite number of terms and small F. Another obvious condition for their applicability is smallness of the velocity in comparison with $V_{\mathrm{min\,ph}}$, so that localized steady-state solutions no longer exist for $c>V_{\mathrm{min\,ph}}$. The necessary conditions for the asymptotic expansion (54) can therefore be written in the form

$$\frac{\gamma^2}{gh^3}\ll 1, \quad \frac{\gamma}{h}\ll(\sigma g)^{1/4}. \tag{58}$$

5. SURFACE WAVE BOUND STATES

A number of structural transformations (bifurcations) take place as the velocity c increases in the stationary solution discussed above. To understand the reason for these transformations, as in the preceding section, we consider the case of large depths h ($k_0h\gg 1$), in which the interaction of surfaces with the vortex can be regarded as weak, and the vortex moves at the constant velocity $c\approx c_0=\gamma/2h$ $\times(<V_{\mathrm{min\,ph}})$. In this approximation the surface is almost flat and, accordingly, the difference between the conformal depth and the ordinary depth is small.

We transform to a reference frame comoving with the vortex. In this frame the interaction between the freestream flow with velocity c and the dipole pair (the vortex and its image) results in the formation along the surface of a time-invariant, highly inhomogeneous flow moving with the velocity (42):

$$V(x)=c_0\frac{3h^2-x^2}{h^2+x^2}. \tag{59}$$

This flow has velocity $-c_0$ at infinity and $3c_0$ at the center (at $x=0$), which is three times the freestream velocity V_∞.

Inasmuch as $V(x)$ varies slowly in comparison with the characteristic surface wavelength ($\sim k_0^{-1}$), we are justified in writing the semiclassical equation of motion for a surface wave packet having the carrier wave number k and center coordinate at the point x:

$$\dot{k}=-\frac{\partial\omega(k,x)}{\partial x}, \tag{60}$$

$$\dot{x}=\frac{\partial\omega(k,x)}{\partial k}. \tag{61}$$

Here

$$\omega(k,x)=kV(x)+\sqrt{|k|(g+\sigma k^2)} \tag{62}$$

is the wave frequency, which undergoes an inhomogeneous Doppler shift because of the dependence of the flow velocity on x. During the motion of the packet, $\omega(k,x)$ remains time-invariant.

Closed level lines $\omega(k,x)=\text{const}$ in the phase plane correspond to finite periodic motion of the wave packet. Such motion is quantized according to the well-known Bohr–Sommerfeld rule:

$$\oint_{\omega(k,x)=\omega_n} k\,dx=2\pi(n+\alpha_n), \tag{63}$$

where n labels the levels, α_n is a number of the order of unity, and ω_n is the eigenfrequency of the level.

The specific form of $V(x)$ leads to the conclusion that bound states with near-zero frequency emerge well before $V_{\text{min ph}}$ is reached. This is an important consideration from the standpoint of the behavior of the steady-state solution as the velocity varies. Beginning with the velocity $c=V_{\text{min ph}}/3$, a finite phase trajectory exists at zero frequency for negative values of k. Two reflected infinite trajectories at zero frequency for positive k also occur for $V_{\text{min ph}}/3$. In terms of bound states this means that the frequency of the first level passes through zero at a velocity close to $V_{\text{min ph}}$. And since the vortex interacts with bound states in spite of its exponential weakness, there is a resonant increase in the contribution of the given level to the surface profile, the whole effect changing sign after the frequency passes through zero. Nonlinearity prevents the resonant mode from increasing without bound, and leads to bifurcation. We are essentially dealing with two different solutions before and after the zero crossing, as long as we trace only solutions for which the surface perturbation has a small amplitude.

With a further increase in the velocity of the vortex, the frequency of the second bound state passes through zero, and so on.

The nonlinear dynamics produced in connection with the stated problem in the system of discrete levels is still not entirely clear and requires separate investigation. Various scenarios for elaborating the behavior as the velocity increases, for example, the scenario of steady-state motion becoming unstable and evolving into an unsteady state after the emergence of several bound states, may seem entirely probable, but still needs to be verified.

We note that the inhomogeneous Doppler effect culminating in such fundamental results drops out of sight in Ref. 12 and in the relevant section of the previously cited book.[11]

6. ASCENT OF A POINT VORTEX

As stated, it is impossible for a vortex to be in steady-state motion under the condition $c>V_{\text{min ph}}$, owing to the Cherenkov radiation of waves. The back influence of radiation on the vortex causes the vertical location of the vortex to vary with time, and the vortex rises. The radiated waves are situated in different parts of the spectrum, one for gravity waves and one for capillary waves. A capillary wave is radiated in the forward direction, and a gravity wave moves in the backward direction. Without radiation, despite the attraction of the vortex to its image, the distance from the vortex to the surface remains constant on the average. We recall that for two-dimensional flows without a free boundary, a dipole pair of point vortices executes stable motion at constant velocity in a direction perpendicular to the dipole.

The existence of Cherenkov radiation has the effect of reducing the distance from the surface by virtue of energy conservation, i.e., the radiated waves carry off positive energy, reducing the interaction energy of the vortex with its image (at the same time increasing its absolute value).

Consequently, the vortex exhibits unsteady dynamics in the given situation. However, under two conditions indicated below, the motion can be regarded as quasi-steady, because the amplitudes of the radiated surface waves are exponentially small in this case, and the ascent of the vortex is slow in such a regime. This quasi-steady-state property refers to the undulating surface zone in a time interval when the wave fronts have already moved far away from the vortex. We note that if the vortex were maintained in uniform motion by some external force, the surface could be stationary relative to the vortex in the undulating zone. During slow ascent the surface profile also changes slowly. From the standpoint of calculating the Cherenkov wave amplitudes, the two regimes of vortex motion—strictly uniform motion in the presence of an external force and quasiuniform motion with slow ascent for zero external force—differ very little. An appreciable difference shows up only when total energy and momentum conservation are violated in strictly steady flow.

We now summarize the conditions for radiation from the vortex to be considered weak. First of all, it is required that the Cherenkov wavelength be small in comparison with the depth h (semiclassical condition). This requirement is satisfied when

$$c\gg V_{\text{min ph}}, \tag{64}$$

whereupon

$$k_2h\approx 4gh^3/\gamma^2=4/F\gg 1. \tag{65}$$

To calculate the Cherenkov radiation, we use the steady-state Bernoulli equation (43):

$$\frac{V^2}{2|z_u|^2}+gy-\sigma\frac{1}{1+|\hat{k}|y}\frac{\partial}{\partial u}\left(\frac{y_u}{|z_u|}\right)=\frac{c^2}{2}.$$

Assuming that surface perturbations are small (by virtue of the smallness of F) and assuming

$$|z'(u)|^2\approx 1+2|\hat{k}|y,\quad c\approx c_0=\gamma/2h, \tag{66}$$

after linearizing (43) we have

$$\hat{L}y(u)=\frac{1}{2}(V^2(u)-c_0^2),\quad V(u)=c_0\frac{3h^2-u^2}{h^2+u^2}. \tag{67}$$

Here, in contrast with Eq. (45), the operator $\hat{L}$ takes the inhomogeneous Doppler effect into account:

$$\hat{L}=V^2(u)|\hat{k}|-g+\sigma\frac{d^2}{du^2}.$$

To solve this equation, we make use of the fact that the function $y(u)$ represents the imaginary part of the function $A(u)=ha(u/h)=ha(q)$, which is analytic in the lower half-plane. An equation for $a(q)$ can be obtained from Eq. (67) by applying the projector $P^{(-)}$. In dimensionless variables it has the form

$$-v^2(q)a'+\frac{\lambda}{i}a-\frac{\epsilon}{i}a''=f(q,d_1,d_2). \tag{68}$$

Here, by definition,

$$V^2=c_0^2v^2,\quad v^2(q)=\left(\frac{q^2-3}{q^2+1}\right)^2=1-\frac{4}{(q-i)^2}-\frac{4}{(q+i)^2},$$

$$f(q,d_1,d_2)=4\left[\frac{1}{(q-i)^2}+\frac{d_2}{q+i}-\frac{\overline{d_2}}{q-i}+\frac{d_1}{(q+i)^2}-\frac{\overline{d_1}}{(q-i)^2}\right], \tag{69}$$

where $\lambda=4/F\gg1$ is a large parameter, and $\epsilon=4h\sigma/\gamma^2\ll1$ is a small parameter. It is readily apparent that terms proportional to $d_1=a'(-i)$ and $d_2=a''(-i)$ remove the singularities of the solution in the lower half-plane. It follows from the asymptotic expansion of the solution in F that the quantities d_1 and d_2 are small ($d_1,d_2\sim F$) for small F. Consequently, their contribution in the leading approximation must be negligible in comparison with the free term in f.

After the stated simplifications we have a linear, inhomogeneous, second-order ordinary differential equation with asymptotic conditions at infinity that stipulate the absence of a gravity wave in the limit $q\to+\infty$ and the absence of a capillary wave in the limit $q\to-\infty$ (radiation condition). Analyzing the overall picture described at the beginning of the present section, it is easy to see that these asymptotic properties of the steady-state solution, along with free unsteady waves, are what ensure the propagation of wave fronts away from the vortex.

To solve the inhomogeneous equation, we first analyze the properties of the solutions of the homogeneous equation (Eq. (68) without the right-hand side), since the desired solution of Eq. (68) can be expressed in terms of the homogeneous-case solutions by the method of variation of constants. When the semiclassical condition is satisfied, the homogeneous solutions $a_1(q)$ and $a_2(q)$ can be found by a procedure similar to that used in quantum mechanics[23]:

$$a_{1,2}(q)\approx C_g^{1,2}\frac{\exp\left\{-i\int^q\kappa_g(x)dx\right\}}{\sqrt{\lambda-\epsilon\kappa_g^2}}+C_c^{1,2}\frac{\exp\left\{-i\int^q\kappa_c(x)dx\right\}}{\sqrt{\lambda-\epsilon\kappa_c^2}}. \tag{70}$$

Here the coordinate dependence of the wave numbers $\kappa_g(q)$ and $\kappa_c(q)$ is given by the expression

$$\kappa_{g,c}(q)=\frac{v^2(q)\mp\sqrt{v^4(q)-4\lambda\epsilon}}{2\epsilon}. \tag{71}$$

At infinity we have $\kappa_g\to\kappa_2$ and $\kappa_c\to\kappa_1$. These equations are valid far from the ''turning points,'' where $\kappa_g=\kappa_c$, and the roots in the denominator vanish. In our case there are four such points, two of which distinguish the ''classically forbidden'' region near $q=-\sqrt{3}$, while the other two apply near $q=+\sqrt{3}$. Each of these two wave numbers has an imaginary part, which gives rise to the exponential behavior of $a_1(q)$ and $a_2(q)$. Accordingly, there are three classically allowed regions, and in a typical situation the wave amplitudes in them differ sharply, so that their ratio in transit across the forbidden region is exponentially large, and the exponent of the exponential function contains only a number of the order of $\lambda(c/V_{\text{min ph}})$. A similar result can be obtained simply by estimating the integral in the forbidden region:

$$\int\ \operatorname{Im}\{\kappa(q)\}dq\sim\int\frac{\sqrt{x^4-\lambda\epsilon}}{\epsilon}dx\sim\lambda(\lambda\epsilon)^{-\frac{1}{4}}\sim\lambda(c/V_{\text{min ph}}). \tag{72}$$

In the vicinity of the ''turning points'' themselves, of course, Eqs. (70) are invalid, because in fact the homogeneous solutions do not have any singularities there. Inasmuch as we do not use the corresponding asymptotic representation, we merely note that they can be expressed approximately in terms of the Airy function, as in quantum mechanics. The only difference is that once the first derivative term in Eq. (68) is removed, we obtain the Schrödinger equation with a complex potential. The ''force'' in the vicinity of the turning point also has an imaginary part, and the Airy function is therefore written with a complex argument.

To solve the inhomogeneous equation with the radiation conditions at infinity, it is useful to choose the functions $a_1(q)$ and $a_2(q)$ in such a way that the solution $a_1(q)$ will contain only a capillary wave far ahead, and the solution $a_2(q)$ will contain only a gravity wave far behind:

$$a_1(q)\to\exp(-i\kappa_1q),\quad q\to+\infty, \tag{73}$$

$$a_2(q)\to\exp(-i\kappa_2q),\quad q\to-\infty. \tag{74}$$

The desired solution can then be written in the form

$$a(q)=-\frac{i}{\epsilon}\left(a_1(q)\int_{-\infty}^{q}f\frac{a_2(x)}{W(x)}dx+a_2(q)\int_{q}^{+\infty}f\frac{a_1(x)}{W(x)}dx\right), \tag{75}$$

where

$$W(q)=a_1a_2'-a_2a_1'\propto\exp\left(-\frac{i}{\epsilon}\int^q v^2(x)dx\right)$$

is the Wronskian for the two given solutions. It is obvious that the surface wave amplitudes are given by the equations

$$|a_g(-\infty)|=\left|-\frac{i}{\epsilon}\int_{-\infty}^{+\infty}f\frac{a_1(q)}{W(q)}dx\right|,$$

$$|a_c(+\infty)| = \left| -\frac{i}{\epsilon}\int_{-\infty}^{+\infty} f\,\frac{a_2(q)}{W(q)}\,dx \right|. \tag{76}$$

We now assume that we are not in the vicinity of a resonance at a quasi-discrete level of the operator $\hat{L}$. This is the second quasi-steady-state condition. The functions a_1 and a_2 are then constructed in such a way that the amplitude is a maximum for the first function when $q<-\sqrt{3}$ and for the second function when $q>+\sqrt{3}$ (resonance would correspond to the middle part). Moreover, for $c \gg V_{\text{min ph}}$ only one of the two waves involved in each of the solutions a_1 and a_2 [see Eq. (70)] makes the main contribution to the integrals, specifically the one that oscillates with the smaller wave number after division by the Wronskian. It is readily verified that this situation is equivalent to the possibility of determining the amplitude $a_g(q)$ from the solution of the first-order equation without capillarity:

$$-v^2(q)a' + \frac{\lambda}{i}a = \frac{4}{(q-i)^2}.$$

Here the choice of points from which the gravity wave amplitude begins to derive its amplitude is dictated by Eq. (76). For example, after straightforward calculations the expression for the backward-radiated wave reduces to the form

$$|a_g(-\infty)| = \left| \frac{i}{\lambda}\int_{-\infty}^{-\sqrt{3}} f'(q,0,0)\exp(i\lambda P(q))\,dq \right|, \tag{77}$$

where

$$P(q) = \int \frac{dq}{v^2(q)} = q + \frac{8}{3}\frac{q}{(3-q^2)} + \frac{8}{3\sqrt{3}}\ln\left(\frac{q-\sqrt{3}}{q+\sqrt{3}}\right).$$

Next, in the above integral (77) we move the contour of integration into the upper half-plane and integrate by parts, eliminating the pole at the point $q=+i$. We apply the steepest descent method to the resulting integral. The saddle point coincides with $q=i$. As a result, we have

$$|a_g(-\infty)| \approx \left| \frac{i+1/\sqrt{3}}{2} \right| \left(\frac{12}{\lambda}\right)^{1/3} \Gamma\left(\frac{1}{3}\right) \times \exp\left(-\lambda\left(\frac{5}{3} + \frac{16\pi}{9\sqrt{3}}\right)\right). \tag{78}$$

The gravity wave amplitude immediately to the right of the point $+\sqrt{3}$ is given by an analogous expression with the limits of integration from $\sqrt{3}$ to $+\infty$. Consequently, after repeating the calculations we find that its modulus is the same. In the given limit $c \gg V_{\text{min ph}}$, after forward reflection and conversion to a capillary wave, the amplitude of the given wave changes by the factor $\sqrt{\lambda/\epsilon\kappa_c^2}$. This result can be obtained by an approach similar to that used in the derivation of semiclassical boundary conditions in quantum mechanics.[23]

The rates of ascent of the vortex is determined by energy conservation:

$$\pi\gamma^2\dot{h}/h \approx -(c/2)\left(\overline{g y^2(-\infty)} + \overline{\sigma y'^2(+\infty)}\right). \tag{79}$$

Here we have made use of the fact that when conditions (64) are satisfied, the gravity wave has the following velocity in the vortex frame:

$$\left.\frac{\partial\omega}{\partial k}\right|_{k=k_1} - c \approx -\frac{c}{2},$$

and accordingly the capillary wave velocity is approximately $c/2$. When the indicated amplitude variation of the second wave is taken into account after its conversion to a capillary wave, we find that the two waves make identical contributions to the rate of ascent of the vortex.

The equation for the vortex trajectory $h(x)$ in the interval between neighboring resonant quasilevels finally has the form

$$\frac{dh}{dx} = \frac{\dot{h}}{c} = -\frac{\Gamma^2(1/3)}{2\pi}\left(\frac{gh^3}{3\gamma^2}\right)^{1/3} \times \exp\left(-\frac{8gh^3}{\gamma^2}\left(\frac{5}{3} + \frac{16\pi}{9\sqrt{3}}\right)\right). \tag{80}$$

Hence the vortex ascends by virtue of Cherenkov radiation. The rate of ascent increases as h decreases, and the radiated power increases accordingly. Inasmuch as γ remains constant in this process, the ascent of the vortex causes it to accelerate simultaneously in the horizontal direction.

The rate of ascent (80) differs from the value given in Ref. 12, not only in the factor preceding the exponential, but also in the exponent itself. The exponents agree only in order of magnitude.

In the case of resonance at a quasidiscrete level, the amplitude of the homogeneous solutions is a maximum in the region of classically finite motion. The amplitudes of the Cherenkov waves must now be estimated using the subbarrier transfer coefficient, which is of the order of $\exp[-C_1\lambda(c/V_{\text{min ph}})]$. The result

$$|a^{res}_{g,c}| \sim \exp\left(\lambda\left(C_1\frac{c}{V_{\text{min ph}}} - C_2\right)\right)$$

shows that the amplitudes of these waves are not small and, accordingly, the quasi-steady-state regime, like the condition for validity of the linear approximation, fails in the vicinity of resonance. Investigation of the transition of the system through a quasilevel requires an analysis of the time-dependent equations with allowance for nonlinearity and poses a difficult task at this stage. We speculate that the quasisteady ascent regime alternates with short-time, sharp energy spikes as resonances are transited.

This trend persists at least until

$$\frac{\gamma^2}{gh^3} < 1, \quad \frac{\gamma}{h} > (\sigma g)^{1/4}. \tag{81}$$

When the depth h is less than or commensurate with the gravity wavelength, the interaction of a vortex with the surface can no longer be calculated in the linear approximation, the behavior of the system becomes highly nonlinear, and it then becomes extremely difficult to decide whether the vortex will continue to rise. Most likely the question can be answered only by numerically integrating the time-dependent equations (33), (34), and (35).

7. THREE-DIMENSIONAL DYNAMICS

In this section we discuss one aspect of the three-dimensional dynamics of a vortex filament: its instability against kink perturbations due to the presence of a surface.

We consider a vortex filament parallel to the surface. Let the relative deformation of the fluid surface be small, so that interaction of the vortex filament with free waves can be disregarded. Accordingly, we ignore the potential energy of the surface in calculating the Hamiltonian and calculate the kinetic energy of the fluid according to an equation in which the velocity is expressed in terms of the scalar potential. Then, integrating by parts, we reduce the equation to half the integral over the surface spanning two filaments. Allowance must be made for the fact that the scalar potential acquires an increment $2\pi\gamma$ when it skirts any one filament. As a result, we have

$$\mathscr{H}=\frac{\pi\gamma}{2}\int_S(\mathbf{V}\,d\mathbf{S}). \tag{82}$$

This integral diverges logarithmically for an infinitely thin filament. For a filament of finite thickness ε_0 the integral can be cut off at a length $\sim\varepsilon_0$.

In the long-wavelength limit, where the characteristic radius of curvature R_0 of the filament is large in comparison with $Y\,(Y'^2\ll 1)$, the Hamiltonian (82) can be approximately written in the form

$$\mathscr{H}_{\text{appr}}=\pi\gamma^2\int\sqrt{1+X'^2}\ln\left(\frac{Y}{\varepsilon_0}\right)dz. \tag{83}$$

The equations for the vortex coordinates $X(z,t)$ and $Y(z,t)$ can be obtained with the aid of (22) by varying the Hamiltonian (83):

$$\dot{X}=\frac{\gamma}{2}\,\frac{\sqrt{1+X'^2}}{Y}, \tag{84}$$

$$\dot{Y}=\frac{\gamma}{2}\left(\frac{X'\ln(Y/\varepsilon_0)}{\sqrt{1+X'^2}}\right)'. \tag{85}$$

In this approximation the relative error in Eq. (84) is of the order of Y/R_0, and in the equation for Y it is of the order of $1/\ln(Y/\varepsilon_0)$. An important feature here is that the values of X'^2 and the range of relative variations of Y are not assumed to be small, so that Eqs. (84) and (85) describe arbtirary deviations of the filament configuration from a straight line, not just small ones.

For small deviations these equations (after linearization) describe instability with a growth rate that depends linearly on $|k|$:

$$\Gamma(k)=|k|\frac{\gamma}{2Y_0}\sqrt{\ln(Y_0/\varepsilon_0)}.$$

This instability is none other than the long-wavelength limit of the Crow instability[17] for symmetric perturbations of two antiparallel vortex filaments.

The physical reason for this instability is the time lead of perturbed parts of the filament close to the surface relative to more distant parts. As a result, the bend in the vortex filament as seen from above (projected onto the surface) exhibits a pronounced focusing character (in this sense the given instability is similar to the Kadomtsev–Petviashvili instability[24,25]). This qualitative picture carries through to the nonlinear stage, when the filament can no longer be regarded as almost straight. As a result of the development of Crow instability, the closest parts of the filament are greatly stretched and rapidly approach the surface, where dissipation, which is actually always present because of viscosity and because of Cherenkov wave radiation in the presence of a free boundary, causes the filament to break and results in the formation of vortex half-rings that begin and end at the surface of the fluid.

8. CONCLUSION

In the paper we have succeeded in clarifying only the simplest aspects of the dynamics of a vortex–surface system. We have left untouched such important problems as the highly nonstationary motion of a point vortex in two dimensions, and the stability of various flow regimes, particularly in three dimensions—for example, the evolution of the system after a filament reaches the surface in three-dimensional space. Some of these problems can be solved by numerical modeling, while others require further analytic study.

This work was supported by the Russian Fund for Fundamental Research (Grant 97-01-00093). E. A. Kuznetsov was supported by grants from the International Association for the Promotion of Cooperation with Scientists from the Independent States of the Former Soviet Union (INTAS 96-0413) and the North Atlantic Treaty Organization (NATO Lineage Grant OUTR.LG 970583); V. P. Ruban was supported by a Landau Scholarship Grant.

APPENDIX

Here we show how the dynamical equations (35), (34), and (33) can be obtained from the variational principle in conformal variables. Variation of the Lagrangian (38) with respect to $\psi(u)$ yields the equation

$$(1/2i)(z_t\overline{z_u}-\overline{z_t}z_u)=|\hat{k}|\psi. \tag{86}$$

Multiplying by $1/|z_u|^2$ and applying the projector $\hat{P}^{(-)}$, we obtain

$$\frac{z_t}{z_u}=\hat{P}^{(-)}\left(\frac{2i|\hat{k}|\psi}{|z_u|^2}\right), \tag{87}$$

which coincides with Eq. (33).

Next, varying the Lagrangian with respect to W and $\overline{W}$, we can make a substitution therein according to the kinematic relation (86). This operation results in the substitution

$$\int \Phi_0(u)\frac{z_t\overline{z_u}-\overline{z_t}z_u}{2i}du\to 2\pi\gamma\left(\frac{\Psi(W)-\overline{\Psi}(\overline{W})}{2i}\right),$$

whereupon the vortex equation of motion (35) is easily obtained by variation with respect to W:

$$Z'(W)(\overline{Z'}(\overline{W}))\dot{\overline{W}}+\overline{Z}_t(\overline{W}))$$
$$=\frac{-i\gamma}{W-\overline{W}}-\frac{i\gamma}{2}\frac{Z''(W)}{Z'(W)}+\Psi'(W). \tag{88}$$

We now derive Bernoulli's equation. To prevent the calculations from becoming too cumbersome, for now we exclude surface tension from the discussion. Varying with respect to $\overline{z(u)}$ and invoking the condition for analyticity of this function in the upper half-plane, in the first step of the calculations we have the equation

$$P^{(-)}\left[\left(\phi_t+g\frac{(z-\overline{z})}{2i}\right)z_u-\phi_u z_t-\frac{\gamma^2}{2}\frac{1}{\overline{Z'}(\overline{W})}\frac{1}{(u-\overline{W})^2}\right.$$
$$\left.+\frac{\gamma}{i}\frac{(Z'(W)\dot{W}+Z_t(W))}{(u-\overline{W})}\right]=0. \tag{89}$$

We then use Eqs. (87) and (88) to eliminate the time derivatives $z_t(u)$ and $\dot{W}$ from the foregoing equation. At this point we note that the relation $P^{(-)}f=0$ is neither more nor less than a statement of the analyticity of f in the upper half-plane. It is still valid, therefore, after f is multiplied by $\overline{z_u}$:

$$P^{(-)}(f\overline{z_u})=0.$$

Making use of this property and taking into account the identities

$$P^{(-)}\left[\frac{\overline{z_u}}{(u-\overline{W})}\right]=\frac{\overline{Z'}(\overline{W})}{(u-\overline{W})}, \tag{90}$$

$$P^{(-)}\left[\frac{\overline{z_u}}{(u-\overline{W})^2}\right]=\frac{\overline{Z''}(\overline{W})}{(u-\overline{W})}+\frac{\overline{Z'}(\overline{W})}{(u-\overline{W})^2} \tag{91}$$

from Eq. (89) we obtain

$$P^{(-)}\left(\left(\psi_t+\Phi_{0t}+g\frac{(z-\overline{z})}{2i}-(\psi_u+\Phi_{0u})\hat{P}^{(-)}\right.\right.$$
$$\left.\times\left(\frac{2i|\hat{k}|\psi}{|z_u|^2}\right)\right)|z_u|^2-\frac{\gamma^2}{2}\frac{1}{(u-\overline{W})^2}$$
$$\left.+\frac{\gamma}{i}\frac{1}{(u-\overline{W})}\left(\frac{i\gamma}{\overline{W}-W}+\overline{\Psi}'(\overline{W})\right)\right)=0. \tag{92}$$

To continue, we must now sum the latter equation with its complex conjugate. Second-order terms in γ are gathered into the expression

$$-\frac{\gamma^2}{2}\frac{(W-\overline{W})^2}{|u-W|^4},$$

and the zeroth-order terms are assumed to coincide with those obtained previously[16] for purely potential flow:

$$|z_u|^2\left(\psi_t-\psi_u\hat{H}\left(\frac{\hat{H}\psi_u}{|z_u|^2}\right)-\frac{(\hat{H}\psi_u)^2-(\psi_u)^2}{2|z_u|^2}+gy\right).$$

We finally consider first-order terms. We first simplify the following first-degree equation in γ:

$$-|z_u|^2\Phi_{0u}\hat{H}\left(\frac{\hat{H}\psi_u}{|z_u|^2}\right)-\hat{H}(\Phi_{0u}\hat{H}\psi_u)-i\gamma\left(\frac{\overline{\Psi}'(\overline{W})}{(u-\overline{W})}\right.$$
$$\left.-\frac{\Psi'(W)}{(u-W)}\right). \tag{93}$$

Here, the second term is equal to

$$-\gamma\left(P^{(-)}\left(\left(\frac{1}{u-W}-\frac{1}{u-\overline{W}}\right)\left(\frac{\Psi'(u)-\overline{\Psi}'(u)}{2i}\right)\right)\right.$$
$$\left.+\text{c.c.}\right)=\frac{i\gamma}{2}\left(\frac{\overline{\Psi}'(\overline{W})-\Psi_u(u)}{u-\overline{W}}+\frac{\Psi_u(u)-\Psi'(W)}{u-W}\right.$$
$$\left.-\text{c.c.}\right)=\frac{\overline{\Psi}_u(u)+\Psi_u(u)}{2}i\gamma\left(\frac{1}{u-W}-\frac{1}{u-\overline{W}}\right)$$
$$+i\gamma\left(\frac{\overline{\Psi}'(\overline{W})}{(u-\overline{W})}-\frac{\Psi'(W)}{(u-W)}\right).$$

The last term is canceled by the third term of (93), so that all first-order terms can be written in the form

$$i\gamma\left(\frac{-\dot{W}}{u-W}+\frac{\dot{\overline{W}}}{u-\overline{W}}\right)|z_u|^2+i\gamma\left(\frac{1}{u-W}-\frac{1}{u-\overline{W}}\right)$$
$$\times\left(\psi_u-\left|z_u\right|^2\hat{H}\left(\frac{\hat{H}\psi_u}{|z_u|^2}\right)\right).$$

Adding the foregoing expressions, we finally obtain Bernoulli's equation (34) in conformal variables.

*)E-mail: kuznetso@itp.ac.ru

1)The equivalence of this equation and Eq. (43) is verified by approximately the same procedure as set forth in the Appendix for Eq. (34).

---

[1] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., rev., Pergamon Press, Oxford–New York (1987).
[2] V. E. Zakharov, Prikl. Mekh. Tekh. Fiz. , No. 2, 86 (1968).
[3] V. E. Zakharov and N. E. Filonenko, Dokl. Akad. Nauk SSSR **170**, 1292 (1966) [Sov. Phys. Dokl. **11**, 881 (1967)].
[4] E. A. Kuznetsov and P. M. Lushnikov, Zh. Éksp. Teor. Fiz. **108**, 614 (1995) [JETP **81**, 332 (1995)].
[5] E. A. Kuznetsov and M. D. Spector, Zh. Éksp. Teor. Fiz. **71**, 262 (1976) [Sov. Phys. JETP **44**, 136 (1976)].
[6] E. A. Kuznetsov, M. D. Spector, and V. E. Zakharov, Phys. Rev. E **48**, 1283 (1994); Phys. Lett. A **182**, 387 (1993).
[7] A. I. D'yachenko, V. E. Zakharov, and E. A. Kuznetsov, Fiz. Plazmy **22**, 916 (1996) [Plasma Phys. Rep. **22**, 829 (1996)].
[8] V. E. Zakharov and N. N. Filonenko, Prikl. Mekh. Tekh. Fiz. No. 5, 62 (1968).
[9] V. E. Zakharov, in *Nonlinear Waves and Weak Turbulence* (*Advances in Math. Sci.*, Series 2), Vol. 182, V. E. Zakharov (ed.), Am. Math. Soc., Providence, R.I. (1998), p. 167.

[10] M. V. Keldysh and M. A. Lavrent'ev, in *Proceedings of the Conference on the Theory of Wave Resistance* [in Russian], Izd. TsAGI, Moscow (1937), p. 31.
[11] N. E. Kochin, I. A. Kibel', and N. V. Roze, *Theoretical Hydrodynamics* [in Russian], Part 1, Fizmatgiz, Moscow (1963).
[12] E. A. Novikov, Izv. Akad. Nauk SSSR Fiz. Atm. Okeana **17**, 956 (1981).
[13] H. L. Lamb, *Hydrodynamics*, 6th ed., Cambridge Univ. Press, Cambridge (1932).
[14] V. E. Zakharov and E. A. Kuznetsov, Usp. Fiz. Nauk **167**, 1137 (1997).
[15] E. A. Kuznetsov and V. P. Ruban, JETP Lett. **67**, 1076 (1998).
[16] A. I. Dyachenko, E. A. Kuznetsov, M. D. Spector, and V. E. Zakharov, Phys. Lett. A **221**, 73 (1996).
[17] S. C. Crow, AIAA J. **8**, 2172 (1970).
[18] A. V. Kats and V. M. Kontorovich, Fiz. Nizk. Temp. **23**, 120 (1997) [Low Temp. Phys. **23**, 89 (1997)].
[19] E. A. Kuznetsov and A. V. Mikhailov, Phys. Lett. A **77**, 37 (1980).
[20] P. J. Morrison and J. M. Greene, Phys. Rev. Lett. **45**, 790 (1980).
[21] D. Lewis, J. Marsden, R. Montgomery, and T. Ratiu, Physica D **18**, 391 (1986).
[22] A. Rouhi and J. Wright, Phys. Rev. E **48**, 1850 (1993).
[23] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, New York–Oxford (1974).
[24] B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Sov. Phys. Dokl. **15**, 539 (1970)].
[25] E. A. Kuznetsov and J. J. Rasmussen, Phys. Rev. E **51**, 4479 (1995).

Translated by James S. Wood

Passive scalar in a large-scale velocity field

I. Kolokolov

Budker Institute of Nuclear Physics, 630090 Novasibirsk, Russia; Novosibirsk State University, 630090 Novosibirsk, Russia

V. Lebedev*)

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia; Department of Physics of Complex Systems, Weizmann Institute of Science, 76100 Rehovot, Israel

M. Stepanov

Institute of Automation and Electrometry, Russian Academy of Sciences, 630090 Novasibirsk, Russia; Novosibirsk State University, 630090 Novosibirsk, Russia

(Submitted 6 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 920–939 (March 1999)

We consider advection of a passive scalar $\theta(t,\mathbf{r})$ by an incompressible large-scale turbulent flow. In the framework of the Kraichnan model all PDF's (probability distribution functions) for the single-point statistics of θ and for the passive scalar difference $\theta(\mathbf{r}_1)-\theta(\mathbf{r}_2)$ (for separations $\mathbf{r}_1-\mathbf{r}_2$ lying in the convective interval) are found. *© 1999 American Institute of Physics.* [S1063-7761(99)01003-3]

INTRODUCTION

We treat advection of a passive scalar field $\theta(t,\mathbf{r})$ by an incompressible turbulent flow; the role of the scalar can be played by temperature or by pollutant density. The velocity field is assumed to contain motions from some interval of scales restricted from below by L_v. A steady situation with a permanent random supply of the passive scalar is considered. We wish to establish statistics of the passive scalar θ for scales that are less than both the scale L_v and the pumping scale L, and larger than the diffusion scale r_{dif} (for definiteness we assume that $L<L_v$). Such a convective interval of scales exists if the Peclet number $\mathrm{Pe}=L/r_{\mathrm{dif}}$ is large enough; we will assume this condition. Since all scales from the convective interval are assumed to be smaller than L_v, we will discuss advection by a large-scale turbulent flow. The problem is of physical interest for dimensionalities $d=2,3$, but formally it can be treated for an arbitrary dimensionality d of space. Below we will treat d as a parameter. In particular, all expressions will be true for a space of high dimensionality d.

Description of the small-scale statistics of a passive scalar advected by a large-scale solenoidal velocity field is a special problem in turbulence theory. This problem was treated consistently from the very beginning and some rigorous results have been obtained, which is quite unusual for a turbulence problem. Batchelor (see Ref. 1) examined the case of an external velocity field being so slow that it does not change during the time of the spectral transfer of the scalar from the external scale to the diffusion scale. Then Kraichnan (see Ref. 2) obtained plenty of results in the opposite limit of a velocity field delta-correlated in time. The pair correlation function of the passive scalar $\langle\theta(\mathbf{r})\theta(0)\rangle$ was found to be proportional to the logarithm $\ln(L/r)$, and the pair correlation function of the passive scalar difference $\langle[\theta(\mathbf{r})-\theta(0)]^2\rangle$ was found to be proportional to $\ln(r/r_{\mathrm{dif}})$ in both cases. The assertions are really correct for any temporal statistics of the velocity field (see Refs. 3 and 4). Thus we are dealing with the logarithmic case which is substantially simpler than cases with power-like correlation functions usually encountered in turbulence problems (see Refs. 5–7).

Now about high-order correlation functions of the passive scalar. As long as all distances between the points are much less than L, the $2n$-point correlation functions of θ are given by their reducible parts (that is, are expressed via products of the pair correlation function) up to $n\sim\ln(L/r)$, where r is either the smallest distance between the points or r_{dif} depending on which is larger (see Ref. 4). The reason for such Wick decoupling is simply the fact that reducible parts contain more logarithmic factors (which are considered as the large ones) than non-reducible parts do. Consistent calculations of the fourth-order correlation function of the passive scalar at $d=2$ (see Ref. 8) confirm the assertion. Therefore, e.g., the single-point PDF of θ has a Gaussian core (that describes the first moments with $n<\ln\mathrm{Pe}$) and a non-Gaussian tail (that describes moments with $n>\ln\mathrm{Pe}$). The tail appears to be exponential (see Refs. 3 and 4). The same is true of the passive scalar difference $\Delta\theta=\theta(\mathbf{r})-\theta(0)$, where instead of $\ln\mathrm{Pe}$ we should take $\ln(r/r_{\mathrm{dif}})$. The tails do not depend on $\ln\mathrm{Pe}$ or on $\ln(r/r_{\mathrm{dif}})$, and contain only coefficients that depend on the statistics of the advecting velocity.

Correlation functions of the passive scalar can be written as averages of integrals of the pumping along Lagrangian trajectories (see, e.g., Ref. 9). For example, the pair correlation function $\langle\theta(\mathbf{r})\theta(0)\rangle$ is proportional to the average time needed for two points moving along Lagrangian trajectories to run from the distance r to the distance L. Generally, correlation functions of a passive scalar are determined by spectral transfer via evolution of Lagrangian separations up to the scale L. For the large-scale velocity field, the Lagrangian dynamics is determined by the stretching matrix $\sigma_{\alpha\beta} = \nabla_\beta v_\alpha$ and, consequently, the statistics of the matrix determines correlation function of the passive scalar. For example, the coefficient of the logarithm in the pair correlation function of the passive scalar is $P_2/\tilde{\lambda}$ (see Refs. 1–4) where P_2 is the pumping rate of θ^2 and $\tilde{\lambda}$ is Lyapunov exponent that is the average of the largest eigenvalue of the matrix $\hat{\sigma}$. The coefficients in the exponential tails are more sensitive to the statistics of $\hat{\sigma}$; specifically, they depend on the dimensionless parameter $\tilde{\lambda}\tau$ (see Ref. 4) where τ is the correlation time of $\hat{\sigma}$. The motion of the fluid particles in the random velocity field resembles in some respects random walks, but one should remember that correlation lengths of both the advecting velocity and of the pumping are much larger than scales from the convective interval we are interested in. Thus the situation is opposite to one usually encountered in solid state physics, where, e.g., random potential is short-range correlated in space.

Since $\ln(L/r)$ is really not very large, it is of interest to find all PDF's for the single-point statistics of θ and for the passive scalar difference $\Delta\theta$. It is possible to do this for the Kraichnan short-correlated case $\tilde{\lambda}\tau \ll 1$ when the statistics of $\hat{\sigma}$ can be regarded to be Gaussian. An attempt to do this was made in Refs. 10 and 11 in terms of the statistics of the main eigenvalue of the matrix $\hat{\sigma}$. Unfortunately, the scheme works only for the dimensionality $d=2$ where the matrix $\hat{\sigma}$ has a single eigenvalue. This was noted in Ref. 12 where also the correct coefficient in the exponential tails for an arbitrary dimensionality of space d was found. Here, we develop a scheme enabling one to obtain all PDF's for arbitrary d. The scheme is also interesting from a methodological point of view. For example, its modification enables one to calculate the statistics of local dissipation (see Ref. 13).

The paper is organized as follows. In Sec. 1 we find a path integral representation for the simultaneous statistics of the passive scalar. In Sec. 2 we analyze the generating functional for correlation functions of the passive scalar in the convective interval of scales. Using different approaches we obtain the functional and establish the applicability conditions of our consideration. In Sec. 3 we find explicit expressions for the single-point PDF and for the PDF of the passive scalar difference. In the Conclusion we briefly discuss the results obtained.

1. GENERAL RELATIONS

The dynamics of the passive scalar θ advected by the velocity field $\mathbf{v}$ is described by Eq.

$$\partial_t \theta + \mathbf{v}\nabla\theta - \kappa\nabla^2\theta = \phi. \qquad (1.1)$$

Here, the term with the velocity $\mathbf{v}$ describes the advection of the passive scalar, the next term is diffusive (κ is the diffusion coefficient), and ϕ describes a pumping source of the passive scalar. Both $\mathbf{v}(t,\mathbf{r})$ and $\phi(t,\mathbf{r})$ are assumed to be random functions of t and $\mathbf{r}$. We regard the statistics of the velocity and source to be independent. Therefore, all correlation functions of θ are to be treated as averages over both statistics.

A. Simultaneous statistics

The source ϕ is believed to possess Gaussian statistics and to be δ-correlated in time. The statistics is entirely characterized by the pair correlation function

$$\langle \phi(t_1,\mathbf{r}_1)\phi(t_2,\mathbf{r}_2)\rangle = \delta(t_1-t_2)\chi(|\mathbf{r}_1-\mathbf{r}_2|), \qquad (1.2)$$

where we assume that the pumping is isotropic. The function $\chi(r)$ is assumed to have a characteristic scale L, which is the pumping length. We will be interested in the statistics of the passive scalar on scales much smaller than L.

Simultaneous correlation functions of the passive scalar θ can be represented as coefficients in the expansion over y of the generating functional

$$\mathcal{Z}(y) = \left\langle \exp\left\{ iy\int d\mathbf{r}\beta(\mathbf{r})\theta(0,\mathbf{r})\right\}\right\rangle, \qquad (1.3)$$

where β is a function of the coordinates and angular brackets denote averaging over both the statistics of the pumping ϕ and the statistics of the velocity $\mathbf{v}$. The generating functional $\mathcal{Z}(y)$ contains complete information about the simultaneous statistics of the passive scalar ϑ. Specifically, knowing $\mathcal{Z}(y)$ one can reconstruct the simultaneous PDF of the passive scalar; the problem is discussed in Sec. 3.

If characteristic scales of β in (1.7) are much larger than the diffusion scale r_{dif}, then it is possible to neglect diffusion when treating the generating functional (1.3). Then the left-hand side of Eq. (1.1) describes simple advection, and it is reasonable to consider a solution of Eq. in terms of Lagrangian trajectories $\varrho(t)$ introduced by Eq.

$$\partial_t \varrho = \mathbf{v}(t,\varrho). \qquad (1.4)$$

We label the trajectories with $\mathbf{r}$, which are the positions of the Lagrange particles at $t=0$: $\varrho(0,\mathbf{r})=\mathbf{r}$. Next, introducing $\tilde{\theta}(t,\mathbf{r})=\theta(t,\varrho)$, we rewrite Eq. (1.1) as $\partial_t\tilde{\theta}=\phi$, which leads to

$$\theta(0,\mathbf{r}) = \int_{-\infty}^{0} dt\,\phi(t,\varrho). \qquad (1.5)$$

Here we have taken into account that at $t=0$ the functions θ and $\tilde{\theta}$ coincide. Starting with (1.5) and exploiting Gaussian pumping statistics, we can average the generating functional (1.3) explicitly over the statistics. The result is

$$\mathcal{Z}(y) = \left\langle \exp\left[-\frac{y^2}{2}\int_{-\infty}^{0} dt\,U\right]\right\rangle, \qquad (1.6)$$

$$U = \int d\mathbf{r}_1 d\mathbf{r}_2 \beta(\mathbf{r}_1)\beta(\mathbf{r}_2)\chi(|\varrho_1-\varrho_2|), \qquad (1.7)$$

where angular brackets mean averaging over the statistics of the velocity field only.

Being interested in the single-point statistics of θ we should take $\beta(\mathbf{r})=\delta(\mathbf{r})$. But this is impossible since we have neglected diffusion. We take $\beta(\mathbf{r})=\delta_\Lambda(\mathbf{r})$ instead, where the function $\delta_\Lambda(\mathbf{r})$ tends to zero at $\Lambda r > 1$ fast enough, and is normalized by the condition

$$\int d\mathbf{r}\,\delta_\Lambda(\mathbf{r}) = 1.$$

Then the generating functional (1.6) will describe the statistics of an object

$$\theta_\Lambda=\int d\mathbf{r}\,\delta_\Lambda(\mathbf{r})\,\theta(\mathbf{r}), \tag{1.8}$$

smeared over a spot of size Λ^{-1}. If $r_{\rm dif}\Lambda \ll 1$, then the statistics of the object is not sensitive to diffusivity. On the other hand, if $\Lambda L \gg 1$, then knowing the correlation functions of θ_Λ, we can reconstruct single-point statistics due to the logarithmic character of the correlation functions. To obtain single-point correlation functions one should substitute simply $\Lambda \to r_{\rm dif}^{-1}$ into the correlation functions of θ_Λ. The above inequalities $\Lambda r_{\rm dif} \ll 1$ and $\Lambda L \gg 1$ are compatible because of $\mathrm{Pe} \gg 1$. If we are interested in the statistics of the passive scalar differences in points with a separation $\mathbf{r}_0$ (where $r_0 \gg r_{\rm dif}$) then instead of $\delta_\Lambda(\mathbf{r})$ we should take

$$\beta(\mathbf{r})=\delta_\Lambda(\mathbf{r}-\mathbf{r}_0/2)-\delta_\Lambda(\mathbf{r}+\mathbf{r}_0/2). \tag{1.9}$$

Then the generating functional (1.6) will describe the statistics of an object

$$\Delta\theta_\Lambda=\theta_\Lambda(\mathbf{r}_0/2)-\theta_\Lambda(-\mathbf{r}_0/2). \tag{1.10}$$

Again, correlation functions of the passive scalar differences can be found from correlation functions of $\Delta\theta_\Lambda$ after the substitution $\Lambda \to r_{\rm dif}^{-1}$.

B. Path integral

Below, we treat advection of the passive scalar by a large-scale velocity field, that is, we assume that the velocity correlation length L_v is larger than the scales from the convective interval. Then for the scales one can expand the difference

$$v_\alpha(\mathbf{r}_1)-v_\alpha(\mathbf{r}_2)=\sigma_{\alpha\beta}(t)(r_{1\beta}-r_{2\beta}), \quad \sigma_{\alpha\beta}=\nabla_\beta v_\alpha. \tag{1.11}$$

Here $\sigma_{\alpha\beta}(t)$ can be treated as an $\mathbf{r}$-independent matrix field. Then Eq. (1.4) leads to

$$\partial_t(\varrho_{1,\alpha}-\varrho_{2,\alpha})=\sigma_{\alpha\beta}(t)(\varrho_{1,\beta}-\varrho_{2,\beta}). \tag{1.12}$$

A formal solution of Eq. (1.12) is

$$\varrho_{1,\alpha}-\varrho_{2,\alpha}=W_{\alpha\beta}(r_{1,\beta}-r_{2,\beta}),$$

$$\partial_t\hat W=\hat\sigma\hat W, \quad \hat W=\mathscr{T}\exp\left(-\int_t^0 dt\ \hat\sigma\right), \tag{1.13}$$

where $\mathscr{T}$ denotes antichronological ordering. Note that $\det\hat W=1$; this property is a consequence of $\mathrm{Tr}\,\hat\sigma=0$ and the initial condition $\hat W=1$ at $t=0$. The Lagrangian difference in (1.7) is now rewritten as

$$|\varrho_1-\varrho_2|=\sqrt{(r_{1\alpha}-r_{2\alpha})B_{\alpha\beta}(r_{1\beta}-r_{2\beta})}, \quad \hat B=\hat W^T\hat W, \tag{1.14}$$

where the subscript T denotes a matrix transpose. Note that $\det\hat B=1$ since $\det\hat W=1$.

The generating functional $\mathscr{Z}(y)$ (1.6) can be explicitly calculated in the Kraichnan case (see Ref. 2) when the statistics of the velocity is δ-correlated in time. Then the velocity statistics is Gaussian and is entirely determined by the pair correlation function, which in the convective interval is written as

$$\langle v_\alpha(t_1,\mathbf{r}_1)v_\beta(t_2,\mathbf{r}_2)\rangle=\delta(t_1-t_2)[\mathscr{V}_0\delta_{\alpha\beta}-\mathscr{K}_{\alpha\beta}(\mathbf{r}_1-\mathbf{r}_2)], \tag{1.15}$$

$$\mathscr{K}_{\alpha\beta}(\mathbf{r})=D(r^2\delta_{\alpha\beta}-r_\alpha r_\beta)+\frac{(d-1)D}{2}\delta_{\alpha\beta}r^2. \tag{1.16}$$

Here $\mathscr{V}_0$ is a huge r-independent constant and D is a parameter characterizing the amplitude of the strain fluctuations. The structure of expression (1.16) is determined by the assumed isotropy and spacial homogeneity, and by the incompressibility condition $\nabla\mathbf{v}=0$. Then the statistics of $\hat\sigma$ is Gaussian and is determined by the pair correlation function, which can be found from Eqs. (1.15) and (1.16):

$$\langle\sigma_{\alpha\beta}(t_1)\sigma_{\mu\nu}(t_2)\rangle=D[(d+1)\delta_{\alpha\mu}\delta_{\beta\nu}-\delta_{\alpha\nu}\delta_{\beta\mu}-\delta_{\alpha\beta}\delta_{\mu\nu}]\delta(t_1-t_2). \tag{1.17}$$

Note that the correlation function (1.17) is $\mathbf{r}$-independent, as it should be. We see from (1.17) that the parameter D characterizes the amplitude of $\hat\sigma$ fluctuations.

Averaging over the statistics of $\hat\sigma$ can be replaced by a path integral over unimodular matrices $\hat W(t)$ with a weight $\exp(i\mathscr{S})$. The effective action $\mathscr{S}=\int dt\mathscr{L}_0$ is determined by (1.17):

$$i\mathscr{L}_0=-\frac{1}{2d(d+2)D}[(d+1)\mathrm{Tr}(\hat\sigma^T\hat\sigma)+\mathrm{Tr}\,\hat\sigma^2]. \tag{1.18}$$

Then the generating functional (1.7) can be rewritten as the following functional integral over unimodular matrices

$$\mathscr{Z}(y)=\int\mathscr{D}\hat W\exp\left[\int_{-\infty}^0 dt\left(i\mathscr{L}_0-\frac{y^2}{2}U\right)\right], \tag{1.19}$$

$$U=\int d\mathbf{r}_1 d\mathbf{r}_2\beta(\mathbf{r}_1)\beta(\mathbf{r}_2)\chi\times[\sqrt{(r_{1\alpha}-r_{2\alpha})B_{\alpha\beta}(r_{1\beta}-r_{2\beta})}]. \tag{1.20}$$

Here, we should substitute $\hat\sigma=\partial_t\hat W(\hat W)^{-1}$ and recall the boundary condition $\hat W=1$ at $t=0$.

Some words about the ''potential'' U (1.7) figuring in (1.20). The characteristic value of $\mathbf{r}_1-\mathbf{r}_2$ in the integral (1.7) is of order Λ^{-1} for $\beta(\mathbf{r})=\delta_\Lambda(\mathbf{r})$. Since we assume $\Lambda L \gg 1$, then for single-point statistics $U\approx P_2$, where $P_2=\chi(0)$, if B is not very large. In particular, it is correct at moderate times $|t|$, since $\hat B=\hat 1$ at $t=0$. With increasing $|t|$ the argument of χ in (1.20) grows and U tends to zero when the argument of χ becomes greater than L. For the passive scalar difference when β is determined by (1.9) the situation is a bit more complicated. Then U is a difference of two contributions. The first contribution behaves as for single-point statistics. The second contribution contains χ with the argument determined by $\mathbf{r}_1-\mathbf{r}_2\approx\pm\mathbf{r}_0$. Then at $t=0$ the meaning of the second contribution is determined again by P_2, but it vanishes with increasing $|t|$ earlier than the first contribution.

The path integral representation (1.19) indicates that we reduced our problem to the quantum mechanics with d^2-1

degrees of freedom. Nevertheless to solve the problem we should perform an additional reduction of the degrees of freedom. The conventional way to do this is passing to eigenvalues, say, of the matrix $\hat{B}$ figuring in (1.20) (see, e.g., Ref. 14) and excluding angular degrees of freedom. Just this way was used by Bernard, Gawedzki and Kupiainen (see Ref. 12). Then the authors using known facts about the quantum mechanics associated with the eigenvalues (see, e.g., Ref. 15) have found the coefficient in the exponential tail of the single-point PDF of θ. Unfortunately this way is not very convenient to find the whole PDF. To do this we will use a special representation of the matrix $\hat{W}$ in the spirit of the nonlinear substitution introduced by Kolokolov (see Ref. 16). That is the subject of the next subsection.

C. Choice of parametrization

To examine the generating functional $\mathcal{T}(y)$ we use a mixed rotational-triangle parametrization

$$\hat{W}=\hat{R}\hat{T}, \quad \hat{B}\hat{T}^T\hat{T}, \tag{1.21}$$

where $\hat{R}$ is an orthogonal matrix and $\hat{T}$ is a triangular matrix: $T_{ij}=0$ for $i>j$. The parametrization (1.21) is the direct generalization of the $2d$ substitution suggested in Ref. 17. Note that $\det\hat{T}=1$ since $\det\hat{W}=1$. Note also that the matrix $\hat{B}$ introduced by (1.14) does not depend on $\hat{R}$, as is seen from (1.21). That is a motivation to exclude the matrix $\hat{R}$ from consideration, integrating over the corresponding degrees of freedom in the path integral (1.19). A Jacobian appears in the integration. To avoid an explicit calculation of the Jacobian, which needs a discretization over time and then an analysis of an infinite matrix (see Ref. 10), we use an alternative procedure described below.

Let us examine the dynamics of the matrix $\hat{T}$. It is determined by the equation

$$\partial_t T_{ij}=\Sigma_{ii}T_{ij}+\sum_{i<k\leqslant j}(\Sigma_{ik}+\Sigma_{ki})T_{kj}, \tag{1.22}$$

following from Eqs. (1.13) and (1.21). Here we used the notations

$$\hat{\Sigma}=\hat{R}^T\hat{\sigma}\hat{R}. \tag{1.23}$$

Next introducing the quantities

$$T_{ii}=\exp(\rho_i), \quad T_{ij}=\exp(\rho_i)\eta_{ij}, \ \text{ if } \ i<j, \tag{1.24}$$

we rewrite Eq. (1.22) as

$$\partial_t\rho_i=\Sigma_{ii}, \tag{1.25}$$

$$\partial_t\eta_{ij}=(\Sigma_{ij}+\Sigma_{ji})\exp(\rho_j-\rho_i)+\sum_{i<k<j}(\Sigma_{ik}+\Sigma_{ki}) \times\exp(\rho_k-\rho_i)\eta_{kj}. \tag{1.26}$$

Comparing (1.13) with (1.21), one can find the following expression for $\hat{A}=\hat{R}^T\partial_t\hat{R}$:

$$A_{ij}=\Sigma_{ij} \ \text{ if } \ i>j, \quad A_{ij}=-\Sigma_{ji} \ \text{ if } \ i<j. \tag{1.27}$$

One can easily check that the irreducible pair correlation function of Σ_{ij} has the same form as for σ_{ij} [see Eq. (1.17)]:

$$\langle\Sigma_{ij}(t_1)\Sigma_{mn}(t_2)\rangle=D[(d+1)\delta_{im}\delta_{jn}-\delta_{in}\delta_{jm} -\delta_{ij}\delta_{mn}]\delta(t_1-t_2). \tag{1.28}$$

Furthermore, the average value of Σ_{ij} is nonzero (see Ref. 10):

$$\langle\Sigma_{ij}\rangle=-D\frac{d(d-2i+1)}{2}\delta_{ij}. \tag{1.29}$$

Nonzero averages of Σ_{ij} are related to Lyapunov exponents (not only the first one), see Ref. 18 (for our model see also Ref. 19). To obtain (1.29) one should take into account that the matrix $\hat{R}$ propagates backward in time since $\hat{R}=1$ is fixed at $t=0$ and we treat negative t. Solving Eq. $\hat{A}=\hat{R}^T\partial_t\hat{R}$ for $\hat{R}$ on a small interval τ we get

$$\hat{R}(t-\tau)\approx\hat{R}(t)\left[1-\int_{t-\tau}^{t}dt'\hat{A}(t')\right].$$

Then with the same accuracy we get from Eq. (1.23)

$$\hat{\Sigma}(t-\tau)\approx\hat{R}^T(t)\hat{\sigma}(t-\tau)\hat{R}(t) -\left[\hat{\Sigma}(t-\tau),\int_{t-\tau}^{t}dt'\hat{A}(t')\right]. \tag{1.30}$$

The average value of $\hat{\Sigma}$ arises from the second term on the right-hand side of (1.30). The explicit form of the average can be found using

$$\left\langle\Sigma_{ij}(t-\tau)\int_{t-\tau}^{t}dt'\Sigma_{mn}(t')\right\rangle =\frac{D}{2}[(d+1)\delta_{im}\delta_{jn}-\delta_{in}\delta_{jm}-\delta_{ij}\delta_{mn}]. \tag{1.31}$$

Here we utilized Eq. (1.28) and replaced the integral

$$\int_{t-\tau}^{t}dt'\,\delta(t-\tau-t')$$

by 1/2. The reason is that the correlation function of $\hat{\sigma}$ actually has a finite correlation time, and therefore $\delta(t)$ (representing this correlation function) should be replaced by a narrow function symmetric under $t\to -t$. Then we will get 1/2. Expressing $\hat{A}$ via $\hat{\Sigma}$ from (1.27) in (1.30) and calculating its average using (1.31) we get the answer (1.29).

The expressions (1.25), (1.26), (1.28), and (1.29) entirely determine the stochastic dynamics of ρ_i and η_{ij}. Using the conventional approach (see Refs. 20–24) correlation functions of these degrees of freedom can be described in terms of a path integral over ρ_i,η_{ij} and over auxiliary fields which we denote by m_i and μ_{in} $(i<n)$. This integral should be taken with the weight $\exp(i\int dt\mathcal{L})$, where the Lagrangian is

$$
\begin{aligned}
\mathscr{L}=&\sum_{a=1}^{d} m_a\left[\partial_t\rho_a+D\frac{d(d-2a+1)}{2}\right]+\frac{iD}{2}\left[d\sum_a m_a^2\right.\\
&\left.-\left(\sum_a m_a\right)^2\right]+iDd\sum_{i<j}\exp(2\rho_j-2\rho_i)\mu_{ij}^2\\
&+2iDd\sum_{i<k<j}\mu_{ij}\mu_{ik}\exp(2\rho_k-2\rho_i)\,\eta_{kj}\\
&+\sum_{i<j}\mu_{ij}\partial_t\eta_{ij}+iDd\sum_{i<k<m,n}\mu_{im}\mu_{in}\eta_{km}\eta_{kn}\\
&\times\exp(2\rho_k-2\rho_i). \qquad (1.32)
\end{aligned}
$$

Since the matrix $\hat{B}$ in accordance with (1.21) does not depend on $\hat{R}$ it is enough to know the statistics of ρ_a and η_{ij} to determine the average (1.6). Therefore, instead of (1.19) we get

$$
\mathscr{J}(y)=\int\mathscr{D}\rho\mathscr{D}\eta\mathscr{D}m\mathscr{D}\mu\,\exp\left[\int_{-\infty}^{0}dt\left(i\mathscr{L}-\frac{y^2}{2}U\right)\right]. \qquad (1.33)
$$

Here U is determined by (1.20), where the matrix $\hat{B}$ is determined by Eqs. (1.21) and (1.24).

Thus we obtained the expression for the generating functional (1.3) in terms of the functional (path) integral which is convenient for the analysis presented in the subsequent section.

2. GENERATING FUNCTIONAL

Here we calculate the generating functional (1.3) for a single-point statistics of θ that is of the object (1.8) corresponding to $\beta(\mathbf{r})=\delta_\Lambda(\mathbf{r})$, and also the statistics of the difference that is of the object (1.10) corresponding to (1.9). The starting point for the subsequent consideration is the expression (1.33). There are different ways to examine $\mathscr{J}(y)$. We will describe two schemes leading to the same answer but carrying in some sense complementary information. We also believe that consideration of the different schemes is useful from a methodological point of view. A modification of the second scheme is presented in the Appendix.

A. Saddle-point approach

The first way to obtain the answer for the generating functional (1.3) is by using the saddle-point approximation for the path integral (1.33). The inequalities justifying the approximation are $\Lambda L\gg 1$ for the object (1.8) and $\Lambda r\gg 1$ for the object (1.10).

As we will see, large values of the differences $\rho_i-\rho_k$ $(i<k)$ will be relevant for us. Then fluctuations of η and μ are suppressed and it is possible to neglect the fluctuations. Therefore we can omit the integration over η and μ in (1.33), substituting $\eta=\mu=0$ into (1.32). After that we obtain a reduced Lagrangian:

$$
\begin{aligned}
\mathscr{L}_r=&\sum_{a=1}^{d} m_a\left[\partial_t\rho_a+D\frac{d(d+1-2a)}{2}\right]\\
&+\frac{iD}{2}\left[d\sum_a m_a^2-\left(\sum_a m_a\right)^2\right]. \qquad (2.1)
\end{aligned}
$$

Now, to obtain $\mathscr{J}(y)$ one should integrate the exponent in (1.33) (with $\mathscr{L}_r$) over ρ_a and m_a. To examine (2.1) it is convenient to pass to new variables $\phi_a=O_{ab}\rho_b$ and $\tilde{m}_a=O_{ab}m_b$, where $\hat{O}$ is an orthogonal matrix. We make the following transformation:

$$
\phi_1=\sqrt{\frac{3}{d(d^2-1)}}[(d-1)\rho_1+(d-3)\rho_2+\ldots+(1-d)\rho_d],
$$

$$
\phi_2=\ldots,\ldots,\qquad \phi_d=\frac{1}{\sqrt{d}}[\rho_1+\rho_2+\ldots+\rho_d]. \qquad (2.2)
$$

Then the expression (2.1) will be rewritten as

$$
\begin{aligned}
i\mathscr{L}_r=&i\sum_{a=1}^{d}\tilde{m}_a\partial_t\phi_a-\frac{Dd}{2}\sum_{a=1}^{d-1}\tilde{m}_a^2\\
&+i\frac{Dd}{2}\sqrt{\frac{d(d^2-1)}{3}}\tilde{m}_1. \qquad (2.3)
\end{aligned}
$$

The Lagrangian (2.3) is a sum over different degrees of freedom. The dynamics of ϕ_1 is ballistic, whereas the dynamics of ϕ_a for $d>a>1$ is purely diffusive. The condition $\det\hat{T}=1$ means $\phi_d=0$, correspondingly the dynamics of ϕ_d determined by the Lagrangian (2.3) is trivial: $\partial_t\phi_t=0$. We will see that times determining the main contribution to the generating functional are large enough that $\phi_1\gg\phi_a$ for the relevant region. Therefore, the potential U (1.20) depends essentially only on ϕ_1, and it is possible to integrate explicitly over, ϕ_a and $\tilde{m}_a$ for $a>1$. After that we are left with only one degree of freedom, which is described by the Lagrangian

$$
i\mathscr{L}_1=i\tilde{m}_1\left(\partial_t\phi_1+\frac{Dd}{2}\sqrt{\frac{d(d^2-1)}{3}}\right)-\frac{Dd}{2}\tilde{m}_1^2. \qquad (2.4)
$$

Neglecting all ϕ_a for $a>1$ and inverting transformation (2.2) we obtain

$$
\rho_1\approx\sqrt{\frac{3(d-1)}{d(d+1)}}\phi_1,\qquad \rho_a\approx\frac{d-2a+1}{d-1}\rho_1. \qquad (2.5)
$$

We will see below that the characteristic value $\phi_1\gg 1$. Therefore the characteristic value of e^{ρ_1} is much larger than other e^{ρ_a}, and we conclude that the potential U depends really only on ρ_1. For the case of the single-point statistics, the characteristic value of the difference $\mathbf{r}_1-\mathbf{r}_2$ in (1.20) is Λ^{-1}. Then it follows from (1.21) and (1.24) that the potential U decreases from P_2 to zero near the point $\rho_1=\ln(L\Lambda)$, which is near the point $\phi_1=\phi_\Lambda$, where

$$
\phi_\Lambda=\sqrt{\frac{d(d+1)}{3(d-1)}}\ln(L\Lambda). \qquad (2.6)
$$

For the difference the potential increases from zero to $2P_2$ at $\phi_1=\phi_R$, where

$$\phi_R=\sqrt{\frac{d(d+1)}{3(d-1)}}\ln\frac{L}{r_0}, \tag{2.7}$$

and then decreases from $2P_2$ to zero near $\phi_1=\phi_\Lambda$. The expressions (2.6) and (2.7) determine the characteristic values of ϕ_1, which are actually large, since $L\Lambda\gg 1$ or $L/r_0\gg 1$; this justifies our conclusions.

Now we can employ the saddle-point approximation:

$$\ln\mathscr{Z}(y)\approx\int_{-\infty}^{0}dt\left(i\mathscr{L}_1-\frac{y^2}{2}U\right)\Bigg|_{\text{inst}}, \tag{2.8}$$

where we should substitute solutions of the extremal conditions, which we will call instantonic equations. The instantonic equations, which can be found from extremal conditions for $i\mathscr{L}_1-y^2U/2$, are

$$\partial_t\phi_1+\frac{Dd}{2}\sqrt{\frac{d(d^2-1)}{3}}=-iDd\tilde{m}_1, \tag{2.9}$$

$$\partial_t\tilde{m}_1=i\frac{y^2}{2}\frac{\partial U}{\partial\phi_1}. \tag{2.10}$$

Eqs. conserve the ''energy''

$$-i\frac{Dd}{2}\tilde{m}_1\sqrt{\frac{d(d^2-1)}{3}}+\frac{Dd}{2}\tilde{m}_1^2+\frac{y^2}{2}U. \tag{2.11}$$

The conservation law is satisfied since $i\mathscr{L}_1-y^2U/2$ does not explicitly depend on t. The ''energy'' (2.11) is equal to zero, since as $t\to-\infty$ the value of $\tilde{m}_1$ should tend to zero. This property can be treated as the extremal condition when $i\mathscr{L}_r-y^2U/2$ is varied over the initial value of ϕ_1. Equating the ''energy'' (2.11) to zero, we can express $\tilde{m}_1$ via ϕ_1. Next, since (2.11) is zero, the saddle-point value of $\mathscr{Z}(y)$ (2.8) can be written as $i\int d\phi_1\tilde{m}_1$, where the integral over ϕ_1 goes from zero to infinity.

Substituting the expression for $\tilde{m}_1$ in terms of ϕ_1 into $i\int d\phi_1\tilde{m}_1$, we get for the single-point statistics

$$\ln\mathscr{Z}(y)\simeq\frac{d(d+1)}{6}\left[1-\sqrt{1+\frac{12y^2P_2}{Dd^2(d^2-1)}}\right]\ln(L\Lambda). \tag{2.12}$$

Note that the expression (2.12) has (as a function of y) two branch points $y=\pm iy_{\text{sing}}$, where

$$y_{\text{sing}}^2=\frac{Dd^2(d^2-1)}{12P_2}. \tag{2.13}$$

The same procedure can be done for the passive scalar difference, or, more precisely, for the object (1.10). Taking into account the presence of the jumps (2.6) and (2.7) in the potential U, we get an answer slightly different from (2.12):

$$\ln\mathscr{Z}(y)\simeq\frac{d(d+1)}{6}\left[1-\sqrt{1+\frac{24y^2P_2}{Dd^2(d^2-1)}}\right]\ln(r_0\Lambda), \tag{2.14}$$

$$y_{\text{sing}}^2=\frac{Dd^2(d^2-1)}{24P_2}. \tag{2.15}$$

Note that (2.14) does not depend on the pumping scale L, but still depends on the cutoff Λ.

The characteristic value of ϕ_1 is determined by the quantity (2.6) which is much larger than unity. Then it follows from (2.5) that $\exp(2\rho_j-2\rho_i)\ll 1$, $i>j$, (excluding a short initial stage of evolution) and we see from (1.32) that fluctuations of the fields η are suppressed in comparison, say, with ρ_a. This justifies neglecting the fields η and μ leading to the reduced Lagrangian (2.1). Next, the dynamics of ϕ_a for $a>1$ is diffusive, and it follows from (2.3) that the characteristic value of ϕ_a can be estimated to be $\sqrt{Dd|t|}$. As follows from (2.3), $\partial_t\phi_1\sim Dd^{5/2}$, and we find from (2.6) the instantonic lifetime

$$t_{It}=D^{-1}d^{-2}\ln(L\Lambda), \tag{2.16}$$

which determines times producing nonzero contributions to the effective action. At $|t|\sim t_{It}$, the characteristic values of ϕ_a for $a>1$ are of order $\sqrt{\ln(L\Lambda)/d}$, and we conclude that

$$\frac{\phi_a}{\phi_1}\sim\frac{1}{d\sqrt{\ln(L\Delta)}}\ll 1 \tag{2.17}$$

at times $|t|\sim t_{It}$. The inequality (2.17) justifies passing to the Lagrangian (2.4). The same arguments can be applied to the generating functional for the passive scalar difference; the only modification is in the substitution $\ln(L\Lambda)\to\ln(r_0\Lambda)$.

There are also additional applicability conditions for the results (2.12) and (2.14). To establish the conditions, one should go beyond the main order of the saddle-point approximation. It will be more convenient for us to develop an alternative scheme, which enables one to find the conditions more simply. That is the subject of the next subsection.

B. Schrödinger equation

Here we present another way to get the answers (2.12) and (2.14). As before, we start with the path integral representation (1.33) for the generation functional $\mathscr{Z}(y)$.

Unfortunately it is impossible to get a closed equation for $\mathscr{Z}(y)$. To avoid the difficulty we introduce an auxiliary quantity

$$\Psi(t,y,\rho_0,\eta_0)=\int\mathscr{D}\rho\mathscr{D}\eta\mathscr{D}m\mathscr{D}\mu\exp\left[\int_{-t}^{0}dt'\times\left(i\mathscr{L}-\frac{y^2}{2}U\right)\right]\Bigg|_{\rho(-t)=\rho_0,\,\eta(-t)=\eta_0}. \tag{2.18}$$

It follows from the definition (2.18) that

$$\mathscr{Z}(y)=\lim_{t\to\infty}\int\prod d\rho_a d\eta_{ij}\Psi(t,y,\rho,\eta). \tag{2.19}$$

Eq. for the function Ψ can be obtained from the expression (1.32) and the definition (2.18):

$$\partial_t\Psi=\frac{Dd}{2}\Bigg[\sum_{i=1}^{d}\frac{\partial^2}{\partial\rho_i^2}-\frac{1}{d}\left(\sum_{i=1}^{d}\frac{\partial}{\partial\rho_i}\right)^2 -\sum_{i=1}^{d}(d-2i+1)\frac{\partial}{\partial\rho_i}+2\sum_{i<j}\exp(2\rho_j-2\rho_i) \times\frac{\partial^2}{\partial\eta_{ij}^2}+4\sum_{i<k<j}\exp(2\rho_k-2\rho_i)\frac{\partial}{\partial\eta_{ij}} \times\frac{\partial}{\partial\eta_{ik}}\eta_{kj}+2\sum_{i<k<m,n}\exp(2\rho_k-2\rho_i) \times\frac{\partial}{\partial\eta_{im}}\frac{\partial}{\partial\eta_{in}}\eta_{km}\eta_{kn}\Bigg]\Psi-\frac{y^2U}{2}\Psi. \tag{2.20}$$

We see that Eq. (2.20) for Ψ resembles the Schrödinger equation. The initial condition for the equation can be found directly from the definition (2.18):

$$\Psi(t=0,y,\rho,\eta)=\prod\delta(\rho_a)\delta(\eta_{ij}). \tag{2.21}$$

The value of $\mathcal{J}$, in accordance with (2.19); is determined by the integral of Ψ over η and ρ. This integral is equal to unity at $t=0$, and then varies with increasing time t due to $U\neq 0$, since only the term with U in (2.21) breaks the conservation of the integral. Thus, to find $\mathcal{J}$ we must establish the evolution of the function Ψ from $t=0$ to large t.

Below we concentrate on the single-point statistics. The scheme can obviously be generalized for the passive scalar difference.

Let us first describe the evolution qualitatively. The initial condition (2.21) shows that at $t=0$ the function Ψ is concentrated at the origin. Then it undergoes spreading in all directions, except for $\rho_1+\ldots+\rho_d$, since the operator on the right-hand side of (2.20) commutes with $\rho_1+\ldots+\rho_d$. This is a consequence of the condition $\det\hat{T}=1$ (to be satisfied), which implies that during evolution $\rho_1+\ldots+\rho_d=0$. This means that a solution of (2.20) is $\Psi\propto\delta(\rho_1+\ldots+\rho_d)$. The function Ψ is smeared diffusively with time, and also moves as a whole in some direction, which is determined by the term with the first derivative in (2.20). The rate of ballistic motion is

$$\langle\partial_t\rho_i\rangle=D\frac{d(d-2i+1)}{2}. \tag{2.22}$$

Therefore Ψ describes a cloud, the center of which moves according to the law

$$\rho_i=D\frac{d(d-2i+1)}{2}t. \tag{2.23}$$

Effective diffusion coefficients for the η's decrease with increasing t, since in accordance with (2.23) the differences $\rho_k-\rho_i$, figuring in (2.20), are negative and grow in absolute value. Therefore diffusion over η stops when the characteristic values of $\rho_i-\rho_k$ becomes greater than unity. Note that the ''frozen'' values of η do not depend on y, since U can be considered uniform during the initial stage of evolution. After that the η's are frozen, diffusion continues only over the ρ's. If the cloud is inside the region where $U\simeq P_2$, then evolution of the cloud is not influenced by U. After a period of time t_{lt} (2.16), the cloud reaches a barrier, where the potential U decreases from P_2 to 0. The subsequent history depends on the value of y. For moderate y the cloud passes this barrier and continues to move at the same rate. After this, the integral of Ψ will not change in time, and its value will determine the generating functional $\mathcal{J}(y)$. Naive estimates yield $\ln\mathcal{J}(y)=-y^2t_{lt}/2$, which reproduces the pair correlation function of θ.

Special consideration is needed if $|y|\gg y_{\text{sing}}$, or if y is close to $\pm iy_{\text{sing}}$, where y_{sing} is defined by (2.13). Just this region determines the PDF's and is consequently of special interest. Note that $y=\pm iy_{\text{sing}}$ corresponds to the appearance of a bound state near the pumping boundary (where U decreases from P_2 to zero). If $y\gg y_{\text{sing}}$, then the front of the cloud reaches the jump of the potential much earlier than t_{lt}. The remainder of the cloud (inside the potential well) is damped due to the term with y, and does not contribute to $\mathcal{J}(y)$. If $|y|\gg y_{\text{sing}}$ then $\mathcal{J}(y)\gg\exp(-y^2t_{lt}/2)$; the asymptotics of $\mathcal{J}(y)$ is actually exponential in the case. If $|y\pm iy_{\text{sing}}|\ll y_{\text{sing}}$ then the cloud stays near the pumping boundary for a long time, that is the shape of Ψ inside the region $U\simeq P_2$ varies in time comparatively slowly. Furthermore, a part of Ψ percolates out to the region where $U\simeq 0$, and the integral of Ψ grows with increasing $|t|$. As y approaches iy_{sing}, this stage lasts longer. One can say that the back of the cloud Ψ gives the right answer for $\mathcal{J}(y)$. The important point is that if y is not very close to iy_{sing} then during the time Ψ leaves the potential, the width of Ψ in terms of diffusive degrees of freedom is much less than $\ln L\Lambda$. This means that the function Ψ is really narrow, which justifies our consideration.

For a quantitative analysis it is convenient to pass to the variables ϕ_i (2.2). Since the η-dependence of Ψ is frozen after the initial evolution, it is possible to obtain an equation for the integral of Ψ over η:

$$\tilde{\Psi}(\phi_1,\ldots,\phi_{d-1})=\int d\phi_d\prod d\eta_{ij}\Psi, \tag{2.24}$$

where we also included an integration over ϕ_d to remove the factor $\delta(\rho_1+\ldots+\rho_d)$. Eq. for the function (2.24) is

$$\partial_t\tilde{\Psi}=\frac{Dd}{2}\left[\sum_{i=1}^{d-1}\frac{\partial^2}{\partial\phi_i^2}-\sqrt{\frac{d(d^2-1)}{3}}\frac{\partial}{\partial\phi_1}\right]\tilde{\Psi}-\frac{y^2\tilde{U}}{2}\tilde{\Psi}, \tag{2.25}$$

where $\tilde{U}$ is function of ϕ_a only which can be found by substituting into U the ''frozen'' values of η's. Qualitatively $\tilde{U}$ has the same structure as U itself. One can conclude from (2.25) that the cloud described by $\tilde{\Psi}$ moves ballistically in the ϕ_1 direction and spreads along other directions. We are going to treat the situation when the cloud remains narrow during the relevant part of the evolution. Then one can integrate $\tilde{\Psi}$ over all ϕ_i, $i>1$ in a similar way as in the case with η's, and get a $1d$ equation for

$$\bar{\Psi}(\phi_1)=\int\prod_2^{d-1}d\phi_i\tilde{\Psi}.$$

The function Ψ satisfies Eq.

$$\partial_t\overline{\Psi}=\frac{Dd}{2}\left[\frac{\partial}{\partial\phi_1}-\sqrt{\frac{d(d^2-1)}{3}}\right]\frac{\partial}{\partial\phi_1}\overline{\Psi}-y^2U2\,\overline{\Psi}. \tag{2.26}$$

The initial condition for Eq. (2.26) is $\overline{\Psi}(t=0)=\delta(\phi_1)$. The potential $\overline{U}$ is obtained from $\widetilde{U}$ by the substitution $\phi_a\to 0$ for $a>0$. In fact, for the direction (2.23) the potential $\overline{U}$ depends only on ρ_1. The barrier is reached when $\rho_1\simeq\ln L\Lambda$. Passing to the variables ϕ_i, we conclude that the potential $\widetilde{U}$ diminishes from P_2 at $\phi_1<\phi_\Lambda$ to zero at $\phi_1>\phi_\Lambda$, where ϕ_Λ is defined by (2.6).

The character of the solution of Eq. (2.26) can be analyzed semiqualitatively in terms of the width l of $\overline{\Psi}$ over ϕ_1 and its amplitude h. When $\overline{\Psi}$ reaches the pumping boundary, it stops there for a period of time. Then the width l and the amplitude h are governed by the equations

$$\frac{dl}{dt}=-Dd\lambda+\frac{Dd}{l},\quad \frac{dh}{dt}=-\frac{Ddh}{l^2}-\frac{y^2P_2h}{2}, \tag{2.27}$$

where $\lambda=\sqrt{d(d^2-1)/12}$, $Dd\lambda$ is the rate of cloud motion along the ϕ_1 direction (when $U=\mathrm{const}$), and Dd is the diffusion coefficient for the ϕ_1 direction. One can estimate from the first equation the width $l\sim 1/\lambda$. Then from the second equation the height h decreases or grows in time depending on y. The characteristic y where the regime changes is of the order $|y_{\mathrm{sing}}|^2\sim Dd\lambda^2/P_2$. We show this by consistent calculations.

Equation (2.26) can be solved analytically, e.g., by the Laplace transform over time t. Taking the Laplace transform, one gets

$$p\overline{\Psi}(p)-\delta(\phi_1)=\frac{Dd}{2}\left[\frac{\partial}{\partial\phi_1}-\sqrt{\frac{d(d^2-1)}{3}}\right]\frac{\partial}{\partial\phi_1}\overline{\Psi}(p)-\frac{y^2}{2}\overline{U}(\phi_1)\overline{\Psi}(p). \tag{2.28}$$

We are interesting in the bound state described by this equation. Solutions for $\Psi(p)$ in the intervals $(-\infty,0)$, $(0,\phi_\Lambda)$, (ϕ_Λ,∞) are exponential, and must be matched. The function $\Psi(p)$ as a function of p has two branch points at

$$p_1=-\frac{Dd^2(d^2-1)}{24}-\frac{y^2P_2}{2},\quad p_2=-\frac{Dd^2(d^2-1)}{24}, \tag{2.29}$$

coming from the regions $\phi_1<\phi_\Lambda$ and $\phi_1>\phi_\Lambda$, respectively. When one of these branch points passes $p=0$, Ψ starts to grow exponentially in time. This happens when y passes $\pm iy_{\mathrm{sing}}$, moving along the imaginary axis.

The value of the generating functional is determined in accordance with (2.19) by the large-time behavior of $\Psi(t)$. This means that we should be interested in the behavior of $\Psi(p)$ at small p. The function $\int d\phi_1\Psi(p)$ in (2.19) has a pole at $p=0$ related to the asymptotic behavior

$$\overline{\Psi}(p)\propto\exp\left(-\frac{2p}{Dd}\sqrt{\frac{3}{d(d^2-1)}}\,\phi_1\right),$$

at $\phi_1>\phi_\Lambda$ and small p; the behavior can be found from (2.28). The residue of $\int d\phi_1\Psi(p)$ at the pole determines $\mathscr{J}(y)$. To find the residue we must analyze the behavior of $\Psi(p)$ at $0<\phi_1<\phi_\Lambda$. At small p there are two contributions to Ψ, proportional to

$$\exp\left\{\left(\sqrt{\frac{d(d^2-1)}{12}}\pm\sqrt{\frac{d(d^2-1)}{12}+\frac{y^2P_2}{Dd}}\right)\phi_1\right\}, \tag{2.30}$$

as follows from (2.28) at $p=0$. Therefore the residue, which is determined by the integral $\int d\phi_1\Psi(p)$ over the region $\phi_1>\phi_\Lambda$, is proportional to

$$\exp\left\{\left(\sqrt{\frac{d(d^2-1)}{12}}+\sqrt{\frac{d(d^2-1)}{12}+\frac{y^2P_2}{Dd}}\right)\phi_\Lambda\right\}. \tag{2.31}$$

Substituting (2.6) here, we reproduce (2.12).

Let us now establish the applicability condition for the above procedure. The expression (2.31) implies that the exponent with the minus sign in (2.30) makes a negligible contribution to $\Psi(p)$ at $\phi_1=\phi_\Lambda$. The condition is satisfied if

$$|y^2+y_{\mathrm{sing}}^2|\,\phi_\Lambda^2\gg\frac{Dd}{P_2}.$$

Substituting (2.6) and (2.13) here, we obtain

$$\left|\frac{y\pm iy_{\mathrm{sing}}}{y_{\mathrm{sing}}}\right|\gg(d^4\ln^2L\Lambda)^{-1}. \tag{2.32}$$

For y close to $\pm iy_{\mathrm{sing}}$ one must be careful, since then the subtle analytic structure of $\mathscr{J}(y)$ will be relevant. As an analysis for $d=2$ shows $\mathscr{J}(y)$ has a system of poles along the imaginary semiaxis starting from $\pm iy_{\mathrm{sing}}$, and the parameter $(d^4\ln^2L\Lambda)^{-1}$ determines the separation between the poles. The poles correspond to bound states. The assertion about the cut made in the previous subsection is related to the restrictions of the saddle-point approximation which cannot feel this fine pole structure; it yields the cut, which is a picture averaged over the interpole distances. This averaged picture is acceptable at the condition (2.32).

Note that the same criterion (2.32) justifies our assumption that the cloud described by Ψ is narrow during the relevant part of the evolution. Namely, the duration of the part is determined by the time $t_{\mathrm{exit}}=p_1^{-1}$ [see (2.29)]. This is the time that the cloud stays near the barrier. For y close to $\pm iy_{\mathrm{sing}}$, the time can be estimated to be $t_{\mathrm{exit}}^{-1}\sim P_2|y_{\mathrm{sing}}||y\mp iy_{\mathrm{sing}}|$. Then the diffusive width $\sqrt{Ddt_{\mathrm{exit}}}$ of Ψ in the directions ϕ_a for $a>1$ is much less than ϕ_Λ precisely if (2.32) is satisfied. In principle the diffusive dynamics at $d>2$ could modify the noted fine pole structure of $\mathscr{J}$; this problem requires additional investigation.

The same procedure can be done for the passive scalar differences. The cloud Ψ should pass the region $\rho_1<\ln(L/r_0)$ before it reaches the potential. Then it enters the region $\overline{U}=2P_2$ with some finite diffusive width. One can note, however, that this is irrelevant. The only characteristics of the potential that are needed are its value (here $2P_2$ instead of P_2) and the length of the path inside it [which is $\Delta\rho_1=\ln(r_0\Lambda)$ instead of $\ln(L\Lambda)$]. The evolution of Ψ goes in the same way as in the case of single-point statistics. Again, we get (2.14) and the criterion analogous to (2.32).

In this subsection we presented an analysis based on the dynamical equation (2.20) for the auxiliary object Ψ. The results obtained can be reproduced also in alternative language: for this we must introduce another auxiliary object, the equation for which is stationary. The corresponding scheme, which might be interesting from a methodological point of view, is sketched in the Appendix.

3. CALCULATION OF PDF

In this section we calculate the PDF's $\mathcal{P}$ for the objects (1.8) and (1.10). The most convenient way to do so is by using the relation

$$\mathcal{P}(\vartheta)=\int\frac{dy}{2\pi}\exp(-iy\vartheta)\mathcal{Z}(y), \tag{3.1}$$

where ϑ is

$$\vartheta=\int d\mathbf{r}\beta(\mathbf{r})\theta(0,\mathbf{r}). \tag{3.2}$$

Let us recall that knowing $\mathcal{P}(\vartheta)$, one can also restore the moments of ϑ:

$$\langle|\vartheta|^n\rangle=\int d\vartheta|\vartheta|^n\mathcal{P}(\vartheta). \tag{3.3}$$

The generating functional in (3.1) is determined by (2.12) or (2.14). Being interested in the main exponential dependence of the PDF's for the objects (1.8) and (1.10), we can forget about preexponents. Then

$$\mathcal{P}(\vartheta)=\int\frac{dy}{2\pi}\exp(-iy\vartheta+q[1-\sqrt{1+y^2/y_{\text{sing}}^2}]), \tag{3.4}$$

where for the single-point statistics and for the statistics of the passive scalar difference respectively

$$y_{\text{sing}}^2=\frac{Dd^2(d^2-1)}{12P_2},\quad y_{\text{sing}}^2=\frac{Dd^2(d^2-1)}{24P_2}, \tag{3.5}$$

$$q=\frac{d(d+1)}{6}\ln(L\Lambda),\quad q=\frac{d(d+1)}{6}\ln(r_0\Lambda). \tag{3.6}$$

Since both q defined by (3.6) are regarded to be much larger than unity, the integral (3.4) can be calculated in the saddle-point approximation. The saddle-point value is

$$y_{\text{sp}}=i\frac{y_{\text{sing}}}{1+q^2/y_{\text{sing}}^2\vartheta^2}. \tag{3.7}$$

Then

$$\ln\mathcal{P}(\vartheta)\simeq q\left(1-\sqrt{1+\frac{y_{\text{sing}}^2\vartheta^2}{q^2}}\right). \tag{3.8}$$

This expression leads to the exponential tail

$$\ln\mathcal{P}(\vartheta)\simeq -y_{\text{sing}}|\vartheta|, \tag{3.9}$$

realized at $|\vartheta|\gg q/y_{\text{sing}}$. The coefficient y_{sing} in (3.9) determined by (2.13) is in agreement with the result obtained in Ref. 12.

The expression (3.8) enables one to find the following averages in accordance with (3.3):

$$\langle\theta_\Lambda^2\rangle=\frac{2P_2}{d(d-1)D}\ln(L\Lambda),$$

$$\langle(\Delta\theta_\Lambda)^2\rangle=\frac{4P_2}{d(d-1)D}\ln(r_0\Lambda). \tag{3.10}$$

The expressions (3.10) can also be obtained by direct expansion of $\mathcal{Z}(y)$ from (2.12) or (2.14). The universal tail (3.9) is realized if

$$\theta_\Lambda\gg\sqrt{\langle\theta_A^2\rangle}d\ln(L\Lambda),\quad \Delta\theta_\Lambda\gg\sqrt{\langle(\Delta\theta_\Lambda)^2\rangle}d\ln(r_0\Lambda). \tag{3.11}$$

Since both logarithms are assumed to be large, we conclude that there exists a relatively wide region where the statistics of ϑ is approximately Gaussian; the region is determined by the inequalities inverse to (3.11).

Let us discuss the applicability conditions of the expression (3.8). First, if one calculates the passive scalar PDF by the saddle point method, then the position of the saddle point is determined by (2.32) if

$$\vartheta\ll d^2\sqrt{\frac{P_2}{D}}\ln^2(L\Lambda). \tag{3.12}$$

The applicability domain of the saddle-point method overlaps the region of validity of (2.12) for the generation function $\mathcal{Z}(y)$. The above inequalities are correct for θ_Λ; for $\Delta\theta_\Lambda$ one must replace $\ln(L\Lambda)$ with $\ln(r_0\Lambda)$. Second, fluctuations of y have to be small compared to the distance between y_{sp} and y_{sing}. This gives the same criterion (3.12).

Let us stress that though formally our procedure is incorrect at $\vartheta\gtrsim d^2\sqrt{P_2/D}\ln^2(L\Lambda)$ the answer will be the same: the PDF will be determined by the exponential tail (3.9). The point is that the character of the integral (3.1) at such extremely large ϑ will be determined by the position of the singular point of $\mathcal{Z}(y)$ nearest to the real axis. This is just iy_{sing}, leading to (3.9). To conclude, only the character of the preexponent in $\mathcal{P}(\vartheta)$ is changed at $\vartheta\sim d^2\sqrt{P_2/D}\ln^2(L\Lambda)$, whereas the principal exponential behavior of $\mathcal{P}(\vartheta)$ remains unchanged there.

4. CONCLUSION

The single-point statistics of the passive scalar θ and the statistics of its difference $\Delta\theta$ are traditional objects which carry essential information about correlation functions of the passive scalar in the convective interval. We examined the passive scalar in the large-scale turbulent flow, where the correlation functions logarithmically depend on scale. Since the logarithms are actually not very large, it is useful to have all the PDF's of θ and $\Delta\theta$. That was the main purpose of our investigation, which was performed in the context of the Kraichnan model. The single-point PDF for the passive scalar and the PDF for the passive scalar differences can be obtained from (3.8) if we substitute $\Lambda\rightarrow r_{\text{dif}}^{-1}$ where r_{dif} is the diffusive length. Though both the advecting velocity and the pumping force in the Kraichnan model are considered δ-correlated in time, we hope that our results are universal, that is, are true in the limit when the size of the convective interval tends to infinity for arbitrary temporal behavior of the velocity and pumping. The reason is that the spectral transfer time grows with increasing convective interval, and in the limit is much larger than the correlation times of the velocity and pumping.

We believe also that the analytic scheme proposed in our work could be extended for other problems related to the passive scalar statistics. Note as an example Ref. 13 where a modification of the scheme enabled one to find the statistics of the passive scalar dissipation. It is also useful for investigating the large-scale statistics (on scales larger that the pumping length) of the passive scalar see Ref. 25. We also hope that it is possible to go beyond the case of the large-scale velocity field using a perturbation technique of the type proposed in Refs. 26–28.

We are grateful to E. Balkovsky, M. Chertkov, G. Falkovich, K. Gawedzki and M. Olshanetsky for useful discussions. This work was supported in part by the Einstein and Minerva Centers at the Weizmann Institute, by grants from the Minerva Foundation, Germany and the Israel Science Foundation, by the Russian Fund for Fundamental Research (I. K., M. S., Grant 98-02-17814), by Soros Foundation (M. S., Grant a98-674) and by INTAS (M. S., Grant 96-0457) within the ICFPM program.

APPENDIX

Here we present an alternative way to obtain the results (2.12) and (2.14). We use an auxiliary quantity.

$$\Xi(y,\rho_0,\eta_0)=\int \mathscr{D}\rho\mathscr{D}\eta\mathscr{D}m\mathscr{D}\mu \times\exp\left[\int_{-\infty}^{0} dt\left(i\mathscr{L}-\frac{y^2}{2}U\right)\right]\Bigg|_{\rho(0)=\rho_0,\,\eta(0)=\eta_0}, \tag{A1}$$

so

$$\mathscr{J}(y)=\Xi(y,0,0). \tag{A2}$$

The function Ξ can be also defined as

$$\Xi(y,\rho_0,\eta_0)=\lim_{t\to\infty}\int\prod d\rho_a d\eta_{ij}\Psi(t,y,\rho,\eta), \tag{A3}$$

where Ψ is governed by Eq. (2.20) with initial condition $\Psi(t=0,y,\rho,\eta)=\delta(\rho-\rho_0)\delta(\eta-\eta_0)$. The equation for Ξ can be found from Eqs. (1.32) and (A1):

$$\left[\sum_{i=1}^{d}\frac{\partial^2}{\partial\rho_i^2}-\frac{1}{d}\left(\sum_{i=1}^{d}\frac{\partial}{\partial\rho_i}\right)^2+\sum_{i=1}^{d}(d-2i+1)\frac{\partial}{\partial\rho_i}\right.$$

$$+2\sum_{i<j}\exp(2\rho_j-2\rho_i)\frac{\partial^2}{\partial\eta_{ij}^2}+4\sum_{i<k<j}\exp(2\rho_k$$

$$-2\rho_i)\eta_{kj}\frac{\partial}{\partial\eta_{ij}}\frac{\partial}{\partial\eta_{ik}}+2\sum_{i<k<m,n}\exp(2\rho_k$$

$$\left.-2\rho_i)\eta_{km}\eta_{kn}\frac{\partial}{\partial\eta_{im}}\frac{\partial}{\partial\eta_{in}}\right]\Xi-\frac{y^2U}{Dd}\Xi=0. \tag{A4}$$

The boundary condition for Eq. (A4) follows from the definition (A1): for large enough ρ_i,η_i the potential $U=0$ at $t=0$ and also remains zero at finite times t. Therefore the integral (A1) must be equal to unity in the case. Thus $\Xi(y,\rho,\eta)$ must tend to unity where $\rho,\eta\to\infty$.

Let us rewrite Eq. (A4) in terms of the variables (2.2):

$$(\hat{\Gamma}_1+\gamma(\Xi_1+\xi)=0,\quad \Xi=\Xi_1+\xi, \tag{A5}$$

$$\hat{\Gamma}_1=\frac{\partial^2}{\partial\phi_1^2}+\sqrt{\frac{d(d^2-1)}{3}}\frac{\partial}{\partial\phi_1}-\frac{y^2U}{Dd}, \tag{A6}$$

$$\hat{\gamma}=\sum_{i=2}^{d-1}\frac{\partial^2}{\partial\phi_i^2}+2\sum_{i<k}\exp(2\rho_k-2\rho_i)\frac{\partial^2}{\partial\eta_{ik}^2}$$

$$+4\sum_{i<k<n}\exp(2\rho_k-2\rho_i)\eta_{kn}\frac{\partial}{\partial\eta_{in}}\frac{\partial}{\partial\eta_{ik}}$$

$$+2\sum_{i<k<m,n}\exp(2\rho_k-2\rho_i)\eta_{km}\eta_{kn}\frac{\partial}{\partial\eta_{im}}\frac{\partial}{\partial\eta_{in}}. \tag{A7}$$

Here U as a function of ϕ_1 is equal to P_2 inside a region restricted by ϕ_Λ^- and ϕ_Λ^+ (where $\phi_\Lambda^\pm$ are functions of variables $\phi_2,\ldots,\phi_d,\eta$) and tends to zero outside the region. We solve Eq. (A5) using perturbation theory over $\hat{\gamma}$, ξ. Then the zero-order equation is

$$\hat{\Gamma}_1\Xi_1=0. \tag{A8}$$

Equation (A8) can easily be solved at $\phi_\Lambda^-<\phi_1<\phi_\Lambda^+$; the answer is

$$\Xi_1\simeq\frac{2\lambda}{\sqrt{\lambda^2+\frac{y^2P_2}{Dd}}+\lambda}\exp\{-(\sqrt{\lambda^2+y^2P_2/Dd}-\lambda)$$

$$\times(\phi_\Lambda^+-\phi_1)\}, \tag{A9}$$

where $\lambda=\sqrt{d(d^2-1)/12}, Dd\lambda$ is the rate of the cloud motion along the φ_1 direction. The result (A9) can be obtained using the inequality $\sqrt{\lambda^2+y^2P_2/Dd}\ln L\Lambda\gg 1$. The derivative $\partial\Xi_1/\partial\phi_1=0$ at $\phi_1<\phi_\Lambda^-$. However, $\Xi_1\neq 1$ in this region. This is due to the following fact: this region corresponds to the evolution of Ψ when its initial position is to the left of potential U [see (A3)]. During evolution, cloud Ψ passes the region of U and its integral over ρ, η changes. Then Ξ is not equal to 1. Only when the distance between the initial position and potential is of order $\ln^2 L\Lambda$ will the diffusion of the cloud lead to smallness of the part of Ψ that passes the potential U, and Ξ becomes closer to unity. Thus, function Ξ has a long tail from the potential pointing toward negative ϕ_1, where it is not equal to 1. The procedure of finding Ξ from Eq. (A8) corresponds to the geometrical optics approximation (taking into account only derivatives in propagation direction; this allows one to get the fact of propagation). This tail of Ξ in this approximation is none other than the shadow of potential U. Higher orders of perturbation theory over the transverse derivatives correspond to diffraction corrections.

Now let us consider the correction ξ. Eq. for it looks like $(\Gamma_1+\hat{\gamma}\xi=-\hat{\gamma}\Xi_1$. Again let us neglect $\hat{\gamma}$ on the left-hand side and solve the equation. Ξ_1 is some exponential function with scale of the order 1. Then $\hat{\gamma}\Xi_1\sim\Xi_1$. Note that $\hat{\gamma}\Xi_1$ is almost equal to zero at $\phi_1>\phi_\Lambda^+$. To estimate ξ one must construct the Green function $G(\phi_1|\phi_0)$ for operator Γ_1:

$$G(0|\phi_0)\simeq\frac{1}{2\lambda}\exp\left(-\left(\sqrt{\lambda^2+\frac{y^2P_2}{Dd}}-\lambda\right)\phi_0\right)\times\left(1-C\exp\left(-2\sqrt{\lambda^2+\frac{y^2P_2}{Dd}}(\phi_\Lambda^+-\phi_0)\right)\right), \tag{A10}$$

where

$$C=(\sqrt{\lambda^2+y^2P_2/Dd}-\lambda)/(\sqrt{\lambda^2+y^2P_2/Dd}+\lambda).$$

The unity in the parentheses in (A10) gives the correction for Ξ, which has the same exponential factor as Ξ_1. Thus ξ does not change the answer, to logarithmic accuracy. The second term in the parentheses gains while ϕ_0 is close to ϕ_Λ^+. This is due to the nonzero width of the cloud Ξ and to the dependence of t_{lt} on other variables. Again, it does not change the exponent.

To get $\mathcal{J}$ from Ξ we in accordance with (A2) have to substitute zero values of ρ and η into Ξ. Then $\phi_1=0$ and $\phi_\Lambda^+=\phi_\Lambda$ where ϕ_Λ is defined by (2.6). Substituting the values into (A9) we reproduce (2.12). The case of the passive scalar differences can be considered in a similar way.

---

*⁾E-mail: lebede@cpd.landau.ac.ru; lwlebede@wicc.weizmann.ac.il

---

[1] G. K. Batchelor, J. Fluid Mech. **5**, 113 (1959).
[2] R. Kraichnan, Phys. Fluids **10**, 1417 (1967); Phys. Fluids **11**, 945 (1968); J. Fluid Mech. **47**, 525 (1971); J. Fluid Mech. **67**, 155 (1975); J. Fluid Mech. **64**, 737 (1974).
[3] B. I. Shraiman and E. D. Siggia, Phys. Rev. E **49**, 2912 (1994).
[4] M. Chertkov, G. Falkovich, I. Kolokolov, and V. Lebedev, Phys. Rev. E **51**, 5609 (1995).
[5] G. K. Batchelor, Theory of Homogeneous Turbulence, Cambridge University Press, New York (1953).
[6] A. Monin and A. Yaglom, Statistical Fluid Mechanics, MIT Press, Cambridge (1975).
[7] U. Frisch, Turbulence: the Legacy of A. N. Kolmogorov, Cambridge University Press, New York (1995).
[8] E. Balkovsky, M. Chertkov, I. Kolokolov, and V. Lebedev, JETP Lett. **61**, 1049 (1995).
[9] G. Falkovich and V. Lebedev, Phys. Rev. E **50**, 3883 (1994).
[10] M. Chertkov, A. Gamba, and I. Kolokolov, Phys. Lett. A **192**, 435 (1994).
[11] G. Falkovich, I. Kolokolov, V. Lebedev, and A. Migdal, Phys. Rev. E **54**, 4896 (1996).
[12] D. Bernard, K. Gawedzki, and A. Kupiainen, J. Stat. Phys. **90**, 519 (1998).
[13] A. Gamba and I. Kolokolov, E-prints archive, chao-dyn/9808001.
[14] M. L. Mehta, Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York (1967).
[15] M. A. Olshanetsky and A. M. Perelomov, Phys. Rep. **94**, 313 (1983).
[16] I. V. Kolokolov, Ann. of Phys. **202**, 165 (1990); Zh. Éksp. Teor. Fiz. **103**, 2196 (1993) [Zh. Eksp. Teor. Fiz. **76**, 1099 (1993)].
[17] M. Chertkov, G. Falkovich, and I. Kolokolov, Phys. Rev. Lett. **80**, 2121 (1998).
[18] O. N. Dorokhov, Zh. Éksp. Teor. Fiz. **85**, 1040 (1983) [Sov. Phys. JETP **58**, 606 (1983)].
[19] A. Gamba and I. Kolokolov, J. Stat. Phys. **85**, 489 (1996).
[20] H. W. Wyld, Ann. Phys. **14**, 143 (1961).
[21] P. C. Martin, E. Siggia, and H. Rose, Phys. Rev. A **8**, 423 (1973).
[22] C. de Dominicis, J. de Phys. **37**, c01-247 (1976).
[23] H. Janssen, Z. Phys. B **23**, 377 (1976).
[24] C. de Dominicis and L. Peliti, Phys. Rev. B **18**, 353 (1978).
[25] E. Balkovsky, G. Falkovich, and V. Lebedev, submitted to Phys. Fluids (1999).
[26] B. I. Shraiman and E. D. Siggia, CRAS **321**, Ser. II, 279 (1995).
[27] A. Pumir, B. I. Shraiman, and E. D. Siggia, Phys. Rev. E **55**, 1263 (1997).
[28] E. Balkovsky, G. Falkovich, and V. Lebedev, Phys. Rev. E **55**, 4881 (1997).
[29] M. Chertkov, A. Gamba, and I. Kolokolov, Phys. Lett. A **192**, 435 (1994).
[30] G. Falkovich, I. Kolokolov, V. Lebedev, and A. Migdal, Phys. Rev. E **54**, 4896 (1996).
[31] D. Bernard, K. Gawedzki, and A. Kupiainen, J. Stat. Phys. **90**, 519 (1998).
[32] A. Gamba and I. Kolokolov, E-prints archive, chao-dyn/98 08 001.
[33] M. L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York (1967).
[34] M. A. Olshanetsky and A. M. Perelomov, Phys. Rep. **94**, 313 (1983).
[35] I. V. Kolokolov, Ann. of Phys. **202**, 165 (1990); Zh. Éksp. Teor. Fiz. **103**, 2196 (1993) [JETP **76**, 1099 (1993)].
[36] M. Chertkov, G. Falkovich, and I. Kolokolov, Phys. Rev. Lett. **80**, 2121 (1998).
[37] O. N. Dorokhov, Zh. Éksp. Teor. Fiz. **85**, 1040 (1983) [Sov. Phys. JETP **58**, 606 (1983).
[38] A. Gamba and I. Kolokolov, J. Stat. Phys. **85**, 489 (1996).
[39] H. W. Wyld, Ann. Phys. **14**, 143 (1961).
[40] P. C. Martin, E. Siggia, and H. Rose, Phys. Rev. A **8**, 423 (1973).
[41] C. de Dominicis, J. de Phys. **37**, c01-247 (1976).
[42] H. Janssen, Z. Phys. B **23**, 377 (1976).
[43] C. de Dominicis and L. Peliti, Phys. Rev. B **18**, 353 (1978).
[44] E. Balkovsky, G. Falkovich, and V. Lebedev, submitted to Phys. Fluids (1999).
[45] B. I. Shraiman and E. D. Siggia, CRAS **321**, Ser. 11, 279 (1995).
[46] A. Pumir, B. I. Shraiman, and E. D. Siggia, Phys. Rev. E **55**, 1263 (1997).
[47] E. Balkovsky, G. Falkovich, and V. Lebedev, Phys. Rev. E **55**, 4881 (1997).

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

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Helical ordering of hydrogen bonds between pairs of DNA bases

V. L. Golo

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

Yu. M. Evdokimov and S. G. Skuridin

Institute of Molecular Biology, Russian Academy of Sciences, 117984 Moscow, Russia

E. I. Kats*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia

(Submitted 12 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 940–950 (March 1999)

The interaction between hydrogen bonds and conformational elastic degrees of freedom has been investigated using the simplest model of a double-strand DNA molecule. The hydrogen bonds are described in terms of two-level quantum systems. After excluding conformational degrees of freedom, one has effective interaction among two-level systems. In the ground state of an ideal double helix, hydrogen bonds in a DNA molecule also have a helical order induced by conformational degrees of freedom. The pitch of the hydrogen-bond helix (and even its sign under certain conditions) is different from that of the basic helix pitch and, generally speaking, is incommensurate with the latter. This effect can, possibly, lead to an inversion of the sign of the circular dichroism in spectral bands, which was detected in some experiments.
[S1063-7761(99)01103-8]

1. INTRODUCTION

According to Watson and Crick's classic model, the DNA molecule is a double helix. In the so-called B-conformation of DNA (see the monograph by Alberts et al.[1] or the review by Vologodskii *et al.*[2]), this double helix is right-handed and formed of two polymer chains, each of which has the shape of a right-handed spiral. Nitrous bases, whose sequence encodes the entire genetic information, form the inside content of the double helix, and its surface is formed by the glycophosphate frames of the polymer chains. Each repeated component of the chain (a nitrous base plus a glycophosphate) is called a nucleotide. There is a strict rule of complementation between nucleotides of two chains, which is controlled mostly by the steric correspondence between bases. This structure of the DNA molecule allows one to model it in most cases as an anisotropic elastic thread associated with the axial line of the double helix.[3,4]

An essential fact for the DNA configuration is that bonds between neighboring nucleotides are rigid and covalent, with energy about 60 kcal/mol,[1,2] whereas the energy of bonds between polynucleotide chains is at least one order of magnitude lower. The two chains are bound in the double helix mostly by the interaction between bases belonging to different strands, and hydrogen bonds serve, as it were, for ''identification'' of nucleotides in complementary pairs. Obviously, the variables characterizing one anisotropic elastic thread are insufficient for modeling the double-strand DNA configuration. A minimal complication of the anisotropic elastic-thread model suggested in our previous publication[5] was equipping the elastic line with vector field **Y**, which defined additional (to those in the anisotropic thread model) degrees of freedom in the DNA double helix.

One can conclude from the above statements that the role of hydrogen bonds in stabilizing the DNA configuration is, in a sense, secondary. This does not mean, however, that they are inessential in the biological functioning of the DNA molecule. Rather the opposite is true, for it is precisely the hydrogen bonds that enable realization of the rule of complementation between bases of different pairs. Therefore, studying the ordering and other properties of the system of hydrogen bonds in the DNA molecule is a biophysical problem of paramount importance, and it constitutes the main subject of this paper.

Hydrogen bonds have been extensively discussed in the literature (see, for example, the review by Marechal[6] and references therein). An important circumstance for our study is that a hydrogen bond has an anisotropic configuration, because three atoms forming a hydrogen bond are shaped in a linear asymmetrical structure. Keeping in mind the problem of ordering of hydrogen bonds in the DNA molecule mentioned above, we can say in other words that, at each site n of a DNA molecule (a position of bonds between complementary pairs), there is a certain direction $\mathbf{H}_n$ which defines a favorable orientation of a hydrogen bond. In contrast to the case of covalent bonds, however, the inherent energy of a hydrogen bond is relatively low (characteristic energies of hydrogen bonds[7] are usually no more than 3 kcal/mol). Therefore, interaction between hydrogen bonds and other degrees of freedom in DNA (along with thermal fluctuations) can notably modify their properties (for example, deform the linear structure of these bonds) or even break them. In the following section, we will formulate a simple model illustrating this phenomenon, namely, interaction between hydrogen bonds and conformational degrees of freedom in the DNA molecule.

This model will be analyzed in Sec. 3. The ground state of the system will be described. It turns out that hydrogen bonds have a helical order in the ground state (like base pairs in the double helical configuration of DNA), but the pitch of this hydrogen-bond helix (and, possibly, its sign) differs from that of the double-helix pitch (10 b.p.≡35 Å) and, generally speaking, these two pitches are incommensurate.

In the final section, we will discuss feasibility of detecting helical ordering of hydrogen bonds and its biophysical aspects.

2. DNA MODEL TAKING INTO ACCOUNT HYDROGEN BONDS

As was mentioned above, the basic component of the DNA model used in our analysis is the conformation of the DNA axial line, determined by Kirchhoff's ''angular velocity'' $\mathbf{\Omega}$ (see Ref. 8, and also 3 and 4). The vector $\mathbf{\Omega}$ (or the equivalent skew-symmetric matrix) is a natural set of variables describing the conformation of an elastic thread (the DNA axial line). Thus, the molecule's conformation (i.e., vector $\mathbf{\Omega}$) is determined by the elastic moduli of the DNA moleclue (or, what is the same, by persistent lengths of bend and twist). The characteristic values of these elastic moduli are well known[1,2] and correspond to a persistent length of order of 10^3 Å. A DNA molecule whose length is of order of the persistent length is almost rigid, i.e., $\Omega=\text{const}$ over this length.

The next element of our model is vector field $\mathbf{Y}$ that frames the axial line conformation. Introduction of this field is dictated by the double-strand structure of the DNA molecule, and this field locally determines the relative location of the strands and separation between them.

This vector field is naturally associated with an energy which includes the kinetic energy of relative motion of strands with respect to one another and the potential energy of the ideal double-helix deformation. In the simplest approximation of the lowest order, the energy of field $\mathbf{Y}$ can be expressed as

$$\mathscr{H}_{\text{el}}=\sum_n \frac{1}{2\rho}\mathbf{P}_n^2+\sum_n \frac{K}{2}(\nabla\mathbf{Y}_n)^2. \tag{1}$$

Here we are using the discrete model of DNA, in which the sites n are defined as positions of glycophosphate bases, $\mathbf{P}_n$ is the generalized momentum corresponding to the relative displacement $\mathbf{Y}_n$ of strands of the double helix, ρ is the reactive factor, which has the sense of the mass density in this relative motion, and K is the elastic modulus characterizing elastic strains in the double-strand structure (or field $\mathbf{Y}$).

Energy (1) has the standard form of the energy of harmonic oscillations, the only, but essential, difference from a conventional one-dimensional oscillator being the presence of the covariant derivative $\nabla\mathbf{Y}_n=(\partial\mathbf{Y}_n/\partial s)-[\mathbf{\Omega}\mathbf{Y}_n]$, where s is the curvilinear coordinate along the conformation of the axial line.

The emergence of the covariant derivative instead of the conventional one results from the fact that the deformation energy of field $\mathbf{Y}$ [expressed in the immobile (laboratory) reference frame in Eq. (1)] should include both changes in field $\mathbf{Y}$ with respect to the natural (local) reference frame defined by vector $\mathbf{\Omega}$ and changes in the local frame itself.

In addition to field $\mathbf{Y}_n$, there are hydrogen bonds at each site, whose main function is (as was noted in the previous section) ''identification'' of complementary pairs of bases. In the roughest approximation, a linear hydrogen bond can be only in two states (corresponding to two proton positions), which will be dubbed in what follows open (the proton is close to one of participating atoms, and the bond is effectively broken) and closed (the proton is located between the two atoms and effectively binds them). A natural ''tool'' for describing the two-level system modeling the hydrogen bond is the triad of Pauli matrices $\mathbf{s}_n$ for spin 1/2. This approach allows us to express the Hamiltonian of the system of hydrogen bonds in DNA as follows:

$$\mathscr{H}_h=\varepsilon\sum_n \mathbf{H}_n\mathbf{s}_n+\gamma\sum_n \mathbf{Y}_n\mathbf{s}_n. \tag{2}$$

The first term on the right corresponds to the inherent energy of a hydrogen bond at site n (ε is the energy needed to break the hydrogen bond, $\mathbf{H}_n$ describes the unperturbed equilibrium alignment of the hydrogen bond), and the second corresponds to interaction between hydrogen bonds and field $\mathbf{Y}$ of displacements in the DNA double-strand configuration. The value of ε is well known (it is of order of 3 kcal/mol), and parameter γ, which is controlled by interaction between hydrogen bonds and nitrous bases, can be several times higher, but, obviously, lower than the elastic strain energy of field $\mathbf{Y}$, which is determined purely by covalent bonds. In what follows, this hierarchy of energy scales will be used explicitly in determination of the ground state.

In studying properties of the model formulated above, one can use techniques developed in the theory of excitons.[9,10] It is convenient to switch to Fourier components in Eqs. (1) and (2):

$$\mathbf{Y}_q=\frac{1}{\sqrt{N}}\sum_n e^{inq}\mathbf{Y}_n,$$

where N is the number of sites (base pairs) on the section under consideration (of order of the persistent length). Moreover, we assume that the double-helix conformation of the DNA molecule is ideal, i.e.

$$\mathbf{\Omega}\equiv(0,0,\Omega)=\text{const}.$$

After the Fourier transform, Hamiltonian (1), (2) takes the form

$$\mathscr{H}=\sum_q\left[\frac{1}{2\rho}\mathbf{P}_q\mathbf{P}_{-q}+\frac{K}{2}(\hat{B}_q^{+}\hat{B}_q)\mathbf{Y}_q\mathbf{Y}_{-q}\right.$$
$$\left.+\varepsilon\mathbf{H}_q\mathbf{S}_{-q}+\gamma\mathbf{Y}_q\mathbf{S}_{-q}\right] \tag{3}$$

with the notation

$$\hat{B}_q=\begin{bmatrix} 2i\sin(q/2) & -\exp(iq/2)\Omega & 0 \\ \exp(iq/2)\Omega & -2i\sin(q/2) & 0 \\ 0 & 0 & -2i\sin(q/2) \end{bmatrix} \tag{4}$$

($\hat{B}_q^+$ is the Hermitian conjugate matrix).

The first two terms on the right of Eq. (3) express the elastic energy of the conformation degrees of freedom in a double-strand DNA molecule. In order to calculate the spectrum of elastic oscillations, one must diagonalize this Hamiltonian part. This can be done with the help of the canonical transformation

$$S^{-1}(\hat{B}_q^+\hat{B}_q)S\equiv\mathrm{diag}\,(\lambda_1,\lambda_2,\lambda_3), \tag{5}$$

where the matrix

$$S=\begin{bmatrix} 1/\sqrt{2} & i/\sqrt{2} & 0 \\ i/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

satisfies the condition $S^+=S^{-1}$.

Equations (5) and (6) yield three characteristic eigenmodes of the double-strand DNA molecule:

$$\lambda_{1,2}=4\sin^2\frac{q}{2}+\Omega^2\pm4\Omega\sin\frac{q}{2}\cos\frac{q}{2}, \qquad \lambda_3=4\sin^2\frac{q}{2}. \tag{7}$$

Operator $\hat{S}$ transforms all variables to eigenmodes:

$$\mathbf{Y}_q=\hat{S}\boldsymbol{\xi}_q,\quad \mathbf{P}_q=\hat{S}^*\boldsymbol{\pi}_q,\quad \mathbf{S}_q=\hat{S}\boldsymbol{\eta}_q,\quad \mathbf{H}_q=\hat{S}\mathbf{h}_q, \tag{8}$$

where, as follows from Eq. (6), $S^*=S^+$ (S^* is the complex conjugate matrix), thus we obtain Hamiltonain (3) in terms of eigenmodes:

$$\mathcal{H}=\sum_q\sum_{i=1}^{3}\left[\frac{1}{2\rho}(\pi_q^{+i}\pi_q^i+\rho^2\omega_i^2(q)\xi_q^{+i}\xi_q^i) + \frac{\gamma}{2}(\xi_q^{+i}\eta_q^i+\eta_q^{+i}\xi_q^i)+\frac{\epsilon}{2}(h_q^{+i}\eta_q^i+\eta_q^{+i}h_q^i).\right], \tag{9}$$

where

$$\omega_i=\sqrt{K\lambda_i/\rho} \tag{10}$$

is the spectrum of eigenmodes (''phonons'' of the double-strand DNA). Branches of the spectrum determined by Eqs. (7) and (10) are shown in Fig. 1. There is an ''acoustic'' branch (ω_3) corresponding to oscillations of the separation between the strands (length of vector **Y**) and two ''optical'' modes (ω_1 and ω_2) associated with deformation in the ideal double helix (with variations in the orientation of field **Y**). Note that there is a minimum of mode ω_2 at $q=Q$, which can be explained entirely in terms of symmetry properties[1].

3. HELICAL ORDERING OF HYDROGEN BONDS IN DNA

It is convenient to express the elastic (''phonon'') part of Hamiltonian (9) in terms of second quantization[12]:

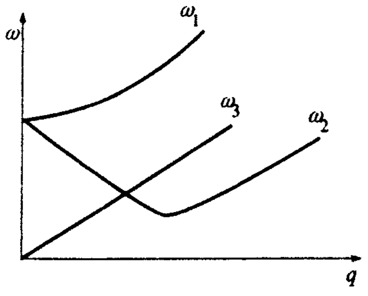

FIG. 1.

$$b_q^i=\frac{\rho\omega_i(q)\xi_q^i+i\pi_q^{+i}}{\sqrt{2\rho\hbar\omega_i(q)}},\qquad b_q^{i+}=\frac{\rho\omega_i(q)\xi_q^{i+}-i\pi_q^i}{\sqrt{2\rho\hbar\omega_i(q)}}. \tag{11}$$

Equations (11) are written in such a form that phonon creation and annihilation operators satisfy the standard commutation relations

$$[b_{q'}^i;b_{q''}^{+j}]=i\hbar\delta_{ij}\delta_{q'q''}$$

with due account of symmetry properties of phonon spectra (10),

$$\omega_1(-q)=\omega_2(q),\qquad \omega_3(q)=\omega_3(-q).$$

The energy of ''bare'' phonons has the form

$$\mathcal{H}_0=\sum_q\sum_{i=1}^{3}\hbar\omega_i(q)b_q^{+i}b_q^i, \tag{12}$$

and the interaction Hamiltonian

$$\mathcal{H}_{\mathrm{int}}=\frac{\epsilon}{2}\sum_q\sum_{i=1}^{3}(h_q^{+i}\eta_q^i+\eta_q^{+i}\eta_q^i)+\gamma\sqrt{\frac{\hbar}{2\rho}} \times\sum_q\sum_{i=1}^{3}\frac{1}{\sqrt{\omega_i(q)}}(b_q^{+i}\eta_q^i+\eta_q^{+i}b_q^i). \tag{13}$$

As was noted above, the characteristic scale of elastic energy is higher than the energies of hydrogen bonds (ϵ) and their interaction with conformational degrees of freedom (γ), therefore we can use the perturbation theory.

In order to exclude phonons with an accuracy up to the terms of second order, we can use the canonical transformation

$$\tilde{\mathcal{H}}=e^{-\varphi}\mathcal{H}e^{\varphi},$$

where the anti-Hermitian operator φ is expressed as

$$\varphi=\sum_q\sum_{i=1}^{3}[A_q^ib_q^{+i}-b_q^iA_q^{+i}], \tag{14}$$

and operators A_q^i can be found in the first order of the perturbation theory from the condition

$$\mathcal{H}_{\mathrm{int}}+[\mathcal{H}_0;\varphi]=0. \tag{15}$$

Hence, $\tilde{\mathcal{H}}=\mathcal{H}_0+(1/2)[\mathcal{H}_{\mathrm{int}};\varphi]$. With the notation

$$A^i_q \equiv A^{ij}(q)\,\eta^j_q\,, \tag{16}$$

$$H^{lk}(q-q') = \sum_{j=1}^{3} h^{+j}_{q-q'} t^{ij}_k\,, \tag{17}$$

where t^{ij}_k are structural constants of pseudospin operators under transformation $\hat{S}$ [Eq. (8)]:

$$[\eta^i_{q'}\,;\eta^j_{q-q'}] = \sum_k t^{ij}_k\,\eta^k_q\,, \tag{18}$$

the calculation of operator φ [or matrix $A^{ik}(q)$] reduces to solving the integral equation

$$A^{ik}(q) = -\frac{\gamma}{\sqrt{2\rho}}\frac{\delta^{ik}}{(\omega_i(q))^{3/2}} + \frac{\epsilon}{2\hbar\omega_i(q)} \times \sum_{q'}\sum_{l=1}^{3} H^{lk}(q-q')A^{ie}(q'). \tag{19}$$

Note that the kernel of this integral equation is a function of both the conformation of hydrogen bonds (H^{lk}) and elastic properties of the double-strand DNA structure $[\omega_i(q)]$.

With an arbitrary relation between parameters γ and ϵ, Eq. (19) is difficult to solve. Physically, the simplest case is when the first term of Hamiltonain (13) is much smaller than the second. In the zeroth-order approximation, i.e., when ϵ is neglected, we have the following effective Hamiltonian of hydrogen bonds:

$$\mathscr{H}_{\text{eff}} = -\frac{\gamma^2}{4K}\sum_{mn} T^i_{mn}\,\eta^i_m\,\eta^i_n\,, \tag{20}$$

where

$$T^3_{mn} = -\frac{4}{N}\sum_q \frac{\cos[(m-n)q]}{\lambda_3(q)}, \tag{21}$$

$$T^1_{mn} = T^2_{mn} = \frac{1}{N}\sum_q \left(\frac{1}{\lambda_1(q)} + \frac{1}{\lambda_2(q)}\right)\cos[(m-n)q].$$

Note that the ‘‘exchange’’ constant in effective Hamiltonian (20) is independent of the mass density ρ (and, by the way, of Planck’s constant $\hbar$). In our opinion, these two properties are quite natural. Indeed, the effective energy of hydrogen bonds is a thermodynamic characteristic of the system, which should not depend on the purely dynamical parameter ρ. The physical cause of this energy is ‘‘magnetostriction,’’ i.e., interaction between hydrogen bonds and phonons, which has a classical nature, although the hydrogen bonds have been described in terms of the quantum theory (two-level systems).

For simplicity, let us consider only the ‘‘$X-Y$’’ configuration of pseudospin operators describing hydrogen bonds, i.e., we set $\eta^z \equiv 0$. In the continuous limit, the calculation of $T^{1,2}_{mn}$ is reduced to the integrals

$$\int_{-\pi}^{\pi}\frac{d\varphi}{2\pi}\,\frac{\cos[(m-n)\varphi]}{4\sin^2(\varphi/2)+\Omega^2 \pm 4\sin(\varphi/2)\cos(\varphi/2)\Omega},$$

which can be easily performed by calculating the required residues. As a result of this simple procedure, we find

$$T^1_{mn} = T^2_{mn} = -\frac{\cos[(m-n)\alpha]}{\Omega^2(1+\Omega^2)(m-n)/2}, \tag{22}$$

where

$$\alpha = \cos^{-1}\frac{1}{\sqrt{1+\Omega^2}}. \tag{23}$$

The effective exchange interaction is very similar to the so-called RKKY (Ruderman, Kittel, Kashuya, and Yoshida) interaction between spins mediated by conduction-band electrons, which is well known in the physics of metals. Like the RKKY interaction, that between hydrogen bonds oscillates with the distance and its amplitude drops following a power law (in this sense, the interaction is long-range). The long-range nature of this interaction allows one to use the mean-field approximation, which leads to a helical structure in the case of the RKKY interaction. Therefore, on the basis of the well-known results for the RKKY interaction, we can suggest a helical configuration of hydrogen bonds and apply a variational technique, which yields an upper limit for the ground-state energy. Thus, in order to minimize energy (20), we use the following trial function:

$$\eta^x_n = \cos\psi_n\,, \quad \eta^y_n = \sin\psi_n\,, \quad \eta^z_n \equiv 0. \tag{24}$$

Substituting function (24) in (20) and taking into account Eqs. (22) and (23), we obtain

$$\mathscr{H}_{\text{eff}} = -\frac{\gamma^2}{4K}\sum_{m\neq n}\cos^{|m-n|}\alpha\cos[(m-n)\alpha] \times\cos(\psi_m-\psi_n). \tag{25}$$

The condition for minimizing the energy of Hamiltonian (25) by varying ψ_m is equivalent to the equation for $\psi_m \equiv m\beta$

$$\sum_n \cos^{|m-n|}\alpha\cos[(m-n)\alpha]\sin[(m-n)\beta] = 0. \tag{26}$$

If the molecule is sufficiently long, one can neglect the boundary conditions and transform Eq. (26) to the following transcendental equation:

$$-\frac{\sin(\alpha+\beta)}{[1-2\cos\alpha\cos(\alpha+\beta)+\cos^2\alpha]^2} + \frac{\sin(\alpha-\beta)}{[1-2\cos\alpha\cos(\alpha-\beta)+\cos^2\alpha]^2} = 0. \tag{27}$$

A solution of Eq. (27) can be found numerically. Figure 2 plots the value of $\mathscr{H}_{\text{eff}}$ given by Eq. (25) as a function of angle β. It clearly shows two symmetric and fairly narrow minima, where $|\beta-\alpha| \ll 1$. Under this condition, we can easily find the solution of Eq. (27) analytically:

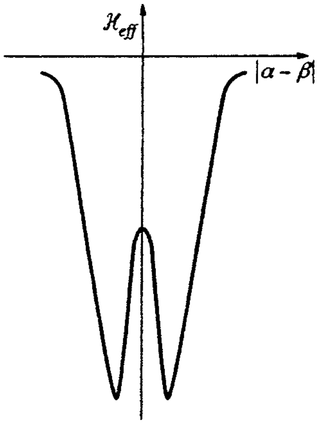

FIG. 2.

$$\beta \approx \alpha - \frac{\sin(2\alpha)(1-\cos\alpha)^4}{[1-2\cos\alpha\cos(2\alpha)+\cos^2\alpha]^2}. \tag{28}$$

Thus, the configuration given by Eqs. (24)–(28) really corresponds to at least a local minimum in the energy determined by Hamiltonian (20). We have also shown that, owing to the interaction among hydrogen bonds, they have a helical ordering given by Eq. (24) in the DNA ground state, and the pitch of this helix is different from (and, generally speaking, is incommensurate with) the double-helix pitch. Moreover, in the approximation of the lowest order, the energies of the left-hand and right-hand helices of hydrogen bonds are equal, i.e., the state is degenerate. Therefore, a domain structure of hydrogen bonds is feasible, for example, a right-handed DNA helix can contain alternating domains with left- and right-handed helical ordering of hydrogen bonds.

4. CONCLUSIONS

Let us summarize the main results of this paper.

1. The spectrum of conformational modes in the double-strand DNA structure with one acoustic and two optical modes (one of which is softened at a finite wave vector) has been calculated. The presence of the two optical modes is a direct consequence of the molecular structure with two helices.

2. The interaction between hydrogen bonds and conformational degrees of freedom results in an effective long-range interaction between hydrogen bonds. The latter orders the hydrogen bonds (described in terms of two-level systems) so that their axes of anisotropy form a helical configuration.

3. We have calculated the pitch of the hydrogen bond helix as a function of the model parameters which determine the energies of interaction between hydrogen bonds and conformational degrees of freedom and of elastic deformations in the double-strand DNA structure. In the limiting case, when the self-energy of a hydrogen bond determined by parameter ϵ can be neglected, the pitch of the hydrogen-bond helix depends only on Kirchhoff's ''angular velocity'' Ω, i.e., there is a universal dependence on the double-helix pitch [Eq. (28)].

Note that the results of the present work can be tested experimentally. For example, the dynamics of conformational degrees of freedom is related to so-called breathing modes introduced by Manning.[13] The softening of one optical mode demonstrated by our calculations should, undoubtedly, affect the stability of various DNA conformations. Experimental detection of such an effect would be, certainly, of great interest.

In our opinion, however, attention should be focused on the predicted helical ordering of hydrogen bonds. In the simplified model investigated in our work, the left- and right-handed helices are degenerate, but even the slightest modification of the model (such as inclusion of the self-energy of a hydrogen bond, or taking into account the fact that there are two hydrogen bonds in the A–T complementary pair and three in the G–C pair) should lift this degeneracy. Therefore, depending on the ratio between the numbers of G–C and A–T pairs, the sign of the pitch of the hydrogen-bond helix may differ. Note that such an inversion of the hydrogen-bond helix pitch does not require a restructuring of the entire double-strand DNA configuration.

This result of our study gives rise to two questions that may be significant in the biophysical context:

1. Since the hydrogen bond is only one part of the electronic interaction between nitrous bases of DNA, the natural question is about the plausibilit y of a helical configuration characterizing the interaction between these bases.

2. Must the sign of the pitch of the DNA structure coincide with that of the helix characterizing the electronic interaction between bases?

These questions demand a more complex model and go beyond the scope of the reported investigation. Nonetheless, in view of our results concerning hydrogen bonds, affirmative answers to both these questions seem quite plausible. It seems also that it is necessary to revise observations of so-called unusual polynucleotides,[14] in which a left-handed helix can transform to a right-handed one,[15] and which demonstrated a tendency of forming in DNA segments either the *A*-conformation (left-handed in optical measurements) when the double helix is rich in G–C pairs, or the optically right-handed *B*-conformation in a DNA molecule rich in A–T pairs.[16]

All these phenomena are largely associated with observations of circular dichroism inversion in optical spectra of DNA solutions. Effects of this sort have been interpreted in terms of geometrical (or, more exactly, stereochemical) models that described transformations of the entire DNA conformation from the right- to left-handed helix, given a certain quantity of defects, namely kinks, but without breaking the complementation rule.

Our results, however, indicate that there may be an alternative physical mechanism on which an interpretation of the circular dichroism inversion in optical spectra can be based. This mechanism does not require a restructuring of the entire double helix. It is known[17] that circular dichroism spectra are determined by the electronic structure of the DNA interior. Assuming that electronic states may have properties similar to those of hydrogen bonds described in this paper and form helical configurations, we can suggest a tentative interpretation of the circular dichroism inversion detected in experiments. The detailed study of these problems is beyond the scope of this paper and requires a considerable modification of our simplified model.

Interaction between these orbitals and conformational degrees of freedom may be described in terms of a model similar to that suggested in Sec. 2 for hydrogen bonds. In conformity with the results of Sec. 3, the ground state of electronic orbitals should have a helical ordering of their anisotropy axes. Moreover, neither the pitch nor the sign of this helix should necessarily coincide with those of the DNA double-strand configuration.

A direct experimental confirmation (or disproof) of the above statements, like the conclusions of this work based on our model, would be of great interest from the viewpoint of both a fundamental understanding of the structure and properties of DNA molecules, and their biophysical functions in different physico-chemical conditions. A simple test of our theory would be, for example, an observation of an alternating circular dichroism sign in optical bands depending on the hydrogenation degree[17] of DNA or synthetic polypeptides (since hydrogenation has direct effect on hydrogen bonds, thus it modifies the generalized field $\mathbf{H}_n$ in our model).

This work was partially supported by grants from the Russian Fund for Fundamental Research. One of the authors (E. K.) is grateful to I. E. Dzyaloshinskiĭ for fruitful discussions of helicoidal structures.

---

*) E-mail: kats@cpd.landau.ac.ru

1) Phonon spectra like those expressed by Eq. (10) were first found in Ref. 11.

---

[1] B. Alberts, D. Bray, J. Lewis, M. Raft, K. Noberts, and J. Watson, *Molecular Biology of the Cell*, Garland Publ., New York (1989).
[2] A. V. Vologodskii, S. D. Leven, K. V. Klenin *et al.*, Annu. Rev. Biophys. Biomol. Struct. **23**, 609 (1992).
[3] J. F. Marko and E. D. Siggia, Macromolecules **27**, 981 (1994).
[4] V. L. Golo and E. I. Kats, JETP Lett. **60**, 679 (1994).
[5] V. L. Golo and E. I. Kats, JETP Lett. **62**, 627 (1995).
[6] Y. Marechal, La Recherche **209**, 482 (1989).
[7] S. Lewin, *Displacement of Water and Its Control of Biochemical Reactions*, Academic Press, New York (1974).
[8] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, New York (1986).
[9] V. M. Agranovich, *Theory of Excitons* [in Russian], Nauka, Moscow (1968).
[10] A. S. Davydov, *Theory of Molecular Excitons* [in Russian], Nauka, Moscow (1968).
[11] H. Cappelmann and W. Beim, Z. der Phys. **209**, 276 (1968).
[12] E. M. Lifshitz and L. P. Pitaevskiĭ, *Statistical Physics*, Part 2, Pergamon Press, New York (1980).
[13] G. Manning, Biopolymers **22**, 689 (1983).
[14] W. K. Olson, Proc. Natl. Acad. Sci. USA **74**, 1775 (1977).
[15] F. M. Pohl and T. M. Jovin, J. Mol. Biol. **67**, 375 (1972).
[16] R. D. Wells, R. W. Blakesley, S. C. Hardies *et al.*, Crit. Rev. Biochem. **4**, 305 (1977).
[17] V. Bloomfield, D. M. Crothers, and I. Tinoco, *Physical Chemistry of Nucleic Acids*, Harper and Row, San Francisco (1974).

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Effect of a magnetic field on the yield point of NaCl crystals

V. I. Al'shits, N. N. Bekkauer, A. E. Smirnov*) and A. A. Urusovskaya*)

Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russia

(Submitted 18 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 951–958 (March 1999)

A constant magnetic field is found to have a substantial effect on the macroplasticity of NaCl crystals when they are being actively strained at a constant rate $\dot{\varepsilon}=\text{const}$ during magnetic treatment. We have measured the dependence of the yield point σ_y on the magnetic induction $B=0-0.48$ T and the strain rate $\dot{\varepsilon}=1.4\times10^{-5}-1.4\times10^{-3}$ s^{-1}. It is shown that this magnetic effect has a threshold character and is observed only for $B>B_c$, where B_c grows with increasing $\dot{\varepsilon}$ as $\sqrt{\dot{\varepsilon}}$. The lower the strain rate $\dot{\varepsilon}$, the larger the relative decrease in the yield point $\sigma_y(B)/\sigma_y(0)$ at fixed field $B>B_c$. At small enough strain rates $\dot{\varepsilon}$ the threshold field B_c ceases to depend on $\dot{\varepsilon}$ and goes constant. A theoretical model is proposed which is in good agreement with the observed regularities. The model is based on the competition between thermally activated and magnetically stimulated depinning of dislocations from paramagnetic impurity centers. © *1999 American Institute of Physics.* [S1063-7761(99)01203-2]

1. INTRODUCTION

The magnetoplastic effect was first discovered in NaCl crystals at the microplastic level in the form of dislocation shifts in samples located in a constant magnetic field in the absence of a mechanical load.[1] Subsequent studies by a number of independent groups[2–10] have made it possible to link this phenomenon with depinning of dislocations from paramagnetic impurities in a magnetic field. The macroscopic dislocation shifts of tens and hundreds of microns, arising as a result of such depinning under the action of long-range internal stresses, are in essence a manifestation of magnetically stimulated relaxation of the dislocation structure. The exact physical mechanism of the elementary act of depinning of a dislocation from a paramagnetic impurity center is still not completely clear. It appears that the discussion here is about spin-dependent transitions in a magnetic field destroying the energy barriers.[2–10] Processes of this type play a significant role in many phenomena.[11] In particular, it is customary to link the effect of a magnetic field on the rate of chemical reactions with them.[12] The hypothesis of spin-dependent processes in dislocations in a magnetic field and the consequences flowing from it are in good agreement with numerous experimental data and are discussed in detail in the publications cited above.[2–10] We will not dwell on this point here. We note only that the magnetic polarization of dislocation nuclei in NaCl and KCl crystals was experimentally discovered[13] long before the magnetoplastic effect.

One of the important accomplishments of recent years making it possible to put the magnetoplastic effect to practical use was the discovery of its macroplastic manifestations. It turns out that in alkali–halide crystals a magnetic field substantially alters the work-hardening coefficient,[14] the yield point,[15,16] and even the microhardness.[17] Thus, for example, according to Ref. 15, in LiF crystals strained at a low enough rate ($\dot{\varepsilon}\leqslant5\times10^{-5}$ s^{-1}) in a constant magnetic field exceeding some threshold ($B\geqslant0.3$ T), the yield point is halved, the extent of all stages on the stress-strain curve is appreciably foreshortened, and the work-hardening coefficient in stage II is decreased while in stage III, in contrast, it grows.

In the present paper we present a similar study in NaCl crystals. We show that in these crystals the magnetic sensitivity of the $\sigma(\varepsilon)$ cures is manifested over a wider range of strain rates and for a lower threshold B_c. This makes it possible to substantially increase the information content of the data on the influence of a magnetic field on the macroplasticity of the crystals and suggests a theoretical interpretation. It may be remarked that the indicated interpretation has a ''kinematic'' character and is based on the competition of thermally activated and magnetically stimulated processes of depinning of dislocations from point defects. What is important for us here is not the micromechanism of magnetic depinning but the very existence of such depinning, convincingly established in numerous experiments on the level of individual dislocations together with the empirical dependence of the time needed for such depinning on the magnetic field.

2. EXPERIMENTAL TECHNIQUE

The objects of study were nominally pure NaCl crystals grown at LOMO without any intentional impurities. The total impurity content of the crystals did not exceed 10^{-3} wt. %. Two types of samples were investigated: 1) unannealed samples, cleaved along the cleavage plane from room-temperature aged crystals, and 2) annealed samples, cleaved from γ-irradiated crystals. Typical dimensions of the samples were roughly $2.5\times2.5\times8$ mm^3.

Mechanical compression tests were performed on the same setup as in Ref. 15, with corundum (Al_2O_3) anvils. A magnetic field was created by placing a permanent magnet

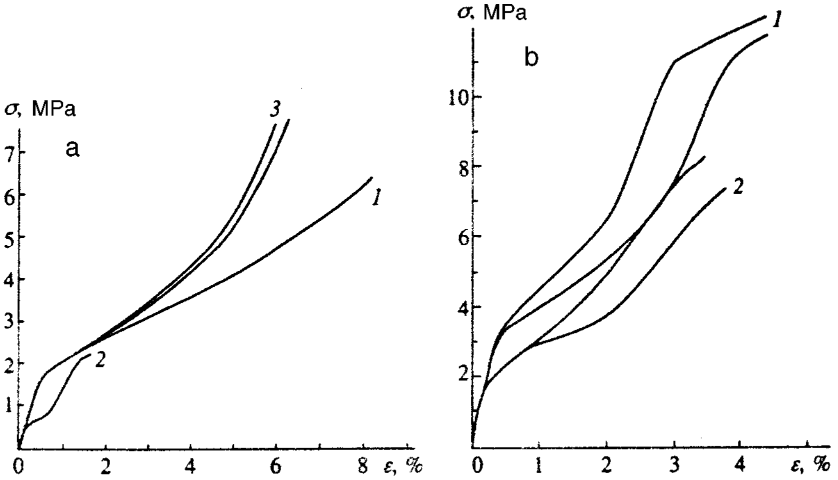

FIG. 1. Stress–strain curves of annealed (a) and unannealed (b) crystals of NaCl, obtained by compression in magnetic fields $B=0$ (*1*), 0.48 T (*2*), and 0.2 T (*3*); $\dot{\varepsilon}=5\times10^{-4}\ \mathrm{s}^{-1}$.

with movable poles over the sample. The maximum magnetic induction was 0.48 T. The shape of the stress–strain curves $\sigma(\varepsilon)$ was studied, as well as the dependence of the yield point σ_y on the magnetic induction B in the field interval 0–0.48 T for strain rates $\dot{\varepsilon}$ from $1.4\times10^{-5}\ \mathrm{s}^{-1}$ to $1.4\times10^{-3}\ \mathrm{s}^{-1}$. The stress–strain curves of NaCl crystals are characterized by a vague. For this reason, the yield point σ_y was estimated to be the stress at which the curve $\sigma(\varepsilon)$ departs from its initial rectilinear dependence.

Of course, each stress–strain curve was recorded on a separate sample and reflects not only the general properties, but also individual properties of the tested crystal. For this reason, a certain spread in the values of the yield point σ_y measured on a series of samples under identical conditions is inevitable. Usually, this spread is significantly less when working with annealed samples characterized by a more homogeneous and stabilized initial dislocation structure (see curves *3* in Fig. 1(a)). Unfortunately, because of the paucity of annealed samples at our disposal, we were not able to duplicate all the unannealed measurements in them under identical conditions. To make up for this, on the more ''dangerous'' (from the foregoing standpoint) unannealed crystals, each measurement of the stress–strain curve $\sigma(\varepsilon)$ was repeated on several samples for fixed values of the parameters $\dot{\varepsilon}$ and B. The corresponding spread in the values of the yield point σ_y, reflected in Fig. 2(b) in the form of experimental error bars, is appreciably less than the observed magnetic effects even for the unannealed crystals.

3. EXPERIMENTAL RESULTS

Figure 1 shows stress–strain curves for compression at the strain rate $\dot{\varepsilon}=5\times10^{-4}\ \mathrm{s}^{-1}$ of annealed (a) and unannealed (b) samples in the absence of a magnetic field (curves *1*) and in a magnetic field $B=0.48$ T (curves *2*). Figure 1(a) also shows curves (curves *3*) measured on two different samples for $B=0.2$ T, demonstrating the good reproducibility of the $\sigma(\varepsilon)$ curves for the annealed samples. As can be seen from Fig. 1(b), the latter is significantly poorer for the unannealed samples. As expected, the annealed samples are softer, i.e., their initial (at $B=0$) yield point is almost half that of unannealed samples. The former are accordingly

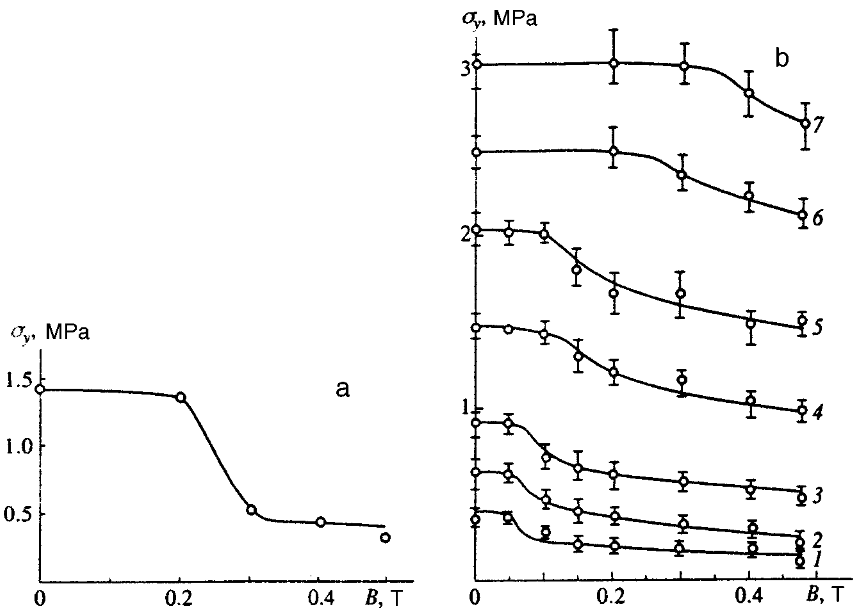

FIG. 2. Dependence of the yield point on magnetic induction for annealed (a, $\dot{\varepsilon}=5\times10^{-4}\ \mathrm{s}^{-1}$) and unannealed (b) crystals of NaCl compressed at strain rates $\dot{\varepsilon}=1.4\times10^{-5}$ (*1*), 2×10^{-5} (*2*), 3×10^{-5} (*3*), 7.3×10^{-5} (*4*), 1.75×10^{-4} (*5*), 5×10^{-4} (*6*), $1.4\times10^{-3}\ \mathrm{s}^{-1}$ (*7*).

more sensitive to a magnetic field than the latter. As the induction B is increased from 0 to 0.48 T, the yield point of the annealed samples falls by roughly a factor of two, whereas for the unannealed samples at this strain rate it falls by only 20–25%. In addition, all stages of plastic flow of the annealed samples in a magnetic field are appreciably foreshortened in the stress as well as the strain, i.e., a magnetic field hastens the changeover of strain stages in these crystals. On the other hand, as can be seen from Fig. 1(b), in the deformation of unannealed samples in a magnetic field the foreshortening of the stages in the stress–strain curves is clear-cut only on the stress scale (in particular, in terms of the decrease in the yield point).

Figure 2 shows the behavior of the yield point as a function of magnetic field for annealed (a) and unannealed (b) crystals. In annealed crystals this dependence has a clear-cut three-stage character. In the first stage, for B less than some threshold value B_c the yield point σ_y remains constant and exhibits no sensitivity to the magnetic field. In the second stage, within a very narrow range above threshold ($B \geqslant B_c$), an abrupt decrease in the yield point is observed. And finally, in the third stage, for $B > B_c'$, within the limits of error of the experiment, the yield point σ_y remains essentially constant, although it does tend to decrease slowly.

In unannealed crystals, the decrease in the yield point with increasing B for $B > B_c$ is smoother. The three-stage character of the dependence of the yield point on the field, $\sigma_y(B)$, is clear-cut only at low strain (compression) rates. Studies in these crystals at different strain rates in the range from $1.4\times10^{-5}\ \mathrm{s}^{-1}$ to $1.4\times10^{-3}\ \mathrm{s}^{-1}$ revealed a substantial increase in the magnetic effect on the yield point at lower strain rates. It follows from Fig. 2(b) that in this case a significant decrease in both the magnetic threshold B_c (roughly sixfold) and the ratio $\sigma_y(B_m)/\sigma_y(0)$ is observed ($B_m = 0.48$ T).

4. DISCUSSION

As we have already noted, independent data based on studies of the microplasticity of the same crystals indicate that the role of a magnetic field in their plastification reduces to depinning of dislocations from point defects. Let us consider the active deformation of a crystal with strain rate $\dot{\varepsilon} = \mathrm{const}$ in a magnetic field B in terms of a simple model, assuming that the yield point σ_y for $B=0$ is limited by the same magnetically sensitive impurity centers coexisting with less plentiful obstacles not affected by a magnetic field. For a prescribed magnetic field B the depinning time τ_B of the dislocations from obstacles of the first kind as a result of spin-dependent transitions is proportional to B^{-2} (Ref. 18), i.e.,

$$\tau_B = \alpha B^{-2}. \tag{1}$$

According to Ref. 18, such transitions, and correspondingly such depinning, are possible only in magnetic fields exceeding some threshold $B_c^{(0)}$. On the other hand, even at $B > B_c^{(0)}$, magnetic depinning of dislocations is in no way manifested in the macroplasticity if τ_B (1) substantially exceeds the thermally activated dislocation detachment time

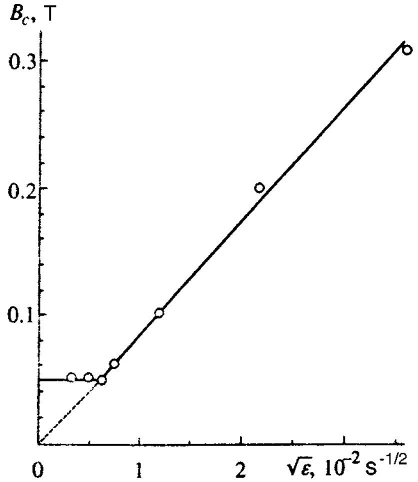

FIG. 3. Comparison of the theoretical dependence $B_c(\sqrt{\dot{\varepsilon}})$ (solid line) and the experimental points according to the data of Fig. 2(b).

$$\tau_{th} = \tau_0 \exp\left[\frac{U(\sigma)}{kT}\right] \tag{2}$$

from the same obstacle under the stress $\sigma = \sigma_y$ determined by the strain rate

$$\dot{\varepsilon} = \dot{\varepsilon}_0 \exp\left[-\frac{U(\sigma)}{kT}\right]. \tag{3}$$

In formulas (2) and (3) T is the temperature, k is Boltzmann's constant, and $U(\sigma)$ is the activation energy for overcoming the magnetically sensitive impurity centers under the stress σ. Thus, the kinetics of deformation dictates the threshold of magnetic sensitivity of the macroplasticity, since the condition $\tau_B < \tau_{th}$, taking relations (1)–(3) into account, is equivalent to

$$B^2 > \frac{\alpha}{\tau_0 \dot{\varepsilon}_0} \dot{\varepsilon}. \tag{4}$$

In other words, a pronounced magnetic influence on the stress–strain curves $\sigma(\varepsilon)$ in the model should be observed only at $B > B_c$, where

$$B_c = \max\{B_c^{(0)}, k\sqrt{\dot{\varepsilon}}\}, \quad k = \sqrt{\alpha/\tau_0 \dot{\varepsilon}_0}. \tag{5}$$

Another more convenient way of writing out the dependence of the threshold field B_c on the strain rate $\dot{\varepsilon}$, equivalent to formula (5), is

$$B_c = \begin{cases} B_c^{(0)}, & \dot{\varepsilon} < \dot{\varepsilon}_m, \\ k\dot{\varepsilon}^{1/2}, & \dot{\varepsilon} > \dot{\varepsilon}_m, \end{cases} \tag{6}$$

where $\dot{\varepsilon}_m = (B_c^{(0)}/k)^2$. Figure 3 compares the function $B_c(\sqrt{\dot{\varepsilon}})$ that follows from the given model with the experimental points obtained by processing the measured $\sigma_y(B)$ at a series of strain rates $\dot{\varepsilon}$ of unannealed samples [Fig. 2(b)]. The good agreement of the predicted type of dependence (5), (6) with experiment favors the physical scheme proposed above.

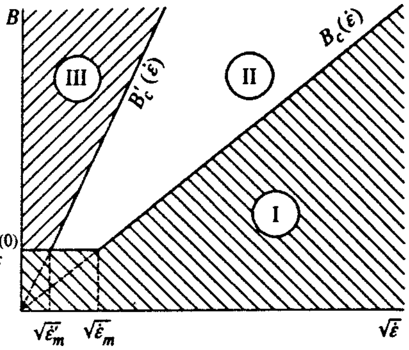

FIG. 4. The dependences $B_c(\dot{\varepsilon})$ and $B'_c(\dot{\varepsilon})$, and the zones I, II, and III of the physical parameters $\{B,\dot{\varepsilon}\}$ corresponding to the three stages of variation of the yield point with increasing magnetic induction.

According to the latter, when the magnetic field exceeds the threshold value B_c, pinning of the dislocations at obstacles of the first type, in accordance with formula (1), falls rapidly with increasing B, and in a high enough field the crystal begins to behave as if it were purged of this type of impurity and the resistance to motion on the part of the dislocations were the result only of the less plentiful magnetically insensitive defects. This happens when the magnetic induction is taken through the second threshold $B=B'_c$, which corresponds to $\tau_B=\tau'_{th}$, where

$$\tau'_{th}=\tau'_0 \exp\left[\frac{U'(\sigma)}{kT}\right] \tag{7}$$

is the thermally activated dislocation detachment time from obstacles of the second type. The quantity B'_c is given by expressions analogous to (5) and (6) with the replacements

$$k\rightarrow k'=\sqrt{\alpha/\tau'_0\dot{\varepsilon}'_0}, \quad \dot{\varepsilon}_m\rightarrow\dot{\varepsilon}'_m=(B_c^{(0)}/k')^2.$$

Figure 4 shows the corresponding dependences $B_c(\dot{\varepsilon})$ and $B'_c(\dot{\varepsilon})$, as well as the three stages of variation in the yield point with increasing magnetic induction B (see Fig. 2). In the first stage, in region I ($B<B_c$), thermally activated detachment from all defect types dominates. In the third stage, in region III ($B>B'_c$) the yield point is limited by thermally activated detachment only from obstacles of the second type, which are insensitive to a magnetic field. Here the dislocations ''do not notice'' the first type of impurity, as it were. An actual dependence of the yield point σ_y on the magnetic field in this model should be observed only in the intermediate region II, in the second stage. As follows from Fig. 4, the width of that magnetic field range should remain minimal and independent of the strain rate as long as $\dot{\varepsilon}<\dot{\varepsilon}'_m=(B_c/k')^2$. For strain rates $\dot{\varepsilon}$ in the range $\dot{\varepsilon}'_m<\dot{\varepsilon}<\dot{\varepsilon}_m=(B_c/k)^2$ the width of region II increases with increasing strain rate $\dot{\varepsilon}$. In that case, the lower left-hand edge of the range remains unchanged, while the position of the right-hand edge varies as $\sqrt{\dot{\varepsilon}}$. Finally, at the largest strain rates $\dot{\varepsilon}>\dot{\varepsilon}_m$, the positions of both edges of the intermediate region vary as $\sqrt{\dot{\varepsilon}}$, and its width can be estimated as $\Delta B\simeq(k'-k)\sqrt{\dot{\varepsilon}}$. The expansion with increasing strain rate $\dot{\varepsilon}$ of the region ΔB in which the main falloff of the yield point σ_y occurs correlates qualitatively with the proposed model.

We note in passing that the width of the step in the dependence $\sigma_y(B)$ can of course be limited by other factors not taken into account by the present simple scheme. One of these factors is the presence of several kinds of magnetically sensitive barriers. Note that even in that case, when the impurity composition of these obstacles is homogeneous, a spread in the dimensions of the complexes and/or dislocation segment lengths between complexes is sufficient to produce a dispersion in the threshold fields and activation parameters. It is clear from general considerations that the indicated spread is smaller in annealed crystals. This possibly also explains the difference in the corresponding widths of region II in the experimental curves of $\sigma_y(B)$ shown in Figs. 2(a) and 2(b).

ACKNOWLEDGMENTS

One of us (V. I. A.) is grateful to the Russian Fund for Fundamental Research for partial financial support (Grant No. 97-02-16327).

---

*⁾E-mail: public@mechan.incr.msk.su

[1] V. I. Al'shits, E. V. Darinskaya, T. M. Perekalina, and A. A. Urusovskaya, Fiz. Tverd. Tela (Leningrad) **29**, 467 (1987) [Sov. Phys. Solid State **29**, 265 (1987)].

[2] V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Izv. VUZov, Chernaya Metallurgiya **10**, 85 (1990).

[3] M. I. Molotskiĭ, Fiz. Tverd. Tela (Leningrad) **33**, 3112 (1991) [Sov. Phys. Solid State **33**, 1760 (1991)].

[4] V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Fiz. Tverd. Tela (Leningrad) **34**, 155 (1992) [Sov. Phys. Solid State **34**, 81 (1992); Fiz. Tverd. Tela (St. Petersburg) **35**, 320 (1993) [Phys. Solid State **35**, 162 (1993)].

[5] Yu. I. Golovin and R. B. Morgunov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1384 (1993) [Phys. Solid State **35**, 700 (1993)].

[6] V. I. Alshits, E. V. Darinskaya, and E. A. Petrzhik, Mater. Sci. Eng., A **164**, 322 (1993).

[7] V. I. Alshits, E. V. Darinskaya, O. L. Kazakova, E. Yu. Mikhina, and E. A. Petrzhik, J. Alloys Compd. **211/212**, 548 (1994).

[8] M. I. Molotskii, R. E. Kris, and V. Fleurov, Phys. Rev. B **51**, 12531 (1995).

[9] V. I. Al'shits, E. V. Darinskaya, and O. L. Kazakova, Zh. Éksp. Teor. Fiz. **111**, 615 (1997) [JETP **84**, 338 (1997)].

[10] V. I. Alshits, E. V. Darinskaya, O. L. Kazakova, E. Yu. Mikhina, and E. A. Petrzhik, Mater. Sci. Eng., A **234**, 617 (1997).

[11] Ya. B. Zel'dovich, A. L. Buchachenko, and E. L. Frankevich, Usp. Fiz. Nauk **155**, 3 (1988) [Sov. Phys. Usp. **31**, 385 (1988)].

[12] A. L. Buchachenko, R. Z. Sagdeev, and K. Z. Salikhov, *Magnetic and Spin Effects in Chemical Reactions* [in Russian], Nauka, Novosibirsk (1978).

[13] E. J. Sharp and D. A. Every, Phys. Rev. **158**, 511 (1967).

[14] Yu. I. Golovin and R. B. Morgunov, JETP Lett. **61**, 596 (1995).

[15] A. A. Urusovskaya, V. I. Al'shits, A. E. Smirnov, and N. N. Bekkauer, JETP Lett. **65**, 497 (1997).

[16] A. A. Urusovskaya, A. E. Smirnov, and N. N. Bekkauer, Izv. Ross. Akad. Nauk, Ser. Fiz. **61**, 937 (1997).

[17] Yu. I. Golovin, R. B. Morgunov, D. V. Lopatin, and A. A. Baskakov, Phys. Status Solidi A **160**, R3 (1997).

[18] V. I. Al'shits, E. V. Darinskaya, O. L. Kazakova, E. Yu. Mikhina, and E. A. Petrzhik, JETP Lett. **63**, 668 (1996).

Translated by Paul F. Schippnick

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Novel effects of electron–phonon interaction in quasi-two-dimensional structures located in a magnetic field

O. V. Kibis*)

Novosibirsk State Technical University, 630092 Novosibirsk, Russia

(Submitted 1 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 959–969 (March 1999)

This paper studies the interaction of electrons and acoustic phonons in a quasi-two-dimensional system with an asymmetric quantizing potential in a magnetic field that is parallel to the structure's plane. It is demonstrated that the electron–phonon interaction in such a system generates an emf when there is a standing acoustic wave, as well as when the structure is heated uniformly. These phenomena are macroscopic manifestations of a universal quantum effect, which amounts to an emf being generated by any isotropic perturbation of any electron system in which the energy depends asymmetrically on the velocity v, i.e., $\varepsilon(v)\neq\varepsilon(-v)$. *© 1999 American Institute of Physics.* [S1063-7761(99)01303-7]

1. INTRODUCTION

We consider a quasi-two-dimensional (2D) electron system in $\{x,y,z\}$ coordinates, with the x axis perpendicular to the plane of the 2D layer. We direct the y axis along the magnetic field, so that $\mathbf{H}=(0,H_y,0)$, and choose the vector potential to be in the form $\mathbf{A}=(H_yz,0,0)$. Then the electron Hamiltonian is

$$\hat{\mathscr{H}}=\frac{1}{2m}\left[\left(\hat{p}_x+\frac{eH_yz}{c}\right)^2+\hat{p}_y^2+\hat{p}_z^2\right]+U(z), \tag{1}$$

where m is the effective electron mass, e the absolute value of the electron charge, and $U(z)$ the quantizing potential of the 2D system, and the electron wave function is

$$\psi_k=C\varphi(k_x,z)\exp(ik_xx+ik_yy)\exp(-i\varepsilon_kt/\hbar), \tag{2}$$

where $C=(L_xL_y\int_{-\infty}^{\infty}|\varphi(k_x,z)|^2dz)^{-1/2}$ is the normalization constant. Here L_x and L_y are the dimensions of the 2D system along the x and y axes, $\mathbf{k}$ is the electron wave vector, and ε_k is the electron energy. Substituting (2) into the Schrödinger equation with Hamiltonian (1) yields

$$-\frac{\hbar^2}{2m}\frac{\partial^2\varphi(k_x,z)}{\partial z^2}+\left[\frac{\hbar^2k_x^2}{2m}+\frac{\hbar eH_yk_xz}{mc}+\frac{(eH_yz)^2}{2mc^2}+U(z)-\varepsilon(k_x)\right]\varphi(k_x,z)=0, \tag{3}$$

where the energy $\varepsilon(k_x)=\varepsilon_k-\hbar^2k_y^2/2m$. To analyze the effects of interest, we use the model of a triangular quantizing potential,

$$U(z)=\begin{cases}\infty, & z<0,\\ eE_zz, & z\geqslant 0,\end{cases} \tag{4}$$

employed in calculations of the energy spectrum of electrons in inversion layers on semiconductor surfaces, where E_z is the absolute value of the electric field at the surface. We examine the electron system in the quantum limit, where the electrons fill the states only in the lower electron subband, provided that $(d_0/l_H)^4\ll 1$, where $d_0=(3\hbar^2\pi^2/16meE_z)^{1/3}$ is the thickness of the 2D layer (the average distance from the boundary $z=0$ to the electron) at $H_y=0$, and $l_H=(\hbar c/eH_y)^{1/2}$ the magnetic length. In this case the solution of Eq. (3) with potential (4) for the lower electron subband has the form[1]

$$\varepsilon(k_x)=\left[\frac{\hbar^2}{2m}\right]^{1/3}\left[\frac{9\pi}{8}\left(eE_z+\frac{\hbar eH_yk_x}{mc}\right)\right]^{2/3}+\frac{\hbar^2k_x^2}{2m}, \tag{5}$$

$$\varphi(k_x,z)=\begin{cases}\mathrm{Ai}\left(\left[\dfrac{2meE_z}{\hbar^2}+\dfrac{2eH_yk_x}{\hbar c}\right]^{1/3}\left[z-\dfrac{\varepsilon(k_x)-\hbar^2k_x^2/2m}{eE_z+\hbar eH_yk_x/mc}\right]\right), & z\geqslant 0,\\ 0, & z<0,\end{cases} \tag{6}$$

where $\mathrm{Ai}(\zeta)$ is the Airy function. Equation (5) implies that at $H_y\neq 0$ an asymmetric electron energy spectrum emerges, i.e.,

$$\varepsilon(v_x)\neq\varepsilon(-v_x), \tag{7}$$

where $v_x=(1/\hbar)[\partial\varepsilon(k_x)/\partial k_x]$ is the electron velocity along the x axis.

What is the physical reason for such asymmetry [Eq. (7)]? A magnetic field parallel to the plane of the 2D system cannot make the electron travel along a cyclotron orbit; at most it can slightly alter the wave function. As the electron travels at a velocity v_x, a Lorentz force acts on it in the direction $\langle -x\rangle$, with the result that the maximum of the electron wave function is shifted in the direction $\langle -z\rangle$. If the electron velocity changes to $-v_x$, the Lorentz force reverses direction and the maximum of the electron wave function shifts in the direction $\langle z\rangle$. Hence in an asymmetric potential $U(z)\neq U(-z)$ the electron energy $\varepsilon(v_x)\neq\varepsilon(-v_x)$. Due to

1063-7761/99/88(3)/6/$15.00

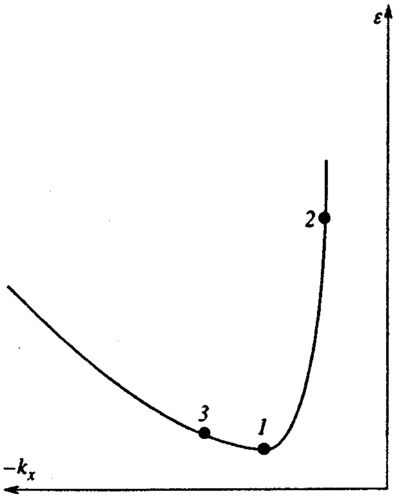

FIG. 1. Change in electron state initiated by a force F_x (*1*→*2* transition) and by a force $-F_x$ (*1*→*3* transition).

the asymmetry (7), the dynamic properties of the electron system differ in the directions $\langle x\rangle$ and $\langle -x\rangle$.

We show that this is the case by discussing a specific example. Suppose that initially the electron had a velocity $v_{x1}=0$ and was in state *1*, which corresponds to the minimum of the energy subband (5) depicted in Fig. 1. Under a force F_x the electron is transferred, in time *t*, into state *2*, while under $-F_x$ it is transferred (in the same time *t*) into state *3*, so that $k_{x2}-k_{x1}=k_{x1}-k_{x3}=F_x t/\hbar$. Since the electron energy spectrum $\varepsilon(k_x)$ is asymmetric with respect to the directions $\langle k_x\rangle$ and $\langle -k_x\rangle$, the velocity acquired by the electron in the first transition is not equal to that acquired in the second, $v_{x2}\neq v_{x3}$. Thus, the linear momentum transferred from the external force to the electron, equal to the product of the electron mass and the velocity v_x, proves to be different for forces in the directions $\langle x\rangle$ and $\langle -x\rangle$. This implies that when an external agent that is isotropic with respect to the directions $\langle x\rangle$ and $\langle -x\rangle$ acts on the electron system, the linear momenta transferred to the electron system differ in these directions, and this leads to an electron drift along the x axis.[1] In other words, any isotropic perturbation of any electron system with an asymmetric electron spectrum, $\varepsilon(v_x)\neq\varepsilon(-v_x)$, generates an emf along the x axis. The effect of anisotropic momentum transfer to an electron system subjected to an external isotropic force[1] is of a universal nature and exists irrespective of the way in which the electron system is perturbed. The goal of the present work is to study this effect for the case in which the interaction of electrons and acoustic phonons constitutes the perturbation of the electron system.

2. SPATIAL ASYMMETRY OF THE ELECTRON–PHONON INTERACTION

We treat a crystal as a three-dimensional elastic continuum and describe the interaction of electrons and acoustic phonons by the deformation potential method, where the electron energy shift $\Delta\varepsilon$ under an arbitrary deformation u_{ij} of the crystal is given by the expression $\Delta\varepsilon=\Xi\Sigma_i u_{ii}$, in which Ξ is the deformation potential constant. Then the probability of electrons absorbing a phonon with the wave vector $\mathbf{q}=(q_x,q_y,q_z)$ is

$$W_a(\mathbf{q})=n(\mathbf{q})\sum_k\sum_{k'}\frac{1}{\hbar^2}|\langle\psi_{k'}|\tilde{U}(\mathbf{q})|\psi_k\rangle|^2 \times f(\varepsilon_k)[1-f(\varepsilon_{k'})], \tag{8}$$

where $f(\varepsilon_k)$ is the Fermi–Dirac distribution function, $n(\mathbf{q})$ is the phonon occupation number, and

$$\tilde{U}(\mathbf{q})=\left(\frac{\hbar q}{2V\rho v_l}\right)^{1/2}\Xi\,\exp(iq_x x+iq_y y+iq_z z)\,\exp(-iv_l qt), \tag{9}$$

where V is the crystal volume, ρ the crystal density, and v_l the longitudinal speed of sound. If in (8) we go from summation over electronic states to integration over the wave vector, we obtain an expression for the probability per unit time of absorbing a photon per unit surface area of the 2D system:

$$\begin{aligned}w_a(\mathbf{q})={}&\frac{\Xi^2 qn(\mathbf{q})}{V\pi\rho v_l}\int d^2\mathbf{k}\int d^2\mathbf{k}'\left(\int_{-\infty}^{\infty}|\varphi(k_x,z)|^2dz\right.\\&\left.\times\int_{-\infty}^{\infty}|\varphi(k'_x,z)|^2dz\right)^{-1}[1-f(\varepsilon_{k'})]f(\varepsilon_k)\left|\right.\\&\left.\times\int_{-\infty}^{\infty}\varphi(k_x,z)\exp(iq_z z)\varphi(k'_x,z)\,dz\right|^2\delta(\varepsilon_{k'}-\varepsilon_k\\&-\hbar v_l q)\,\delta(k'_x-k_x-q_x)\,\delta(k'_y-k_y-q_y),\end{aligned} \tag{10}$$

where integration with respect to $\mathbf{k}$ and $\mathbf{k}'$ is carried out within the first Brillouin zone. The subsequent analysis is carried out for the energy spectrum and wave function of the electron, which are given by (5) and (6). Expanding (10) in a power series in the magnetic field and keeping only the linear term in the expansion, we find that for $(q_z d_0)^2\ll 1$ formula (10) becomes

$$w_a(\mathbf{q})=w_{a0}(\mathbf{q})+w_{aH}(\mathbf{q}), \tag{11}$$

where

$$\begin{aligned}w_{a0}(\mathbf{q})={}&\frac{\Xi^2 qn(\mathbf{q})}{V\pi\rho v_l}\int d^2\mathbf{k}\int d^2\mathbf{k}'\,\delta(k'_x-k_x-q_x)\\&\times\delta(k'_y-k_y-q_y)\,\delta(\varepsilon_{k'0}-\varepsilon_{k0}-\hbar v_l q)\\&\times[1-f(\varepsilon_{k'0})]f(\varepsilon_{k0})\end{aligned} \tag{12}$$

is the probability (10) in the absence of a magnetic field,

$$\begin{aligned}w_{aH}(\mathbf{q})={}&\left(\frac{\Xi^2 qn(\mathbf{q})}{V\pi\rho v_l}\right)\left(\frac{\hbar H_y}{15mcE_z}\right)(q_z d_0)^2\int d^2\mathbf{k}\int d^2\mathbf{k}'\\&\times\delta(k'_x-k_x-q_x)\,\delta(k'_y-k_y-q_y)(k_x+k'_x)\\&\times\delta(\varepsilon_{k'0}-\varepsilon_{k0}-\hbar v_l q)[1-f(\varepsilon_{k'0})]f(\varepsilon_{k0})\end{aligned} \tag{13}$$

is the term linear in the magnetic field, and ε_{k0} is the electron energy ε_k in the absence of a magnetic field. Equations (11)–(13) imply that for $(d_0/l_H)^4\ll 1$ and $(q_z d_0)^2\ll 1$ the difference in probabilities is

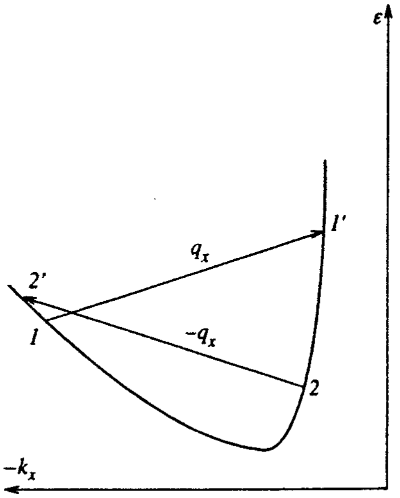

FIG. 2. Structure of electron transitions with absorption of photons whose wave vectors are $\mathbf{q}$ and $-\mathbf{q}$.

$$w_a(\mathbf{q})-w_a(-\mathbf{q})=2w_{aH}(\mathbf{q}). \tag{14}$$

The situation in which $q_z=0$ warrants a separate discussion. Here all linear terms in the expansion of the probability (10) in powers of the magnetic field vanish, so that we are forced to include higher-order terms in our discussion. Allowing for terms in this expansion that are cubic in the magnetic field, we find that for $\mathbf{q}=(2k_F,0,0)$ and $(d_0/l_H)^4\ll 1$ at absolute zero $(T=0)$ and low electron concentrations (the Fermi velocity $v_F\approx v_l$ and $(k_Fd_0)^2\ll 1$) the probability difference is

$$w_a(-\mathbf{q})-w_a(\mathbf{q})=\left(\frac{2}{\pi}\right)^2\frac{6v_l}{c}\left(\frac{2d_0}{3l_H}\right)^4\left(\frac{H_y}{E_z}\right)w_{a0}(\mathbf{q}), \tag{15}$$

where k_F is the length of the Fermi wave vector of the electron. But at high electron concentrations $(v_F\gg v_l)$ we have

$$w_a(-\mathbf{q})-w_a(\mathbf{q})=\left(\frac{2}{\pi}\right)^2\left(\frac{v_F}{c}\right)\left(\frac{2d_0}{3l_H}\right)^4\left(\frac{H_y}{E_z}\right)w_{a0}(\mathbf{q}). \tag{16}$$

The relations for the phonon emission probability $w_e(\mathbf{q})$ can be obtained from Eqs. (10)–(16) by formally replacing the subscript a with e, the wave vector $\mathbf{k}'$ with $\mathbf{k}$, the wave vector $\mathbf{k}$ with $\mathbf{k}'$, and the phonon occupation number $n(\mathbf{q})$ with $n(\mathbf{q})+1$.

The fact that (14)–(16) fail to vanish immediately suggests that the electron–phonon interaction is asymmetric: identical phonons with oppositely directed wave vectors interact differently with electrons. But what is the physical reason for such asymmetry? When an electron absorbs a photon, it goes from one state of the subband (5) into another. Here the initial electron state k_x and the final electron state k_x' both conserve energy and the length of the wave vector, as indicated in Fig. 2. Since the electron subbands $\varepsilon(k_x)$ are asymmetric in the directions $\langle k_x\rangle$ and $\langle -k_x\rangle$, the wave functions $\varphi(k_x,z)$ of the initial and final electronic states change when the sign of the component q_x of the wave vector changes. From Eqs. (5) and (6) it follows that $\varphi(k_{x1},z)\neq\varphi(k_{x2},z)$ and $\varphi(k_{x1}',z)\neq\varphi(k_{x2}',z)$. Hence the absolute value of the matrix element of the electron–phonon interaction potential (9),

$$|\langle\psi_{k'}|\tilde{U}(\mathbf{q})|\psi_k\rangle|\propto\left|\int_{-\infty}^{\infty}\varphi(k_x',z)\exp(iq_zz)\,\varphi(k_x,z)\,dz\right|,$$

turns out to differ for the absorption of phonons with wave vector components q_x and $-q_x$. Accordingly, for $q_x\neq 0$ the probabilities of the interaction of electrons and phonons with $\mathbf{q}$ and $-\mathbf{q}$ are different. This qualitative analysis is fairly general and suggests that there is a difference between the interactions of an electron system with an asymmetric energy spectrum and any elementary quasiparticles (photons, acoustic phonons, etc.) moving in the directions $\langle x\rangle$ and $\langle -x\rangle$ (see Refs. 1 and 2).

We see that the spatial asymmetry of the electron–phonon interaction is a purely quantum phenomenon. Below we examine specific physical situations in which this quantum phenomenon leads to macroscopic effects.

3. EMF GENERATION IN A STANDING ACOUSTIC WAVE

Since both absorption and emission of phonons are accompanied by a change in the momentum of the electron system, different probabilities of the interaction of electrons and phonons with $\mathbf{q}$ and $-\mathbf{q}$ lead to different momentum transfer to the electrons from acoustic waves propagating in opposite directions. This leads directly to an anomaly in the acoustoelectric effect, with the emf of phonon drag of electrons differing for waves with oppositely directed wave vectors. In particular, an emf of phonon drag of electrons is generated in the presence of a standing acoustic wave, which is a superposition of acoustic waves with equal amplitudes and oppositely directed wave vectors.[2] This phenomenon is a special case of the effect of anisotropic momentum transfer to electrons under an external isotropic perturbation discussed earlier in Ref. 1, where the standing wave is the perturbation isotropic with respect to two opposite directions that acts on the electron system.

A traveling acoustic wave can be generated in a crystal by applying a periodic force to the crystal's boundary. The wave decays upon leaving the boundary, as it imparts energy to electrons. The intensity $I(r)$ of a wave propagating in the direction of vector $\mathbf{r}$ therefore takes the form $I(r)=I(0)\exp(-\alpha r)$, where α is the wave absorption coefficient. Hence, two traveling waves propagating in opposite directions form a standing wave only if absorption is low, i.e., if the amplitudes of the traveling waves can be assumed constant (for all practical purposes), which corresponds to the condition

$$\alpha L\ll 1, \tag{17}$$

where L is the size of the 2D structure in the direction of wave propagation.

Suppose that condition (17) is met and that there is a standing acoustic wave in the system, a wave that is a superposition of longitudinal acoustic waves with amplitudes u_0 and wave vectors $\mathbf{q}$ and $-\mathbf{q}$. If electron transitions initiated by the electron–wave interaction occur within the first Brillouin zone and the temperature is zero ($T=0$), the phonon drag emf's of electrons generated in the x and y directions in the standing wave, $\mathscr{E}_x$ and $\mathscr{E}_y$, are given by[2]

$$\mathscr{E}_x=\frac{\hbar L_x q_x}{en_s}[w_a(-\mathbf{q})-w_a(\mathbf{q})],$$
$$\mathscr{E}_y=\frac{\hbar L_y q_y}{en_s}[w_a(-\mathbf{q})-w_a(\mathbf{q})], \tag{18}$$

where n_s is the electron concentration per unit surface area of the 2D system. When $T\neq 0$, the presence of a standing wave leads not only to absorption but also to stimulated emission of phonons by electrons, in view of which Eqs. (18) become

$$\begin{aligned}\mathscr{E}_x&=\frac{\hbar L_x q_x}{en_s}[w_a(-\mathbf{q})-w_a(\mathbf{q})+w_e(\mathbf{q})-w_e(-\mathbf{q})],\\ \mathscr{E}_y&=\frac{\hbar L_y q_y}{en_s}[w_a(-\mathbf{q})-w_a(\mathbf{q})+w_e(\mathbf{q})-w_e(-\mathbf{q})].\end{aligned} \tag{19}$$

If the wave vector $\mathbf{q}=(q_x,0,q_z)$, Eqs. (13), (14), and (19) yield

$$\mathscr{E}_x=(q_z d_0)^2\frac{4v_l}{15c}\frac{q}{q_x}\frac{H_y}{E_z}\mathscr{E}_{x0}(\mathbf{q}),\qquad \mathscr{E}_y=0, \tag{20}$$

where

$$\mathscr{E}_{x0}(\mathbf{q})=\frac{\hbar L_x q_x[w_{e0}(\mathbf{q})-w_{a0}(\mathbf{q})]}{en_s} \tag{21}$$

is the ordinary emf of phonon drag of electrons in a traveling acoustic wave with a wave vector $\mathbf{q}$ in the absence of a magnetic field. For a wave vector $\mathbf{q}=(2k_F,0,0)$ at absolute zero $T=0$ and low electron concentrations ($v_F\approx v_l$), Eqs. (12), (15), and (19) yield

$$\mathscr{E}_x=-\left(\frac{2}{\pi}\right)^2\frac{6v_l}{c}\left(\frac{2d_0}{3l_H}\right)^4\frac{H_y}{E_z}\mathscr{E}_{x0}(\mathbf{q}),\qquad \mathscr{E}_y=0, \tag{22}$$

where the emf (21) in explicit form is

$$\mathscr{E}_{x0}(\mathbf{q})=-\sqrt{v_l(2v_F-v_l)}\left(\frac{2\Xi u_0 m}{\hbar}\right)^2\frac{2L_x}{e\hbar},\qquad v_F\geqslant v_l/2.$$

In the same case but at high electron concentrations ($v_F\gg v_l$), Eqs. (12), (16), and (19) yield

$$\mathscr{E}_x=-\left(\frac{2}{\pi}\right)^2\frac{v_F}{c}\left(\frac{2d_0}{3l_H}\right)^4\frac{H_y}{E_z}\mathscr{E}_{x0}(\mathbf{q}),\qquad \mathscr{E}_y=0, \tag{23}$$

where the emf (21) in explicit form is

$$\mathscr{E}_{x0}(\mathbf{q})=-\sqrt{2v_l v_F}\left(\frac{4\Xi u_0 m}{\hbar}\right)^2\frac{2L_x}{e\hbar}.$$

Equation (23) implies that for 2D inversion layers on a silicon surface with $d_0\approx 5\times 10^{-7}$ cm and $H_y\sim 10^5$ G we have $|\mathscr{E}_x/\mathscr{E}_{x0}(\mathbf{q})|\sim 10^{-3}$.

4. EMF GENERATION VIA SPATIALLY UNIFORM HEATING OF THE 2D SYSTEM

We now discuss the interaction of electron and phonon systems that are in a nonequilibrium state because of uniform heating. We consider low temperatures, where electron–phonon transfer processes outside the first Brillouin zone can be ignored. This means that the exchange of momentum between the electron and phonon systems generates an emf along the x axis given by[3]

$$\mathscr{E}_x=\frac{\hbar L_x}{en_s}\sum_{\mathbf{q}} q_x[w_e(\mathbf{q})-w_a(\mathbf{q})]. \tag{24}$$

At temperatures much lower than $\hbar v_l/k_B d_0$ the main contribution to electron–phonon interaction is provided by long-wavelength phonons, whose wave vectors satisfy $(q_z d_0)^2\ll 1$. In this case, for magnetic fields that satisfy $(d_0/l_H)^4\ll 1$ we can use the expansion (11), whereupon (24) takes the form

$$\begin{aligned}\mathscr{E}_x=\frac{\hbar\Xi^2 L_x}{8\pi^4\rho v_l e n_s}\frac{\hbar H_y}{15mcE_z}(q_z d_0)^2\int d^2\mathbf{k}\int d^2\mathbf{k}'\int d^3\mathbf{q}\\ \times(k_x+k_x')q_x q([1-f(\varepsilon_{k'0})]f(\varepsilon_{k0})[n(\mathbf{q})+1]\\ -[1-f(\varepsilon_{k0})]f(\varepsilon_{k'0})n(\mathbf{q}))\,\delta(k_x'-k_x+q_x)\\ \times\delta(k_y'-k_y+q_y)\,\delta(\varepsilon_{k'0}-\varepsilon_{k0}+\hbar v_l q),\end{aligned} \tag{25}$$

where integration with respect to $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{q}$ is carried out over the first Brillouin zone. We examine the situation in which the phonon energy distribution is given by the Bose–Einstein function

$$n(\mathbf{q})=\left(\exp\frac{\hbar v_l q}{k_B T}-1\right)^{-1},$$

where T is the temperature of the phonon system, and the electron energy distribution is given by the Fermi–Dirac function

$$f(\varepsilon_k)=\left(\exp\frac{\varepsilon_k-\varepsilon_F}{k_B T_e}+1\right)^{-1},$$

where $T_e=T+\Delta T$ is the temperature of the electron system. For $|\Delta T/T|\ll 1$, $2mv_l^2\ll k_B T\ll\hbar v_l/d_0$, and a nondegenerate electron gas, Eq. (25) becomes

$$\mathscr{E}_x=\left(\frac{k_B T}{v_l}\right)^5\left(\frac{2\Xi d_0}{\hbar^3}\right)^2\frac{mL_x}{\pi\rho ce}\frac{\Delta T}{T}\frac{H_y}{E_z}. \tag{26}$$

For $|\Delta T/T|\gg 1$ at $T=0$, $2mv_l^2\ll k_B\Delta T\ll\hbar v_l/d_0$, and a nondegenerate electron gas, Eq. (25) becomes

$$\mathscr{E}_x=\left(\frac{k_B\Delta T}{v_l}\right)^5\left(\frac{2\Xi d_0}{\hbar^3}\right)^2\frac{mL_x}{5\pi\rho ce}\frac{H_y}{E_z}. \tag{27}$$

From (26) and (27) it follows that for negative ΔT we have $\mathscr{E}_x<0$, for positive ΔT we have $\mathscr{E}_x>0$, and at $\Delta T=0$ we have $\mathscr{E}_x=0$. In physical terms, when ΔT is negative, energy is transferred from the phonon system to the electron system, and this is accompanied by phonon absorption by the electrons. Since the phonon absorption probability (11) is such that $w_a(\mathbf{q})>w_a(-\mathbf{q})$ for positive q_x, such energy transfer is accompanied by momentum transfer to the electron system in the direction $\langle x\rangle$, by virtue of which $\mathscr{E}_x<0$. When ΔT is positive, energy is transferred from the electron system to the phonon system, and this is accompanied by phonon emission. Since the phonon emission probability is such that $w_e(\mathbf{q})>w_e(-\mathbf{q})$ for positive q_x, such energy transfer is accompanied by momentum transfer to the electron system in the direction $\langle -x\rangle$, by virtue of which $\mathscr{E}_x>0$. Finally, at $\Delta T=0$ the electron and phonon systems are in thermodynamic equilibrium and, in accordance with detailed balance, the emission and absorption probabilities are equal for every phonon, so that there is no momentum transfer from one system to the other and $\mathscr{E}_x=0$. Thus, an increase in the phonon temperature T (e.g., due to uniform heating of the crystal) or an increase in the electron temperature T_e (e.g., due to heating of the electron gas by an electric field) generates an emf of phonon drag of electrons.[3] This phenomenon is a special case of the effect of anisotropic momentum transfer to electrons under an external isotropic perturbation discussed in Ref. 1, where spatially uniform heating acts as an isotropic perturbation of the electron system.

When the crystal lattice is heated $(\Delta T<0)$, the temperatures T and T_e become equal in a time interval equal to the characteristic time of energy relaxation in the electron–phonon system due to phonon absorption by electrons, with the result that the emf disappears. Hence the drift of electrons (an effect that is described by (26) when $\Delta T<0$) that emerges when the crystal lattice is uniformly heated is a time-dependent effect and can be detected if the lattice temperature changes rapidly enough.

The situation is different when the electron gas is heated $(\Delta T>0)$ by an electric field $E_\parallel=\sqrt{E_x^2+E_y^2}$, which ensures that an electric current flows in the plane of the 2D system, since in this case the emf's specified by (26) and (27) are time-independent at fixed values of T and $E_\parallel$. This feature (the fact that the emf is time-independent) can be explained by the fact that the energy transferred from the electron system to the crystal lattice during photon emission (Joule heat) is completely absorbed by the heat bath that ensures the constancy of the temperature T of the crystal lattice, so that the temperature difference ΔT does not vary with time.

We now discuss in greater detail the effects that emerge when the electron gas is heated by an electric field. The x axis is directed along the electric field, so that $E_\parallel=|E_x|$. The electric current flowing in the x direction is

$$j_x=\sigma(E_x+\mathscr{E}_x/L_x), \tag{28}$$

where σ is the conductivity in the plane of the 2D system. The scalar quantity ΔT depends only on the absolute value of the electric field, $|E_x|$, so a change in field direction does not alter the emf's specified by (26) and (27), and thus $\mathscr{E}_x(E_x)=\mathscr{E}_x(-E_x)$. Hence Eq. (28) implies that $j_x(E_x)\neq j_x(-E_x)$, and the electric current is therefore anisotropic. Suppose that the external electric field is directed along the y axis, so that $E_\parallel=|E_y|$. In this case the electron gas is heated by the current flowing in the y direction, and the emf generated by heating, $\mathscr{E}_x$, is perpendicular to the external electric field, so that a transverse emf emerges.

The electron system can be heated not only by a constant electric field but also by an alternating electric field. In particular, the electron gas becomes heated and an emf $\mathscr{E}_x$ is detected when a plane electromagnetic wave propagating in the z direction is incident on the 2D structure, with the result that a photoinduced emf appears. Thus, the generation of an emf in the event of spatially uniform heating of a 2D system leads to new kinetic effects that differ from a phenomenological standpoint, but are of the same microscopic origin.

For 2D inversion layers at the surface of the silicon, $\Delta T\sim 10\,\mathrm{K}$ at $E_\parallel\sim 10\,\mathrm{V\,cm^{-1}}$ (see Ref. 4). Hence an estimate of (27) at $E_\parallel\sim 10\,\mathrm{V\,cm^{-1}}$, $d_0\approx 5\times 10^{-7}\,\mathrm{cm}$, and $H_y\sim 10^4\,\mathrm{G}$ yields $\mathscr{E}_x/L_x\sim 10^{-2}\,\mathrm{V\,cm^{-1}}$.

5. CONCLUSION

In a quasi-two-dimensional system with an asymmetric quantizing potential $U(z)\neq U(-z)$ located in a magnetic field $\mathbf{H}=(0,H_y,0)$ parallel to the plane of the system there emerges an asymmetric electron energy spectrum $\varepsilon(v_x)\neq\varepsilon(-v_x)$, where v_x is the electron velocity in the x direction. Because of this asymmetry an anomaly emerges in the electron–phonon interactions, i.e., the probabilities of interaction of electrons with identical phonons whose wave vectors $\mathbf{q}$ and $-\mathbf{q}$ point in opposite directions turn out to differ when $q_x\neq 0$. This feature of the electron–phonon interaction is purely a quantum phenomenon, which leads to a situation in which the momenta transferred from the phonon system to the electron system prove to be different for the two opposite directions. As a result, new macroscopic effects emerge, such as the generation of an emf in a standing acoustic wave and the generation of an emf when the quasi-two-dimensional system is heated uniformly in space. These phenomena are particular manifestations of the universal effect of the anisotropic momentum transfer to the electrons in a system that is under an isotropic perturbation,[1] an effect that leads to the emergence of an emf under any isotropic perturbation of any electron system with an asymmetric energy spectrum.

APPENDIX

In calculating the integrals of Airy functions we use the following supplemental relationships:

$$\int \mathrm{Ai}^2(\zeta)\,d\zeta=\zeta\mathrm{Ai}^2(\zeta)-\mathrm{Ai}'^2(\zeta),$$

$$\int \zeta\mathrm{Ai}^2(\zeta)\,d\zeta=\frac{1}{3}[\zeta^2\mathrm{Ai}^2(\zeta)-\zeta\mathrm{Ai}'^2(\zeta)+\mathrm{Ai}(\zeta)\mathrm{Ai}'(\zeta)],$$

$$\int \zeta^2\mathrm{Ai}^2(\zeta)\,d\zeta=\frac{1}{5}[\zeta^3\mathrm{Ai}^2(\zeta)-\zeta^2\mathrm{Ai}'^2(\zeta)+2\zeta\mathrm{Ai}(\zeta)\mathrm{Ai}'(\zeta)-\mathrm{Ai}^2(\zeta)],$$

whose validity can be verified via direct differentiation, with allowance for the fact that the Airy function $\mathrm{Ai}(\zeta)$ is the solution of the differential equation $\mathrm{Ai}''(\zeta)=\zeta\mathrm{Ai}(\zeta)$. Using these relationships, we find that for $(q_z d_0)^2\ll 1$ and $(d_0/l_H)^4\ll 1$ the integrals of the Airy function in (10) assume the form

$$\left|\int_{-\infty}^{\infty}\varphi(k_x,z)\exp(iq_z z)\,\varphi(k_x',z)\,dz\right|^2$$
$$\times\left(\int_{-\infty}^{\infty}|\varphi(k_x,z)|^2dz\int_{-\infty}^{\infty}|\varphi(k_x',z)|^2dz\right)^{-1}$$
$$=1-\frac{1}{5}(q_zd_0)^2+\frac{\hbar H_y}{15mcE_z}(q_zd_0)^2(k_x+k_x')$$
$$+O[(q_zd_0)^4]+O\left[\left(\frac{d_0}{l_H}\right)^4\right].$$

---

*)E-mail: Oleg.Kibis@nstu.ru

[1] O. V. Kibis, JETP Lett. **66**, 588 (1997).
[2] O. V. Kibis, Phys. Lett. A **237**, 292 (1998); **244**, 574 (1998).
[3] O. V. Kibis, Phys. Lett. A **244**, 432 (1998).
[4] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Nature of the low-temperature anomalies in the physical properties of the intermediate-valent compound SmB_6

N. E. Sluchanko,*) A. A. Volkov, V. V. Glushkov, B. P. Gorshunov, S. V. Demishev,
M. V. Kondrin, A. A. Pronin, and N. A. Samarin

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

Y. Bruynseraede and V. V. Moshchalkov

Laboratorium voor Vaste-Stoffysika en Magnetisme, Katholieke Universiteit Leuven, Celestijnenlaan 200 D,
B-3001 Leuven, Belgium

S. Kunii

Department of Physics, Tohoku University, Sendai 980-77, Japan

(Submitted 21 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 970–978 (March 1999)

The transport properties (Hall coefficient, thermopower, and resistivity) of high-quality single-crystal samples of the classical mixed-valent compound SmB_6 are investigated over a broad temperature range (1.6−300 K) in magnetic fields up to 45 T for the first time following the quasioptical measurements in the 0.6−4.5 meV frequency range [B. Gorshunov, N. Sluchanko, A. Volkov *et al.*, submitted to Phys. Rev. B (1998)]. Measurements in the intrinsic conduction region permit determination of the gap width $E_g \approx 20$ meV and evaluation of the behavior of the mobility and concentration of light and heavy charge carriers, as well as the temperature dependence of the carrier relaxation time, in samarium hexaboride. The results of experimental investigations in the ''impurity'' conduction region ($E_{\mathrm{ex}} \approx 3.5$ meV) are discussed within the Kikoin–Mishchenko exciton–polaron model of charge fluctuations. Arguments supporting the formation of a metallic state with an electron–hole liquid in SmB_6 at liquid-helium temperatures are presented.
[S1063-7761(99)01403-1]

1. INTRODUCTION

Although the first detailed investigation of samarium hexaboride was performed almost 30 years ago[2] and more than 100 studies devoted to this very interesting material have been published, we are still lacking a widely accepted, fully consistent physical framework that accounts for the unusual properties of this intermediate-valent compound of Sm ($v(\mathrm{Sm}) \approx 2.6$). It is especially difficult to interpret low-temperature anomalies in the physical parameters of SmB_6, which include not only dependence of the characteristics of the object of investigation on the method used to prepare the surface of the sample,[3] but also on its history.[4]

The investigations of SmB_6 performed by different investigators (see, for example, the review in Ref. 5) enable us to classify this compound as a narrow-gap semiconductor. The width of the gap in the spectrum of elementary excitations of SmB_6 found by various experimental methods ranges from 3−5 to 10−15 meV.[5,6]

Recent direct measurements of the low-temperature dynamic conductivity $\sigma(\omega)$ and the dielectric constant $\varepsilon(\omega)$ in the energy range 0.6−4.5 meV performed in Ref. 1 permitted establishment of the gap width $E_g \approx 19 \pm 2$ meV in the spectrum of electron states. In addition, within the simple phenomenological semiconductor approach, the properties of samarium hexaboride at liquid-helium temperatures may be due, according to Ref. 1, to the existence of an additional narrow band of impurity (donor) states 3.5 meV below the bottom of the conduction band.

In such a situation, in our opinion, important information on the origin and structure of the gap states can be provided by detailed precision measurements of the Hall coefficient $R_H(H,T)$, the resistivity $\rho(T)$, and the thermopower $S(T)$ performed on the same single crystal as the quasioptical measurements in Ref. 1. Note that despite the considerable number of studies of these parameters in SmB_6 reported in the literature (see, for example, Refs. 2 and 7–10, as far as we know there has not hitherto been a comparison of experimental data on $R_H(H,T)$, $\rho(T)$, and $S(T)$ obtained on a single, high-purity ($RRR > 10\,000$) single-crystal sample of samarium hexaboride.

Thus, the purpose of the present study was to obtain detailed experimental information on the behavior of $R_H(H,T)$, $\rho(T)$, and $S(T)$ at low temperature and to analyze it within the phenomenological semiconductor and other approaches.

Single-crystal samples from the same ingot as in Ref. 1 were used in the investigation. Special attention was focused on the preparation of the sample surface[3] and the contacts for performing the resistance measurements. The procedure for the thermopower measurements was similar to the method employed in Refs. 11 and 12, and the procedure for the gal-

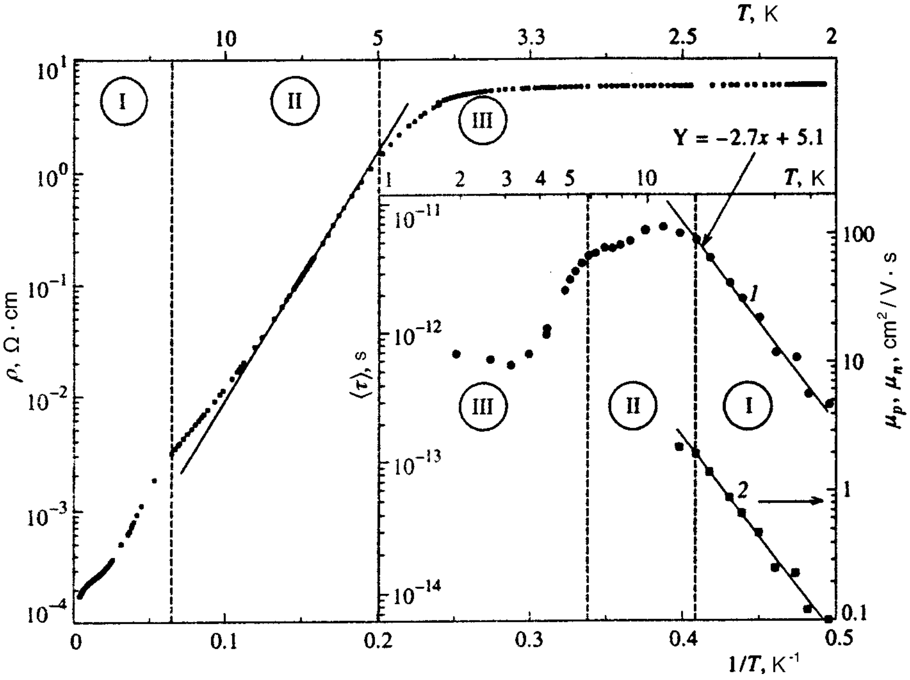

FIG. 1. Temperature dependence of the resistivity of one of the SmB$_6$ samples. Inset — temperature dependence of the mobility and relaxation time (see the text) of light and heavy charge carriers.

vanomagnetic measurements was similar to the method described in Ref. 13. In addition, an automated experimental apparatus of our own original design, which permits control of the rotation of the sample in a constant magnetic field using a stepper motor, was used to investigate the Hall coefficient.

2. EXPERIMENTAL RESULTS

The temperature dependence of the resistivity of one of the three SmB$_6$ samples investigated in the present work is presented in Fig. 1 and conforms, on the whole, to the results in Refs. 5 and 14. An increase in resistivity close to the activation value is observed below 70 K, and between 6 and 14 K the experimental data can be approximated by the expression $\rho \sim \exp(T_0/T)$ with $T_0 \approx 44-46$ K for different samples of SmB$_6$.

Measurements of the Hall coefficient $R_H(H_0,T)$ carried out in the range of magnetic fields $H_0 \leqslant 8$ T are presented in Fig. 2. Three characteristic segments of the $R_H(T)$ curve can be identified in the range below liquid-nitrogen temperature (I–III in Fig. 2). The plots of the temperature dependence of the Hall coefficient (Fig. 2) in the ranges $14-50$ K (I) and $6-14$ K (II), which closely conform to the activation dependence

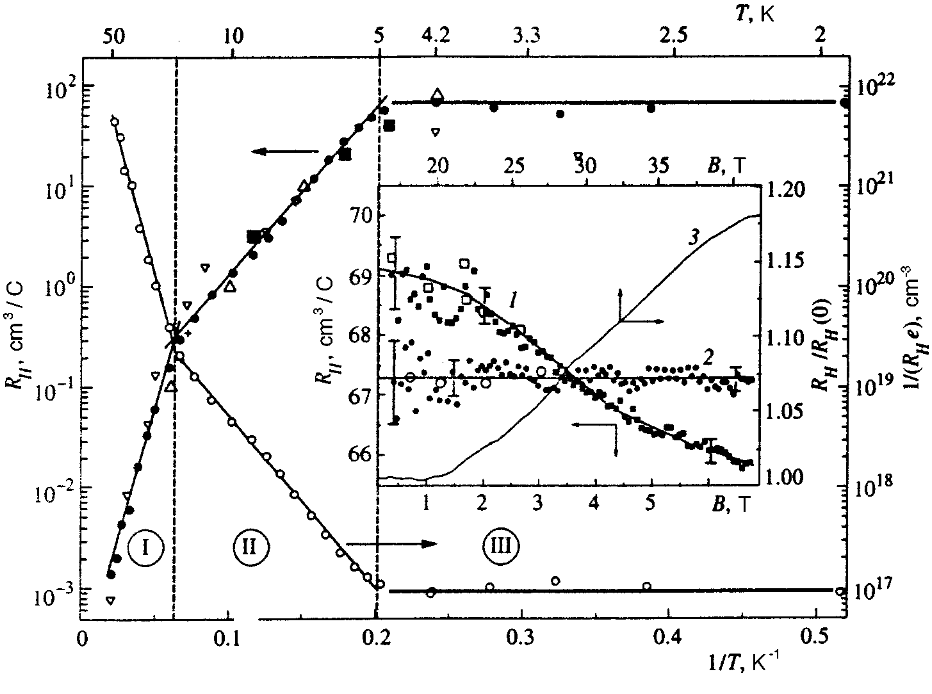

FIG. 2. Temperature dependence of the Hall coefficient (1) and $(R_H e)^{-1}$ (2) (see the text): ● — measurements carried out in the present work in magnetic fields up to 1 T; ■ — measurements performed at $H=5$ T; ▽, △ — data from Refs. 7 and 10, respectively. The inset shows field dependences of the Hall coefficient: *1* — $T=4.2$ K, *2* — $T=2$ K (in constant fields up to 7 T), *3* — $T=4.2$ K (in pulsed fields up to 45 T).

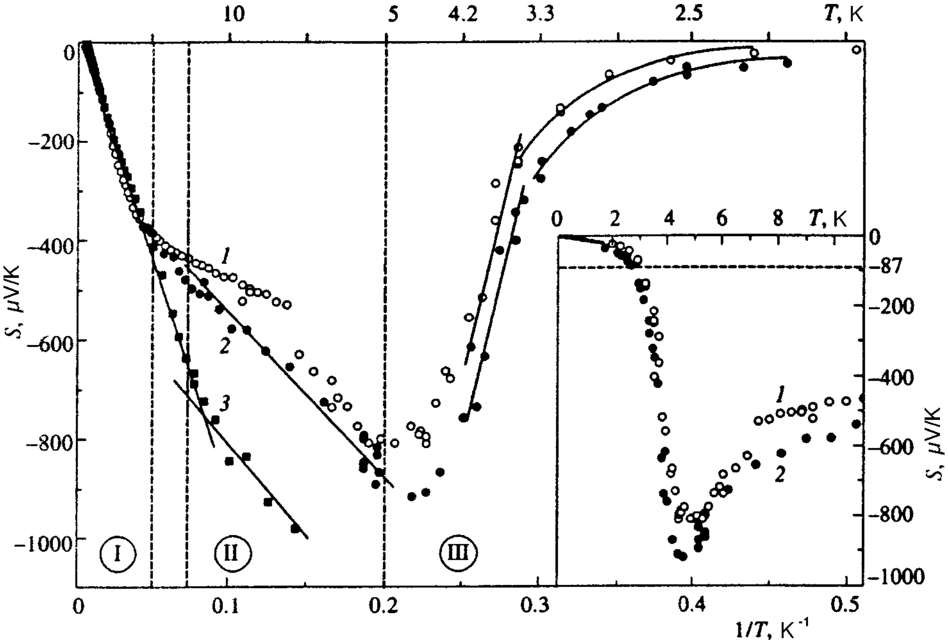

FIG. 3. Temperature dependence of the thermopower for two SmB_6 samples obtained by different methods. Inset — plot of $S(T)$ at liquid-helium temperatures.

$$|R_H| \sim \exp(T_0/T), \tag{1}$$

provide the estimates $T_0^{\mathrm{I}} \approx 120$ K and $T_0^{\mathrm{II}} \approx 45$ K. We note that the absolute value and behavior of the Hall coefficient $R_H(T)$ obtained in the present work are consistent with the results in Refs. 7,8, and 10 (see Fig. 2). The most appreciable differences are observed at liquid-helium temperatures (region III in Fig. 2). The spread of values of R_H ($T \leqslant 5$ K) obtained by different investigators is probably due to the appreciable difference between the quality of the SmB_6 single crystals that we investigated ($RRR \geqslant 10\,000$) and the samples measured in Refs. 7,8, and 10. In addition, along with the determination of the activation parameters from the Hall measurements performed on the SmB_6 single crystals, the use of the apparatus with rotation of the sample in the magnetic field revealed the nonlinear course of $R_H(H)$ in the vicinity of liquid-helium temperature (see the inset in Fig. 2). The degree of nonlinearity of the Hall coefficient was less than 5% in magnetic fields up to 7 T. The nonlinearity of $R_H(H)$ is not observed at temperatures in the vicinity of 2 K (curve *2* in the inset in Fig. 2). At the same time, systematic limitations measurements of the field dependence of the Hall voltage arise in temperature ranges II and III of the exponential variation of $R_H(T)$, due to the influence on measurement accuracy of temperature instabilities that amount to less than 0.05 K.

The temperature dependence of the thermopower for two SmB_6 samples, obtained at temperatures below 100 K by different methods,[11] is shown in Fig. 3. As in the case of $\rho(T)$ and $R_H(T)$ (Figs. 1 and 2), three characteristic ranges of the thermopower can be identified on the $S(T)$ curves (I–III in Fig. 3): activation segments that are nearly linear in the coordinates and for which

$$S \sim T_0/T \tag{2}$$

(I–II) and a low-temperature region (III), which corresponds to a rapid drop in the absolute value of $S(T)$ with decreasing temperature. The values of T_{0S}^{I} and T_{0S}^{II} estimated from the results in Fig. 3 are 115 and 40 K, respectively. The rapid drop in thermopower in region III with decreasing temperature is accompanied by a transition to a power-law dependence of $S(T)$ with fairly low absolute values in the range $S \leqslant 86.2\ \mu\mathrm{V/K} = k_B/e$, which are characteristic of a system of the metallic type.

3. DISCUSSION OF RESULTS

In accordance with the generally accepted interpretation[5] (see also Refs. 15 and 16), the spectrum of elementary excitations of SmB_6 contains an indirect gap due to hybridization of the *f* and *d* states, which is responsible for the appearance of low-temperature anomalies in the physical characteristics of samarium hexaboride. Let us dwell in detail on an analysis of the experimental results in Figs. 1–3 in each of the temperature ranges I–III.

a. Intrinsic conduction region $T \geqslant 14$ K (I)

Following Ref. 1, we utilize the phenomenological semiconductor approach to interpret the experimental results in Figs. 1–3 in the intrinsic conduction region (I). The intrinsic thermopower of a semiconductor can be described using the expression

$$S = \frac{k_B}{e} \left\{ \frac{b-1}{b+1} \frac{E_g}{2k_BT} + \frac{3}{4} \ln \frac{m_n}{m_p} \right\}, \tag{3}$$

where $b = \mu_n/\mu_p$; μ_n, m_n, μ_p, and m_p are the electron and hole mobilities and effective masses, respectively; k_B is the Boltzmann constant; and e is the charge of the electron. Since T_{0S}^{I} and E_g are related by

$$T_{0S}^{\mathrm{I}} = \frac{E_g}{2} \frac{b-1}{b+1}, \tag{4}$$

and the value found from the Hall measurements, $T_{0H}^{\mathrm{I}} = E_g/2 = 120$ K, enables us to estimate the mobility ratio $b = \mu_n/\mu_p \approx 50$. As a result, because of the fairly small relative value of μ_p it is possible to evaluate the behavior of the temperature dependence of the electron mobility from the data in Figs. 1 and 2 using the simplest model that allows for one group of carriers:

$$\mu_n = R_H/\rho. \tag{5}$$

The temperature dependence of the electron Hall mobility in the conduction band μ_n has a significantly nonmonotonic character (see the inset in Fig. 1). In the intrinsic conduction region (I) there is a power-law dependence of the form $\mu_n \sim T^{-2.7}$, which probably corresponds to a combination of the contributions from scattering on acoustic ($\mu_n \sim T^{-3/2}$) and optical lattice modes. We note that a similar dependence $\mu_n \sim T^{-2.6}$ is observed in electronic silicon in the temperature range $300\mathrm{K} < T \leqslant 400$ K.[17]

Next, let us utilize the results in Refs. 5 and 1 to estimate the carrier relaxation time in region I. Let us also take into account that the value $m_f = m_p \approx 1000 \pm 500 m_0$ found in Ref. 5 characterizes the plasma resonance of the carriers right at Sm centers and is thus an estimate of the effective mass of the valence-band carriers. The value $m^* = m_n \approx 100 m_0$, obtained in the Drude model from an analysis of quasioptical spectra,[1] corresponds, in turn, to thermally excited carriers, i.e., conduction electrons. Substituting the set of values of the parameters $b = \mu_n/\mu_p$, μ_n, m_n, and m_p into the expression $\mu = e\langle\tau\rangle/m$, we find $\langle\tau_p\rangle \approx (1/5)\langle\tau_n\rangle$. We emphasize that in the presence of fast $[\sim 10^{-14}$ s (Ref. 5)] charge fluctuations, significantly heavier charge carriers, i.e., $4f$ holes ($m_p/m_n \approx 10$), are characterized by considerably smaller values of $\langle\tau_p\rangle$ than are the conduction electrons. We also note that the temperature dependence of $(R_H e)^{-1}$ in region I corresponds to the variation of the concentration of intrinsic carriers to within the multiplier $(1-b)/(1+b) = 0.95$ (curve *2* in Fig. 2).

b. Temperature range 6–14 K (II)

Activation behavior of all the parameters $R_H(H,T)$, $\rho(T)$, and $S(T)$ with similar values of the activation energy $E_{\mathrm{ex}} = T_0^{\mathrm{II}} = 40 \pm 5\mathrm{K} \approx 3.5$ meV is observed in this temperature range (Figs. 1–3), obviously due to the exponential decrease in the concentration of conduction electrons $n = n_0 \exp(-E_{\mathrm{ex}}/k_B T)$ (curve *2* in Fig. 2). Estimates of the mobility and relaxation time of ''light'' carriers in the approximation (5) lead to the conclusion that there is a transition to impurity scattering in region II (see the inset in Fig. 2), in accordance with the results of Ref. 1. Using the values $m_n \approx 100 m_0$ and $E_{\mathrm{ex}} \approx 3.5$ meV, we estimate the localization radius of the impurity states:

$$a^* = \hbar/\sqrt{2m^* E_{\mathrm{ex}}} \approx 3\,\text{Å}. \tag{6}$$

Such a small value of a^* probably suggests leaning toward interpreting the low-temperature properties of SmB_6 in terms of the Kikoin–Mishchenko exciton–polaron model.[18] We note that unlike the approach in Refs. 19–21, which treats SmB_6 as a metallic Kondo system with low carrier concentration, the idea of an exciton semiconductor has provided a framework for proposing an explanation for the additional local modes in the vibrational spectrum of SmB_6,[18,22] as well as for performing a detailed analysis of the collective low-frequency excitations in the vicinity of $\Delta E \approx 14$ meV.[23] In developing the approach in Ref. 18 for application to the analysis of the experimental data in Figs. 1–3, we must assume that the states at $E_{\mathrm{ex}} \approx 3.5$ meV correspond to the formation of excitons of small radius ($a_{\mathrm{ex}} \approx 2-4$ Å) in the vicinity of Sm centers as a consequence of fast valence fluctuations of electrons in the samarium $4f$ subshell. Simple estimates of the Bohr radius of an exciton $a_{\mathrm{ex}} \approx 2\varepsilon m_0 a_B/m^*$ (a_B is the Bohr radius) and its energy $E_{\mathrm{ex}} \approx m^* e^4/4\varepsilon^2\hbar^2$ with consideration of the values $m^* \approx 100 m_0$ and $\varepsilon \approx 600$ (Ref. 1) lead to similar values of the parameters: $a_{\mathrm{ex}} \leqslant 6$ Å and $E_{\mathrm{ex}} \geqslant 2$ meV.

c. Temperature range $T \leqslant 6$ K (III)

The emergence of a maximum in $R_H(T)$ and $S(T)$ in this temperature range and the tendency for saturation of the resistivity $\rho(T)$ (Figs. 1–3) can be analyzed within the approximation of several groups of charge carriers. However, the nonlinearity of the field dependence of $R_H(H,T = 4.2\,\mathrm{K})$ in fields up to 7 T (curve *1* in the inset in Fig. 2) can probably be explained only as a manifestation of the weak-field ($\mu_n H \ll 1$) asymptote

$$R_H \sim \frac{1}{ne}(1 - a_1 \mu_n^2 H^2) \tag{7}$$

and the strong-field ($\mu_n H \gg 1$) asymptote

$$R_H \sim \frac{1}{ne}\left(1 + \frac{a_2}{\mu_n^2 H^2}\right). \tag{8}$$

In this case, reducing the temperature to 2 K is accompanied by an appreciable decrease in the mobility μ_n (curve *1* in the inset in Fig. 1), which in turn significantly reduces the quadratic term in (7) and shifts the strong-field asymptote (8) outside the experimental range $H \leqslant 8$ T. As a result, the value of $R_H(H, T = 2\mathrm{K})$ is constant to within the measurement error (curve *2* in the inset in Fig. 2).

To detect the contribution of heavy carriers to the Hall coefficient $R_H(H,T)$, measurements were also carried out in this study in pulsed fields up to 45 T at liquid-helium temperature. We note that the variations of $R_H(H, T = 4.2\mathrm{K})$ (curve *3* in the inset in Fig. 2) observed at $H \geqslant 20$ T correspond to the contribution of *p*-type heavy carriers, but consideration of only the light carriers is fully justified in fields up to 8 T.

Analyzing the results in region III in Figs. 1–3, we also note that the thermal expansion coefficient vanishes[24] and that appreciable variation of the elastic modulus C_{11}, [4] a change in the NMR spin-lattice relaxation regime,[25] a maximum of the quadratic negative magnetoresistance,[26] and other effects are observed in the vicinity of $T_m \approx 5$ K. Such significant changes in the thermodynamic and kinetic characteristics of SmB_6 are probably evidence of a phase transition in the electronic subsystem with $T_m \approx 5$ K. In our opinion, one of the possible reasons for such a transition may be the condensation of excitons with the formation of electron–hole droplets of submicron radius in SmB_6. Let us briefly recount plausible arguments favoring the proposed interpretation.

Unlike classical semiconductors with electron–hole droplets, in which the charge carriers and excitons are generated as a result of photoexcitation, in SmB_6 the fast charge fluctuations at each Sm center are sources of carriers in the conduction band and at exciton levels near liquid-helium temperatures. Polaron effects[18] significantly facilitate the passage of excitons into the condensed phase. In addition, because intrinsic defects and impurities in SmB_6 distort the crystal lattice in the immediate vicinity of impurity centers, they should probably have a decisive effect on the formation of electron–hole droplets. Consequently, the doping level and the character of the distribution of impurities in samarium hexaboride crystals are the main factors governing the emergence of anomalies in the low-temperature properties of SmB_6.

Returning to the results in Figs. 1–3, we note that the maximum of the Hall coefficient $R_H(T)$ can be interpreted within Shklovskiĭ's model[27] for an inhomogeneous medium, i.e., a dielectric matrix containing metallic inclusions. Under the assumption of a dynamically inhomogeneous medium containing electron–hole droplets moving along voltage and temperature gradients[28] we can obtain a qualitative explanation for the plots of $\rho(T)$ in Fig. 1 and of the thermopower in Fig. 3.[29]

Finally, based on the arguments in Ref. 28, we present order-of-magnitude estimates of the principal parameters of the condensed phase and the region for its existence. For example, for the critical concentration and binding energy of the particles in the condensed phase we have $n_c \sim a_{\mathrm{ex}}^{-3} \approx 3.5 \times 10^{22}\ \mathrm{cm}^{-3}$ and $E_1 \sim E_{\mathrm{ex}} \sim 10 k_B T \approx 3-5$ meV. These estimates agree quite well with the underlying assumptions. We also note that to within the accuracy of the experiment and the current estimates of the binding energy E_1, the collective low-frequency excitations in SmB_6 with $\Delta E \approx 14$ meV detected in inelastic neutron scattering[23] and Raman scattering[30] experiments can be associated with the formation of electron–hole droplets in the dielectric matrix at low temperatures. At the same time, we stress once again that additional research is needed to confirm the hypotheses advanced regarding the existence of electron–hole droplets in the low-temperature phase of SmB_6 within the approach developed in Ref. 28.

We are indebted to M. Dressel and A. Loidl for numerous useful discussions. This research was carried out with financial support from the Russian Fund for Fundamental Research (Projects 98-02-17163 and 96-15-96929) and the ''Fundamental Spectroscopy,'' ''Microwave Physics,'' and ''Fullerenes and Atomic Clusters'' programs of the Ministry of Science and Technology, as well as with financial support from INTAS (Grants 96-451 and 94-4435).

*)E-mail: nes@lt.gpi.ru

---

[1] B. Gorshunov, N. Sluchanko, A. Volkov *et al.*, Phys. Rev. B **59**, 1808 (1999).
[2] V. C. Nickerson, R. M. White, K. N. Lee *et al.*, Phys. Rev. B **3**, 2030 (1971).
[3] A. Kebede, M. C. Aronson, C. M. Buford *et al.*, Physica B **223–224**, 256 (1996).
[4] S. Nakamura, T. Goto, M. Kasaya *et al.*, J. Phys. Soc. Jpn. **60**, 4311 (1991).
[5] P. Wachter, in *Handbook on the Physics and Chemistry of Rare Earths*, Vol. 19, K. A. Gschneidner, Jr., L. Eyring, G. H. Lander, and G. R. Choppin (eds.), Elsevier (1994).
[6] T. Namba, H. Ohta, M. Motokawa *et al.*, Physica B **186–188**, 440 (1993).
[7] J. W. Allen, B. Battlogg, and P. Wachter, Phys. Rev. B **20**, 4807 (1979).
[8] S. von Molnar, T. Theis, A. Benoit *et al.*, in *Valence Instabilities*, P. Wachter and H. Boppart (eds.), North-Holland (1982), p. 389.
[9] T. Kasuya *et al.*, J. Phys. Colloq. **40**, C5-308 (1979).
[10] J. C. Cooley, M. C. Aronson, Z. Fisk, and P. C. Canfield, Phys. Rev. Lett. **74**, 1629 (1995).
[11] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev *et al.*, Phys. Rev. B **56**, 10 816 (1997).
[12] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, M. V. Kondrin, N. A. Samarin, V. V. Brazhkin, and V. V. Moshchalkov, Zh. Éksp. Teor. Fiz. **113**, 339 (1998) [JETP **86**, 190 (1998)].
[13] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev *et al.*, Phys. Rev. B **51**, 1112 (1995).
[14] J. Roman, V. Pavlik, K. Flachbart *et al.*, Physica B **230–232**, 715 (1997).
[15] N. F. Mott, *Metal–Insulator Transitions*, Taylor and Francis, London; Barnes and Noble, New York (1974).
[16] R. M. Martin and J. W. Allen, J. Appl. Phys. **50**, 7561 (1979).
[17] A. Anselm, *Introduction to Semiconductor Theory*, Prentice–Hall, Englewood Cliffs, NJ (1981), p. 504.
[18] K. A. Kikoin and A. S. Mishchenko, J. Phys.: Condens. Matter **7**, 307 (1995).
[19] T. Kasuya, Europhys. Lett. **26**, 283 (1994).
[20] T. Kasuya, Europhys. Lett. **26**, 277 (1994).
[21] T. Kasuya, J. Phys. Soc. Jpn. **63**, 397 (1994).
[22] K. A. Kikoin and A. S. Mishchenko, J. Phys.: Condens. Matter **2**, 6491 (1990).
[23] P. A. Alekseev, J-M. Mignot, J. Rossat-Mignod *et al.*, J. Phys.: Condens. Matter **7**, 289 (1995).
[24] D. Mandrus, J. L. Sarao, A. Lacerda *et al.*, Phys. Rev. B **49**, 16 809 (1994).
[25] O. Pena, D. E. MacLaughlin, M. Lysak *et al.*, Sol. St. Comm. **40**, 539 (1981).
[26] J. C. Cooley, M. C. Aronson, A. Lacerda *et al.*, Phys. Rev. B **52**, 7322 (1995).
[27] B. I. Shklovskiĭ, Zh. Éksp. Teor. Fiz. **72**, 288 (1977) [Sov. Phys. JETP **45**, 152 (1977)].
[28] *Electron–Hole Droplets in Semiconductors*, C. D. Jeffries and L. V. Keldysh (eds.), North-Holland, Amsterdam–New York; Elsevier, New York (1983).
[29] S. V. Demishev, Yu. V. Kosichkin, N. E. Sluchanko, A. G. Lyapin, Usp. Fiz. Nauk **64**, 195 (1994) [Phys. Usp. **37**, 185 (1994)].
[30] P. Nyhus, S. L. Cooper, Z. Fisk, and J. Sarao, Phys. Rev. B **55**, 12 488 (1997).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Statistical thermodynamics of the formation of an infinite cluster of thermally reversible chemical bonds

I. Ya. Erukhimovich*) and A. V. Ermoshkin

M. V. Lomonosov Moscow State University, 117234 Moscow, Russia

(Submitted 6 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 979–990 (March 1999)

A systematic ‘‘mean-field’’ treatment of the thermodynamic equilibrium formation of an infinite cluster of bonds in a system of identical monomers capable of forming from $n=0$ to $n>2$ reversible chemical bonds with one another is proposed within the Cayley-tree approximation. For this purpose the difference between the symmetry of the monomers appearing in ‘‘point-to-point’’ and closed bond paths, respectively, is taken into account on the basis of an analysis of the structure of the infinite cluster. Minimization with respect to the distribution of such monomers yields a nontrivial solution corresponding to a lower free energy than the classical solution, which does not allow for the symmetry difference indicated. In addition, it is shown that the classical solution corresponds to the free-energy maximum when the infinite cluster is formed and that the formation of the infinite cluster is a first-order phase transition. The possible form of the phase diagrams of the systems considered is analyzed. *© 1999 American Institute of Physics.* [S1063-7761(99)01503-6]

1. INTRODUCTION

First-order phase transitions with the coexistence of dilute and condensed phases are known to occur as a result of both the attraction (a negative potential energy of the pair interaction) and association (the formation of thermally reversible bonds between specific chemical groups) of molecules of the substance. Another characteristic and firmly established property of some associating systems is the so-called sol-gel transition, upon which an infinite cluster of bonded particles (the gel fraction) appears on a background of a large number of finite cluster particles joined by labile bonds (the sol fraction). (Such systems, among which silicate melts and water are especially interesting, are also called weak gels.[1]) Until recently, however, there has been a debate as to whether the sol-gel transition, like the percolation transition, is a purely geometric phenomenon, which can be observed only in a computer experiment and in the dynamic behavior of weak gels, or whether observable thermodynamic features of the latter are also associated with it.

In several publications devoted to the thermodynamics of weak gels,[2–7] association was analyzed within the mean-field approximation, according to which the extent of conversion, i.e., the fraction of the chemical groups participating in the formation of labile saturated bonds, depends only on the total density of these groups. In such a treatment the sol-gel transition is a purely geometric phenomenon, which does not lead to thermodynamic singularities of the system. However, as was shown in Ref. 8, the formation of an infinite cluster of bonds (even labile bonds) leads to the appearance of a new order parameter, which, as will be described in detail below, describes the nontrivial internal structure of the infinite cluster rather than its bond concentration. The appearance of this new order parameter and the spontaneous violation of the symmetry associated with it, which can be interpreted as violation of the identity of the particles appearing in elements of different topology in the infinite cluster, naturally entails corresponding thermodynamic features. In particular, in the approximation proposed in Ref. 8 for analyzing these features near the sol-gel transition, the latter was a second-order phase transition.

In the present work the analysis in Ref. 8 of the features of the thermodynamics of a simple system of associating particles, which can bind to form an infinite cluster, is generalized to include arbitrary extents of conversion. This permits significant refinement of the description of the character of both the sol-gel transition itself and the phase diagrams of weak gels. The ensuing presentation is organized in the following manner. In Sec. 2 we present a systematic ‘‘mean-field’’ derivation of an expression for the contribution of the formation of labile bonds between associating particles to their free energy. A detailed analysis of the features of the thermodynamic behavior of weak gels associated with the formation of an infinite cluster of labile bonds and a comparison with the results of the more approximate treatment performed in Ref. 8 are presented in Sec. 3. Section 4 discusses the features of phase diagrams typical of weak gels in both the absence and presence of the infinite cluster.

2. STRUCTURE OF THE THERMODYNAMICS OF AN EQUILIBRIUM INFINITE CLUSTER AND ITS FREE ENERGY

To fix ideas, we consider a system of N monomers A_n confined to a volume V, each containing n identical chemical groups A, which are capable of forming A–A bonds in the reversible chemical reaction $A+A=A_2$. The structure of the clusters appearing in such a system can be described in an obvious way[8–10] by means of graphs with vertices of order $l \leqslant n$, which correspond to monomers with l reacting groups A. For $n<3$ the densities ρ_2 and ρ_1 of the groups which

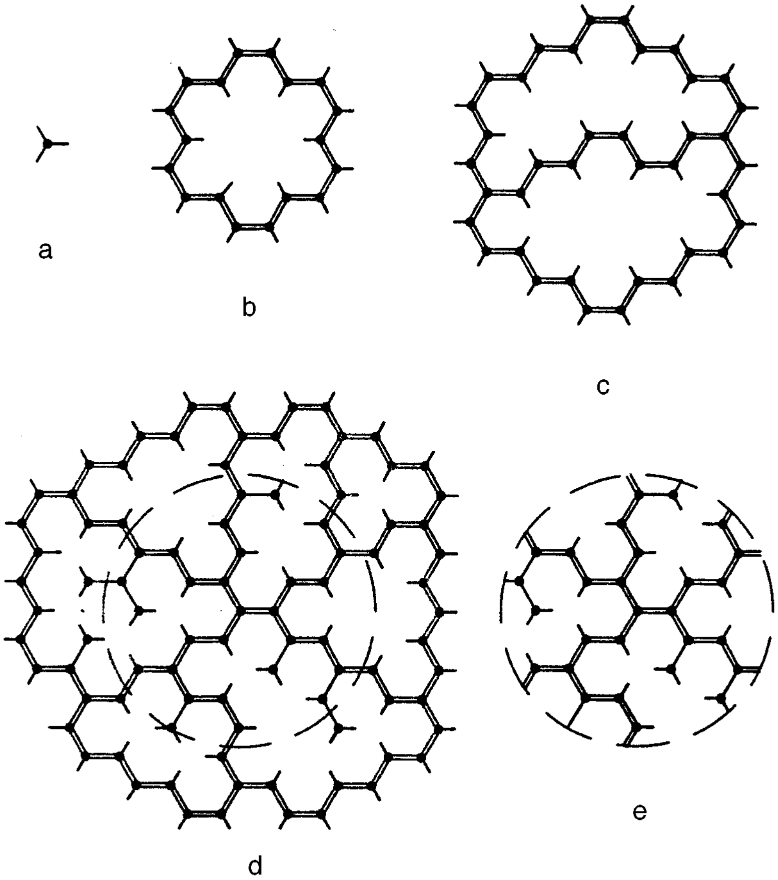

FIG. 1. Typical structure of the blocks appearing in equilibrium polymer systems of trifunctional monomers: a) individual monomers (''bare'' vertices); b, c) simple cyclic blocks; d, e) a cyclic block of complex structure and a fragment of it that appears tree-like within a small window. The inclusion of the latter in a system of closed bond paths is intimated only by the presence of local coloring (a double line corresponds to a green bond, and a single line corresponds to a red bond).

have reacted and have not reacted during this reaction, respectively, are determined by the equilibrium constant k appearing in the law of mass action:[11]

$$\rho_2 = k\rho_1^2. \tag{1}$$

For $n>2$ this law remains valid only in the absence of an infinite cluster of bonds. When an infinite cluster is present, the correlation between the bonds, which is unavoidable in this case, leads to significant modification of the relation (1).

The generally accepted starting assumption of the theory of polymer systems with an assigned temperature T and distribution of the monomer density $\rho(\mathbf{r})$ (which is, generally speaking, nonuniform and nonequilibrium) is the representation of their free energy in the form of the sum[12,13]

$$F(\{\rho(\mathbf{r})\},T) = F^*(\{\rho(\mathbf{r})\},T) + F_{\rm str}(\{\rho(\mathbf{r})\},T), \tag{2}$$

where $F_{\rm str}$ is the entropic structural contribution associated with a) the arrangement of the monomers in space, and b) the formation and distribution of labile bonds between the monomers, and F^* is the energy contribution of the ''system of unconnected monomer units'' that would exist in the absence of bonds. The latter term is often specified by the phenomenological expression corresponding to the description of a system of unconnected monomer units as a lattice liquid:[14]

$$F^*(\{\rho(\mathbf{r})\},T) = \int [T(1-\phi(\mathbf{r})) \times \ln(1-\phi(\mathbf{r})) - 2\phi^2(\mathbf{r})] \frac{dV}{v}, \tag{3}$$

where the temperature and free energy are measured in units of T_c (T_c is the critical temperature of the lattice liquid), and v and $\phi = v\rho$ are the excluded volume and the volume fraction of the monomers, respectively. Thus, our task is to calculate $F_{\rm str}$.

The starting point for our analysis is the seemingly trivial fact that an infinite cluster without cycles (a Cayley tree or a Bethe lattice), unlike a finite cluster, cannot exist in finite-dimensional space. In other words, although the structure of the infinite cluster can be represented in the form of a Cayley tree at large enough scales, it must include not only ''bare'' vertices of order $1 \leq l \leq n$, but also ''effective'' vertices of arbitrary order and complexity, which have the form of the 1-irreducible blocks depicted in Fig. 1.

Such a description of the structure of the infinite cluster[8,9] is similar to the droplet model of the infinite cluster,[15,16] but, unlike the latter, we focus our attention not on the self-similar structure of the 1-irreducible blocks, but on the possibility of distinguishing between ''internal'' bonds, from which these blocks are constructed, and ''external'' bonds, which join blocks in an effective Cayley tree. A nontrivial procedure that enables us to realize this possibility can be described as follows.[8] We choose a finite window and color all the chemical groups A within that window. If a bond between them belongs to at least one closed path of bonds lying completely within the window, we color them green; if not (or if a group has not reacted) we color it red. As the size of the window is increased, at large enough scales L some of the initially red groups appear in a closed path and should be recolored green. Thus, as L increases, the fraction of green groups grows and reaches a certain limiting value in the limit $L\to\infty$, where all the ''internal'' bonds are colored green and the ''external'' bonds are colored red.

The coloring procedure just described does not alter the statistical weights and symmetry indices of the various diagram realizations of the structure of the infinite cluster, but it is very important for finding the correct procedure for an approximate ''mean-field'' calculation followed by summation of the contributions of these diagrams to the entropic structural term F_{str}, which is similar in many respects to the calculation of the contributions of high-order diagrams to the Gell-Mann–Low function in Ref. 17.

In fact,[8] when

$$n\rho a^3 \gg 1, \tag{4}$$

where $n\rho$ is the total density of chemical groups in the system and a is the characteristic scale (length) of a chemical bond formed as a result of the reaction of two such groups, the typical blocks determining the structure of the infinite cluster include ''bare'' vertices (Fig. 1a) and very complex blocks (Fig. 1d gives a some idea of the structure of such complex blocks). The contribution of comparatively simple blocks (see Figs. 1b and 1c) can be neglected in the limit (4).

Therefore, when a certain finite volume of a system (a ''window'') that is large compared to the bond length a is considered, those bonds that belong to such a large block and that are located in that window can look like an ordinary Cayley tree (see Fig. 1d). Nevertheless, the true nature of such a quasi-Cayley tree (i.e., its inclusion in the system of closed bond paths) is intimated by the presence of local coloring, which indicates the number i of green groups belonging to each monomer in an assigned realization of the infinite cluster (i runs through all integer values in the range $0\leqslant i\leqslant n$ with the exception of $i=1$, since a monomer cannot be included in a closed path consisting of only one group that has reacted).

The next step is to postulate that a fairly faithful description of the thermodynamics of the systems under consideration can be obtained, if the calculation of the total structural free energy F_{str} is confined to a calculation of contributions to the free energy that correspond to the formation and all possible recombinations of the bonds within windows that are not too large, followed by summation of these contributions. Of course, within such a ''mean-field'' description of the infinite cluster we ignore effects resulting from the correlation of the structure of neighboring windows (i.e., the hierarchical structure of the blocks), but, on the other hand, we can take into account effects due to the difference between ''internal'' and ''external'' bonds described above.

Indeed, just the fact that a path might be closed, even somewhere far beyond the window under consideration, alters the combinatorial behavior of the corresponding functional groups. This change in combinatorial behavior at the level of the proposed ''mean-field'' description can also be taken into account by including vertices (monomers) with different colors in the treatment and appropriately altering the symmetry index. In this case we should first calculate the foregoing contributions to the free energy for an arbitrary distribution of the densities ρ_i of monomers with i green and $n-i$ red chemical groups within the window, and then determine the thermodynamic equilibrium values of these densities by minimizing the total free energy. Therefore, the desired ''mean-field'' expression for F_{str} takes the form

$$F_{\mathrm{str}}(\{\rho\},T)=V\ \min F(\{\rho_i\},\Gamma_r), \tag{5}$$

where the expression for the free energy per unit volume of the system with assigned values of the densities and of the fraction Γ_r of red groups that have reacted is

$$\frac{F(\{\rho_i\},\Gamma_r)}{T}=-\sum_{i\neq 1}\rho_i\ln\frac{e}{i!(n-i)!\rho_i\lambda^3}+\rho_r[\Gamma_r\ln\Gamma_r+(1-\Gamma_r)\ln(1-\Gamma_r)]-\tilde{S}(\rho_g)-\tilde{S}(\rho_r\Gamma_r). \tag{6}$$

Here, the first term, in which λ is the thermal wavelength of the monomers and the summation is carried out over all permissible values of i, describes the entropic contribution of the arrangement of the monomers of an assigned color in space with consideration of their symmetry index. The second term in (6) describes the entropy of selecting the $N_r\Gamma_r$ red groups that have reacted from their total number N_r. (The extent of conversion of the red groups Γ_r will also be called the extent of external conversion.) Finally, the entropic energy contributions of all possible arrangements a) of the internal bonds between green groups and b) of the external bonds between red groups that have reacted are described by the third and fourth terms, respectively, where the expression for the function $\tilde{S}(\rho)$ obtained in Ref. 8 has the form $\tilde{S}(\rho)=(\rho/2)\ln(k\rho/e)$.

The expression (6) naturally generalizes both the traditional approach, which follows from it under the assumption that the fraction of bonds belonging to closed paths is strictly equal to zero in the mean-field approximation,

$$\rho_i\equiv\begin{cases}\rho_0, & i=0,\\ 0, & i\geqslant 2,\end{cases}\qquad \rho_r\equiv n\rho_0\equiv n\rho,\qquad \rho_g\equiv 0 \tag{7}$$

(here and in the following ρ is the mean total monomer density), and the approximation proposed in Ref. 8. In this approximation only the first nontrivial density in (6), ρ_2, which plays the role of a new order parameter, is retained and derived by minimization:

$$\rho_i=\begin{cases}\rho_0, & i=0,\\ \rho_2, & i=2, \quad \rho_2+\rho_0=\rho,\\ 0, & i>2,\end{cases}$$

$$\rho_r=n\rho_0+(n-2)\rho_2,\qquad \rho_g=2\rho_2. \tag{8}$$

Minimization of the free energy (6) in the classical solution (7) leads to the following expression for F_{str}:

$$\frac{F_{\rm str}^{\rm Flory}}{VT}=\rho\ln\frac{\rho\lambda^3 n!}{e}+n\rho f_{\rm Flory}(\Gamma),$$

$$f_{\rm Flory}(\Gamma)=\frac{\Gamma}{2}+\ln(1-\Gamma), \tag{9}$$

where the dependence of the extent of conversion Γ on ρ is specified by the law of mass action (1), rewritten in the form

$$\tilde{\rho}=k\rho=\Gamma/(1-\Gamma)^2. \tag{10}$$

A detailed analysis of the approximation (8) was performed in Ref. 8, and we do not dwell on it here. The extremum of the free energy (6) is attained without additional assumptions and with consideration only of the conservation laws

$$\sum_m m\rho_m=\rho_g, \tag{11}$$

$$\sum_m (n-m)\rho_m=\rho_r, \tag{12}$$

$$\sum_m \rho_m=\rho \tag{13}$$

in the following thermodynamic equilibrium distribution of the monomers according to color:

$$\rho_i\lambda^3=z\frac{\Phi^{n-i}\Psi^i}{i!(n-i)!}, \quad \Psi=\exp\lambda, \quad \Phi=\exp\nu, \quad z=\exp\mu, \tag{14}$$

where λ, ν, and μ are the Lagrange multipliers (chemical potentials) corresponding to the conservation laws (11), (12), and (13), respectively. Here both the parameters of the distribution of the monomers according to color z, Φ, and Ψ and the structural characteristics of the infinite cluster are amenable to a comparatively simple one-parameter representation.

In order to obtain this representation we note that minimization of the structural free energy (4) with respect to Γ_r leads to the law of mass action for the equilibrium reaction leading to the formation of ''external'' bonds between red functional groups:

$$\tilde{\rho}_r=k\rho_r=\Gamma_r/(1-\Gamma_r)^2. \tag{15}$$

The minimization of $F_{\rm str}$ with respect to ρ_r (with consideration of the preceding expression) and ρ_g gives the relations

$$1/(1-\Gamma_r)=\Phi, \tag{16}$$

$$\tilde{\rho}_g=k\rho_g=\Psi^2. \tag{17}$$

Plugging the expression (14) for ρ_i into Eqs. (11) and (12) and performing the required summation, we obtain the expressions for $\tilde{\rho}_r$ and $\tilde{\rho}_g$ in the following form:

$$\tilde{\rho}_r=\frac{\tilde{z}\Phi^n}{(n-1)!}[(1+\delta)^n-(n-1)\delta], \tag{18}$$

$$\tilde{\rho}_g=\frac{\tilde{z}\Phi^n}{(n-1)!}\delta[(1+\delta)^{n-1}-1], \tag{19}$$

where

$$\tilde{z}=kz/\lambda^3, \tag{20}$$

and the quantity

$$\delta=\Psi/\Phi, \tag{21}$$

which is positive by definition, provides the desired parametrization.

In fact, equating the expressions for the ratio ρ_r/ρ_g obtained from formulas (15), (16), and (17), on the one hand, and (18) and (19), on the other, we obtain an explicit expression for the external extent of conversion Γ_r as a function of δ:

$$\Gamma_r(\delta)=\frac{(1+\delta)^n-(n-1)\delta}{[(1+\delta)^{n-1}-1]/\delta}. \tag{22}$$

The substitution of (22) into (16) and (21) specifies the functions $\Phi(\delta)$ and $\Psi(\delta)$, and subsequent consideration of formulas (15)–(17) and (21) leads to an explicit expression for the dependence of the reduced total density $\tilde{\rho}$ on δ:

$$\tilde{\rho}(\delta)=\tilde{\rho}_r+\tilde{\rho}_g=\Phi(\Phi-1)+\Psi^2=\frac{\Gamma_r(\delta)+\delta^2}{[1-\Gamma_r(\delta)]^2}. \tag{23}$$

The function (23) parametrically assigns any of the structural characteristics whose dependence on δ can be obtained using the expressions presented above as a function of $\tilde{\rho}$.

In particular, the dependence of the total extent of conversion Γ (the fraction of all the chemical groups that have reacted and participate in the formation of either external or internal bonds) on $\tilde{\rho}$ in the system under consideration is specified jointly by (23) and the expression

$$\Gamma(\delta)=\frac{\Gamma_r^2(\delta)+\delta^2}{\Gamma_r(\delta)+\delta^2}. \tag{24}$$

Equations (22)–(24) comprise the desired eneralization of the law of mass action (10) with consideration of the alteration of the symmetry of the monomers included in closed (and thereby correlated) bond paths belonging to the infinite cluster.

It is also useful to present the expression for the value of the structural free energy achieved in the thermodynamic equilibrium distribution (14):

$$\frac{F_{\rm str}}{VT}=\rho\ln\frac{\rho\lambda^3 n!}{e}+n\rho\tilde{f}(\delta),$$

$$\tilde{f}(\delta)=\frac{\Gamma(\delta)}{2}+\ln[1-\Gamma_r(\delta)]-\frac{\ln[(1+\delta)^n-n\delta]}{n}, \tag{25}$$

where the functions $\Gamma_r(\delta)$ and $\Gamma(\delta)$ are defined by formulas (22) and (24), respectively.

3. THERMODYNAMIC FEATURES OF THE FORMATION OF AN INFINITE CLUSTER OF LABILE BONDS

Thus, we have shown in the preceding section that consideration of the presence of highly cyclized blocks in the infinite cluster, which was first done in Ref. 8 in a prelimi-

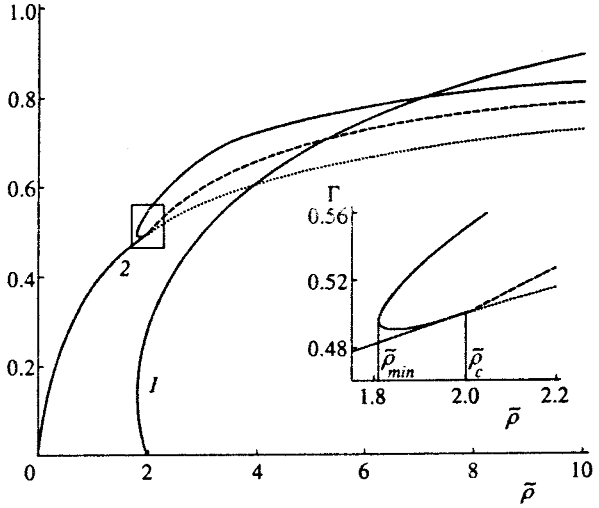

FIG. 2. Dependence of the distribution parameter δ and the extent of conversion Γ (curves *1* and *2*, respectively) on the reduced density $\tilde{\rho}$ for trifunctional monomers. In the family of curves *2* the solid line corresponds to the solution (7) at $\tilde{\rho}<\tilde{\rho}_c$ and the solution (14), and the dashed and dotted lines show the values of the extent of conversion for the classical approximation (7) and the approximation (8) at $\tilde{\rho}>\tilde{\rho}_c$.

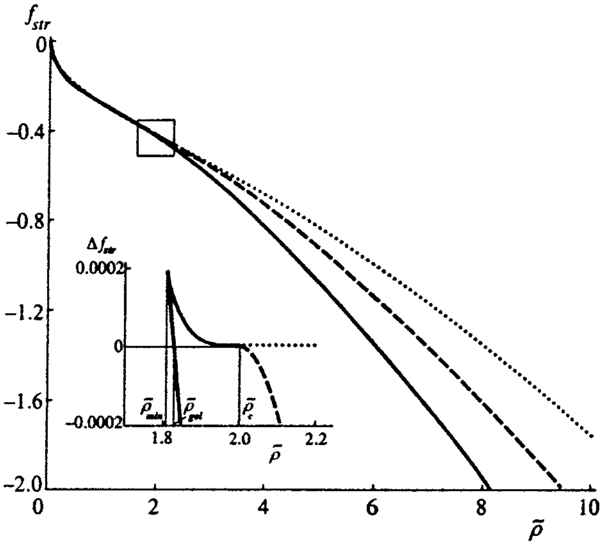

FIG. 3. Dependence of the structural part of the free energy of a system on the reduced density $\tilde{\rho}$ for trifunctional monomers. The solid, dashed, and dotted lines have the same meanings as for curves *2* in Fig. 2. Inset — values of the free energies calculated relative to the classical dependence corresponding to the solution (7).

nary study and in the present work in a systematic mean-field approximation, does, in fact, lead to the appearance of a new order parameter and, thereby, to significant modification of the law of mass action and the thermodynamics of weak gels. We first analyze the character of the emerging features for the dependences of the distribution parameter δ and the extent of conversion Γ on the reduced density $\tilde{\rho}$, which are defined by (23) and (24).

As is shown in Fig. 2, physically meaningful solutions of $\delta(\tilde{\rho})$ exist only for $\tilde{\rho}>\tilde{\rho}_{\min}$, where $\tilde{\rho}_{\min}$ is the minimum of the function (23). In the interval $(\tilde{\rho}_{\min}, \tilde{\rho}_c)$, where $\tilde{\rho}_c=\tilde{\rho}(0)$, both $\Gamma(\tilde{\rho})$ and $\delta(\tilde{\rho})$ are double-valued functions. In other words, in this interval the structural free energy (25) of the system has two extrema, the free-energy maximum corresponding to small values of Γ and the minimum corresponding to large values. An additional minimum of the structural free energy, which exists in this interval, lies on the boundary of its domain of definition $\delta=0$, i.e., on the classical solution (7). Thus, both the classical solution and the solution (14), which corresponds to a cyclized infinite cluster, exist in the interval $(\tilde{\rho}_{\min}, \tilde{\rho}_c)$, but one of these solutions is metastable.

The two extrema corresponding to the solution (14) merge at $\tilde{\rho}=\tilde{\rho}_{\min}$ and vanish as $\tilde{\rho}$ decreases further. Therefore, at $\tilde{\rho}<\tilde{\rho}_{\min}$ a single minimum of the function (25) is achieved only by the classical solution (7). Conversely, at $\tilde{\rho}>\tilde{\rho}_c$ the maximum of the function (25) merges with the classical solution, so that the classical solution ceases to exist even as a metastable solution in this region.

As follows from (22), (23), and (10), the values of the reduced density and the extent of conversion on the boundary for the existence of the classical solution (7) are specified by the expressions

$$\tilde{\rho}_c=\tilde{\rho}_c(0)=\frac{n-1}{(n-2)^2}, \quad \Gamma_c=\frac{1}{n-1}, \tag{26}$$

which correspond to the classical condition for the sol-gel transition established back in Refs. 18 and 19.

As for the values of the reduced density and the extent of conversion on the boundary for the existence of the nontrivial solution (14), a numerical calculation shows that the ratio $\tilde{\rho}_{\min}/\tilde{\rho}_c$ depends weakly on n and is roughly equal to 0.9.

Let us now examine the dependences of the values of the structural free energy of the system corresponding to the solutions (7), (8), and (14) on the reduced density $\tilde{\rho}$ (Fig. 3). It can be seen that at $\tilde{\rho}>\tilde{\rho}_c$ the largest value of the structural free energy corresponds to the classical solution and the smallest value corresponds to the solution (14) obtained in the present work. Moreover, the solution (14) already becomes thermodynamically advantageous at the point $\tilde{\rho}_{\text{gel}}<\tilde{\rho}_c$, at which the structural free energy curves corresponding to the solutions (7) and (14) intersect. This point would be the sol-gel transition point if this transition occurred without altering the value of the monomer density.

4. ANALYSIS OF PHASE DIAGRAMS

The analysis performed in the preceding section is somewhat oversimplified. The fact is that, as we show forthwith, the sol-gel transition is always accompanied by a separation into two phases of different density. To analyze the phase diagrams of the system under consideration we take advantage of the generally known fact that the conditions for equilibrium between two one-component phases at an assigned temperature T, i.e., the equality between their pressures P and chemical potentials μ, can be represented in the form

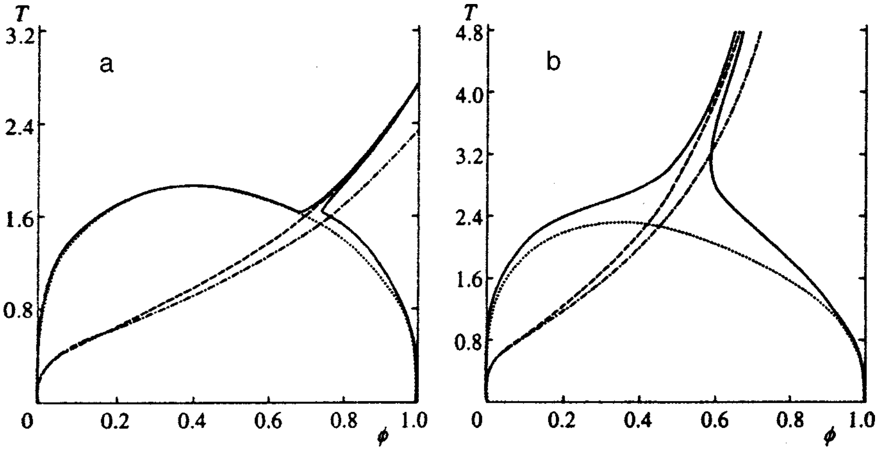

FIG. 4. Typical phase diagrams of trifunctional monomers in T vs. ϕ coordinates, where T is the temperature of the system and ϕ is the volume fraction of the polymer. The dotted lines are diagrams corresponding to the classical approximation. The dashed and dot-dashed lines are the sol-gel transition lines in our approximation and the classical approximation, respectively: a) phase diagram with a triple point ($g_0=-1$, $E=-1.4$); b) phase diagram without a triple point ($g_0=-0.5$, $E=-2$).

$$\left.\frac{\partial f}{\partial\rho}\right|_{\rho=\rho_1}=\left.\frac{\partial f}{\partial\rho}\right|_{\rho=\rho_2},$$
$$f(T,\rho_2)=f(T,\rho_1)+\left.\frac{\partial f}{\partial\rho}\right|_{\rho=\rho_1}(\rho_2-\rho_1) \tag{27}$$

for the presence of a common tangent to the specific free energy.

$$f(T,\rho)=F(V,T,N)/V. \tag{28}$$

In (27) ρ_1 and ρ_2 are used to denote the particle densities in the coexisting phases, which are equal to the abscissas of the points of contact of the plot of (28) with the common tangent.

The critical points on the phase diagram correspond to the vanishing of the common tangent

$$\frac{\partial^2 f}{\partial\rho^2}=0,\quad \frac{\partial^3 f}{\partial\rho^3}=0, \tag{29}$$

and the number of critical points determines the topology of the phase diagram of the system. However, in this paper we do not dwell on a detailed investigation (see, for example, Ref. 14) of the topology of the phase diagrams and a comparison of the phase portraits for the system under consideration, but merely present several typical diagrams for it.

For simplicity, we assume that the temperature dependence of the chemical equilibrium constant $k(T)$ is determined by an ordinary activation mechanism:

$$k=k_0\exp(-E/T), \tag{30}$$

where k_0 is the normalization factor, which has the dimensions of volume, and $E<0$ is the bond energy, which, like all quantities with the dimensions of energy, can be measured in units of the critical temperature T_c of the system of unconnected monomer units. Thus, the system under consideration is characterized by two dimensionless parameters, viz., $g_0=k_0/v$ and E, whose variation leads to different types of phase diagrams.

Typical phase diagrams constructed numerically in the approximation (14) proposed above are shown in Fig. 4, which takes into account the presence of a complex cyclized structure in the infinite cluster and leads to the expression (25) for the structural part of the free energy, and in the classical Flory approximation (7), which ignores this structure and leads to the expression (9). This figure also shows curves corresponding to the sol-gel transition in the new and classical approximations. One characteristic feature of the phase diagrams constructed in the approximation that we developed is the presence of a phase-separation ''corridor'' in the vicinity of the sol-gel transition. In some (but not all) cases there may be a triple point (Fig. 4a), at which three phases with differing density coexist, an infinite cluster of bonds being present in the densest of these phases (the behavior of the free energy of the system above the triple point is shown in Fig. 5).

If the bond energy $E>0$ (such a situation can be realized in a description of the competitive inhibition of chemical bonds between the polyfunctional A_n monomers under consideration due to the formation of bonds between the latter and monofunctional B_1 monomers in an A_n+B_1 system[14]), the phase diagrams of such a system become more diverse. Scrutinizing them, however, is beyond the scope of the present work.

5. CONCLUSION

Thus, we have shown that the classical theory of the sol-gel transition in thermally reversible weak gels based on

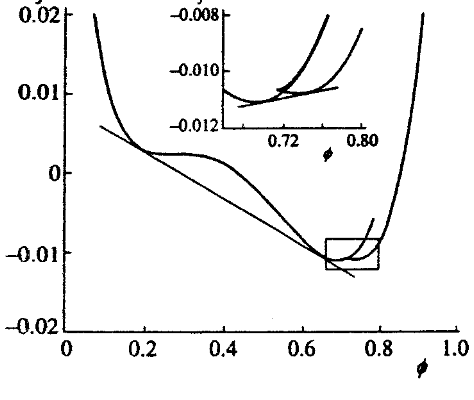

FIG. 5. Typical dependence of the free energy of the system on the volume fraction of the polymer at a temperature above the triple point on the phase diagram shown in Fig. 4a ($T=1.68$).

the Cayley-tree approximation remains valid only in the sol region, i.e, in the absence of an infinite cluster of thermally reversible bonds. When such a cluster (a gel fraction) is present, only the description developed above, which allows for the presence of a complex cyclized structure in the infinite cluster, is correct. Moreover, the formation of the infinite cluster, which is described in our formalism as the transition from the classical solution (7) to the solution (14) with spontaneous violation of the identity of the monomers, is a first-order transition, which is always accompanied by a separation into two phases.

This work was financially supported by INTAS and the Russian Fund for Fundamental Research (INTAS–RFBR Grant No 95-82).

---

*)E-mail: ierukhs@ierukh.phys.msu.su

[1] P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca (1979).

[2] A. Coniglio, H. E. Stanley, and W. Klein, Phys. Rev. Lett. **42**, 518 (1979).

[3] J. C. Wheeler and G. R. Anderson, J. Chem. Phys. **73**, 5778 (1980).

[4] J. S. Walker and C. A. Vause, J. Chem. Phys. **79**, 2660 (1983).

[5] B. A. Veitsman, J. Chem. Phys. J. Phys. Chem. **94**, 8499 (1990).

[6] C. Panayiotou and I. C. Sanchez, J. Chem. Phys. **95**, 10 090 (1991).

[7] A. N. Semenov and M. Rubinstein, Macromolecules **31**, 1373 (1998).

[8] I. Ya. Erukhimovich, Zh. Éksp. Teor. Fiz. **108**, 1004 (1995) [JETP **81**, 553 (1995)].

[9] I. Ya. Erukhimovich, in *Proceedings of the All-Union Conference ''Mathematical Methods for Investigating Polymers''* [in Russian], Akad. Nauk SSSR, Pushchino (1982), p. 52.

[10] S. I. Kuchanov, S. V. Korolev, and S. V. Panyukov, Adv. Chem. Phys. **72**, 115 (1988).

[11] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Vol. 1*, 3rd. ed., Pergamon Press, Oxford–New York (1980)

[12] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **55**, 2408 (1968) [Sov. Phys. JETP **28**, 1280 (1969)].

[13] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules*, American Institute of Physics, New York (1994).

[14] I. Ya. Erukhimovich, Doctoral Dissertation, Moscow State University, Moscow (1994).

[15] A. Coniglio, Phys. Rev. Lett. **46**, 250 (1981).

[16] A. P. Vinogradov and A. K. Sarychev, Zh. Éksp. Teor. Fiz. **85**, 1144 (1983) [Sov. Phys. JETP **58**, 665 (1983)].

[17] L. P. Lipatov, Zh. Éksp. Teor. Fiz. **71**, 2010 (1976) [Sov. Phys. JETP **44**, 1055 (1976)].

[18] P. J. Flory, J. Am. Chem. Soc. **63**, 3083, 3091, 3096 (1941).

[19] W. H. Stockmayer, J. Chem. Phys. **11**, 45 (1943).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Features of the spin fluctuations and superconductivity of $Tl_2Ba_2CaCu_2O_{8-\delta}$ according to ^{63}Cu and ^{17}O NMR data

A. P. Gerashchenko,*) K. N. Mikhalev, S. V. Verkhovskiĭ, Yu. V. Piskunov,
A. V. Anan'ev, and K. A. Okulova

Institute of the Physics of Metals, Russian Academy of Sciences, Ural Branch, 620219 Ekaterinburg, Russia

A. Yu. Yakubovskiĭ and L. D. Shustov

Kurchatov Institute, 123182 Moscow, Russia

(Submitted 18 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 991–1001 (March 1999)

Data on the NMR line shifts, the spin–lattice relaxation rate of ^{63}Cu and ^{17}O nuclei, and the spin–spin relaxation rate of ^{63}Cu are obtained for $Tl_2Ba_2CaCu_2O_{8-\delta}$ ($T_c=112$ K) in the normal and superconducting states. The hyperfine constants at the copper and oxygen atoms in a CuO_2 plane are estimated from an analysis of the temperature dependence of the Knight shift. The temperature-dependent behavior of the long- and short-wavelength parts of the dynamic spin susceptibility is discussed by modeling an antiferromagnetic Fermi liquid. The possible relation between the characteristics of the spin-fluctuation spectrum and the superconducting transition temperature is analyzed for the oxide investigated. [S1063-7761(99)01603-0]

1. INTRODUCTION

Studies of high-T_c superconductivity carried out during the past 10 years have revealed some special features of the electronic properties of high-T_c superconducting cuprates that are not observed in low-temperature superconductors. It was established that the superconducting transition temperature is achieved when there is a certain optimal hole concentration n_h^{opt} in the CuO_2 layers.[1] An analysis of the electric field gradient at the Cu and O atoms in copper layers performed for metal oxides[2–5] revealed a close relationship between T_c and both the total hole concentration in the copper planes and the relative populations of the valence orbitals of the Cu ($3d_{x^2-y^2}$) and O ($2p_\sigma$) atoms. A decrease in the population $n_{3d_{x^2-y^2}}$ of the $3d_{x^2-y^2}$ orbital with a simultaneous increase in n_{2p_σ} is accompanied by an increase in the critical temperature for lightly doped cuprates with a carrier concentration less than n_h^{opt}. The estimates of the single-particle density of states obtained from data on ^{89}Y NMR line shifts[6] does not provide a reasonable explanation for the high values of T_c in the context of the phonon mechanism of Cooper pairing.

Inelastic neutron scattering experiments and studies of the spin–lattice relaxation of ^{63}Cu and ^{17}O nuclei in the metallic phase of high-T_c superconducting oxides revealed the existence of strong antiferromagnetic ($q=Q_{\text{AF}}=\{\pi/a,\pi/a\}$) spin correlations between nearly localized electrons in the $3d_{x^2-y^2}$ orbitals of neighboring copper atoms. Consideration of the features of the spectrum of spin fluctuations in a layer is important in describing the ground state of the conduction band and in analyzing possible nonphonon channels of superconducting pairing. The possibility of carrier pairing in a CuO_2 layer due to the virtual exchange of antiferromagnetic paramagnons was discussed in Refs. 7–9 under the assumption that charge carriers experience attraction near the Fermi surface, which is effective in a layer having a thickness proportional to the characteristic spin-fluctuation energy $\Gamma_{Q_{\text{AF}}}$. In this case the preexponential factor in the BCS formula for T_c can be represented in the form of the product $\Gamma_{Q_{\text{AF}}}\xi^2(1-n_h)$, where ξ is the spin-fluctuation correlation length and n_h is the hole concentration in the CuO_2 layer.

Estimates of spin-fluctuation parameters based on data from NMR experiments have been obtained mainly by modeling an almost antiferromagnetic Fermi liquid. Measurements of the spin–lattice relaxation rate (T_1^{-1}) on ^{63}Cu and ^{17}O nuclei and the contribution of the indirect spin–spin interaction to the damping of the transverse nuclear magnetization of ^{63}Cu ($^{63}T_{2g}$) make it possible to study the behavior of the dynamic spin susceptibility at low frequencies. In Refs. 3 and 4 it was established from an analysis of data for T_1^{-1} in $YBa_2Cu_3O_{6.9}$ ($T_c=94$ K) and $Tl_2Ba_2Ca_2Cu_3O_{10}$ ($T_c=115$ K) that the increase in the superconducting transition temperature is accompanied by displacement of the spin-fluctuation spectrum toward higher energies, attesting to the spin-fluctuation nature of the superconductivity in these metal oxides.

The present work is devoted to an analysis of data on the NMR of ^{63}Cu and ^{17}O nuclei in a sample of $Tl_2Ba_2CaCu_2O_{8-\delta}$ ($T_c=112$ K) oriented by a magnetic field for the purpose of obtaining information on the variation of the characteristic energy, the correlation length, and other parameters of the spectrum of spin fluctuations in this compound with temperature.

2. EXPERIMENTAL METHOD

NMR measurements were performed on a single-phase ceramic sample of $Tl_2Ba_2CaCu_2O_{8-\delta}$ with $T_c = 112$ K, which will be abbreviated as Tl2212 below. A detailed description of the procedure for the synthesis and $^{17}O-^{16}O$ isotopic substitution was given in Ref. 10.

The temperature of the transition to the superconducting state was determined by observing the appearance of a diamagnetic response when the ac magnetic susceptibility was measured. For Tl2212 the value $T_c = 112$ K corresponds to the maximum on the $T_c(n_h)$ phase diagram with the optimal carrier concentration n_h^{opt} in a CuO_2 layer. The polycrystalline sample was mixed with an epoxy resin and oriented in a 9-T magnetic field.

The NMR investigations were carried out with a pulsed NMR spectrometer in the temperature range 10–300 K. The measurements for ^{63}Cu nuclei were carried out in a magnetic field $B_0 = 9$ T, and the measurements for ^{17}O nuclei were carried out at $B_0 = 8$ T. The method for recording the NMR spectra consisted of excitation of the solid-echo signal followed by complex Fourier transformation of the second half of the echo. To eliminate distortions of the spectra due to transients in the resonant circuit, a sequence that alternated $(0-180°)$ the phase of the first rf pulse was employed. When spectra with a width exceeding the frequency band excited by the rf pulse were recorded, we summed the array of Fourier signals accumulated at a set of uniformly spaced spectrometer frequencies. The ^{63}Cu NQR spectra were obtained in a similar manner. The quadrupole frequency ν_Q was determined from the peak of the NQR line.

The components $K_{\alpha\alpha}$ of the magnetic shift tensor of the ^{63}Cu and ^{17}O NMR lines were determined from the position of the peaks of the NMR lines for the $m = 1/2 \rightarrow -1/2$ transition with consideration of the quadrupole corrections $\nu_{ab,c}$ to the resonant frequency shift in second-order perturbation theory.

The line shifts were determined relative to the position ν_0 of the NMR lines of ^{63}Cu in metallic copper $[^{63}K(\mathrm{Cu_{met}}) = 0.23\%]$ and of ^{17}O in H_2O.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. NMR line shifts, hyperfine fields at ^{63}Cu and ^{17}O nuclei, and homogeneous contribution of the spin susceptibility χ_s $(q=0)$

3.1.1. ^{63}Cu *NMR line shift*

The temperature dependence of the magnetic shift ^{63}K (Fig. 1), which includes the orbital shift $^{63}K_{\mathrm{orb}}$ and the spin contribution $^{63}K_s(T)$, was obtained from the total shift of the ^{63}Cu $(I = 3/2)$ NMR line after subtraction of the quadrupole corrections calculated for the case of axial symmetry of the electric field gradient tensor:[11]

$$\nu_{ab} = \frac{1}{16}\left[I(I+1) - \frac{3}{4}\right]\frac{\nu_Q^2}{\nu_0}, \qquad \nu_c = 0. \tag{1}$$

To within the measurement error, the quadrupole frequency $\nu_Q = 17.35(20)$ MHz did not depend on temperature over the entire temperature range 10–300 K.

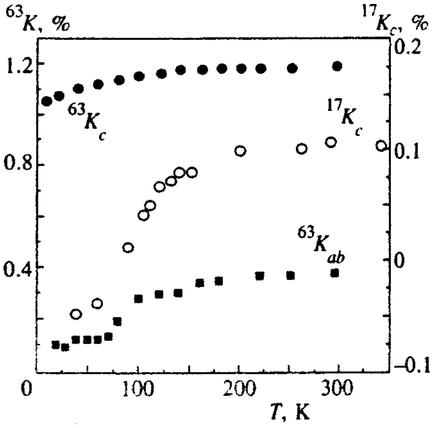

FIG. 1. Temperature dependence of magnetic shifts (a,b,c — principal crystallographic axes).

In the normal state region the components of the NMR shifts decrease monotonically as the temperature is reduced. There is a stronger temperature dependence for $^{63}K_{ab}$ than for $^{63}K_c$. The transition to the superconducting state is accompanied by a sharp decrease in the magnitude of the shift. Such behavior of the shift on Cu nuclei is typical of all superconducting oxides studied, and results from elimination of the spin contribution (the Knight shift), which is proportional to the homogeneous spin susceptibility.

An additional diamagnetic contribution K_{dia} to the NMR shift appears at temperatures below T_c due to the distribution of the magnetic fields within the sample created by the vortex structure of the magnetic field. According to our estimates for a 9-T magnetic field at 10 K, $K_{ab,\mathrm{dia}} < 0.005\%$, which is less than the measurement error in the magnitude of the shift. When the spin contribution $K_{ab,s}$ to the NMR line shifts was determined, we assumed that $K_{ab,s} = 0$ at $T = 10$ K. It follows from this assumption, in accordance with Fig. 1, that the orbital contribution $K_{ab,\mathrm{orb}} = 0.1\%$. Its value is assumed to remain unchanged over the entire temperature range of the normal and superconducting states. In this case the difference $K_{ab}(T = T_c) - K_{ab}(T = 10\mathrm{K})$ determines the spin contribution $K_{ab,s}$.

3.1.2. ^{17}O *NMR line shift*

Figure 2 presents the ^{17}O $(I = 5/2)$ NMR spectra of an oriented sample of Tl2212 recorded over a broad frequency range, which includes the NMR lines of all transitions. The spectra were obtained at 120 K for the case where the **c** axis of the crystallites is oriented parallel (a) and perpendicularly (b) to $\mathbf{B}_0$. They are similar in form to the spectra presented for magnetically oriented powders of thallium ceramics (Tl2201) with one CuO_2 plane.[12] We retained the same notation for the oxygen satellite lines corresponding to different sites as in the preceding paper on Tl2212 (Ref. 10). Our subsequent discussion of the ^{17}O NMR spectra pertains to lines that exhibit a strongly temperature-dependent positive shift and are assigned to oxygen atoms in CuO_2 layers (O1 sites).

To determine the quadrupole frequencies ν_Q and the asymmetry parameter η of the electric field gradient tensor,

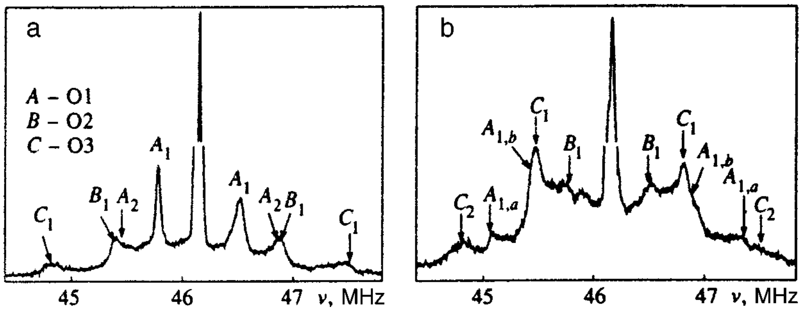

FIG. 2. ^{17}O NMR spectra at $T=120$ K for $\mathbf{B}_0\|\mathbf{c}$ (a) and $\mathbf{B}_0\perp\mathbf{c}$ (b).

we used the positions of the features of the satellite lines of the $\pm 3/2 \leftrightarrow \pm 1/2$ transitions for various orientations of the external magnetic field $\mathbf{B}_0$ relative to the **c** axis of the crystallites. For example, when $\mathbf{B}_0\|\mathbf{c}$, peak A_1 of a site for which the principal axis of the electric field gradient tensor lies in the ab plane should correspond to the frequency ν_1, which is specified by the expression[11]

$$\nu_1=\nu_0(1+K_y)-\frac{\nu_Q}{2}(1+\eta)+\frac{5\nu_Q^2}{16\nu_0}\left[1+\frac{2}{3}\eta+\frac{1}{9}\eta^2\right]. \tag{2}$$

When $\mathbf{B}_0\perp\mathbf{c}$, peak A_{1b} appears for crystallites with a principal axis of the electric field gradient tensor perpendicular to $\mathbf{B}_0$ at the frequency

$$\nu_2=\nu_0(1+K_x)+\frac{\nu_Q}{2}(1-\eta)+\frac{5\nu_Q^2}{16\nu_0}\left[1-\frac{2}{3}\eta+\frac{1}{9}\eta^2\right], \tag{3}$$

and peak A_{1a} appears for crystallites with a principal axis of the electric field gradient tensor parallel to $\mathbf{B}_0$ at

$$\nu_3=\nu_0(1+K_z)-\nu_Q+\frac{5\nu_Q^2}{36\nu_0}\eta^2. \tag{4}$$

It was assumed in the analysis of the NMR line shift ^{17}K that the symmetry of the tensor is nearly axial: $^{17}K_x={}^{17}K_y$. For the compound investigated we obtained $^{17}\nu_Q=1.09$ MHz and $\eta=0.33$, which do not depend on temperature. The parameters of the electric field gradient tensor are close to those given for the sites of O atoms in Y123, Tl2201, and Tl2212. The components of the magnetic line shift tensor ^{17}K were determined from the position of the peak of the NMR line of the $1/2 \leftrightarrow -1/2$ transition with allowance for the quadrupole correction ν_c to the shift of the resonant frequency ν_0 (Ref. 11) of oxygen atoms in Cu–O–Cu chains along the **c** axis of the crystal:

$$\nu=\frac{(3+\eta)^2}{144}\left[I(I+1)-\frac{3}{4}\right]\frac{\nu_Q^2}{\nu_0}. \tag{5}$$

Figure 1 presents the temperature dependence of the component $^{17}K_c$ of the magnetic NMR line shift tensor corresponding to the O1 site, which decreases monotonically as the temperature is reduced with an increase in the rate of variation of the shift as T_c is approached.

3.1.3. Magnetic hyperfine coupling constants and orbital shift

The hyperfine coupling constants at copper and oxygen atoms were determined from plots of the parametric functions $^{63}K_c(^{63}K_{ab})$ and $^{17}K_c(^{63}K_{ab})$, which were constructed using data for the normal state above 100 K in order to eliminate the influence of the diamagnetic contribution, which begins to increase significantly at temperatures below T_c. In accordance with the Mila–Rice Hamiltonian,[13] which was proposed to describe the magnetic hyperfine interactions of atoms in a CuO_2 layer, the Knight shift at copper and oxygen atoms is proportional to the homogeneous spin susceptibility $\chi_s(q=0)\equiv\chi_0$ of the layer:

$$^{63}K_{ab,c}=(A_{ab,c}+4B)\chi_0, \tag{6}$$

$$^{17}K_c=2C_c\chi_0. \tag{7}$$

The anisotropy constant $A_{\alpha\alpha}$ takes into account the total contribution of the magnetic hyperfine interaction of the spin of the ^{63}Cu nucleus with the nearly localized electron spin of the $3d_{x^2-y^2}$ orbital. The contribution of the indirect Cu–O–Cu interaction from the four nearest neighbors of each Cu atom is taken into account by the hyperfine constant *B*, which was assumed to be an isotropic quantity in Ref. 13. Finally, the covalency of the $\mathrm{Cu}_{3d_{x^2-y^2}}$ and O_{2p_σ} orbitals is taken into account by the constant *C*, whose value depends on the direction of the magnetic field relative to the crystallographic axes. For the ensuing analysis we assume that the values of $A_{\alpha\alpha}$ in the compound investigated are the same as in $YBa_2Cu_3O_7$ (Ref. 13), and equal $A_{ab}=37\,\mathrm{kOe}/\mu_B$ and $A_c=-165\,\mathrm{kOe}/\mu_B$. The values $B=71\,\mathrm{kOe}/\mu_B$ and $C_c=73\,\mathrm{kOe}/\mu_B$ were obtained from the slopes of the $^{63}K_c(^{63}K_{ab})$ and $^{17}K_c(^{63}K_{ab})$ curves in accordance with (6) and (7). The estimates of the hyperfine constants coincide with the data[4] for $Tl_2Ba_2Ca_2Cu_3O_{10}$ (Tl2223 below) and somewhat exceed the values given for the isostructural compound Bi2212 in Ref. 14.

Taking into account that $^{63}K_{ab,\mathrm{orb}}=0.1\%$, we find $^{63}K_{c,\mathrm{orb}}=1.08\%$ and $^{17}K_{c,\mathrm{orb}}=-0.02\%$ from the parametric dependences. The magnitude of the orbital shift $^{63}K_{ab,\mathrm{orb}}$ is determined by the van Vleck susceptibility of the Cu atoms. The decrease in $^{63}K_{ab,\mathrm{orb}}$ compared to the value for Y123, $^{63}K_{ab,\mathrm{orb}}=1.25\%$, might suggest an additional downward displacement of the energies of the filled d_{xy}, d_{xz}, and d_{yz} orbitals relative to E_F in $Tl_2Ba_2CaCu_2O_{8-\delta}$.

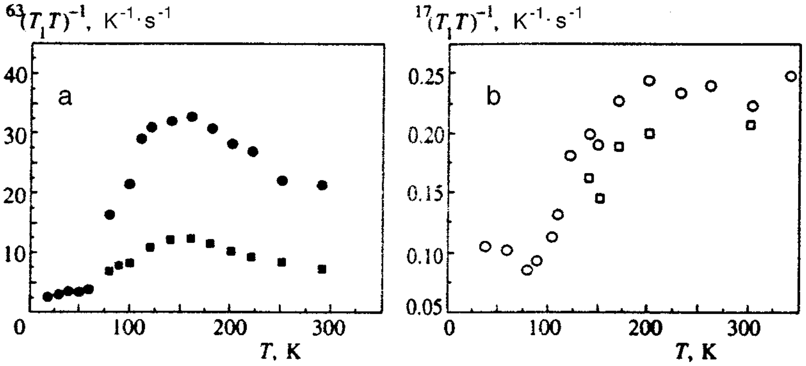

FIG. 3. Temperature dependence of ${}^{63}(T_1T)^{-1}$ (a) and ${}^{17}(T_1T)^{-1}$ (b) for $\mathbf{B}_0\perp\mathbf{c}$ (squares) and $\mathbf{B}_0\|\mathbf{c}$ (circles).

3.1.4. Principal value of the electric field gradient tensor for ^{63}Cu *and* ^{17}O; *population of the orbitals of the* Cu *and* O *atoms in Tl2212*

The quadrupole frequencies ν_Q for oxygen and copper in the compound under investigation, ${}^{63}\nu_Q=17.3(2)$ MHz and ${}^{17}\nu_Q=1.09(5)$ MHz, are close to the values for Tl2223 ($T_c=115$ K), ${}^{63}\nu_Q=17.3$ MHz and ${}^{17}\nu_Q=1.15$ MHz, respectively.[2]

It was shown in Ref. 15 that the electric field gradient $V_{\alpha\alpha}$ at the copper and oxygen atoms in CuO$_2$ layers is produced mainly by electrons in partially filled orbitals of atoms having aspherical symmetry. In this case the quadrupole frequency

$$^{63}\nu_Q=\frac{3e^2Q}{2I(2I-1)}V_{cc}$$

is proportional to the population of the copper $3d_{x^2-y^2}$ orbital, and the change in the quadrupole frequency of the ^{17}O nuclei is determined by the population increment of the O_{2p_σ} orbital: $\Delta\nu_Q\sim\Delta n_{2p_\sigma}$. Zheng *et al.*[2] analyzed the charge distribution in several superconducting cuprates within this scheme. They found that an increase in the superconducting transition temperature is accompanied by a decrease in the population of the copper $3_{d_{x^2-y^2}}$ orbital and an increase in n_{2p_σ}. The experimental values of ν_Q obtained for Tl2212 confirm this empirical rule: the equal values of the quadrupole frequency ${}^{63}\nu_Q$ in Tl2212 and Tl2223 correspond to similar values of T_c, and the larger value of ν_Q for oxygen in Tl2223 (or the larger hole concentration n_{2p_σ}) corresponds to a higher value of T_c. An increase in the population of the oxygen orbitals may signify an increase in the weight of the less localized O_{2p_σ} states in the wave function describing the state of the conduction-band hole carriers along the Y123–Tl2212–Tl2223 series.

3.2. Magnetic relaxation of ^{63}Cu and ^{17}O nuclei and characteristics of the spin-fluctuation spectrum in a CuO$_2$ layer

3.2.1. Spin–lattice relaxation rate of ^{63}Cu *and* ^{17}O *nuclei*

Figure 3 presents the temperature dependence of ${}^{63}(T_1T)^{-1}$ and ${}^{17}(T_1T)^{-1}$ for the orientations of the external magnetic field $\mathbf{B}_0\|\mathbf{c}$ and $\mathbf{B}_0\perp\mathbf{c}$.

When T_1 was measured for ^{63}Cu, the variation of the total intensity of the NMR spectrum $y(t)$ measured within the linewidth was recorded. The experimental file was treated by the least-squares method using the formula

$$y(t)=A+B\exp(-t/T_1)+C_2\exp(-6t/T_1), \qquad (8)$$

where A, B, C, and T_1 are variable parameters.

The temperature dependences of ${}^{63}(T_1T)^{-1}$ for $\mathbf{B}_0\|\mathbf{c}$ and $\mathbf{B}\perp\mathbf{c}$ have a characteristic maximum at $T=150$ K.

The spin–lattice relaxation time T_1 of ^{17}O nuclei ($I=5/2$) was measured using a method that equalizes the populations of energy levels with different values of the magnetic quantum number m. A sample immersed in a magnetic field is subjected to a series of radio-frequency pulses, whose filling frequency varies within the total width of the spectrum according to a definite periodic law, and the amplitude of the spin echo is measured after a time t. It follows from the solution of the system of relaxation equations for the populations of the levels that when the populations of nonequidistant energy levels are equalized, the magnetization $M_z(t)$ obeys a simple exponential function of time. This method significantly reduced the error in the determination of ${}^{17}T_1$. To separate the contributions to the resultant echo signal of ^{17}O atoms in different layers when T_1 was measured, the variation of the intensity of the absorption signal corresponding to the NMR line of the O1 sites in the CuO$_2$ layers was recorded. The experimental results were processed using Eq. (8). The coefficient C of the ‘‘slow’’ exponential function reached a value of 0.8, suggesting a nearly single-exponential functional dependence of the magnetization restoration function.

The temperature dependence of ${}^{17}(T_1T)^{-1}$ for $\mathbf{B}_0\|\mathbf{c}$ decreases monotonically in the normal state with a further sharp drop in the superconducting region. To elucidate the anisotropy of the spin–lattice relaxation rate, measurements were carried out with $\mathbf{B}_0\perp\mathbf{c}$. It was found that the anisotropy constant ${}^{17}r=(T_1)_{ab}/(T_1)_c=1.3$.

3.2.2. Spin–spin relaxation rate of copper nuclei

Figure 4 presents the temperature dependence of ${}^{63}T_{2g}^{-1}$ for $\mathbf{B}_0\|\mathbf{c}$.

When T_{2g} was measured for ^{63}Cu, the variation of the intensity of the NMR line was also recorded. The experimental array was processed with the least-squares method using the formula

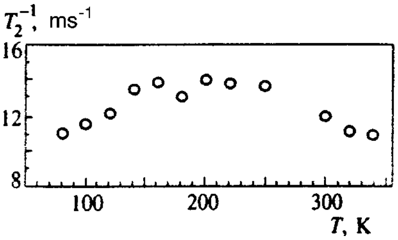

FIG. 4. Temperature dependence of $^{63}T_{2g}^{-1}$ for $\mathbf{B}_0 \| \mathbf{c}$.

$$y(t)=A \exp\left\{-0.5\left(\frac{t}{T_{2g}}\right)^2-\frac{t}{T_{2L}}\right\}, \tag{9}$$

where t is the time between the first pulse and the spin-echo signal and

$$1/T_{2L}=3(1/T_1)_c+(1/T_1)_{ab}. \tag{10}$$

The Gaussian contribution T_{2g}^{-1} to the spin–spin relaxation rate for $\mathbf{B}_0 \| \mathbf{c}$ increases slowly with decreasing temperature, peaks at $T=160$ K, and then decreases. The temperature resulting dependence is similar to the one presented for Tl2223 ($T_c=115$ K) in Ref. 4.

3.2.3. Estimates of the parameters of the spin-fluctuation spectrum in the model of an antiferromagnetic Fermi liquid

Millis, Monien, and Pines[16] proposed a phenomenological model of an antiferromagnetic Fermi liquid to describe the spin correlations in a CuO_2 layer. In this model the values of the dynamic spin susceptibility $\chi_s(q,\omega)$ at the Brillouin-zone center ($q=0$) and on its boundary ($q=Q_{AF}$) are related to one another. This allowed us to discuss NMR data for different atoms within a single spin degree of freedom. The corresponding imaginary part of the susceptibility in the low-frequency limit can be written in the form

$$\chi''(q,\omega\to 0)=\frac{\pi\chi_0\omega}{\Gamma_0}\frac{1+\beta(\xi/a)^4}{[1+\xi^2(q-Q_{AF})^2]^2}. \tag{11}$$

Here ξ is the magnetic correlation length, Γ_0 is the characteristic spin-fluctuation energy for $q=0$, a is the lattice parameter, and β is a parameter that takes into account the increase in fluctuation intensity for $q=Q_{AF}=\{\pi,\pi\}$ relative to the value at the Brillouin-zone center. The expression (11) was obtained using the following coupling conditions for the values of the spin-fluctuation parameters for $q=0$ and $q=Q_{AF}$:

$$\chi_Q=\chi_s\beta^{1/2}(\xi/a)^2, \tag{12}$$

$$\Gamma_Q=(\Gamma_0/\beta^{1/2})(\xi/a)^2\pi. \tag{13}$$

We obtained the anisotropy constant of the spin–lattice relaxation rate for the compound in question: $^{63}r={}^{63}(1/T_1)_{ab}/{}^{63}(1/T_1)_c$. This ratio is approximately constant over the entire temperature range and equals $^{63}r=2.65(0.3)$. The value obtained is greater than the value $^{63}r\approx 1.8$ for Tl2223 (Refs. 3 and 4) and less than the value $^{63}r=3.7$ for Y123 (Refs. 17 and 18). In the limit $\xi \gg a$, ^{63}r is determined by a combination of hyperfine fields.[19]

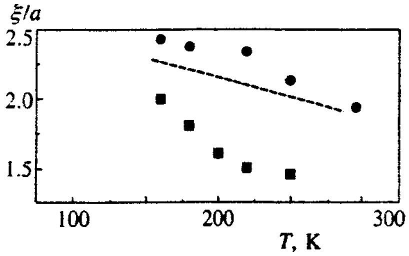

FIG. 5. Temperature dependence of ξ/a for Tl2212 (circles), Tl2223 (squares), and Y123 (dashed line).

The temperature-independent behavior of ^{63}r may attest to the applicability of the approximation of large values of the correlation length ξ.

As was shown in Ref. 20, the Gaussian component $^{63}T_{2g}^{-1}$ of the spin–spin relaxation rate at copper for $\mathbf{B}_0 \| \mathbf{c}$ contains information on the wave-vector-dependent real part of the spin susceptibility. For $\xi > a$ the contribution of the fluctuations for $q=Q_{AF}$ to T_{2g}^{-1} is dominant:[19]

$$\left(\frac{1}{T_{2g}}\right)^2=\frac{0.69({}^{69}\gamma\eta)^4(Ac-4B)^4}{32\pi\eta k_B}\beta\chi_0^2\left(\frac{\xi}{a}\right)^2 \tag{14}$$

(γ is the gyromagnetic ratio). It is convenient to write (14) in a different form:

$$\left(\frac{1}{T_{2g}}\right)^2=\frac{0.69({}^{69}\gamma\eta)^2(Ac-4B)^2}{16\pi\eta k_B}\times\left[\frac{2}{{}^{63}(T_1T)_{ab}}-\frac{1}{{}^{63}(T_1T)_c}\right]\chi_0\eta\Gamma_0. \tag{15}$$

Using (15) and the values of the homogeneous spin susceptibility χ_0, we obtain estimates of the characteristic spin-fluctuation energy $\eta\Gamma_0(T)$, which significantly exceeds the value $\eta\Gamma_0=1.2$ eV for $YBa_2Cu_3O_7$ (Ref. 21) and is comparable to the data for Tl2223 at temperature above 180 K. The quantity $\eta\Gamma_0$ remains roughly constant over a broad temperature range ($\eta\Gamma_0=2.91$ eV). Similar temperature-independent behavior was observed for $YBa_2Cu_3O_7$. Unlike the latter compound, Tl2223 exhibited a significant decrease in $\eta\Gamma_0$ with temperature in the normal-state region. The significant increase in $\eta\Gamma_0$ attests to displacement of the fluctuation spectrum toward higher frequencies in Tl2212, which has a higher value of T_c compared with Y123.

Determination of the correlation length using Eq. (14) requires knowledge of β, which takes into account the increase in the fluctuation intensity for $q=Q_{AF}=\{\pi,\pi\}$ relative to its value at the Brillouin-zone center. In particular, in the previous studies β was assumed to be equal to $\pi^2\approx 10$, mainly for YBaCuO. The value $\beta=60$ was given in Refs. 3 and 4. In the present work we took $\beta=0$, which correspond to the value for $Tl_2Ba_2Ca_2Cu_3O_{10-d}$, which is structurally closely related and has similar superconducting properties.

The values of the correlation length of the antiferromagnetic spin fluctuations are presented in Fig. 5 in units of a, which is equal to the distance between neighboring Cu atoms. The temperature dependence of ξ/a for Tl2212 is very close to the analogous dependence for $YBa_2Cu_3O_7$.[21] Interestingly enough, the value of ξ/a for $Tl_2Ba_2Ca_2Cu_3O_{10}$ near T_c approximates the value obtained for $Tl_2Ba_2CaCu_2O_8$ in the present work.

Monthoux and Pines[8] showed that if Cooper pairing occurs as a result of the virtual exchange of antiferromagnetic paramagnons, the interaction between the electrons in a pair will be proportional to the spin-fluctuation energy. In this case the following expression was proposed for the superconducting transition temperature:

$$T_c=\frac{\Gamma_0}{\pi\beta^{0.5}}0.79(1-n_h)\exp\left(-\frac{1}{\lambda}\right). \tag{16}$$

Here λ is the dimensionless coupling constant (for the system under consideration $0.42<\lambda<0.48$), and n_h is the hole concentration in a layer. The factor $\Gamma_0/\pi\beta^{0.5}$ preceding the exponential in (16) for the compound investigated is 1.5 times the analogous value for $YBa_2Cu_3O_7$ at $T=220$ K. This is fully consistent with the higher value of T_c for Tl2212 when the possible differences in the other quantities (λ and n_h) are taken into account. In our opinion, the estimates of the parameters of the resulting spin-fluctuation spectrum support the nonphonon superconductivity mechanism discussed for YBaCuO in Ref. 8.

In metals with a broad conduction band the relaxation rate is proportional to the temperature. This is because in the limit of short electron-motion correlation times, $\omega_0\tau_e\ll 1$, the density of states of the spin excitations at the NMR frequencies ω_0 is constant, and their number is proportional to the temperature. The spatial dispersions of the spin susceptibility $\chi(q)$ and the quasiparticle damping Γ_q are weak. Their values are completely determined by the density of states at the Fermi level $N(E_F)$. In the paramagnetic state $\chi''(q,\omega)/\omega$ exhibits a Lorentzian frequency dependence:

$$\chi(q)\approx\chi(q=0)\equiv\chi_0=2\mu_B^2N(E_F),$$

$$\Gamma_q\approx\Gamma\propto N^{-1}(E_F).$$

In this case the expression for the relaxation rate takes the form

$$^{17}W=\gamma^2\eta k_BTC^2\sum_q\frac{\chi''(q,\omega_0)}{\omega_0}\approx\gamma^2k_BTC^2 \times\sum_q\frac{\pi\chi(q)}{\Gamma_q}\approx 2\gamma^2\hbar k_BTC^2N^2(E_F). \tag{17}$$

When relaxation processes are analyzed in normal metals, K_s is usually estimated using the Korringa relation:

$$K_s^2TT_1=\frac{4\pi\mu_B^2}{\hbar\gamma^2k_B}\equiv\text{const.} \tag{18}$$

The phenomenological model of a nearly antiferromagnetic Fermi liquid leads to a relation between the spin–lattice relaxation time T_1 and the Knight shift K_s that differs from the Korringa law:

$$K_sT_1T=\text{const.} \tag{19}$$

The experimental data on the Knight shift and the spin–lattice relaxation rate are closely approximated by the relation

$$^{17}K_s^{0.95\pm 0.2}TT_1=\text{const.}$$

This result enables us to conclude that it is better to analyze electron excitations in the copper layers of $Tl_2Ba_2CaCu_2O_{8-\delta}$ in a model that allows for the presence of strong spin correlations between copper atoms in the CuO_2 planes.

4. CONCLUSION

^{63}Cu and ^{17}O NMR data for the normal and superconducting states of the oriented cuprate $Tl_2Ba_2CaCu_2O_{8-\delta}$ have been presented in this paper. Information on charge redistribution in the copper planes in comparison to other high-T_c superconducting oxides have been obtained from an analysis of the ^{63}Cu and ^{17}O quadrupole frequencies. The data confirm the correlation detected between the increase in superconducting transition temperature in high-T_c superconductors and enhancement of the covalent copper–oxygen bonding.

An analysis of the spin–spin and spin–lattice relaxation rates in the model proposed by Millis, Monien, and Pines has revealed a significant increase in the characteristic spin-fluctuation energy in comparison to Y123, while the magnetic correlation length does not vary significantly as the superconducting transition temperature increases. Thus, according to the NMR data, the increase in T_c in $Tl_2Ba_2CaCu_2O_{8-\delta}$ is accompanied by an increase in the spin-fluctuation energy, which is a compelling argument favoring a superconductivity mechanism associated directly with spin fluctuations in the copper layers.

This research was carried out under the auspices of State Programs of the Russian Federation for basic research in the area of the physics of the condensed state (Superconductivity Subdivision, Project No. 961223) and for supporting leading scientific schools (Project No. 96-15-96515).

*)E-mail: Gerashenko@ifm.ural.ru

---

[1] C. Berthier, M. Horvatic, P. Carretta *et al.*, Physica C **235**, 67 (1994).
[2] G. Zheng, Y. Kitaoka, K. Ishida *et al.*, J. Phys. Soc. Jpn. **64**, 2524 (1995).
[3] G. Zheng, Y. Kitaoka, K. Asayama *et al.*, J. Phys. Soc. Jpn. **64**, 3184 (1995).
[4] G. Zheng, Y. Kitaoka, K. Asayama *et al.*, Physica C **260**, 197 (1996).
[5] K. Magishi, Y. Kitaoka, G. Zheng *et al.*, J. Phys. Soc. Jpn. **64**, 4561 (1995).
[6] H. Alloul, J. Appl. Phys. **69**, 247 (1991).
[7] P. Monthoux, A. V. Balatsky, and D. Pines, Phys. Rev. Lett. **67**, 3488 (1991).
[8] P. Monthoux and D. Pines, Phys. Rev. B **49**, 4261 (1994).
[9] T. Moriya, Y. Takahashi, and K. Ueda, J. Phys. Soc. Jpn. **59**, 290 (1990).
[10] A. Trokiner, K. Mikhalev, A. Yakubovskii *et al.*, Physica C **252**, 204 (1995).
[11] R. B. Greel and S. L. Segel, J. Chem. Phys. **60**, 2310 (1972).
[12] S. Kambe, H. Yasuoka, A. Hayashi, and Y. Ueda, Technical Report of ISSP, Ser. A, No. 2551 (1992).
[13] F. Mila and T. M. Rice, Physica C **157**, 561 (1989).
[14] K. Ishida, Y. Kitaoka, K. Asayama *et al.*, J. Phys. Soc. Jpn. **63**, 1104 (1994).
[15] K. Schwarz, C. Ambrosch-Draxl, and P. Blaha, Phys. Rev. B **42**, 2051 (1990).

[16] A. J. Millis, H. Monien, and D. Pines, Phys. Rev. B **42**, 167 (1990).

[17] R. E. Walstedt, W. W. Warren, R. F. Bell, and G. P. Espinosa, Phys. Rev. B **40**, 2572 (1989).

[18] S. E. Barrett, J. A. Martindale, D. J. Durand *et al.*, Phys. Rev. Lett. **66**, 108 (1991).

[19] V. Barzykin, D. Pines, A. Sokol, and D. Thelen, Phys. Rev. B **49**, 1554 (1994).

[20] S. H. Pennington and C. P. Slichter, Phys. Rev. Lett. **66**, 381 (1991).

[21] D. Thelen and D. Pines, Phys. Rev. B **49**, 3528 (1994).

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Spin effects in the magneto-oscillations of the capacitance of a two-dimensional gas of semiconductors with Kane and Dirac spectra

V. F. Radantsev*)

Ural State University, 620083 Ekaterinburg, Russia

(Submitted 1 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1002–1015 (March 1999)

The results of experimental investigations and computer modeling of the magneto-oscillations of the capacitance of surface layers with a two-dimensional gas in the narrow-gap semiconductor HgCdTe with direct and inverted band structure are compared. The structure of the Landau levels is calculated by a model that has a clear physical interpretation and is based on the reduction of matrix equations to Schrödinger-like equations with an effective potential in which the terms responsible for nonparabolic and spinor-type effects are easily separable. An analytic approach is developed for describing magneto-oscillation phenomena in the two-dimensional gas of materials with a quasirelativistic spectrum, and the emergence is noted of new theoretical parameters (compared to the parameters of materials with parabolic bands). Finally, the parameters of level broadening in spin–orbit split subsubbands are determined and the dominant scattering mechanisms are discussed. © *1999 American Institute of Physics.* [S1063-7761(99)01703-5]

1. INTRODUCTION

The lifting of the spin degeneracy of the spectrum of a two-dimensional gas by the electrostatic potential of asymmetric quantum wells due to spin–orbit coupling is a much-discussed problem.[1–12] This effect, which is actually a relativistic effect, is the largest in narrow-gap semiconductors. The most reliable experimental data on the splitting parameters have been obtained by studying magneto-oscillation effects, with the analysis usually based on measurements of the populations of spin-split subsubbands taken from oscillation periods or their Fourier spectra. In this approach all data on the position and amplitude of individual oscillations are ignored, which means, however, that the information about the exact energy position and broadening of Landau levels is lost. Here the commonly used assumption about the semiclassical nature of the magnetic quantization of the two dimensional spectra does not correspond to the Hamiltonian of a system with strong spin–orbit coupling and does not agree with the experimental data, which, in particular, is an indication that for small Landau-level numbers there is an appreciable nonperiodicity of the oscillations in the reciprocal magnetic field. Along with the ambiguity in determining the Fourier frequencies (these strongly and nonmonotonically depend on the range of field strengths being analyzed) this leads to sizable errors in determining the splitting parameters, thus reducing confidence in the faithfulness of the models used.

Methods based on computer simulations of magneto-oscillation effects are more productive. This approach, however, is fraught with difficulties when one deals with materials with a Kane spectrum, difficulties associated with the complexity of self-consistent calculations of the spectrum in a magnetic field that systematically allow for spinor-type effects and with the insufficient development of the theory of magneto-oscillation phenomena in the two-dimensional electron gas of semiconductors with a quasirelativistic spectrum. The faithfulness of the approaches often used in the study of such materials, approaches actually based on the parabolic approximation, at least requires substantiation.

This paper reports on studies of the spin–orbit coupling-related features of the energy spectrum in a magnetic field and of the magneto-oscillation effects in surface quantum wells in Kane semiconductors, and the manifestations of these features in materials with direct and inverted band structure. In addition to the experimental studies of inverted and enriched HgCdTe layers done by the magneto-oscillation capacitive spectroscopy method, a theory for calculating the Landau levels and describing capacitance magneto-oscillation effects is developed (this theory takes into account spinor-type effects). In light of the computer simulation problem solved in this study, one more advantage of capacitive methods comes to the fore, in addition to the fact that such methods are not limited by the magnitude and size of the gap. Being actually the measure of the density of two-dimensional states, the differential capacitance makes it possible, by measuring its absolute value, to extract information not only about the structure of Landau levels but also about the level-smearing parameters and hence about the dominant scattering mechanisms, including the features of such mechanisms in systems with strong spin–orbit coupling. This is all the more important because the standard methods of determining the broadening parameters for narrow-gap materials are often inapplicable due to the irregular nature of the oscillations.

2. BASIC EXPERIMENTAL RESULTS

We measured the magneto-oscillations of the capacitance of $Hg_{1-x}Cd_xTe$-based MOS structures with a positive

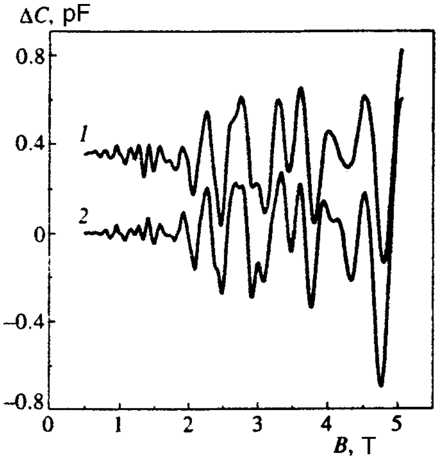

FIG. 1. Experimental (*1*) and computer-simulated (*2*) capacitance magneto-oscillations for the sample S1 at $V_g=2$ V ($N_s=3.1\times10^{12}\,\mathrm{cm}^{-2}$). The capacitance of the oxide in the metal oxide semiconductor structure, C_{oxide}, is 148 pF, and the area S is $9\times10^{-4}\,\mathrm{cm}^2$. For the sake of clarity, curve *1* is shifted upward. The following values of T_D were used in the calculations: 13.5 K for $i=0$, 10 K for $i=1$, and 9 K for $i=2$.

(sample S1, $E_g=+80$ meV, $N_A-N_D=1.8\times10^{16}\,\mathrm{cm}^{-3}$) and negative (sample P1, $E_g=-50$ meV, $N_A-N_D=3\times10^{15}\,\mathrm{cm}^{-3}$) Kane gap E_g with a 700–1000-Å thick anodic oxide acting as the gate dielectric, as functions of the magnetic field strength B and the voltage V_g on the field electrode. All measurements were made at $T=4.2$ K. Examples of the magnetic-field dependence of the oscillations of C and dC/dV_g are shown in Figs. 1 and 2a, while the C vs. V_g dependence at $B=4$ T is depicted in Fig. 2b. Beats are clearly visible in the oscillations, while the splitting of subband lines into doublets is visible in the Fourier spectra, which points to the splitting of subband Fermi surfaces due to spin–orbit coupling.

Despite the fact that the splitting and especially the ratio I_i^-/I_i^+ of the intensities of the Fourier lines corresponding to the lower and upper spin subbands (i is the number of the two-dimensional subband) strongly depend on the range of field strengths being used (for instance, it is clear that a Fourier analysis will reveal no splitting for the range between beat nodes), several characteristic features can be detected. While for low surface concentrations N_s the values of the intensities I_i^+ and I_i^- are close ($I_i^+>I_i^-$ for excited subbands near their edges), as N_s increases the ratio I_i^-/I_i^+ decreases significantly, especially for the ground subband, $i=0$. At extremely high subband concentrations, $N_i>(2–3)\times10^{12}\,\mathrm{cm}^{-2}$, only one Fourier line exists, and this line corresponds, as the analysis of the $N_i^{\pm}$ vs. V_g dependence shows, not to the average subband concentration $(N_i^+ + N_i^-)/2$ (which for regions of N_i with clearly detectable spin–orbit splitting can be determined from the Fourier transform in the range of fields between beat nodes) but to the concentration in the upper spin subsubband, N_i^+.

For HgCdTe with $E_g>0$, spin–orbit splitting in the ground subband does not manifest itself at low concentrations N_i either. However, in this range of concentrations, double-frequency Fourier lines are detectable. This phenomenon is due to the resolution in the oscillations of individual spin components for small Landau-level numbers ($n\leqslant3$) and manifests itself in the oscillations of the C vs. V_g dependence (Fig. 2b). In samples with $E_g<0$, there is not a single regime in which individual spin components manifest themselves in the C vs. V_g and C vs. B oscillations, even for the ground Landau level.

The subband and subsubband concentrations agree, within experimental error, with the values of the Fermi quasimomenta calculated by the method used in Ref. 8. As N_s increases, the relative splittings $\Delta N_i/N_i=(N_i^- - N_i^+)\times(N_i^+ + N_i^-)^{-1}$ tend, within experimental error ($\approx$20%), to their ''ultrarelativistic'' limits of approximately 0.1 for HgCdTe with $E_g>0$ and about 0.18 for HgCdTe with inverted bands (for different subbands these limits are essentially the same).[8]

3. THEORETICAL MODEL

The modeling of capacitance magneto-oscillations requires solving three relatively independent problems: (1) a calculation of the energy position of Landau levels in a self-

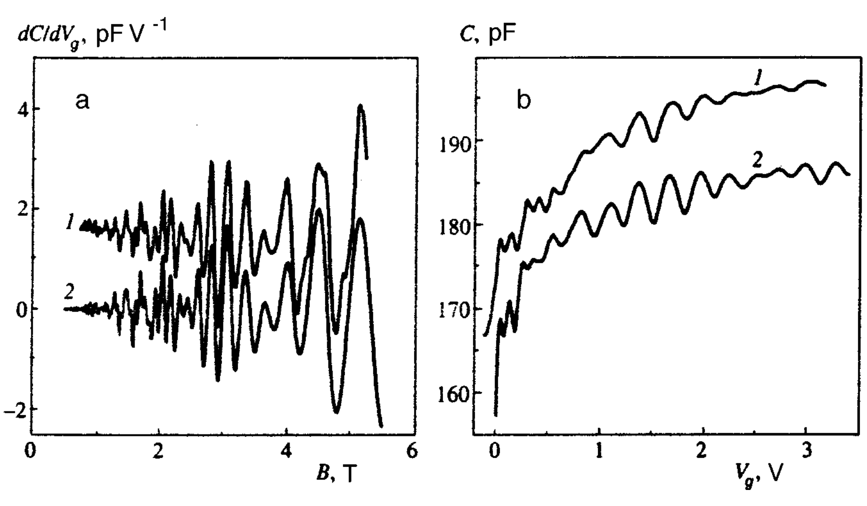

FIG. 2. Experimental (*1*) and computer-simulated (*2*) (a) magneto-oscillations of dC/dV_g at $V_g=8$ V (with $N_1=1.65\times10^{12}\,\mathrm{cm}^{-2}$ and $N_2=4.9\times10^{11}\,\mathrm{cm}^{-2}$) and (b) the C vs V_g curves for a magnetic field $B=4$ T for the sample P1. The capacitance of the oxide, C_{oxide}, is 195 pF, and the area S is $12\times10^{-4}\,\mathrm{cm}^2$. The following values of T_D were used in the calculations: 9 K for $i=1$ and $T_D=6$ K for $i=2$. The experimental C vs. V_g curve is shifted upward by 20 pF.

consistent surface quantum well, (2) a calculation of the density of states based on a model of magnetic-level broadening, and (3) a calculation of the differential capacitance of the space charge region with the spectrum quantized by electric and magnetic fields.

3.1. Spectrum in a magnetic field

The complexity of a theoretical description of the subband spectra in the surface layers of Kane semiconductors is due not only to the matrix nature of the Hamiltonian and the fact that the boundary conditions imposed on the spinor components are ambiguous[13,14] but also to the complexity of the self-consistency procedure in view of the dependence of the wave functions describing the electron-density distribution in the direction of confinement on the two-dimensional momentum, the number of the subband, the spin, and interband tunnel mixing. In the presence of a vector potential, direct self-consistent calculations of the density of states are impossible in view of their computational complexity (to say nothing of trying to model capacitance oscillations as a function of B and especially the depth of the potential well).

We believe it is justified to carry out the calculations of the spectrum within an approach developed in Refs. 15 and 16 for describing vacuum condensate of Dirac electrons near supercharged nuclei and used in Refs. 8 and 17 for analyzing surface quantum wells in Kane and Dirac semiconductors in the absence of a magnetic field. At $B=0$ the results obtained by this approach agree fairly well (including spinor-type effects) with the experimental data and with direct numerical calculations.[9] (As noted in Ref. 9, the difference between the approach of Minkov *et al.*[9] and our approach developed in Ref. 8 concerns a unique case of vacant states near the bottom of a subband for a narrow and shallow unfilled quantum well in HgCdTe with an exceptionally high level of doping, $>10^{19}\,\mathrm{cm}^{-2}$; in standard conditions, including those in which the samples and regimes were studied in this research, the two approaches lead to essentially identical results.)

When the applied magnetic field is parallel to the direction of the confinement potential $V(z)$, the motion in the two-dimensional plane can be quantized, which means that within the six-band model (valid for small values of E_g) the problem of calculating the spectrum $E(B)$ clearly reduces to solving the following matrix equation (here we will not write the gauge-dependent standard expressions that describe the behavior of the wave functions in the plane perpendicular to the magnetic field):

$$\begin{vmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{vmatrix} \begin{pmatrix} f_1^{n-1}(z) \\ f_3^{n-2}(z) \\ f_5^{n}(z) \\ f_2^{n}(z) \\ f_4^{n-1}(z) \\ f_6^{n-1}(z) \end{pmatrix} = 0,$$

$$H_{12}=\begin{vmatrix} 0 & 0 & s_b\hbar\hat{k}_z \\ 0 & 0 & 0 \\ s_b\hbar\hat{k}_z & 0 & 0 \end{vmatrix}, \tag{1}$$

$$H_{11}=\begin{vmatrix} -E_- & \dfrac{\sqrt{3(n-1)}}{2}E_B & \dfrac{\sqrt{n}}{2}E_B \\ \dfrac{\sqrt{3(n-1)}}{2}E_B & -E_+ & 0 \\ \dfrac{\sqrt{n}}{2}E_B & 0 & -E_+ \end{vmatrix},$$

$$H_{22}=\begin{vmatrix} -E_- & \dfrac{\sqrt{3(n+1)}}{2}E_B & -\dfrac{\sqrt{n}}{2}E_B \\ \dfrac{\sqrt{3(n+1)}}{2}E_B & -E_+ & 0 \\ -\dfrac{\sqrt{n}}{2}E_B & 0 & -E_+ \end{vmatrix},$$

where $E_\pm=E-V\pm E_g/2$, $s_b=\sqrt{|E_g|/2m_b}$ and m_b are the Kane velocity and mass, and n is the number of the Landau level. The "magnetic energy" $E_B=\sqrt{2m_b s_b^2\hbar\omega_b}=\sqrt{2}s_b\hbar/\lambda$ ($\hbar\omega_b=\hbar eB/m_bc$ is the cyclotron energy) and the magnetic length $\lambda=\sqrt{c\hbar/eB}$ are actually independent of the band parameters, since the value of s_b for all Kane semiconductors is essentially the same. Reasoning along the same lines as in Ref. 8, i.e., "squaring" the system with respect to the components f_1^{n-1} and f_2^n for electron layers in materials with $E_g>0$ (we call the electrons in such layers s-electrons) or f_5^n and f_6^{n-1} in materials with $E_g<0$ (p-electrons), we can write the subband equations in a form that is the same for s- and p-electrons, i.e., in the form of a system of two Schrödinger-like equations for the envelopes $\varphi_{2.5}^n=f_{2.5}^n/\sqrt{H_n^+}$ and $\varphi_{1.6}^{n-1}=f_{1.6}^{n-1}/\sqrt{H_n^-}$,

$$\begin{vmatrix} \dfrac{\hbar^2\hat{k}_z^2}{2m_b}-(E_{\mathrm{eff}}-U_0-U_B^+-U_R^+) & -iU_{\mathrm{s.o.}}^+-C_g^+\hat{k}_z \\ iU_{\mathrm{s.o.}}^-+C_g^-\hat{k}_z & \dfrac{\hbar^2\hat{k}_z^2}{2m_b}-(E_{\mathrm{eff}}-U_0-U_B^--U_R^-) \end{vmatrix} \begin{pmatrix} \varphi_{2.5}^n \\ \varphi_{1.6}^{n-1} \end{pmatrix}=0, \tag{2}$$

with an effective energy $E_{\text{eff}}=(E^2-m_b^2 s_b^4)/2m_b s_b^2$ and an energy-dependent potential containing a Klein–Gordon term $U_0=(V^2-2EV)/2m_b s_b^2$ and three spinor-type potentials: the ''magnetic'' potential

$$U_B^{\pm}=E_B^2\frac{g^2 nR^{\pm}+3(n\pm 1)}{2m_b s_b^2},$$

the potential responsible for interband tunnel mixing

$$U_R^{\pm}=\frac{s_b^2\hbar^2}{2m_b s_b^2}\left\{\left[\frac{3}{4}\left(\frac{1+L_n^{\pm}}{H_n^{\pm}}\right)^2+\frac{L_n^{\pm}}{H_n^{\pm}E_+}\right]\right.$$
$$\left.\times\left(\frac{dV}{dz}\right)^2+\frac{1+L_n^{\pm}}{H_n^{\pm}}\,\frac{d^2V}{dz^2}\right\},$$

and the spin–orbit term

$$U_{\text{s.o.}}^{\pm}=g\,\frac{s_b\hbar E_B}{4m_b s_b^2}\,\sqrt{nR_n^{\mp}}\left[\left(\frac{1+L_n^{\pm}}{H_n^{\pm}}+\frac{1+L_n^{\mp}}{2H_n^{\mp}}\right.\right.$$
$$\left.\left.\times(R_n^{\pm}-1)\right)\frac{dV}{dz}\right],$$

where

$$C_g^{\pm}=C_g\sqrt{R_n^{\mp}}\,(R_n^{\pm}-1),\qquad C_g=g\,\frac{s_b\hbar E_B\sqrt{n}}{4m_b s_b^2},$$

and $R_n^{\pm}=H_n^{\pm}/H_n^{\mp}$. In the above expressions we must put $g=-1$, $L_n^{\pm}=3E_B^2(n\pm 1)/4E_+^2$, and $H_n^{\pm}=E_- - L_n^{\pm}E_+$ for p-electrons and $g=+1$, $L_n^{\pm}=0$, and $H_n^{\pm}=E_+$ (here $R_n^{\pm}=1$) for n-electrons (the dimensionless parameter $L_n^{\pm}$ and the second term in the expression for $U_B^{\pm}$ emerge because of the contribution of the heavy-hole band). Clearly, an equation for Kane s-electrons but with $g=+2$ also describes the subband Landau levels in the case of a Dirac spectrum. Due to the spin–orbit term $U_{\text{s.o.}}^{\pm}$ (and for p–electrons, also due to the terms linear in $\hat{k}_z$), the system of equations does not (in contrast to the case with $B=0$; see Ref. 8) separate into independent equations for individual spin components.

Below, in calculating the self-consistent potential $V(z)$ and in quantizing the spectrum in the effective potential, we employ, as we did in Ref. 8, the semiclassical approach (clearly, if we were to apply the semiclassical approach directly to Eq. (1), we would lose spinor-type effects, and this is equivalent to employing Klein–Gordon approximation that takes into account only nonparabolic effects). By introducing the substitution

$$\varphi_i^m=C_i^m\exp\left[i\int k_z(z)\,dz\right]$$

and ignoring the terms proportional to $i(z\,dk_z/dz+k_z)^2$ (i.e., by ignoring higher-order terms in the expansion of the action in powers of $\hbar$), we reduce the system (2) to an expression for k_z:

$$k_z^{\pm}=\frac{\sqrt{2m_b s_b^2}}{s_b\hbar}\left(\frac{A}{2}\mp\sqrt{\frac{A^2}{4}-(E_{\text{eff}}-U_0-U_B^{+}-U_R^{+})(E_{\text{eff}}-U_0-U_B^{-}-U_R^{-})+U_{\text{s.o.}}^{+}U_{\text{s.o.}}^{-}}\right)^{1/2}, \tag{3}$$

where

$$A=2(E_{\text{eff}}-U_0)-(U_B^{+}+U_R^{+}+U_B^{-}+U_R^{-})$$
$$-\left(\frac{2m_b s_b^2}{s_b^2\hbar^2}\right)C_g^2(R_n^{+}-1)(R_n^{-}-1),$$

which along with the Bohr–Sommerfeld quantization rule (here it is convenient to pass to integration over V since for self-consistent potentials the Poisson equation directly yields dV/dz as a function of V)

$$\int_{V(z=0)}^{V(k_z=0)}k_z(E,V)\left(\frac{dV}{dz}\right)^{-1}dV=\pi\left(i+\frac{3}{4}\right) \tag{4}$$

determine the magnetic levels $E_n^{\pm}(i,B)$ in the quantum well $V(z)$. As in the case of a zero magnetic field, the term 3/4 for the phase factor in (4) corresponds to an infinitely high potential barrier for s- and p-electrons at the semiconductor–insulator interface, i.e., to vanishing boundary conditions for the corresponding spinor components (the other components do not vanish), while the boundary conditions from the bulk side of the structure are dictated by the presence of an impenetrable potential barrier in the effective potential.[8,15–18]

For Dirac electrons, this can be considered a generalization of the results of Artimovich and Ritus[18] that allows for spin effects and magnetic quantization. In the case of Kane s-electrons we arrive at the results of Ohkawa and Uemura.[19] Clearly, as $n\to\infty$, both $E_B\sqrt{n\pm 1}$ and $E_B\sqrt{n}\to s_b\hbar k_z$ (k_z is the two-dimensional quasimomentum) and Eqs. (3) and (4) describe the subband spectra $E_i^{\pm}(k_s)$ in the absence of a magnetic field (they become the respective formulas of Ref. 8), while at $V=\text{const}$, Eq. (3) describes the Landau subbands $E_n^{\pm}(B,k_z)$ in the bulk of a Kane (Dirac) semiconductor. Note that for p-electrons the terms in (3) proportional to C_g (which originate from the terms in (2) linear in $\hat{k}_z$) do not, generally speaking, produce a contribution appreciable in comparison to $U_{\text{s.o.}}^{\pm}$. However, ignoring these terms in calculations that do not account for spin–orbit coupling leads to significant errors and, in some cases, changes the order in which the spin levels are arranged.

In the ultrarelativistic limit $E_g=0$ in undoped quantum wells, the spectra are scale-invariant with respect to the well depth μ_s (as they are in a vanishing magnetic field). Figure 3 depicts these spectra for s-electrons ($E_g=+0$) and p-electrons ($E_g=-0$) in dimensionless coordinates E/μ_s and $B/B_s=(E_B/\mu_s)^2$ ($B_s=c\mu_s^2/2es_b^2\hbar$). Spin–orbit cou-

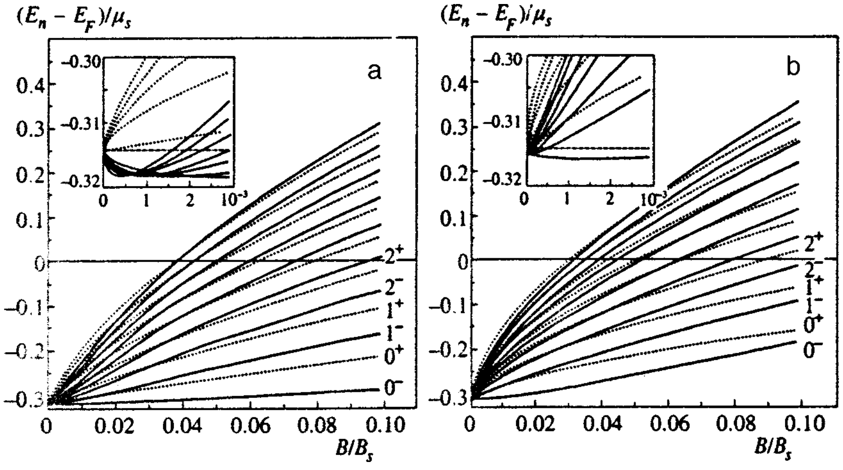

FIG. 3. The Landau levels of the ground subband for (a) *s*- and (b) *p*-electrons in the quasi-ultrarelativistic limit $E_g=0$. Energies are measured from the Fermi level E_F.

pling and interband resonant mixing lead to a transformation of the spectrum so dramatic that the introduction of a g factor (which is largely a nonrelativistic parameter) to describe the spin splitting of a two-dimensional spectrum in a magnetic field loses all meaning even for small-number Landau levels. The magnitudes of the effects for *s*- and *p*-electrons differ substantially both in the low-energy region (see the insets in Fig. 3) and for the states tested in oscillation effects close to the Fermi level, where the contribution of the spin–orbit term for *p*-electrons is almost twice as large. For Dirac electrons, in view of the larger value of the parameter g for *p*-electrons ($g=2$), the spin–orbit splitting at all energies is twice as large as it is for *s*-electrons ($g=1$) and at the Fermi level is close to the splitting for *p*-electrons. In all three cases, the spectra cannot be described by the Bychkov–Rashba formula with a quasimomentum-independent spin–orbit coupling parameter α (see Ref. 1).

3.2. Density of states

The density of states, whose singularities are responsible for the magneto-oscillations of kinetic and thermodynamic characteristics, is determined by the positions of the Landau levels and the level-broadening parameters. The calculation of collision broadening of magnetic levels is a complicated problem even in the one-band approximation.[20,21] We study this problem under the assumption of independent subsubband Landau levels of Gaussian shape,[22,23]

$$D_i^{\pm}(E)=\frac{eB}{c\hbar\sqrt{2\pi^3}}\sum_{n=0}^{\infty}\frac{1}{\Gamma_i^{\pm}}\exp\left[2\left(-\frac{E-E_{ni}^{\pm}}{\Gamma_i^{\pm}}\right)^2\right], \qquad (5)$$

with the broadening parameters $\Gamma_i^{\pm}$ not depending explicitly on n but depending on B (and energy-dependent for the nonparabolic system considered here). To describe magneto-oscillation effects analytically, which is preferable from the viewpoint of obtaining greater physical clarity of the results and establishing the specific features of oscillation phenomena in two-dimensional systems with quasirelativistic spectra in comparison to the case of standard bands,[22] we use analytic approximations that are in good agreement with the numerical results for the subsubband dispersion laws in the form of dispersion relations of the relativistic type,

$$E_i^{\pm}(k)=\sqrt{(s_i^{\pm}\hbar k)^2+[m_{0i}^{\pm}(s_i^{\pm})^2]^2}-m_{0i}^{\pm}(s_i^{\pm})^2, \qquad (6)$$

with the subsubband Kane masses $m_{0i}^{\pm}$ and velocities $s_i^{\pm}$ depending on the surface chemical potential μ_s (these masses and velocities can be found for the calculated subband concentrations and effective masses at the Fermi level[8]). Using an analytic description of the semiclassical spectrum in a magnetic field, a description that corresponds to such a representation (to make the notation simpler we will drop all ± labels),

$$E_{ni}=\sqrt{E_{Bi}^2(n_i+\delta_i)+m_{i0}^2s_i^4}-m_{i0}s_i^2 \qquad (7)$$

($E_{Bi}=\sqrt{2}\,s_i\hbar/\lambda$), instead of solving Eq. (4) with k_z from (3) numerically cannot greatly affect the density of states and the oscillation amplitudes [in calculating the energy positions of the levels one must use the "exact" solutions based on (3) and (4)]. Plugging (7) into (5) and using the Poisson summation formula, we arrive, via direct calculations under the assumption that $\Gamma_i \ll E+m_{0i}s_i^2$ (which is justified even for states near the bottom of subbands), at a "harmonic" representation for the density of states (convenient for describing oscillation effects):

$$\begin{aligned} D_i(E)={}&\frac{E+m_{0i}s_i^2}{2\pi s_i^2\hbar^2}\sqrt{\frac{\pi}{2}} \\ &\times\sum_{j=0}^{\infty}\frac{(-1)^j(2-\delta_{j0})}{(1+j^2\pi^2\gamma_i^4)^{5/4}}(\cos\theta+j\pi\gamma_i\sin\theta) \\ &\times\exp\left[-\frac{2}{1+j^2\pi^2\gamma_i^4}\left(\frac{j\pi\Gamma_i(E+m_{0i}s_i^2)}{E_{Bi}^2}\right)^2\right], \end{aligned} \qquad (8)$$

where

$$\gamma_i=\frac{\Gamma_i}{E_{Bi}},$$

$$\begin{aligned}\theta={}&\frac{1}{2}\arctan(j\pi\gamma_i^2)+j\pi\gamma_i^2 \\ &\times\left[\frac{2(E+m_{0i}s_i^2)^2}{\Gamma_i^2(1+j^2\pi^2\gamma_i^4)^2}-\frac{m_{0i}^2s_i^4+\delta_iE_{Bi}^2}{\Gamma_i^2}\right],\end{aligned}$$

with δ_{j0} the Kronecker delta. Within the Born approximation, the broadening parameters Γ_i are linked to the momentum relaxation times τ_i in zero magnetic field through the relationship $\Gamma_i^2=\sqrt{2/\pi}\,\hbar^2\,\omega_{ci}(E)/\tau_i$ (see Ref. 20), where $\omega_{ci}(E)=eB/cm_{ci}(E)$, with $m_{ci}(E)=m_{0i}+E/s_i^2$, and Eq. (8) becomes simpler (assuming the broadening is not too large, or $\gamma_i^2\ll 1$):

$$D_i(E,B)=D_i(E,0)\left[1+2\sum_{j=1}^{\infty}(-1)^j\times\exp\left(-\frac{j^2\pi}{\omega_{ci}(E)\tau_i}\right)\cos(2\pi j n_i(E))\right],\qquad(9)$$

where $D_i(E,0)=m_{di}(E)/2\pi\hbar^2=(W+m_{0i}s_i^2)/2\pi s_i^2\hbar^2$ is the density of states at $B=0$ (note that in the two-dimensional case the effective mass of the density of states, m_{di}, coincides with the cyclotron mass m_{ci}). The Landau-level number n_i, which in (9) is considered an arbitrary number (not necessarily an integer), is determined at a given energy by Eqs. (3) and (4) or, in the simplest approximation, by Eq. (7). The corresponding cyclotron energies can also be determined from (4) as $\hbar\omega_c(E)=E(n+0.5)-E(n-0.5)$, but the results differ little from those provided by analytic approximations for ω_{ci}. Clearly, in the nonrelativistic limit $s_i\to\infty$, Eq. (9) at $j=1$ becomes the Ando formula.[22]

Direct comparison of calculations based on (5) and (8) suggests that the approximations employed in deriving (9) are meaningful for essentially all regimes and parameters important from the experimental viewpoint, i.e., regimes and parameters corresponding to a sinusoidal shape of the D vs. E dependence (high Dingle temperatures $T_D=\hbar/2k_B\pi\tau$) and to a "nonsinusoidal" shape of the dependence (low Dingle temperatures).

3.3. Capacitance magneto-oscillations

For systems with a multiband spectrum, the differential capacitance of the space charge region with a two-dimensional gas,

$$C_{\text{s.c.}}=e^2\frac{dN_s}{d\mu_s}=e^2\frac{d}{d\mu_s}\sum_{i,\sigma}N_{i\sigma}=\sum_{i,\sigma}C_{i\sigma},\qquad(10)$$

where $N_{i\sigma}$ is the surface concentration in the spin branch σ of the subband i, is defined as the sum of partial subsubband capacitances (as before, we discard the subsubband indices $\sigma=\pm$):

$$C_i=\frac{e^2dN_i}{d\mu_s}=\frac{e^2dN_i}{d\mu_{Fi}}\frac{d\mu_{Fi}}{d\mu_s}=\frac{e^2}{\pi s^2\hbar^2}\frac{d\mu_{Fi}}{d\mu_s}\frac{d}{d\mu_{Fi}}\int D_i(E)f(E-\mu_{Fi})\,dE,\qquad(11)$$

where $f(E-\mu_{Fi})$ is the Fermi–Dirac distribution function, and μ_{Fi} is the subband Fermi energy. Using (9) and ignoring the integrals of odd functions, which is justified at low temperatures ($kT\ll\mu_{F_i}$), we arrive at an expression for the capacitance in a magnetic field:

$$\begin{aligned}\frac{C_i(B)}{C_i(0)}=1-2\sum_j(-1)^j\exp\left(-\frac{j^2\pi}{\omega_{ci}\tau_i}\right)\int_0^\infty dy\\ \times\{\cos(2\pi j n_i)\,[\cos(2\pi j c_{Bi}^2y^2)\\ -(jc_{\tau i}+c_{Ti}y\tan(jb_iy))\sin(2\pi j c_{Bi}^2y^2)]\\ -\sin(2\pi j n_i)\,[\sin(2\pi j c_{Bi}^2y^2)+(jc_{\tau i}\\ +c_{Ti}y\tan(jb_iy))\cos(2\pi j c_{Bi}^2y^2)]\}\frac{\cos(jb_iy)}{2\cosh^2(y/2)},\end{aligned}\qquad(12)$$

where

$$y=\frac{E-\mu_{Fi}}{kT},\quad \omega_{ci}=\omega_{ci}(\mu_{Fi}),\quad b_i=\frac{2\pi kT}{\hbar\omega_{ci}},$$

$$c_{Ti}=K_{mi}\frac{kT}{\mu_{Fi}},\quad c_{\tau i}=K_{mi}\frac{\hbar}{2\tau_i\mu_{Fi}},$$

$$c_{Bi}=\frac{kT}{E_{Bi}}=\frac{kT\lambda}{\sqrt{2}\,s_i\hbar},$$

$$K_{mi}=\left[1+\frac{d(m_{0i}s_i^2)}{d\mu_{Fi}}\right]\left[1+\frac{m_{0i}s_i^2}{\mu_{Fi}}+\frac{d(m_{0i}s_i^2)}{d\mu_{Fi}}\right]^{-1},$$

$$\begin{aligned}C_i(0)=\frac{e^2dN_i}{d\mu_{Fi}}\frac{d\mu_{Fi}}{d\mu_s}=\frac{e^2\mu_{Fi}}{2\pi s_i^2\hbar^2}\left[1+\frac{m_{0i}s_i^2}{\mu_{Fi}}\right.\\ \left.+\frac{d(m_{0i}s_i^2)}{d\mu_{Fi}}\right]\frac{d\mu_{Fi}}{d\mu_s}=e^2\left[D_i(\mu_{Fi},0)\right.\\ \left.+\frac{\mu_{Fi}}{2\pi s_i^2\hbar^2}\frac{d(m_{0i}s_i^2)}{d\mu_{Fi}}\right]\frac{d\mu_{Fi}}{d\mu_s}.\end{aligned}$$

Note that the equivalence between the capacitance in zero magnetic field, $C_i(0)$, and the quantity $D(\mu_F,0)\,e^2\,d\mu_F/d\mu_s$ (actually, the density of states at the Fermi level), which occurs in the parabolic approximation, is violated for the system with a relativistic spectrum considered here. In comparison to the case of a parabolic spectrum, the expression (12) for capacitance magneto-oscillations contains new parameters of the theory: c_T, c_τ, and c_B. The first two parameters reflect the energy dependence of the effective mass of the density of states, m_{di}, and the cyclotron mass m_{ci} in the Dingle exponential factor. Here two effects are important: the fact that the subband spectra are nonparabolic (for a fixed well depth), and the variation of the parameters of the subband spectra, $m_{0i}^{\pm}(\mu_s)$, under modulations of the well depth. The third parameter, c_B (Vshivtsev and Klimenko[24] suggested that such a parameter must exist in the theory; they, however, examined the case of Dirac electrons in a weakly relativistic system, $\mu_F\ll ms^2$, and ignored collision broadening), is actually determined only by the ratio $T/\sqrt{B}$ and is independent of band and subband parameters, since the subsubband Kane velocities $s_i^{\pm}(\mu_s)$ differ little from the universal quantity s_b.

In strong magnetic fields, the phase $2\pi j\, c_B^2 y^2$ is stationary in the range of values of y that provides the leading contribution to the integral, with the result that the integral in (12) can be evaluated:

$$\frac{C_i(B)}{C_i(0)} \approx 1 - 2\sum_j (-1)^j \frac{j\pi b_i}{\sinh(j\pi b_i)} \exp\left(-\frac{j^2\pi}{\omega_{ci}\tau_i}\right) \times[\cos(2\pi j n_i) - j c_{\tau i} \sin(2\pi j n_i)]. \qquad (13)$$

Note the quadratic dependence in the two-dimensional case of the arguments of the Dingle exponentials on the harmonic number j in the expressions (9), (12), and (13) for magneto-oscillations. In the nonrelativistic limit ($s_i \to \infty$) we have $c_{Ti} = c_{\tau i} = c_{Bi} = 0$ (when $c_{Ti} = c_{Bi} = 0$, the transition from (12) to (13) is exact), and at $j = 1$ Eq. (13) becomes the Ando expression for magneto-oscillations in materials with parabolic bands.[20,22] We see that the parabolic approximation is sufficient in describing oscillations in materials with a nonparabolic spectrum (with the rest mass in the amplitude factor replaced by the cyclotron mass at the Fermi level) only if strong magnetic fields are involved, $E_{Bi} > (5\text{–}10)\, kT$, and the smearing of the levels is not too great, $\mu_{Fi} \geqslant \hbar/2\tau_i$.

When there is spin–orbit splitting, the individual spin branches differ in both density of states and cyclotron energy, and the partial oscillations of capacitance for the spin subsubbands differ not only in periods (which leads to experimentally observable beats) but also in amplitudes, including the case in which all relaxation times are the same. Here, although the density of states and hence the partial capacitance for the low-energy branch of the spectrum at $B = 0$ are larger, the oscillation amplitude for this subsubband may be lower due to the large cyclotron-mass values in the parameter b and in the Dingle exponential factor, especially for p-electrons. Thus, the experimentally observed difference in the amplitudes of different spin components of the oscillations (probably Refs. 25 and 5 were the first papers where this difference was pointed out) is natural for systems with strong spin–orbit coupling and does not require using ideas about spin-dependent scattering for its explanation. Clearly, the amplitude ratio depends both on the surface concentration (through effective subband masses) and on the magnetic field, temperature, and broadening parameters.

4. RESULTS OF COMPUTER SIMULATION AND DISCUSSION

When calculating the capacitance of MOS structures, $C(B) = C_{\text{oxide}} C_{\text{s.c.}}(B)\, [C_{\text{oxide}} + C_{\text{s.c.}}(B)]^{-1}$, we used the values of the geometric capacitance of the oxide, C_{oxide}, found from the experimental values of capacitance in the region of large negative band bending, which corresponds to a strong enrichment in holes (the capacitance–voltage characteristics of the structures being studied, including those with $E_g > 0$, up to frequencies ~ 1 MHz of the low-frequency type). The subband capacitances in zero magnetic field and the voltages V_g across the structures (needed in calculations of dC/dV_g) were calculated from the dependence of the subsubband concentrations on the surface potential within the approach developed in Ref. 8. For inversion layers we allowed for the substantial increase with μ_s of the surface density of the hole-uncompensated acceptor charge in the depletion layer (this increase is important for narrow-gapsemiconductors and especially for gapless semiconductors because the thickness of the inversion layer in such materials is close to the overall thickness of the space charge region). Usually the calculated values of the capacitances for the case of zero magnetic field are in good agreement with the experimental values.

Although the structure of the oscillations modeled with only one adjustable parameter, the Dingle temperature, is close to the structures detected in experiments for the same two-dimensional concentration, the exact positions of the peaks in the oscillations and the beat nodes differ. This should come as no surprise since in the six-band model used here the contribution of distant bands to the bulk g factor is ignored, with the position of the spin components of the Landau levels being very sensitive to the value of this factor. As for the beat periods, comparison of the modeled and measured oscillation curves points to a substantial (up to 20% for the ground subband and approximately 10% for excited subbands) theoretical underestimation of the size of spin–orbit splitting in materials with a positive gap, while an analysis based on subsubband populations reveals no discrepancies (due to the large experimental error, as pointed out in Sec. 2). The differences may be related to the contribution of the interface region to spin–orbit splitting[11] (this contribution cannot be consistently calculated in the effective mass method). In both cases, within the Rashba mechanism the corrections are proportional to the electric field strength and can be taken into account phenomenologically by introducing a factor into the spin–orbit term of the effective Hamiltonian. Indeed, the discrepancies are removed by introducing a factor of roughly 1.2 into $U_{\text{s.o.}}$ and a correction (which models the contribution of distant bands to the g factor) to the phase of the oscillations (both the factor and the correction are the same for all concentrations, subband numbers, temperatures, and broadening parameters).

The computer-modeled magneto-oscillations depicted in Figs. 1 and 2 are in good agreement with the experimental data regarding both the position of the oscillations and beat nodes and the absolute values of capacitance. As noted earlier, in our calculations we used two adjustable parameters for p-electrons (T_D and a correction to the oscillation phase) and three adjustable parameters for s-electrons (the additional parameters being the factor of $U_{\text{s.o.}}$ mentioned earlier). In all regimes and for all magnetic-field and temperature ranges in which capacitance magneto-oscillations were observed in the experiments with the specified materials, calculations based on (12) differ very little from those obtained in the approximation (13). Usually the contribution of the sinusoidal term in (13) is also insignificant. The absolute values of the capacitance oscillation amplitudes and their magnetic-field dependence are described satisfactorily by the theory on the assumption that the values of the relaxation-time adjustable parameters τ_i in spin-split subsubbands used in the modeling process are roughly the same. In some cases where up to three beat nodes are observed in experiments, digital filtering and the inverse Fourier transform make it

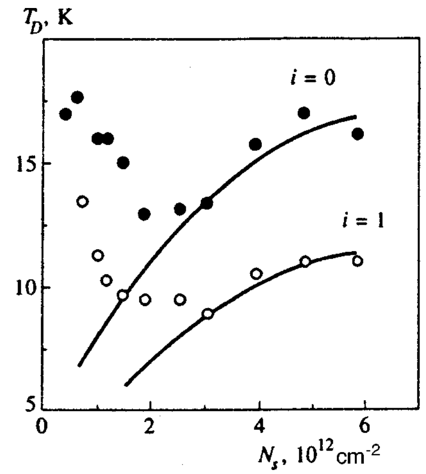

FIG. 4. The Dingle temperature as a function of the two-dimensional concentration for the sample S1. The solid curves represent the results of calculations for the scattering off "bumps" on the surface at $L = 110$ Å and $\Delta = 20$ Å.

possible to reliably separate the "partial" oscillations belonging to individual subsubbands. The relaxation times found from such oscillations for different spin subsubbands coincide, within experimental error, but the amplitudes I^+ and I^- may differ substantially (severalfold). Note that in such cases the values of τ_i found for the magnetic-field dependence of oscillation amplitudes are close to those obtained by fitting the calculated absolute values of the oscillation amplitudes to the experimental values.

The curves representing the T_D vs. N_s dependence with a minimum at $N_s \approx 2 \times 10^{12}\,\mathrm{cm}^{-2}$ (for the sample S1 such curves are depicted in Fig. 4) are similar in shape to such curves for silicon inversion layers. In contrast to the latter, however, the declining section is not related to the screening of Coulomb scattering, since the values of T_D predicted by the theory and corresponding to scattering off the charge built into the oxide and off ionized impurities are lower than the measured values for the studied structures by a factor of at least ten. A further increase in concentration drives T_D up, which is characteristic of scattering off "bumps" on the surface, but the T_D vs. N_s dependence is far from being quadratic. What is more, the dependence is sublinear, which suggests that the effectiveness of this mechanism is suppressed substantially as the Fermi de Broglie wavelength decreases (this is possible at large values of the correlation length L). The best agreement between the values of T_D calculated according to the approach developed in Ref. 26 and the experimental values for the structures under investigation is achieved for an average fluctuation $\Delta = 20$–25 Å and a correlation length $L \approx 100$–120 Å, which is almost ten times the corresponding values for silicon. The values of L are also almost twice the values for binary semiconductors,[26] which should come as no surprise in view of the many defects at the interfaces of the ternary compounds and their oxides. The strong broadening of levels near the subband edges suggests that there is an additional reaction channel and can be related to the interband (intersubband) scattering, whose effectiveness decreases as N_s increases, due to the increase in intersubband energies. Calculating such scattering consistently is a separate problem, whose complexity is due not only to the complexity of the spectrum but also to the need to allow for screening and multielectron effects, since in the systems considered the Fermi wavelengths in all regimes prove to be, as it can be demonstrated on the basis of the results reported in Ref. 8, quantities of the same order of magnitude as the Thomas–Fermi screening lengths and the size of the wave functions in the direction of confinement.

While in the modeled oscillations of structures based on materials with $E_g > 0$ the individual spin components can easily be resolved for small Landau-level numbers, in structures with $E_g < 0$, where the spin–orbit splitting is much larger and, correspondingly, the difference in the cyclotron splitting for distinct spin subbands is greater, such resolution cannot be achieved no matter how realistic the values of the broadening parameters are or how low the temperature is (this fact is fully corroborated by experiments). Note that the deterioration of the resolution of spin splitting with increasing temperature and broadening of levels in oscillations of structures with $E_g > 0$ is due not only to the increase in the smearing of the Fermi step (the "spectral gap") but also, primarily for small values of B and high values of T, to a much more rapid decrease in the oscillation amplitude for the low-energy branch of the spectrum.

In conclusion we will briefly summarize the results. A relatively simple model that allows for a clear physical interpretation has been proposed. The model is used to calculate the energy spectrum of a two-dimensional electron gas in a magnetic field in the surface quantum wells of semiconductors with Kane and Dirac spectra and is based on reducing the initial matrix equations to a Schrödinger-like equation with an effective potential in which the terms responsible for the nonparabolic features, the spin–orbit coupling, and the interband tunnel mixing are easily distinguishable. The model also reveals the specific features of Kane semiconductors with direct and inverted band structure and with a Dirac spectrum. An analytic approach has been developed in order to describe the magneto-oscillation effects in a two-dimensional gas of materials with a quasirelativistic spectrum. The emergence is noted of new theoretical parameters (compared to the parameters of materials with a quadratic spectrum). Comparison of the modeled and measured capacitance magneto-oscillations in HgCdTe-based MOS structures suggests that probably there is a small ($<20\%$) contribution of the interface region to the spin–orbit splitting of the spectrum. As has been discovered in experiments and substantiated in the theory, the oscillation amplitudes referring to two spin–orbit split ladders of Landau levels differ substantially while their collision-broadening parameters are equal. The established average displacements of interfaces and the correlation lengths characterizing the scattering off the "bumps" on the surface exceed many times over their values in silicon and binary semiconductors in view of the many defects at the interfaces of the ternary compounds and their oxides. The short relaxation times near the edges of the two-dimensional subbands and their increase with the two-dimensional concentration N_s in the region of small values of N_s can be related to intersubband scattering.

The work was supported by a grant from the Russian Ministry of General and Vocational Education (Grant No. 97-0-7.1-137).

*⁾E-mail: Victor.Radantsev@usu.ru

---

[1] Yu. A. Bychkov and E. I. Rashba, J. Phys. C **17**, 6039 (1984).
[2] S. I. Dorozhkin and E. B. Ol'shanetskiĭ, JETP Lett. **46**, 502 (1987).
[3] V. F. Radantsev, Zh. Éksp. Teor. Fiz. **96**, 1793 (1989) [Sov. Phys. JETP **69**, 1012 (1989)].
[4] R. Sizmann and F. Koch, Semicond. Sci. Technol. **5**, 5115 (1990).
[5] J. Luo, H. Munekata, F. F. Fang, and P. J. Stiles, Phys. Rev. B **41**, 7685 (1990).
[6] B. Das, S. Datta, and R. Reifenberger, Phys. Rev. B **41**, 8278 (1990).
[7] A. V. Germanenko, G. M. Minkov, V. A. Larionova, O. E. Rut, C. R. Becker, and G. Landwehr, Phys. Rev. B **52**, 17 254 (1995).
[8] V. F. Radantsev, T. I. Deryabina, G. I. Kulaev, and E. L. Rumyantsev, Phys. Rev. B **53**, 15 756 (1996).
[9] G. M. Minkov, A. V. Germanenko, V. A. Larionova, and O. E. Rut, Phys. Rev. B **54**, 1841 (1996).
[10] S. I. Dorozhkin and G. Landwehr, JETP Lett. **64**, 677 (1996).
[11] G. Engels, J. Lange, Th. Schäpers, and H. Lüth, Phys. Rev. B **55**, R1958 (1997).
[12] J. Nitta, T. Akazaki, H. Takayanagi, and T. Enoki, Phys. Rev. Lett. **78**, 1335 (1997).
[13] P. Sobkowicz, Semicond. Sci. Technol. **5**, 183 (1990).
[14] M. J. Godfrey and A. M. Malik, Phys. Rev. B **53**, 16 504 (1996).
[15] Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys. Usp. **14**, 673 (1972)].
[16] A. B. Migdal, V. S. Popov, and D. N. Voskresenskiĭ, Zh. Éksp. Teor. Fiz. **72**, 834 (1977) [Sov. Phys. JETP **45**, 436 (1977)]; A. B. Migdal, *Fermions and Bosons in Strong Fields* [in Russian], Nauka, Moscow (1978).
[17] V. F. Radantsev, Semicond. Sci. Technol. **8**, 394 (1993).
[18] G. I. Artimovich and V. I. Ritus, Zh. Éksp. Teor. Fiz. **104**, 2912 (1993) [JETP **77**, 348 (1993)].
[19] F. Ohkawa and Y. Uemura, J. Phys. Soc. Jpn. **37**, 1325 (1974).
[20] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[21] L. Spies, W. Apel, and B. Kramer, Phys. Rev. B **55**, 4057 (1997).
[22] T. Ando, J. Phys. Soc. Jpn. **27**, 1233 (1974).
[23] R. Gerhardts, Z. Phys. B **21**, 275 (1975).
[24] A. S. Vshivtsev and K. G. Klimenko, Zh. Éksp. Teor. Fiz. **109**, 954 (1996) [JETP **82**, 514 (1996)].
[25] V. F. Radantsev, T. I. Deryabina, L. P. Zverev, G. I. Kulaev, and S. S. Khomutova, Zh. Éksp. Teor. Fiz. **88**, 2088 (1985) [Sov. Phys. JETP **61**, 1234 (1985)].
[26] V. F. Radantsev, T. I. Deryabina, L. P. Zverev, G. I. Kulaev, and S. S. Khomutova, Zh. Éksp. Teor. Fiz. **91**, 1016 (1986) [Sov. Phys. JETP **64**, 598 (1986)].

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Drift in a random walk along self-similar clusters

V. E. Arkhincheev*)

Buryat Science Center, Siberian Branch of the Russian Academy of Sciences, 670047 Ulan-Ude, Russia

(Submitted 2 June 1998; resubmitted 28 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1016–1023 (March 1999)

This paper is a study of the relationship between diffusion and conductivity when the random walks of particles occur via Lévy hops. It shows that because of the unusual nature of Lévy hops the particle mobility is a nonlinear function of the electric field in arbitrarily weak fields. The crossover to ordinary diffusion by introduction of a finite displacement in each step is also discussed. © *1999 American Institute of Physics.* [S1063-7761(99)01803-X]

1. INTRODUCTION

Classical diffusion, in which diffusing particles only wander to nearest neighbors, has been thoroughly studied, and many methods related to the investigation of this phenomenon have been developed. Random walks, in which Brownian particles diffuse not just to nearest neighbors, have been less well-studied, however. To this class of diffusion problems belongs the problem of random walks of particles via Lévy hops. What makes Lévy hops so special is that in each step a particle can travel arbitrarily far, so that the rms displacement per unit time can be infinite.[1] Numerical modeling of diffusion via Lévy hops has shown that the points visited by a diffusing particle form spatially well-separated clusters. A more thorough examination shows that each cluster is in turn a collection of clusters. Thus, we obtain a hierarchy of self-similar clusters.[2] In this way Lévy diffusion constitutes a random walk among self-similar clusters. The probability distribution function in the Fourier representation has the form

$$P(k,t)=\exp(-A|k|^{\mu}t), \tag{1}$$

where A and μ are positive quantities, and $1<\mu<2$. Such distributions are known as Lévy distributions. A detailed description of Lévy flights can also be found in Ref. 3.

The study of Lévy diffusion is a problem in its own right as a microscopic model with unusual diffusion, but it is also of interest in connection with possible applications to problems of hopping conductivity in disordered media,[4] when the hop probability and the hop length are uncorrelated.

The goal of the present work is to study the relationship between diffusion and conductivity when there is Lévy diffusion in the system. When there is ordinary diffusion and the response is linear (Ohm's law), the relationship is known as the Einstein formula. However, in the general case it is not known what this relationship is. In this paper we generalize to the case where in the event of Lévy diffusion the drift velocity proves to be a nonlinear function of the electric field:

$$V\propto E^{\mu-1}. \tag{2}$$

Note that the nonlinearity occurs in arbitrarily weak fields and is a consequence of the unusual nature of the diffusion. The degree of nonlinearity is described by the critical index μ of Lévy hops. In other words, Ohm's law (linear response to the field) is a consequence of the ordinary nature of diffusion, so that for other diffusion, e.g., Lévy hops, Ohm's law is not satisfied. This becomes especially evident when in addition to Lévy hops there is ordinary diffusion, i.e., particles diffuse in the usual way. Then, in accordance with the two limits on the behavior of particles in random walks, the particle velocity has two asymptotic regimes, linear and the nonlinear. We also interpret the results from the standpoint of scaling. The results of this research were previously published in a brief communication.[5]

The nonlinearity of the function in (2) can be explained in the following way. If we combine Eq. (1) and the ordinary expression for the field current, we readily obtain an equation for diffusion along self-similar clusters in an electric field in the form of a continuity equation:

$$[\partial/\partial t+(A|k|^{\mu}+ikV)]N(k,t)=0. \tag{3}$$

Here $N(k,t)$ is the number density of the diffusing particles in the Fourier representation, and the current has a diffusion component and a field component, with the latter being of the ordinary form $J=NV$.

Next we use the well-known ideas developed by Einstein. In equilibrium, the diffusion current J_d is balanced by the field current J_f and the distribution function is of the Boltzmann form:

$$J_d+J_f=0,\quad N=\exp(-U/kT), \tag{4}$$

where U is the potential energy.

Using the series definition of a derivative of fractional order,[6]

$$|k|^{\mu}=\lim_{\varepsilon\to 0}(\Delta^2+\varepsilon)^{\mu}=\sum_{n=0}^{\infty}C_n^{\mu}\left(\frac{\Delta}{\varepsilon}\right)^n, \tag{5}$$

we obtain the general expression for the drift velocity:

$$\mathbf{V}=\exp\left(\frac{U}{kT}\right)\lim_{\varepsilon\to 0}(\Delta^2+\varepsilon)^{(\mu-2)/4}\nabla\exp\left(-\frac{U}{kT}\right). \tag{6}$$

For a uniform electric field $U=-qEx$, and we obtain the result (2).

2. DISCRETE DISTRIBUTION OF LÉVY HOPS

Let us consider the one-dimensional analog of Lévy hops.[1] We denote the probability of finding a particle at the lth site after n steps by $P_n(l)$, and the hop probability distribution in length by $f(l)$:

$$P_{n+1}(l)=\sum_{m=-\infty}^{\infty} f(l-m)P_n(m). \tag{7}$$

For $f(l)$ we take

$$f(l)=\sum_{n=0}^{\infty} a^{-n}(\delta_{l,-b^n}+\delta_{l,b^n}), \tag{8}$$

where $\delta_{n,m}$ is the Kronecker delta. Then the structure function for such a random walk is

$$\lambda(k)=\int f(l)\exp(ikl)\,dl=\sum_{n=0}^{\infty} a^{-n}\cos kb^n. \tag{9}$$

Note that the structure function $\lambda(k)$ satisfies the functional equation

$$\lambda(k)=a\lambda(kb)+\cos k. \tag{10}$$

Hence, as $K\rightarrow 0$, the structure function is a power law with exponent $\mu=\ln a/\ln b$. Nonanalytic behavior of the type $|k|^{\mu}$ as $k\rightarrow 0$ can be obtained with exponent μ by taking Mellin transforms or using the Poisson sum formula (see Ref. 1 for details).

Next we introduce anisotropy into a random walk along self-similar clusters. By virtue of the specific features of Lévy hops, a particle can travel any distance b^n in one step, so that a small anisotropy $1+\alpha$ with $\alpha=qEs/kT$ over a small displacement s can turn out to be exponentially large over large distances b^n. Since in each step the diffusing particle leaves a site, the sum of probabilities of moving along the field, W_+, and against the field, W_-, must equal unity, i.e., $W_+ + W_- = 1$. This leads to an expression for the probabilities of moving along and against the field:

$$W_{\pm}=\frac{(1\pm\alpha)^{b^n}}{(1+\alpha)^{b^n}+(1-\alpha)^{b^n}}. \tag{11}$$

Hence, if diffusion in an electric field occurs via Lévy hops, the structure function is

$$2\lambda(k;E)=\sum_{n=0}^{\infty} a^{-n}[\cos kb^n+i\,\sin(kb^n)\,(W_+-W_-)]. \tag{12}$$

As in ordinary diffusion, the second term contains the drift velocity as $k\rightarrow 0$:

$$V=i\frac{\partial\lambda(k;E)}{\partial t}\Big|_{k\rightarrow 0}=\sum_{n=0}^{\infty}\left(\frac{b}{a}\right)^n\frac{(1+\alpha)^{b^n}-(1-\alpha)^{b^n}}{(1+\alpha)^{b^n}+(1-\alpha)^{b^n}}$$

$$\approx\sum_{n=0}^{\infty}\left(\frac{b}{a}\right)^n\tanh\alpha b^n. \tag{13}$$

Clearly, the drift velocity satisfies the functional equation

$$V(\alpha)=\frac{b}{a}V(\alpha b)+\coth\alpha. \tag{14}$$

Hence, in arbitrarily weak fields, the velocity is a power law in the electric field with exponent $\mu-1$. To calculate this dependence exactly, we use the Poisson sum formula:

$$\sum_{n=0}^{\infty} f(n)=\frac{1}{2}f(0)+\int_0^{\infty} f(t)\,dt+2\sum_{m=1}^{\infty} f(t)\cos(2\pi mt).$$

In the present case, $f(t)=(b/a)^t\tanh\alpha b^t$.

Introducing the variables $t'=t\ \ln b$ and $z=\exp\ t'$, we obtain $f(z)=z^{-\mu}\tanh\alpha z$. Hence

$$V(\alpha)=\frac{\alpha}{2}+\alpha^{\mu-1}\left[\sum_{m=-\infty}^{\infty}\int_1^{\infty}\frac{\tanh z}{z^{\gamma_m}}\,dz+\int_0^{\alpha}\frac{\tanh z}{z^{\gamma_m}}\,dz\right], \tag{15}$$

where $\gamma_m=\mu+2\pi mi/\ln b$.

Clearly, the second term in brackets is small (in parameter α) compared to the first. Thus, in arbitrarily weak electric fields, we obtain for the velocity the nonlinear dependence (2) on the electric field. We also note that the nonanalytic behavior of the structure function as $k\rightarrow 0$ and the nonlinear dependence of the velocity on the electric field in arbitrarily weak fields are asymptotic.

3. CONTINUUM LIMIT OF LÉVY HOPS

We now consider the continuous hop-length distribution and find the particle velocity in an electric field for this case. In the continuum limit of Lévy hops, the distribution in hop length is a power law,

$$f(L)\propto 1/L^{\mu+1}, \tag{16}$$

and the probability of finding a particle at point x after n steps is determined by the integral equation

$$P_{n+1}(x)=\int_{-\infty}^{\infty}\frac{P_n(y)\,\partial y}{|x-y|^{\mu+1}}. \tag{17}$$

Simple transformations lead to

$$P_{n+1}(x)=\int_0^{\infty}[P_n(x+y)+P_n(x-y)]\frac{\partial y}{|x-y|^{\mu+1}}. \tag{18}$$

Now we introduce anisotropy into random walks with a power-law hop-length distribution function by analogy with the discrete case (11) by substituting $|x-y|$ for b^n. Using (11) in the continuum limit and (18), we can easily derive an equation for the particle number density in an electric field:

$$P_{n+1}(x)=\int_0^{\infty}\frac{[(1+\alpha)^tP_n(x+y)+(1-\alpha)^tP_n(x-y)]\,dy}{[(1+\alpha)^t+(1-\alpha)^t]|x-y|^{\mu+1}}. \tag{19}$$

Expanding $P_n(x\pm y)$ in Taylor series and separating even and odd powers of t, we obtain

$$\begin{aligned}P_{n+1}(x)&\\ =\int_0^\infty \sum_{m=0}^{\infty} &\frac{\partial^{2m}P_n(x)}{\partial x^{2m}}\frac{y^{2m}\,dy}{|x-y|^{\mu+1}}\\ +\int_0^\infty \sum_{m=0}^{\infty} &\frac{\partial^{2m+1}P_n(x)}{\partial x^{2m+1}}\frac{y^{2m+1}[(1+\alpha)^t-(1-\alpha)^t]\,\partial y}{[(1+\alpha)^t+(1-\alpha)^t]|x-y|^{\mu+1}}.\end{aligned} \tag{20}$$

The first term on the right-hand side of Eq. (20) corresponds to the diffusion contribution, while the second term contains the field-current contribution. Accordingly, the expression for the field current is

$$J\propto \int_0^\infty \sum_{m=0}^{\infty}\frac{\partial^{2m}P_n(x)}{\partial x^{2m}}y^{2m-\mu}\tanh(\alpha y)\,\partial y. \tag{21}$$

Introducing the variable $z=\alpha y$ and reducing Eq. (21) to dimensionless form, we obtain (2), which represents the nonlinear current–field dependence in arbitrarily weak fields $(\alpha\rightarrow 0)$.

4. TRANSITION FROM ORDINARY DIFFUSION TO LÉVY DIFFUSION

In this section, in addition to Lévy hops, we allow for ordinary diffusion and consider the transition to linear dependence of the velocity on the electric field due to ordinary diffusion. The simplest way to do this is to introduce a finite hop length ξ at each step. What we obtain is a random walk in which ordinary diffusion alternates with Lévy hops. However, due to the superlinear time dependence of the rms displacement for Lévy diffusion, on small scales (times) the main contribution to the random walk is provided by ordinary diffusion, while over long time intervals it is the Lévy hops that contribute most to the random walk. Accordingly, the hop-length distribution function has the form

$$f(l)=\sum_{n=0}^{\infty}a^{-n}[\delta_{l,-(b^n+\xi)}+\delta_{l,(b^n+\xi)}]. \tag{22}$$

Hence the structure function is

$$\lambda(k,\xi)=\sum_{n=0}^{\infty}a^{-n}\cos(kb^n+k\xi). \tag{23}$$

In the limit of small lengths $(b\rightarrow 0)$, this formula turns into the expression corresponding to ordinary diffusion:

$$\lim_{b\rightarrow 0}\lambda(k,\xi)=\frac{a-l}{a}\cos k\xi. \tag{24}$$

Anisotropy can be introduced into random walks by the method described earlier: we replace the hop length b^n with the quantity $b^n+\xi$. Thus, the structure function in an electric field for finite hop lengths is

$$\begin{aligned}\lambda(k,\xi,\alpha)=\sum_{n=0}^{\infty}a^{-n}[&\cos(kb^n+k\xi)+i\\ &\times\sin(kb^n+k\xi)\,(W_+-W_-)].\end{aligned} \tag{25}$$

Accordingly, for the velocity we have

$$V=i\frac{\partial\lambda(k;E)}{\partial t}\Big|_{k\rightarrow 0}=\sum_{n=0}^{\infty}\frac{b^n+\xi}{a^n}\tanh(\alpha b^n+\alpha\xi). \tag{26}$$

To calculate the sum on the right-hand side of Eq. (26) we employ the Poisson method. Clearly, the velocity turns out to be nonlinear in the field (Eq. (2)) in arbitrarily weak fields $(qE\xi/kT\ll 1)$ and linear in strong fields $(qE\xi/kT\gg 1)$:

$$V\propto E\xi^{2-\mu}. \tag{27}$$

Thus, the particle velocity in an electric field has two asymptotic limits in accordance with two diffusion regimes: Lévy hops and ordinary diffusion.

For the particle mobility η we can write the above result in scaling form:

$$\eta\propto\xi^{2-\mu}f(qE\xi/kT), \tag{28}$$

where the scaling function $f(x)$ has the asymptotic behavior

$$f(x)=\begin{cases}1, & x\ll 1,\\ x^{\mu-2}, & x\gg 1.\end{cases} \tag{29}$$

On small scales, where ordinary diffusion dominates $(L_E\ll\xi)$, particle mobility depends only on the homogeneity length ξ, while on large scales, where Lévy hops dominate $(L_E\gg\xi)$, mobility ceases to depend on the homogeneity length and becomes a function of the extent of the field L_E with the same exponent. In other words, in such fields the mobility ''forgets'' the homogeneity scale and becomes a nonlinear function of the electric field.

5. CONCLUSION

Nonlinear properties of inhomogeneous materials have attracted much attention from both theoreticians and experimentalists.[7–12] Theoreticians usually expand the current in powers of the electric field out to a cubic nonlinearity:

$$J=\sigma E+\chi|E|^2E+\cdots. \tag{30}$$

Our results differ substantially from those obtained by such a method. In the microscopic model of Lévy diffusion, we show that in the event of Lévy hops the current proves to be a highly nonlinear function of the electric field because of the unusual regime of diffusion in space, i.e., there is no linear term in the field expansion of the current, Eq. (30). We examine the transition from ordinary diffusion to Lévy hops by introducing a finite displacement length ξ at each step. We show that the problem acquires a new parameter $qE\xi/kT$, that determines whether the particle mobility behaves linearly or nonlinearly. In other words, a new length L_E governed by the electric field emerges in such diffusion problems:[13]

$$L_E=kT/qE\,. \tag{31}$$

To appreciate the significance of this quantity, we consider an ordinary random walk in an external electric field. Let us imagine that the medium is partitioned into blocks of size L_E. We then study the behavior of a particle within a single block. With a probability of order unity the particles leaves such a block when it moves along the field but does not leave the block when it moves against the field. In other words, within a block whose linear size is of order L_E, ordered motion prevails over diffusion. This makes it possible to estimate the particle velocity to be

$$V=L_E/t_E, \tag{32}$$

where t_E is the diffusion time for displacement L_E. For ordinary diffusion $t_E=L_E^2/D$, and we have the well-known Einstein formula

$$V=q^2DE/kT. \tag{33}$$

When Lévy hops dominate, the same estimates yield Eq. (2), and in the two diffusion limits we have (29). Earlier[6] a dependence of type (29) was predicted by a phenomenological description of anomalous diffusion along percolation clusters in the effective-medium approximation. The correlation length of the percolation clusters was taken as the homogeneity length. We also note that in contrast to Lévy hops, anomalous diffusion along percolation clusters is of a sublinear nature, so that in mixed random walks (anomalous diffusion plus ordinary diffusion) the main contribution over short time intervals is provided by anomalous diffusion, while over long time intervals the main contribution is provided by ordinary diffusion. An attempt to detect the nonlinearity by computer modeling of drift along clusters did not succeed,[14] since over the desired range of fields, the electric field in inhomogeneous media induces traps. Such traps are sections of the current paths directed against the electric field. Hence the question of the nonlinear dependence of the velocity on the electric field due to anomalous nature of the diffusion remained unresolved.

The present paper establishes for the first time the nonlinear dependence of the velocity on the electric field for the model of diffusion via Lévy hops. Moreover, the deduced nonlinear behavior of the velocity due to the unusual nature of the dependence (2) was recently corroborated by computer modeling of particle drift in the presence of Lévy diffusion.[15]

As for experimental results, many researchers have observed the nonlinear power dependence of the current in inhomogeneous materials with exponents close to the anomalous-diffusion index and have provided various explanations of this phenomenon (see, e.g., Refs. 11 and 12). In our opinon, nonlinear behavior can indeed be observed, and there is a universal explanation for nonlinear behavior as stemming from the anomalous nature of random walks in inhomogeneous media. Comparisons of experimental and theoretical results require further study, however.

This work was initiated by numerous interesting discussions with É. G. Batyev and É. N. Baskin. The author is also grateful to A. A. Snarskiĭ for kindly supplying offprints of various papers.

*)E-mail: varkhin@bsc.buriatia.ru

---

[1] B. D. Hughes, M. F. Shlesinger, and E. W. Montroll, Proc. Natl. Acad. Sci. USA **78**, 3287 (1981).

[2] B. Mandelbrot, *Fractals: Form, Chance and Dimension*, Freeman, San Francisco (1977).

[3] G. Zimofen and J. Klafter, Physica D **69**, 436 (1993); Phys. Rev. E **47**, 851 (1993).

[4] B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors*, Springer-Verlag, Berlin (1984).

[5] V. E. Arkhincheev, JETP Lett. **67**, 545 (1998).

[6] V. E. Arkhincheev, E. M. Baskin, and E. G. Batyev, J. Non-Cryst. Solids **90**, 21 (1987).

[7] R. Blumenfeld and D. J. Bergman, Phys. Rev. B **43**, 13 682 (1991); P. M. Hui, Phys. Rev. B **49**, 15 344 (1994).

[8] A. A. Snarskiĭ and S. I. Buda, Zh. Tekh. Fiz. **68**(6), 5 (1998) [Tech. Phys. **43**, 619 (1998)].

[9] A. M. Satanin, S. V. Khor'kov, and V. V. Skuzovatkin, Zh. Éksp. Teor. Fiz. **112**, 643 (1997) [JETP **85**, 351 (1997)].

[10] A. M. Dykhne, V. V. Zosimov, and S. A. Rybak, Dokl. Ross. Akad. Nauk **345**, 467 (1995) [Phys. Dokl. **40**, 613 (1995)].

[11] H. Overhoff and W. Beyer, Philos. Mag. B **43**, 433 (1981).

[12] J. Benjamin, C. Adkins, and J. van Clever, J. Phys. C **17**, 559 (1981).

[13] É. G. Batyev, private communication (1987).

[14] V. E. Arkhincheev and É. M. Baskin, Zh. Éksp. Teor. Fiz. **100**, 292 (1991) [Sov. Phys. JETP **73**, 161 (1991)].

[15] V. E. Arkhincheev and A. V. Nomoev, submitted to Physica A (1998).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Imperfection of the Sm sublattice and valence instability in compounds based on SmB_6

E. V. Nefedova,*) P. A. Alekseev,*) E. S. Klement'ev, V. N. Lazukov,*) I. P. Sadikov,*)
M. N. Khlopkin, and M. B. Tsetlin

Kurchatov Institute, 123182 Moscow, Russia

E. S. Konovalova and Yu. B. Paderno

I. N. Frantsevich Institute of Problems in Materials Science, Ukrainian National Academy of Sciences, 252680 Kiev, Ukraine

(Submitted 19 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1024–1038 (March 1999)

The magnetic susceptibility, electrical resistance, specific heat, and thermal expansion coefficient of SmB_6, $Sm_{0.8}B_6$, and $Sm_{1-x}La_xB_6$ ($x=0.1$ and 0.2) are measured in the temperature range $T=4-300$ K. The dispersion curves of the acoustic phonon branches in lanthanum-doped samples are studied. A combined analysis of the results confirms the existence of an activation gap in the electron density of states in both nonstoichiometric and lanthanum-doped compounds. The anomalies in the electronic component of the thermal expansion coefficient are associated to a considerable degree with the temperature variation of the valence and, like the magnetic susceptibility, reflect features of the *f*-electron excitation spectrum. It is found that lanthanum doping does not lead to significant changes in the anomalies in the phonon spectrum of SmB_6. It is established that the homogeneous intermediate-valent state of the samarium ion is fairly stable and is maintained when the perfection of the Sm sublattice is violated. © *1999 American Institute of Physics.* [S1063-7761(99)01903-4]

1. INTRODUCTION

Samarium hexaboride (SmB_6) is known as the first of the valence-instable compounds to be discovered. Its properties have been actively investigated since the 1970's, as samples of increasingly higher quality were obtained. This compound evoked heightened attention because of two characteristic features. First, an unusually narrow gap ($\sim$5 meV) was discovered in the electron density of states of SmB_6. The existence of the gap has been confirmed by numerous experiments on the resistivity, Hall effect, and specific heat,[1–4] as well as optical measurements.[5–7] The nature of the gap in SmB_6 has been discussed from various standpoints. It has been interpreted as a hybridization gap[8] and as a consequence of Wigner crystallization in the metal[9] or the formation of an exciton state in an intermediate-valent phase.[10] It has been theorized that the gap in the electron density of states vanishes when SmB_6 is doped with other rare-earth (RE) ions[11] or when pressure is applied.[12] However, in Ref. 13, where the influence of doping on the hybridization gap was investigated by measuring the electrical resistance, it was discovered that the gap is maintained in all compounds based on SmB_6 having a state with an intermediate valence. Thus, there is no unequivocal answer to the question of whether the gap vanishes or whether the Fermi level is displaced relative to the gap.

The second feature of this compound is that the samarium ions in SmB_6 are in an intermediate-valent state. The valence of the samarium ion at room temperature estimated from the lattice constant, the magnetic susceptibility, the L_{III} absorption edge, and the Mössbauer effect is $\simeq 2.55\pm0.03$.[14–16] It is generally assumed that the valence instability of RE ions is characterized by the presence of charge and spin fluctuations with a characteristic time $\tau \sim 10^{-12}-10^{-13}$ s. Charge fluctuations should influence the lattice excitation spectrum under certain conditions, and spin fluctuations should be manifested in the magnetic excitation spectrum. In fact, the results of the measurements of the phonon dispersion curves of SmB_6 in Ref. 17 reveal overall softening of the acoustic and low-lying optical phonons in comparison to isostructural LaB_6. In addition, distinct anomalies in the longitudinal-acoustic branches along the [110] and [111] directions and an additional mode, which is located in the energy gap between the acoustic and optical branches, were discovered. Some important results were also obtained when the magnetic excitation spectrum of SmB_6 was investigated in Ref. 18 and 19. In particular, a narrow low-energy excitation with unusual properties was discovered at low temperatures ($T<40$ K) in the magnetic component of the inelastic neutron scattering spectrum along with some broad structural peaks associated with intermultiplet transitions. An explanation for the specific features of both the phonon and magnetic excitation spectra based on an excitonic model, where the key factor is the formation of a mixed quantum-mechanical state for each Sm ion, was recently proposed.[20,21] An alternative model, which accounts for the electron density of states and the magnetic excitation spectrum of SmB_6, was developed in Ref. 22. It is based on the concept of a mixed-valent (inhomogeneous) state for the Sm ions in SmB_6. However, neither of these models is free of certain difficulties in substantiating the original approximations. Both the variation of the mean valence and the concentration of free carriers are quite critical for both models. For this reason it would be interesting to examine the transformation of the properties of SmB_6 as the valence state of the samarium ions varies.

The valence of the samarium ions in SmB_6 can be varied by replacing Sm by di- and trivalent ions of other elements. When Sm ions in the RE sublattice of SmB_6 are replaced by La^{3+} ions, the valence of the samarium ion decreases.[23] The kinetic and magnetic properties of a series of samples with the general formula $Sm_{1-x}La_xB_6$ have been studied in some detail.[15] Another method for altering the valence of the Sm ions in SmB_6, but without inserting foreign atoms into the RE sublattice, is to create a deficiency of samarium ions (Sm_xB_6). In this case the crystal structure remains stable over a very broad range of samarium concentrations. According to the results of the magnetic susceptibility measurements, the lattice period, and the L_{III} absorption edge, the valence of the samarium ion in Sm_xB_6 ($x=0.7-0.9$) remains appreciably different from an integer value and shifts toward 3^+ [$v(300\mathrm{K})\simeq 2.63\pm 0.03$].[24–26] The current list of experimental studies of the physical properties of nonstoichiometric samarium hexaboride is very limited. It includes measurements of the lattice period and the magnetic susceptibility at relatively high temperatures ($T>80$ K), whereas the features of the intermediate-valent state are displayed most clearly at lower temperatures. It is noteworthy that the nonstoichiometric compound Sm_xB_6 itself merits special attention because of its stability when there is a large number of vacancies (up to 30%).

The purpose of the present work was to study and jointly analyze magnetic, kinetic (electrical resistance), and thermodynamic (specific heat and thermal expansion coefficient) properties of the same samples of the stoichiometric compound SmB_6, the Sm-deficient compound $Sm_{0.8}B_6$, and the lanthanum-doped compounds $Sm_xLa_{1-x}B_6$ ($x=0.8$ and 0.9) over the broad temperature range 4–300 K ($2<T<45$ K in the case of the specific heat), as well as to study the influence of doping on the features of the phonon spectrum of SmB_6.

2. SAMPLES AND METHOD

Powdered samples were obtained by the high-temperature reduction of Sm_2O_3 by boron with variation of the ratio between the amounts of the oxide and boron in the Institute of Problems in Materials Science of the Ukrainian National Academy of Sciences (Kiev). The single-phase state of all the samples and their correspondence to the cubic structure of CaB_6 were established by x-ray diffraction analysis.

To refine the structural parameters of Sm_xB_6 and La_xB_6 ($x=0.8$ and 1), a full-profile Rietveld analysis was performed using x-ray and neutron diffraction at room temperature in the range of angles $2\theta=20-160°$. The following parameters were varied during the refinement of the structure: the scale factor, the parameters of the background, the shape of the peaks, the lattice period, the parametric position of the boron atom, and the populations of the sites of the Sm and boron atoms. Because of the small range of variation of the momentum transfer, the thermal factors of the RE and boron atoms were taken from the literature[27,28] and were not varied during fitting. The neutron-diffraction data were treated with allowance for absorption by the RE ions and boron atoms. The full-profile analysis showed that violation of the stoichiometry leads to a decrease in the lattice period ($a=4.1344\pm 0.0003$ Å for SmB_6, $a=4.1278\pm 0.0003$ Å for $Sm_{0.8}B_6$, $a=4.1563\pm 0.0003$ Å for LaB_6, and $a=4.1555\pm 0.0003$ Å for $La_{0.8}B_6$), while the positional parameter of the boron atoms remained essentially unchanged. The analysis also confirmed that the ratio between the populations of the sites of the RE and boron atoms is consistent with the structural formulas. The values of the lattice period at $T=300$ K for $Sm_{0.8}La_{0.2}B_6$ and $Sm_{0.9}La_{0.1}B_6$ were 4.1449 ± 0.0003 Å and 4.1366 ± 0.0004 Å, respectively.

Measurements of the temperature dependence of the lattice period at $T=10-300$ K were performed on polycrystalline samples using x-ray diffraction (Cu K_α radiation) on a DRON-3 diffractometer in the range of angles $2\theta=120-160°$ with a closed-cycle helium refrigerator. The temperature dependence of the thermal expansion coefficient α,

$$\alpha=\frac{1}{a}\frac{da}{dT}, \qquad (1)$$

was obtained by differentiating the temperature dependences of the lattice periods after preliminary smoothing by cubic polynomials.

Measurements of the temperature dependences of the specific heat $C(T)$ in the temperature range from 2 to 45 K were performed in an adiabatic calorimeter.[29]

The possible influence of the defect density in the lattice on the specific heat and the thermal expansion coefficient was evaluated on the basis of a comparison of the data obtained for the lanthanum compounds La_xB_6 ($x=0.8$ and 1). The lattice contribution to the thermal expansion was taken into account by subtracting the corresponding dependences for Sm_xB_6 and La_xB_6 ($x=0.8$ and 1), since the latter are structural analogs, in which, however, the La ions have an empty f shell.

The resistance measurements were performed by the four-probe technique at temperatures between 4 and 300 K.

The magnetic susceptibility was measured on a magnetometer with a magnetic field strength equal to 50 Oe. The temperature range for these measurements was 1.5–300 K for SmB_6 and 5–300 K for $Sm_{0.8}B_6$.

The dispersion curves of the acoustic phonons were obtained using single-crystal, doubly isotopic samples of ${}^{154}Sm_{1-x}La_x^{11}B_6$ ($x=0.1$ and 0.22) having volumes of 0.25 cm^3 ($x=0.22$) and 0.1 cm^3 ($x=0.1$), which were cut from rods obtained from the polycrystalline material by the crucibleless float-zone technique in the Institute of Problems in Materials Science of the Ukrainian National Academy of Sciences. The measurements were performed on a 2T1 triaxial crystal spectrometer (Laboratoire Léon Brillouin, Saclay, France), which provides the high flux of monochromatic neutrons onto the samples needed in experiments with such small samples, which absorb neutrons fairly strongly (by virtue of the residual admixtures of strongly absorbing Sm and boron isotopes). The measurements were performed, for the most part, in a regime with constant momentum trans-

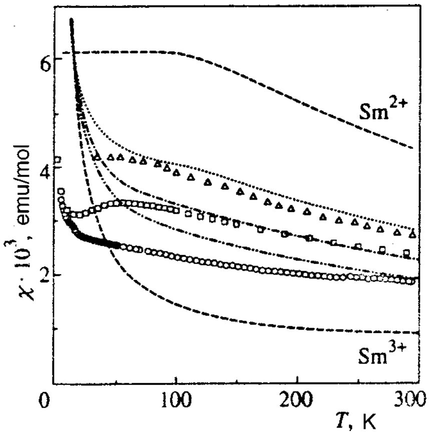

FIG. 1. Temperature dependence of the magnetic susceptibility of samples of SmB_6 (□), $Sm_{0.8}B_6$ (○), and $Sm_{0.75}La_{0.25}B_6$ (△, Ref. 15). Dashed lines — calculated curves for free Sm^{2+} and Sm^{3+} ions; dotted line — calculated curve for $Sm_{0.75}La_{0.25}B_6$ under the assumption of an inhomogeneous state of the Sm ions (see text) $[\chi_{cal}=0.56\chi(Sm^{2+})+0.44\chi(Sm^{3+})]$; dot-dashed curve — calculated curve for SmB_6 $[\chi_{cal}=0.4\chi(Sm^{2+})+0.6\chi(Sm^{3+})]$ under the same assumption; solid curve — calculation for $Sm_{0.8}B_6$ $[\chi_{cal}=0.3\chi(Sm^{2+})+0.7\chi(Sm^{3+})]$ under the same assumption.

fer and fixed final energy $E_f=13.7$ meV ($k_f=2.66$ Å^{-1}) and, in a few cases, in a regime with $E=\mathrm{const}$. Pyrographite (the 002 reflection) served as the monochromator and analyzer, and the higher orders of the reflections were suppressed by a pyrographite filter. The temperature was stabilized by a closed-cycle helium refrigerator.

3. RESULTS

Magnetic susceptibility

Figure 1 presents the temperature dependence of the magnetic susceptibility $\chi(T)$ for the samples of SmB_6 of $Sm_{0.8}B_6$, as well as for $Sm_{0.75}La_{0.25}B_6$ (Ref. 15). The $\chi(T)$ curve for SmB_6 agrees well with the results obtained in Refs. 30 and 31. For all compounds the experimental dependence of the magnetic susceptibility differs from the calculated curve for free samarium ions (Sm^{2+} and Sm^{3+}). It can be seen from the figure that the temperature dependences of the magnetic susceptibility for the samples with an imperfect Sm sublattice vary in accordance with the variation of the valence relative to samarium hexaboride and approximate the curves for the corresponding ions with integer valence. The $\chi(T)$ curves undergo quantitative changes in the temperature range $2<T<100$ K. The characteristic maximum for SmB_6 at $T\simeq 50$ K on the $\chi(T)$ curve vanishes or shifts toward lower temperatures in response to violation of the stoichiometry and shifts toward somewhat higher temperatures in the case of the lanthanum-doped compound.

Electrical resistance

Figure 2 shows the temperature dependence for single-crystal and polycrystalline SmB_6 and nonstoichiometric $Sm_{0.8}B_6$, as well as the results for $Sm_{0.75}La_{0.25}B_6$ from Ref. 13. Both the single crystal and polycrystal of SmB_6 exhibit a sharp increase in resistance as the temperature is reduced.

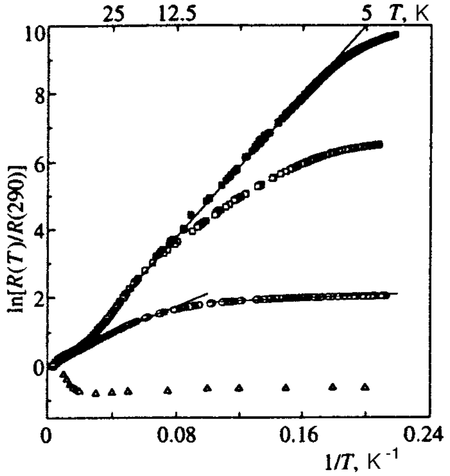

FIG. 2. Temperature dependence of the resistance for SmB_6 [single crystal (■) and polycrystal (□)], $Sm_{0.8}B_6$ (○), and $Sm_{0.75}La_{0.25}B_6$ (△, Ref. 13). Lines — approximation given by (2) (see text).

The temperature curves have linear segments, which are characteristic of a thermally activated type of conduction:

$$R=R_0\exp(-E_g/2k_BT). \tag{2}$$

The segment for the single crystal lies in the temperature range $6<T<20$ K, and the segment for the polycrystal extends over a somewhat smaller range ($8<T<20$ K). The width of the activation gap in the electron density of states obtained from measurements of the resistance on the single crystal is $E_g/2=E_a\approx 50$ K. At $T>35$ K, as well as at $T<5$ K for the single crystal and $T<8$ K for the polycrystal, the temperature dependence of the resistance deviates from a straight line.

The thermally activated type of electrical resistance is maintained as a whole for $Sm_{0.8}B_6$. A linear segment of the resistance curve is observed in the temperature range $15<T<30$ K. The width of the gap in the electron density of states was found to be $E_a\approx 20$ K. Unlike the samples with the stoichiometric composition, the material with $x=0.8$ exhibits an additional linear segment at low temperatures ($T\leqslant 14$ K).

The temperature dependence of the electrical resistance for $Sm_{0.75}La_{0.25}B_6$ has a complex course and is similar on the whole to the dependence for a metal.

Specific heat

Figure 3a presents the data from measuring the temperature dependence of the specific heat $C(T)$ for La_xB_6 and Sm_xB_6 ($x=0.8$ and 1), as well as for $Sm_{0.8}La_{0.2}B_6$. The results for SmB_6 agree well with the data in Ref. 3. As can be seen, violation of the stoichiometry in the valence-unstable samarium hexaboride Sm_xB_6 leads to appreciable changes in the specific heat, while the temperature dependences for the analogous lanthanum-based compounds (La_xB_6, inset in Fig. 3a) essentially coincide.

The electronic contribution to the specific heat, C_{el} (Fig. 3b), was determined as the difference between the total specific heat and its lattice component (C_{lat}, solid line in Fig. 3a). The lattice component was calculated from the phonon

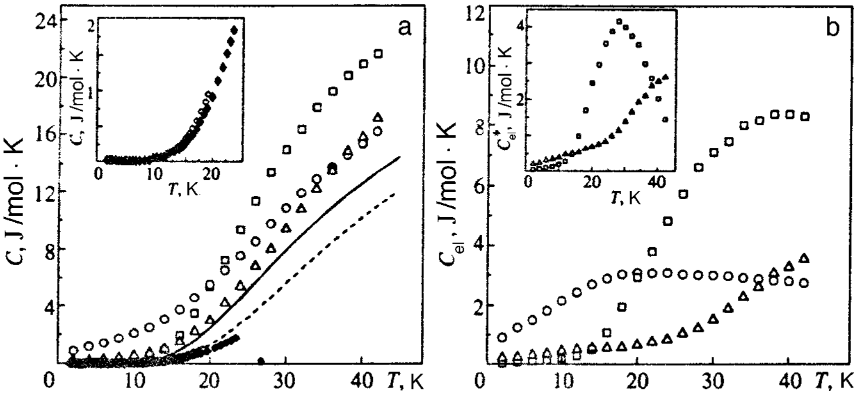

FIG. 3. Temperature dependences of the total specific heat (a) and the electronic contribution to the specific heat (b) for SmB_6 (□), $Sm_{0.8}B_6$ (○), $Sm_{0.8}La_{0.2}B_6$ (△), LaB_6 (◇), $La_{0.8}B_6$ (◆), and LaB_6 (dashed line, Ref. 3). Solid line — lattice component of the specific heat of SmB_6 calculated from the phonon density of states (see text). Inset: a) temperature dependence of the specific heat of La_xB_6 ($x=0.8$ and 1); b) electronic contribution to the specific heat for samples of SmB_6 and $Sm_{0.8}La_{0.2}B_6$ after subtraction of the contribution associated with the *f*-electron excitation spectrum (see text).

density of states for SmB_6, which was obtained from model calculations with consideration of the contribution of the exciton-phonon interaction[20] on the basis of experimentally measured dispersion curves.[17] The electronic contribution to the specific heat of SmB_6 has a maximum at $T\approx 40$ K. The behavior of $C_{el}(T)$ for $Sm_{0.8}B_6$ in the temperature range above 10 K is similar to that for SmB_6; there is a maximum at $T\approx 25$ K, but its amplitude is smaller (Fig. 3b). In contrast to nonstoichiometric samarium hexaboride, the specific-heat maximum for $Sm_{0.8}La_{0.2}B_6$ is probably at a higher temperature ($T>40$ K) than is the specific-heat maximum for SmB_6. The electronic component of the specific heat of the samarium compounds is plotted in C_{el}/T vs. $T^2(T^2)$ coordinates in Fig. 4. At low temperatures $Sm_{0.8}B_6$ exhibits qualitatively different behavior in the temperature dependence of C_{el}/T in comparison to SmB_6: there is a sharp increase in the specific heat with decreasing temperature, which is usually observed for heavy-fermion materials. The estimated value of the Sommerfeld coefficient at $T=2$ K is $\gamma\approx 450$ mJ/mol$\cdot$K^2, which is almost two orders of magnitude greater than the value $\gamma\approx 6$ mJ/mol$\cdot$K^2 for stoichiometric SmB_6. The value of γ also increases somewhat for the lanthanum-doped sample ($\gamma\approx 60$ mJ/mol$\cdot$K^2 at $T=4$ K), but the character of the temperature dependence displayed by SmB_6 is maintained.

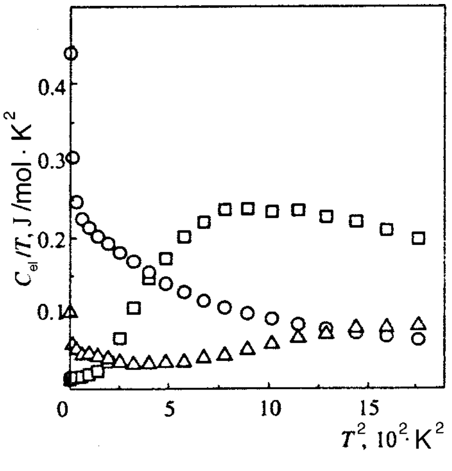

FIG. 4. Temperature dependence of the electronic contribution to the specific heat in C_{el}/T vs. $T^2(T^2)$ coordinates for SmB_6 (□), $Sm_{0.8}B_6$ (○), and $Sm_{0.8}La_{0.2}B_6$ (△).

Thermal expansion

The temperature dependence of the electronic component α_{el} of the thermal expansion coefficient, which was obtained by subtracting the value of $\alpha(T)$ for LaB_6 from the total thermal expansion coefficient for the Sm-containing samples, is presented in Fig. 5. The plot of $\alpha_{el}(T)$ for SmB_6 exhibits a minimum at $T\approx 50$ K, which is well known from the literature.[32] For the lanthanum compounds La_xB_6 ($x=0.8$ and 1) violation of the stoichiometry leads to variation of only the absolute value of the lattice parameter and does not influence the smooth temperature course of the thermal expansion coefficient. Violation of the perfection of the Sm sublattice in samarium hexaboride does not eliminate the minimum in the $\alpha_{el}(T)$ curve but leads to variation of both its position and the area under the $\alpha_{el}(T)$ curve (Fig. 5). In the case of the nonstoichiometric compound $Sm_{0.8}B_6$ the position of the minimum shifts toward lower temperatures ($T\approx 25$ K), and the width of the temperature range where the anomaly exists decreases ($0<T<65$ K). The replacement of Sm by La leads to displacement of the minimum of the anomaly toward higher temperatures ($T\approx 120$ K) and an increase in the area under the $\alpha_{el}(T)$ curve.

Phonon spectra

Figure 6 shows dispersion curves for the longitudinal-acoustic and some transverse-acoustic branches at $T=300$ K

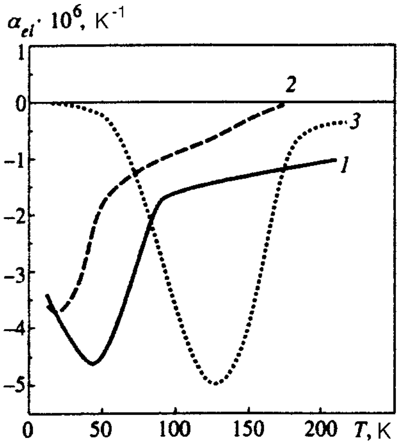

FIG. 5. Temperature dependence of the electronic component of the thermal expansion coefficient for SmB_6 (*1*), $Sm_{0.8}B_6$ (*2*), and $Sm_{0.9}La_{0.1}B_6$ (*3*).

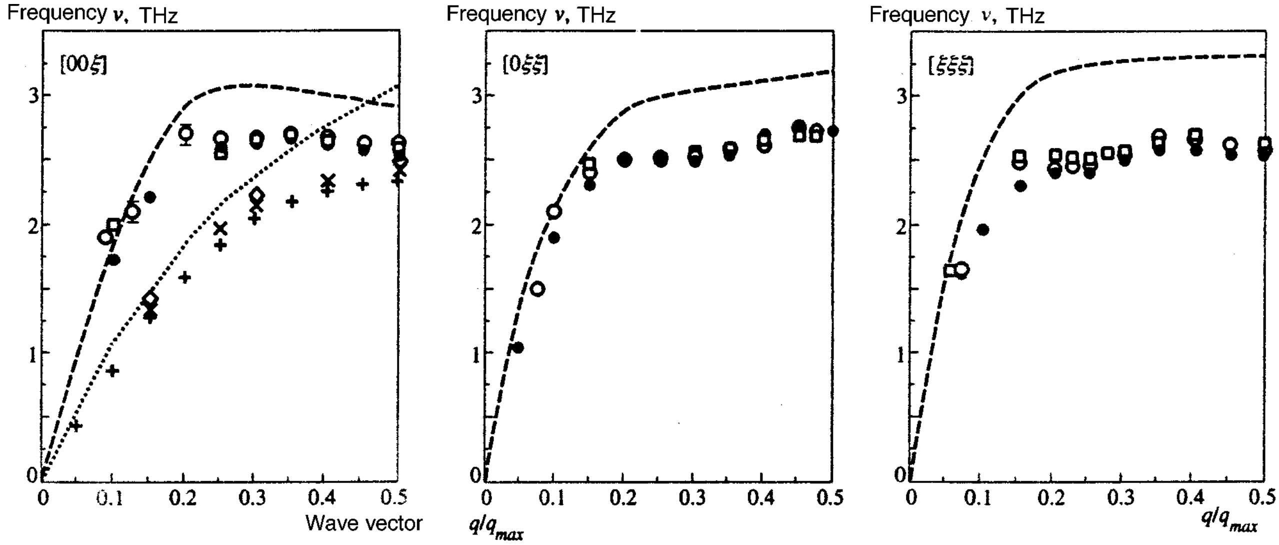

FIG. 6. Dispersion curves for the longitudinal-acoustic (LA) and transverse-acoustic (TA) branches in $Sm_{1-x}La_xB_6$ ($x=0.1$, 0.22), as well as in LaB_6 and SmB_6 (Ref. 17) at $T=300$ K. LA phonons in SmB_6 (●), $Sm_{0.9}La_{0.1}B_6$ (○), and $Sm_{0.78}La_{0.22}B_6$ (□); TA phonons in SmB_6 (+), $Sm_{0.9}La_{0.1}B_6$ (×), and $Sm_{0.78}La_{0.22}B_6$ (◇). Dashed line — LaB_6.

for $Sm_{1-x}La_xB_6$ ($x=0.1$ and 0.22), as well as LaB_6 and SmB_6 (Ref. 17) along the three principal symmetry directions. All the Sm-doped compounds exhibit general softening of the acoustic phonons in comparison to LaB_6, but the changes in their phonon frequencies in comparison to SmB_6 are negligible and amount to about 5% on the average. The anomalies in the longitudinal-acoustic branches along the $[\xi\xi 0]$ and $[\xi\xi\xi]$ directions for a value of the reduced wave vector equal to $q/q_{\max}=0.25$ in SmB_6 are also observed in $Sm_{1-x}La_xB_6$ ($x=0.1$ and 0.22), but they become less pronounced as the La concentration is increased.

Temperature measurements were performed for several phonon frequencies at individual points of the Brillouin zone (Fig. 7). As can be seen from the figure, when the temperature is reduced by 250 K, most of the wave vectors are characterized by changes as large as 0.1 THz in the frequencies of the lattice vibrations. The greatest frequency changes occur in the range $2<T<100$ K.

4. DISCUSSION

An analysis of the magnetic susceptibility measurements supports the hypothesis that an intermediate-valent state exists both in the case of doping and in the case of violation of the stoichiometry. In fact, the decrease in χ in $Sm_{0.8}B_6$ in comparison to SmB_6 over the entire temperature range investigated attests to an increase in the valence of the samarium ions, in agreement with the data in Ref. 23. On the other hand, a decrease in the valence of Sm and an increase in χ are observed in $Sm_{0.75}La_{0.25}B_6$. None of the measured curves coincides with the $\chi(T)$ curves calculated for integer-valent Sm ions (see Fig. 1). In addition, the temperature dependence of $\chi(T)$ does not coincide with the $\chi_{\mathrm{cal}}(T)$ curve obtained under the assumption of an algebraic sum of the contributions $\chi(Sm^{3+})$ and $\chi(Sm^{2+})$ in a ratio corresponding to the mean valence. The temperature range of the most significant deviation from $\chi_{\mathrm{cal}}(T)$ increases with increasing valence of the Sm ions (among the samples measured, this range is broadest for $Sm_{0.8}B_6$).

The results of the direct measurements of $\chi(T)$ for SmB_6 are qualitatively consistent with a calculation of the magnetic susceptibility using the Kramers–Kronig relation from magnetic inelastic neutron scattering spectra.[19] In this case the absolute values of $\chi(T)$ obtained from integration of the neutron spectra are less than the magnetometric values. The disparities between the absolute values are probably due to the strong anisotropy and unusual form factor of the low-energy excitation.[19] The change in the course of $\chi(T)$ for Sm(La)B_6 may be caused by reorganization of its magnetic excitation spectrum in comparison to SmB_6,[33] for which the temperature dependence of the spectral function has been studied in detail. In particular, for SmB_6 the redistribution of the intensity from the region of the inelastic peak with an energy of 14 meV at $T<20$ K into the region of quasielastic scattering at $T>100$ K is responsible for the observed dependence of $\chi(T)$. The growth of the energy of this excitation and the smoothing of its temperature dependence with increasing x in the alloys $Sm_{1-x}La_{x6}$ (Ref. 33) account for the tendency observed in the behavior of $\chi(T)$ for these compounds. As $x\rightarrow 1$, the curves tend to the plot of $\chi(T)$ for the divalent Sm ion. Thus, a simple ''mechanical mixture'' of heterovalent Sm ions is not observed in any of the compounds investigated, in contrast to, for example, the inhomogeneous mixed-valent system Sm_3Te_4.[34] The observed differences between the $\chi(T)$ curves of the imperfect samples can occur mainly because of changes in the mean valence, but with conservation of a homogeneous intermediate-valent state.

Let us consider the results of the resistance measurements (see Fig. 2). The temperature dependence of the resistance $R(T)$ is known to be determined mainly by the behavior of the band electrons, their interaction with localized

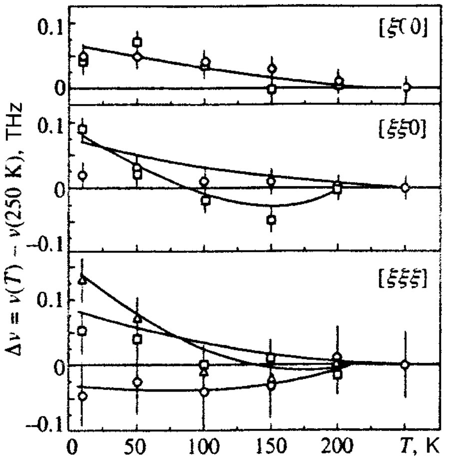

FIG. 7. Temperature dependence of the frequency difference of longitudinal-acoustic phonons in $Sm_{0.9}La_{0.1}B_6$ in the [ξ00], [ξξ0], and [ξξξ] directions for $\xi = 0.35$ (○), 0.45 (□), and 0.25 (△); the curves were drawn by eye.

moments, the presence of impurity levels, etc. As was noted above, samarium hexaboride has a narrow gap ≈ 50 K. The nonlinear portion of the plots of $\ln[R(T)/R(290\mathrm{K})] = f(1/T)$ at $T > 35$ K is most probably due to the temperature dependence of the carrier mobility. The deviation of the temperature dependence from a straight line (Fig. 2) at low temperatures ($T < 6$ K) is due to a change in the type of conduction. Hopping conduction probably becomes decisive at these temperatures. In a polycrystal hopping conduction begins to play an appreciable role at higher temperatures, probably because of the lower degree of purity and the imperfect nature of the crystal, as was recently convincingly demonstrated in Ref. 35. To account for the deviation of $R(T)$ from an exponential law at $T < 3$ K, Bat'ko *et al.*[36] proposed a model based on the appearance of a fine structure in the hybridization gap, which is associated with the presence of a donor or acceptor impurity and lattice defects in the sample.

The character of the temperature behavior of the resistance of nonstoichiometric $Sm_{0.8}B_6$ over the entire temperature range investigated (a linear segment (see Fig. 2) at low temperatures and a sharp decrease in resistance when the temperature is increased) is similar to the behavior of a typical extrinsic semiconductor. There are two segments corresponding to different types of conduction: intrinsic conduction with $E_a \approx 20$ K and impurity conduction due to the high concentration of defects (samarium vacancies), which introduce additional states located several degrees (2–4 K) above the upper edge of the filled band. These states also specify the type of conduction at low temperatures, specifically the hopping type of conduction. The activation gap in the electron density of states is still observed in the samples of $Sm_{0.8}B_6$, but its magnitude is reduced in comparison to the gap in stoichiometric SmB_6.

The introduction of 25% lanthanum into SmB_6, which produces a number of defects close to the number in $Sm_{0.8}B_6$, is sufficient to cause the radical change from a semiconductor type of resistance to a metallic type. It is noteworthy that the change in the character of the $R(T)$ curve of $Sm_{0.75}La_{0.25}B_6$ can be attributed either to disappearance of the gap or the appearance of additional electrons upon doping with La^{3+}.

If there is a gap in the electron density of states, the thermal excitation of electrons through the gap should lead to the appearance of a contribution to the specific heat resembling a Schottky anomaly. The latter can be observed in the electronic contribution to the specific heat of SmB_6 and, in its time, was one of the pieces of evidence for the existence of a gap. Apart from the contribution of the gap to $C_{\mathrm{el}}(T)$, there should also be a contribution associated with the state of the localized f electrons in the Sm ions, which was not analyzed in previous studies. In intermediate-valent samarium hexaboride the f electrons have a ground state, which is a singlet, and a first excited state, which is a triplet.[21] It follows from magnetic inelastic neutron scattering measurements that the characteristic energy of the low-energy excitation corresponding to the transition between these states is approximately 14 meV.[19] Such an excitation spectrum makes a contribution to the specific heat with a maximum at $T \approx 55$ K. After this contribution is taken into account, a component of the specific heat associated only with excitation through the gap remains ($C^*_{\mathrm{el}}(T)$ in the inset in Fig. 3b). Thus, the results of the specific heat measurements convincingly confirm the presence of a gap in the electron density of states. Its width can be roughly estimated using a simple two-level model, and is approximately 60 K. This value is consistent with the width of the gap in the electron density of states obtained from the resistance measurements.

The situation in $Sm_{0.8}La_{0.2}B_6$ is more complicated: the specific heat data are qualitatively similar to the results for SmB_6, and in this respect they differ from the resistance data. It follows from the magnetic inelastic neutron scattering measurements that the characteristic scale of the low-energy excitation associated with the new state of the Sm ion is roughly 25 meV in $Sm_{0.8}La_{0.2}B_6$.[33] The corresponding state makes a contribution to the specific heat at about 110 K, which is significantly higher than the temperatures reached in our experiments. After the contribution from the excited state of the f electrons is subtracted from $C_{\mathrm{el}}(T)$, a contribution to the specific heat which is probably due mainly to the gap remains (see the inset in Fig. 3b). It would seem that despite the course of the temperature dependence of the resistance characteristic of metals, the gap is maintained in the doped sample of $Sm_{0.8}La_{0.2}B_6$. In addition, the lack of any evidence for a gap in the temperature dependence of the resistance may be related to the presence of additional electrons introduced by the La ions, which ''shunt'' the gap in the kinetic measurements. The additional states are manifested by a change in the type of conduction, by the presence of an additional contribution to the specific heat at $T < 15$ K, and by an appreciable increase in the value of γ (see Fig. 4).

The anomaly is also maintained in the plot of $C_{\mathrm{el}}(T)$ for the nonstoichiometric sample (see Fig. 3b). The lack of inelastic neutron scattering data for $Sm_{0.8}B_6$ precludes evaluation of the specific heat component associated with excitations of f electrons. However, an investigation of the magnetic excitations in $Sm_{0.5}Ba_{0.5}B_6$ disclosed the disappearance of the peak characteristic of SmB_6 at 14 meV and the appearance of a low-temperature feature of the magnetic type at about 9 meV, which vanishes when the temperature is raised slightly (at $T=12$ K).[37] Since nonstoichiometry with respect to the samarium ions leads to an increase in the valence of Sm, as does their replacement by Ba ions ($v(Sm_{0.5}Ba_{0.5}B_6)\approx 2.7$), it would seem that in $Sm_{0.8}B_6$ a similar excitation, which is responsible for the intermediate-valent state of the Sm ion, appears at an energy appreciably smaller than 14 meV and produces a contribution to the electronic component of the specific heat only at $T<25$ K. In this case the component $C_{el}(T)$ for $Sm_{0.8}B_6$ at $T=30-35$ K is mostly likely associated with the gap. Thus, the result obtained is consistent with the conclusion following from the resistance measurements. The additional charge carriers that appear in the case of doping by trivalent La ions are not present in $Sm_{0.8}B_6$, and ''shunting'' of the gap does not occur: the gap in the electron density of states is also maintained when there is a 20% samarium deficiency.

Despite the sharp rise observed in the plot of $C/T=f(T^2)$ as $T\rightarrow 0$ (see Fig. 4) for $Sm_{0.8}B_6$, this compound should probably not be classified as a heavy-fermion material. In fact, the value of the magnetic susceptibility (which is an order of magnitude smaller in $Sm_{0.8}B_6$ than in heavy-fermion compounds) and the temperature dependence of the resistance do not correspond to the behavior of heavy-fermion systems. The rise observed can be attributed either to the presence of a low-energy excitation of the samarium ion or to the appearance of additional electron states caused by vacancies in the samarium sublattice in $Sm_{0.8}B_6$. These vacancy-related states can be responsible for the appearance of a segment with a linear dependence of the resistance at $T<10$ K (see Fig. 2). However, it should be noted that an alternative interpretation is possible in principle. More specifically, the possibility of the formation of a heavy-fermion state due to the partial delocalization of f electrons in compounds with a small number of free electrons was advanced in Ref. 38 in connection with the results of investigations of mixed-valent compounds of the type Sm_3X_4, which are also characterized by a large value of γ in the absence of free carriers.

Let us examine the results of the thermal-expansion measurements (see Fig. 5). It is usually assumed that the anomaly in the electronic component $\alpha_{el}(T)$ of intermediate-valent systems is caused by variation of the valence of the RE ion, which is a function of temperature.[39–41] Under such an assumption, in the case of a linear relation between the variation of the valence and the variation of the lattice constant, we can write

$$\frac{d\ln a}{dT}=\frac{d\ln a}{dn_f}\frac{dn_f}{dT}=\text{const}\cdot\frac{dn_f}{dT}, \tag{3}$$

where a is the lattice constant and n_f is the population of the f subshell of the Sm ion. In this case the area S under the anomaly of the electronic component of the thermal expansion coefficient specifies the change in valence as a function of the change in temperature $[\Delta v(\Delta T)]$, and the position of the extremum corresponds to the temperature at which the largest change in the valence related to the effective temperature of the valence fluctuations[39] occurs, i.e.

$$S=\int_{T_1}^{T_2}\text{const}\cdot\frac{dn_f}{dT}dT=\text{const}\cdot\Delta v, \tag{4}$$

$$S=k\Delta v, \tag{5}$$

where k is a proportionality factor. The value $k=0.39$ was obtained on the basis of measurements of the valence of the Sm ion in SmB_6 at $T=295$ and 12 K by x-ray absorption spectroscopy[14] ($\Delta v=0.08$). Knowing the area under the anomaly of $\alpha_{el}(T)$ and assuming that k varies only slightly in response to doping, we can calculate the temperature variation of the valence of the Sm ion in $Sm_{0.8}B_6$ and $Sm_{0.9}La_{0.1}B_6$. The calculation of the corresponding temperature variations of the valence for $Sm_{0.9}La_{0.1}B_6$ gives $\Delta v(300-4\text{K})\approx 0.1$ (the increase in $\Delta v(300-4\text{K})$ in the lanthanum-doped sample is consistent with the result in Ref. 14), and when the stoichiometry is violated ($Sm_{0.8}B_6$), the valence decreases by $\Delta v(300-4\text{K})\approx 0.04$.

In accordance with the Grüneisen relation, the thermal expansion coefficient is proportional to the specific heat. This relationship can be generalized not only to lattice excitations, but also to electron and other excitations. Then, if $C_{el}(T)$ contains a contribution associated with the presence of a gap, it should also be manifested in $\alpha_{el}(T)$. At the same time, the anomalies in the thermal expansion coefficient and the specific heat probably cannot be attributed to the presence of a gap in the electron density of states alone, as was suggested in Ref. 42, since the contribution associated with the f-electron excitation spectrum must be taken into account when the electronic components are considered. In addition, the spectrum of lattice excitations of valence-unstable systems may differ from the spectra for the isostructural compounds devoid of f electrons,[17] which are usually used to take into account the lattice component of the specific heat and the thermal expansion coefficient. Different components of certain properties can be manifested to different degrees. In fact, an examination of the results for all the compounds investigated reveals that, despite the qualitative similarity between the temperature dependences (the existence of extrema and their displacement toward lower temperatures in the nonstoichiometric sample and toward higher temperatures when samarium is replaced by lanthanum, see Figs. 3b and 5), the temperature at the maximum of the electronic component of the specific heat and the temperature at the minimum of $\alpha_{el}(T)$ for the same compound differ significantly. Therefore, $\alpha_{el}(T)$ is determined not just by the contribution from the gap in the systems under consideration. The anomaly in the electronic component $\alpha_{el}(T)$ should also reflect features of the magnetic spectrum of the intermediate-valence compound, primarily changes in the state of the f subshell. In this context we must point out the existence of a correlation between the temperature behavior of the electronic contribution to the thermal expansion coefficient and the magnetic component of the inelastic scattering spectrum of $Sm_{1-x}La_xB_6$. First, the position of the $\alpha_{el}(T)$ anomaly coincides with the temperature at which the low-energy excitation discovered in the magnetic component of the scattering for SmB_6 and $Sm_{0.9}La_{0.1}B_6$ (Ref. 33) is significantly suppressed. Second, both the minimum of $\alpha_{el}(T)$ and the low-energy excitation of SmB_6 are shifted very sharply toward higher energies when only 10% of the Sm atoms are replaced by La. The latter finding is probably attributable to disruption of the coherency in the Sm sublattice, which plays a definite role in shaping the magnetic excitation spectrum.[33] The significant role of the coherency of the RE sublattice was also noted when the thermal expansion of the compounds $Ce_{1-x}(Y,La)_xNi$ was investigated. For example, the position and height of the maximum of the temperature dependence of the electronic component of the thermal expansion coefficient for CeNi deviated from the systematic dependence on concentration. This deviation was associated with the additional contribution to the electronic component of the thermal expansion coefficient due to the Ce–Ce interaction, which is suppressed as a result of doping (when the RE sublattice is disordered.[43]

Thus, the results obtained suggest that the electronic component of the thermal expansion coefficient in intermediate-valence systems is governed not only by the change in valence, which depends on temperature,[39] but also by the spectral characteristics of the ground state of the *f* subshell.

As a final step, let us discuss the measurements of the phonon spectra of $Sm_{1-x}La_xB_6$. As was noted above, lanthanum doping did not lead to significant changes in the phonon spectrum of SmB_6 and the features of the dispersion curves (see Fig. 6). Therefore, it can be stated that the excitonic model[17] proposed for the spectra of lattice vibrations in SmB_6 is probably valid as a whole for diluted systems. It must be emphasized that the exciton state of the samarium ion is extremely stable and does not vanish even after considerable violation of the perfection of the lattice.

The increase in the phonon frequencies with decreasing temperature (see Fig. 7) can be caused either by the presence of a gap in the electron density of states or by variation of the valence of the Sm ion with temperature. The latter cause seems more likely to us. On the one hand, the temperature ranges for variation of the frequencies of the acoustic branches and the existence of anomalies in the macroscopic properties associated with the intermediate-valent state of the samarium ions coincide ($2<T<100$ K). On the other hand, lanthanum doping causes an increase in the lattice constant, which should have led to some softening of the phonon spectrum. As follows from the measurements of samples of $Sm_{1-x}La_xB_6$ at room temperature, doping led to an increase in the frequencies of the acoustic phonons (see Fig. 6). Therefore, the observed changes in the phonon spectrum may be associated with a decrease in valence. In particular, on the basis of this study the increase in the phonon energy with decreasing temperature (Fig. 7) is most probably attributable to the decrease in valence detected in measurements of the L_{III} absorption edge.[14]

5. CONCLUSION

Violation of the perfection of the Sm sublattice leads to significant transformation of the microscopic and macroscopic properties of compounds based on SmB_6. In this case the homogeneous intermediate-valent state of the Sm ion is maintained as a whole in all the compounds investigated, although the mean valence of the Sm ion and the width of the energy gap in the electron density of states vary somewhat. It has been shown in this work that consideration of the reorganization of the *f*-electron excitation spectrum upon variation of the valence is important for understanding the transformation of the physical properties of the compounds investigated.

We express our sincerest thanks to A. S. Mishchenko for fruitful stimulating discussions of this work, J.-M. Mignot and M. Braden for their substantial support of this work, and L. Gurevich for assisting in the measurement of the magnetic susceptibility of nonstoichiometric samarium hexaboride. E.S.K. and P.A.A. thank the administration of the Laboratoire Léon Brillouin (France) for making it possible to perform the neutron experiment. This research was carried out with financial support from the ''Critical Areas in the Physics of Condensed Media'' State Scientific–Technical Program, the Foundation for Supporting Initiatives of a Fundamental and Exploratory Character at the Kurchatov Institute, and the ''Correlation'' Project of the Ministry of Science.

*⁾E-mail: paval@sftt.kiae.su

---

[1] J. C. Nickerson, R. M. White, K. N. Lee *et al.*, Phys. Rev. B **3**, 2030 (1971).

[2] J. W. Allen and R. M. Martin, J. Phys. Colloq. **41**, C5-171 (1980).

[3] T. Kasuya and K. Takegahara, J. Phys. Colloq. **40**, C5-308 (1979).

[4] T. Kasuya, K. Takegahara, Y. Aoki *et al.*, in *Valence Fluctuation in Solids*, L. M. Falikov, W. Hanke, and M. B. Maple (Eds.), North-Holland, Amsterdam (1981), p. 215.

[5] T. S. Al'tshuler, G. G. Khaliullin, D. I. Khomskiĭ, Zh. Éksp. Teor. Fiz. **90**, 2104 (1986) [Sov. Phys. JETP **63**, 1234 (1986)].

[6] B. Batlogg, P. H. Schmidt, and J. M. Rowell, in *Valence Fluctuation in Solids*, L. M. Falikov, W. Hanke, and M. B. Maple (Eds.), North-Holland, Amsterdam (1981), p. 267.

[7] S. von Molnar, T. Theis, A. Benoit *et al.*, in *Valence Instabilities*, P. Wachter and H. Boppart (Eds.), North-Holland, Amsterdam (1982), p. 389.

[8] R. M. Martin and J. M. Allen, J. Appl. Phys. **50**, 7561 (1979).

[9] T. Kasuya, J. Phys. Colloq. **37** C4-26 (1976).

[10] K. A. Kikoin and A. S. Mishchenko, J. Phys.: Condens. Matter **2**, 6491 (1990).

[11] M. Kasaya, H. Kimura, Y. Isikawa *et al.*, in *Valence Fluctuation in Solids*, L. M. Falikov, W. Hanke, and M. B. Maple (Eds.), North-Holland, Amsterdam (1981), p. 251.

[12] V. V. Moshchalkov, I. V. Berman, N. B. Brandt *et al.*, J. Magn. Magn. Mater. **47–48**, 289 (1985).

[13] E. S. Konovalova, Yu. B. Paderno, N. I. Perepelitsa *et al.*, Fiz. Tverd. Tela (Leningrad) **26**, 2138 (1984) [Sov. Phys. Solid State **26**, 1296 (1984)].

[14] J. M. Tarascon, Y. Isikawa, B. Chevalier *et al.*, J. Phys. (Paris) **41**, 1141 (1980).

[15] J. M. Tarascon, Y. Isikawa, B. Chevalier *et al.*, J. Phys. (Paris) **41**, 1135 (1980).

[16] R. L. Cohen, M. Eibschütz, and K. W. West, Phys. Rev. Lett. **24**, 383 (1970).

[17] P. A. Alekseev, A. S. Ivanov, B. Dorner *et al.*, Europhys. Lett. **10**, 457 (1989).

[18] P. A. Alekseev, Physica B **186–188**, 365 (1993).

[19] P. A. Alekseev, J.-M. Mignot, J. Rossat-Mignod *et al.*, J. Phys.: Condens. Matter **7**, 289 (1995).

[20] K. A. Kikoin and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **104**, 3810 (1993) [JETP **77**, 828 (1993)].

[21] K. A. Kikoin and A. S. Mishchenko, J. Phys.: Condens. Matter **7**, 307 (1995).

[22] T. Kasuya, Europhys. Lett. **26**, 277 (1994).

[23] E. S. Konovalova, Yu. B. Paderno, T. Lundstrem *et al.*, Poroshk. Metall. **10**, 78 (1982).

[24] Yu. B. Paderno, E. S. Konovalova, N. L. Baturinskaya *et al.*, Neorg. Mater. **18**, 47 (1982).

[25] M. I. Aĭvazov, S. V. Aleksandrovich, B. A. Evseev *et al.*, Neorg. Mater. **19**, 211 (1983).

[26] M. I. Aĭvazov, S. V. Aleksandrovich, B. A. Evseev *et al.*, Neorg. Mater. **16**, 59 (1980).

[27] A. A. Eliseev, V. A. Efremov, G. M. Kuz'micheva *et al.*, Kristallografiya **31**, 803 (1986) [Sov. Phys. Crystallogr. **31**, 476 (1986)].

[28] V. A. Trounov, A. L. Malyshev, D. Yu. Chernyshov *et al.*, J. Phys.: Condens. Matter **5**, 2479 (1993).

[29] M. N. Khlopkin, N. A. Chernoplekov, P. G. Cheremnykh, Preprint IAE-3549/10, Institute of Atomic Energy, Moscow (1982).

[30] M. Kasaya, J. M. Tarascon, and J. Etourneau, Solid State Commun. **33**, 1005 (1980).

[31] J. Roman, V. Pavlik, K. Flachbart *et al.*, Physica B **230–232**, 715 (1997).

[32] P. A. Alekseev, E. S. Konovalova, V. N. Lazukov *et al.*, Fiz. Tverd. Tela (Leningrad) **30**, 2024 (1988) [Sov. Phys. Solid State **30**, 1167 (1988)].

[33] P. A. Alekseev, J.-M. Mignot, V. N. Lazukov *et al.*, J. Solid State Chem. **133**, 230 (1997).

[34] M. Sugita, S. Kunii, K. Takegahara, N. Sato *et al.*, in *Crystalline Electric Field in f-Electron Magnetism*, Plenum Press, New York (1982), p. 479.

[35] A. Kebede, M. C. Aronson, C. M. Buford *et al.*, Physica B **223–224**, 256 (1996).

[36] I. Bat'ko, P. Farkasovsky, K. Flachbart *et al.*, Solid State Commun. **88**, 405 (1993).

[37] P. A. Alekseev, V. N. Lazukov, R. Osborn *et al.*, Zh. Éksp. Teor. Fiz. **108**, 1064 (1995) [JETP **81**, 586 (1995)].

[38] T. Suzuki, in *Physical Properties of Actinide and Rare Earth Compounds (JJAP, Series 8)*, T. Kasuya, T. Ishii, T. Komatsubara *et al.*, (Eds.) (1993), p. 267.

[39] N. E. Bickers, D. L. Cox, and J. M. Wilkins, Phys. Rev. B **36**, 2036 (1987).

[40] E. Müller-Hartmann, Solid State Commun. **31**, 113 (1979).

[41] G. Gangadhar Reddy and A. Ramakanth, Solid State Commun. **78**, 133 (1991).

[42] D. Mandrus, J. L. Sarrao, A. Lacerda *et al.*, Phys. Rev. B **49**, 16 809 (1994).

[43] V. N. Lazukov, P. A. Alekseev, E. S. Klement'ev *et al.*, Zh. Éksp. Teor. Fiz. **113**, 1731 (1998) [JETP **86**, 943 (1998)].

Translated by P. Shelnitz

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Fluctuation states and optical spectra of solid solutions with a strong isoelectronic perturbation

A. A. Klochikhin

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

S. A. Permogorov and A. N. Reznitskiĭ

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

(Submitted 30 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1039–1062 (March 1999)

We propose an approach to describing the density of fluctuation states in a disordered solid solution with a strong perturbation introduced by isoelectronic substitution in the range of attraction-center concentrations below the threshold of percolation along the sites of a disordered sublattice. To estimate the number of localized states we use the results of lattice percolation theory. We describe a method for distinguishing, within the continuum percolation theory, among the various ''radiating'' states of the fluctuation-induced tail, states that form the luminescence band at weak excitation. We also establish the position of the band of radiating states in relation to the absorption band of the excitonic ground state and the mobility edge of the system. The approach is used to describe the optical spectra of the solid solution $ZnSe_{1-c}Te_c$, which at low Te concentrations can be interpreted as a system with strong scattering. We take into account the exciton–phonon interaction and show that the calculated and observed luminescence spectra of localized excitons are in good agreement with each other.
[S1063-7761(99)02003-X]

1. INTRODUCTION

Wide-gap solid solutions of the II–VI groups are being actively used in manufacturing heterostructures. It is well known that the changes in the optical characteristics of a solid solution in relation to those of the compounds forming that solution is determined chiefly by two effects: the concentration variation of the band gap and the smearing of the gap edge due to the formation of a fluctuation-induced tail of localized states.

In the absence of long-range order, the description of the spectrum of the electronic states of a crystal is an important problem of the physics of disordered systems and has remained in the fore for many years. Reviews of earlier work that cover the main directions of research in this field can be found in Refs. 1–4. A separate case is that of very low concentrations of impurity atoms, $c \ll 1$, where the perturbation at a lattice site exceeds a critical value and leads to formation of split-off states. This problem is examined by Koster and Slater[5] (see also the results of more recent work in Refs. 6 and 7) for electronic states and by Lifshitz[8] for the oscillation problem. For finite concentrations, attempts to solve for the spectrum of impurity states were made in the coherent potential approximation by Elliott *et al.*[9] and by using more complex approaches.[10,11] Disordered solid solutions of crystals form a group of systems for which allowance of the statistics of lattice-site occupation is critical to the understanding of the localization-energy distributions of what is known as the tail fluctuation states. In this connection, considerable attention has been paid to random-potential studies (a potential whose statistics is that of Gaussian white noise[12–14]). This, however, does not correspond to the physical nature of solid solutions.[15] Some aspects of tail fluctuation states with the atoms of two species distributed at random over the lattice sites have been examined in Refs. 16 and 17.

In this paper we develop a systematic approach to calculating the density of fluctuation states of disordered solid solutions to describe the optical absorption and luminescence spectra of such solutions for the case in which the perturbation of electronic states introduced by the substituting atoms cannot be considered small and a localization state can be formed by a few substituting atoms (the strong scattering limit). The case of weak scattering was examined in Refs. 18 and 19.

In Sec. 2 we find the density of localized states by using a three-dimensional two-component model described by the one-band Hamiltonian with diagonal disorder. This type of disorder emerges as a result of random occupation of the sites of a crystal lattice by atoms of two species, e.g., *A* and *B*. We also use the continuum variant of this model. Here we limit ourselves to the region of low concentrations of *A* atoms, which are attraction centers, $c < p_c$, where p_c is the critical concentration in the problem of percolation along the sites of the sublattice of the substituting atoms. When there is strong scattering in solid solutions with diagonal disorder, fluctuation states may be formed by individual substituting atoms and by relatively small clusters of such atoms, which form when the atoms are randomly distributed.

In Sec. 3 we calculate the contour of the phononless luminescence band. Within the given problem, the possibility of optical recombination is limited by the exciton lifetime in relation to the transitions (tunneling) into low-lying states belonging to the tail of fluctuation states, with the transitions accompanied by phonon emission. To describe luminescence we must find the energy distribution of isolated states, i.e., states for which the channels of such relaxation are either closed or barely exist. The restriction here is due to the finiteness of the tunneling radius. Of the total number of tail states only those are isolated for which the distance to neighboring states is greater than this radius. Generally, the problem of finding the energy distribution of spatially isolated states belongs to the quantum percolation theory.[2–4,20–29] Isolated states in lattice models were studied in Refs. 30–36. If the finiteness of the tunneling radius is taken into account, one can employ the continuum theory of percolation over overlapping spheres.[30–36] Using the results of this theory, we can divide fluctuation states into states of single (noninteracting) clusters, states of superclusters (complexes consisting of interacting clusters), and percolation cluster states. We assume all along that the states of single clusters and the ground states of superclusters are spatially isolated.

Knowing the distribution of states among superclusters makes it possible to establish the position of the phononless luminescence band in relation to the peak in the phononless absorption band of the excitonic ground state and the mobility edge.

In Sec. 4 we describe the effect of the interaction of excitons and acoustic and optical phonons on the absorption and luminescence spectra of localized excitons. In the concentration range of interest to us, we assume that an electron interacts with the hole density distribution averaged over the fluctuation-induced well. The preliminary results of allowing for the exciton–phonon interaction can be found in Ref. 37.

Section 5 discusses the results of numerical calculations of the absorption and luminescence spectra for the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ at $c=0.15$, which at such a concentration may serve as a good model of systems with strong scattering. There we also compare the contours of the calculated and observed luminescence bands.

2. SPECTRUM OF FLUCTUATION STATES

In the model developed in this paper we assume that the macroscopic volume of the solid solution consists of N crystal lattice sites occupied at random by two species of atoms, A and B. The average numbers of A and B atoms are, respectively, $N_A=cN$ and $N_B=(1-c)N$, where c is the concentration of A atoms. The one-band Hamiltonian of this system can be written

$$H=-\sum_{\mathbf{n},\mathbf{m}}\Psi_{\mathbf{n}}W_{\mathbf{m}}(\Psi_{\mathbf{n+m}}-\Psi_{\mathbf{n}})+\sum_{\mathbf{n}}E_{\mathbf{n}}\Psi^2_{\mathbf{n}}, \tag{1}$$

where $\Psi_{\mathbf{n}}$ is assumed real. The diagonal element $E_{\mathbf{n}}$ takes the value E_A if the site is occupied by an A atom and the value E_B in the opposite case. In the limits $c=0$ and $c=1$ the Hamiltonian (1) reduces to the Hamiltonian of a perfect crystal consisting of B or A atoms. Here all $E_{\mathbf{n}}$ are replaced by $E_{\mathbf{n}}=E_A$ or $E_{\mathbf{n}}=E_B$. In the plane wave representation, the matrix element of the Hamiltonian in this case becomes $H^{\alpha}_q=\varepsilon_q+E_\alpha$, where α is either A or B, and

$$\varepsilon_q=W_0-W_q \tag{2}$$

is the dispersion of the electron band.

2.1. The Green's function representation

The Green's function of a perfect crystal A or B is

$$G^{\alpha}_{\mathbf{nm}}(\epsilon)=\{\epsilon\mathbf{I}-\mathbf{H}^{\alpha}\}^{-1}_{\mathbf{nm}}=\frac{1}{N}\sum_{\mathbf{q}}\frac{\exp\{i\mathbf{q}\cdot(\mathbf{R}_{\mathbf{n}}-\mathbf{R}_{\mathbf{m}})\}}{\epsilon-\varepsilon_{\mathbf{q}}-E_\alpha}, \tag{3}$$

where $\mathbf{R}_{\mathbf{n}}$ is the radius vector of the lattice site $\mathbf{n}$. When the lattice sites are occupied at random by atoms of two species, the Green's function for each realization depends on the composition and arrangement of the atoms:

$$G_{\mathbf{nm}}(\omega)=\{\omega\mathbf{I}-\mathbf{H}^{vc}-\boldsymbol{\Delta}\}^{-1}_{\mathbf{nm}}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix, and the matrices $\mathbf{H}^{vc}$ and $\boldsymbol{\Delta}$ are

$$H^{vc}_{\mathbf{nm}}=W_{\mathbf{n}-\mathbf{m}}-(W_0-E_G)_{\mathbf{n}}\delta_{\mathbf{nm}},\quad \boldsymbol{\Delta}=\Delta_{\mathbf{n}}\delta_{\mathbf{nm}}, \tag{5}$$

with E_G the edge of the band considered.

The position of the bottom of the band of the solid solution in the virtual crystal approximation is related to the average value of $E_{\mathbf{n}}$ at the site:

$$\langle E\rangle_G=cE_A+(1-c)E_B. \tag{6}$$

The energy ω is also measured from the band bottom E_G, with ω positive for localized states. The diagonal matrix elements of the potential $\boldsymbol{\Delta}$ can be written

$$\Delta_{\mathbf{n}}=E_{\mathbf{n}}-E_G, \tag{7}$$

where the value of $\Delta_{\mathbf{n}}$ also depends on the concentration of the solid solution.

The solution of the eigenvalue problem for the crystal Hamiltonian when the atoms of the two species occupy fixed but disordered positions can be obtained by diagonalizing a matrix of rank N, each row in which has the form

$$\omega_\Lambda\varphi^{\boldsymbol{\Delta}}_\Lambda(\mathbf{n})+\sum_{\mathbf{m}}(H^{vc}_{\mathbf{nm}}-\Delta_{\mathbf{n}}\delta_{\mathbf{nm}})\varphi^{\boldsymbol{\Delta}}_\Lambda(\mathbf{m})=0. \tag{8}$$

The spectrum of the disordered system can be found as a result of averaging, i.e., by summing the spectra corresponding to different realizations with weight factors equal to the probabilities of each realization. Of interest from the standpoint of physics here is the imaginary part of the Green's function averaged over all possible variants of occupation of the sites by atoms.

The expression for the density of states for a fixed set of values of $\Delta_{\mathbf{n}}$ and fixed values of quantum numbers, which we denote by Λ, has the form

$$\rho_{\boldsymbol{\Delta},\Lambda}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau_\Lambda\sum_{\mathbf{n}}|\varphi^{\boldsymbol{\Delta}}_\Lambda(\mathbf{n})|^2 \times\exp\{-i[\omega\mathbf{I}-\mathbf{H}^{vc}-\boldsymbol{\Delta}]_{\Lambda\Lambda}\tau_\lambda\}, \tag{9}$$

where

$$[\omega\mathbf{I}-\mathbf{H}^{vc}-\boldsymbol{\Delta}]_{\Lambda\Lambda}=\sum_{\mathbf{n},\mathbf{m}}\varphi_{\Lambda}^{\boldsymbol{\Delta}}(\mathbf{n})[\omega\delta_{\mathbf{nm}}+H^{vc}_{\mathbf{nm}}-\Delta_{\mathbf{n}}\delta_{\mathbf{nm}}]\varphi_{\Lambda}^{\boldsymbol{\Delta}}(\mathbf{m}). \qquad (10)$$

The total average density of states can be written as a sum over all Λ and all realizations of $\boldsymbol{\Delta}$ with allowance for their weight factors $P_{\boldsymbol{\Delta}}$:

$$\rho(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{\boldsymbol{\Delta}}P_{\boldsymbol{\Delta}}\sum_{\Lambda}d\tau_{\Lambda}\sum_{\mathbf{n}}|\varphi_{\Lambda}^{\boldsymbol{\Delta}}(\mathbf{n})|^{2}\times\exp\{-i[\omega\mathbf{I}+\mathbf{H}^{vc}-\boldsymbol{\Delta}]_{\Lambda\Lambda}\tau_{\Lambda}\}. \qquad (11)$$

Consequently, we must consider all possible variants of lattice occupation and for each variant find the eigenvalues and eigenfunctions. Integration with respect to τ_Λ in (11) can be carried out in general form, which yields an expression for $\rho(\omega)$:

$$\rho(\omega)=\sum_{\boldsymbol{\Delta}}P_{\boldsymbol{\Delta}}\sum_{\Lambda}\sum_{\mathbf{n}}|\varphi_{\Lambda}^{\boldsymbol{\Delta}}(\mathbf{n})|^{2}\delta(\omega-\omega_{\Lambda}^{\boldsymbol{\Delta}}). \qquad (12)$$

Here we denote the eigenvalue of Eq. (8) for a fixed set of $\Lambda_{\mathbf{n}}$ by $\omega_{\Lambda}^{\boldsymbol{\Delta}}$. For bound states the functions $\varphi_{\Lambda}^{\boldsymbol{\Delta}}$ can always be normalized to unity.

The above formula is a rigorous definition of the density of states of the disordered system described by Hamiltonian (1), i.e., a two-component solid solution A_cB_{1-c}.

2.2. Fluctuation states of clusters. Sum rules

In this paper we examine a situation in which the concentration of *A* atoms is lower than the critical value for the problem of percolation along the sites of the anion sublattice: $c<p_c\approx 0.2$. Under this condition, the *A* atoms can form only finite clusters. If the perturbation introduced by a cluster consisting of *n* atoms of the species *A* exceeds a critical value and splits off a state from the bottom of the virtual-crystal band, all higher-order clusters will also split off localized states from the bottom.

For the zeroth approximation we assume that the wave functions of individual clusters do not overlap and examine the resulting spectrum of the split-off states. Within this approach the medium outside a cluster is described by the virtual crystal approximation. In this approximation the spectrum of the density of states is a set of delta-function peaks corresponding to the individual clusters. Next we allow for composition fluctuations in lowest-order perturbation theory. Finally, we apply a variation procedure to determine the most-probable fluctuations. As a result the individual peaks become smeared and the structure in the spectrum of the density of states is smoothed out. At the same time, the structural features of the spectrum can be observed at very low concentrations.

2.2.1. Isolated cluster approximation

For any cluster in the approximation we can write the equation of motion as

$$\{\mathbf{I}+\mathbf{g}^{vc}(\omega)\boldsymbol{\Delta}\}_{\mathbf{nm}}\varphi_{\Lambda}^{\boldsymbol{\Delta}}(\mathbf{m})=0, \qquad (13)$$

where **n** and **m** run through the values corresponding to the coordinates of the sites occupied by clusters, and $\mathbf{g}^{vc}(\omega)$ is the fragment of the Green's function matrix limited to the region that is occupied by the cluster. Using the approach developed in the theory of impurity lattice dynamics,[38] we find the eigenvalues and eigenfunctions of the matrix $[\mathbf{g}^{vc}(\omega)\boldsymbol{\Delta}]$ with rank equal to the cluster size, i.e., we solve the equation

$$\sum_{\mathbf{m}\in\mathscr{C}_{st,\kappa}}[\mathbf{g}^{vc}(\omega)\boldsymbol{\Delta}]_{\mathbf{n},\mathbf{m}}\Phi^{\sigma}(\mathbf{m},\omega)=\lambda^{\sigma}(\omega)\Phi^{\sigma}(\mathbf{n},\omega), \qquad (14)$$

where the sum is over sites occupied by clusters. The resulting eigenfunctions constitute a complete orthonormal set:

$$\sum_{\sigma}\Phi^{\sigma}(\mathbf{m},\omega)\Phi^{\sigma}(\mathbf{n},\omega)=\delta_{\mathbf{nm}}. \qquad (15)$$

Using these eigenfunctions and eigenvalues, we can write the matrix $\{\mathbf{I}+\mathbf{g}^{vc}(\omega)\boldsymbol{\Delta}\}^{-1}$ in the form

$$\{\mathbf{I}+\mathbf{g}^{vc}(\omega)\boldsymbol{\Delta}\}^{-1}_{\mathbf{nm}}=\sum_{\sigma}\frac{\Phi^{\sigma}(\mathbf{n},\omega)\Phi^{\sigma}(\mathbf{m},\omega)}{1-\lambda_{\sigma}(\omega)}. \qquad (16)$$

In the given approximation a localized state splits off if at least one σ has a positive value $\omega=\omega_{\mathrm{loc}}>0$ at which

$$\lambda_{\sigma}(\omega_{\mathrm{loc}})=1. \qquad (17)$$

The state with the nodeless wave function always splits off first; such a state has the deepest value of the localization energy and provides the greatest contribution to optical processes.

The wave functions of the split-off localized states normalized to unity are, respectively,

$$\phi^{\sigma}(\mathbf{n})=\sum_{\mathbf{m}\in\mathscr{C}_{st,\kappa}}G^{vc}_{\mathbf{n},\mathbf{m}}(\omega)\Phi^{\sigma}_{\mathbf{m}}(\omega)\times\left|\sum_{\mathbf{n},\mathbf{m}\in\mathscr{C}_{st,\kappa}}\Phi^{\sigma}_{\mathbf{n}}G'^{vc}_{\mathbf{n},\mathbf{m}}(\omega)\Phi^{\sigma}_{\mathbf{m}}\right|^{-1/2}, \qquad (18)$$

where $G'^{vc}_{\mathbf{n},\mathbf{m}}(\omega)=\partial G^{vc}_{\mathbf{n},\mathbf{m}}(\omega)/\partial\omega$, and the sum is over sites occupied by clusters.

Suppose that the probability that a cluster consisting of *s* atoms of species *A* and having a perimeter of *t* atoms of species *B* appears is

$$g_{st}c^{s}(1-c)^{t}, \qquad (19)$$

where g_{st} is the number of clusters having the same number of atoms of species *A* and the same number of atoms of species *B* at the perimeter, but different spatial configurations. Allowing only for nodeless states, instead of the expression (11) for the total average density of states we can write

$$\rho(\omega)=\sum_{s}\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\Phi^{st,\kappa}_{\mathbf{n}}|^{2}c^{s}(1-c)^{t}\delta(\omega-\omega^{st,\kappa}_{\mathrm{loc}}). \qquad (20)$$

Here we have numbered the eigenfunctions and eigenvalues according to their membership in the cluster. The overall density of the states that have split off from bottom of the band of the virtual crystal per site is

$$\mathcal{N}(0)=\int_0^{E_L}\rho(\omega)\,d\omega=\sum_{st,\kappa}g_{st}\,c^s(1-c)^t=\sum_s\,n_s(c), \tag{21}$$

i.e., it coincides with the total number of clusters per site.[39] To calculate the number of states whose localization energy exceeds ω, we must replace the lower limit of integration in (21) with this quantity and begin summation on the right-hand side of this expression with cluster sizes s that lead to the emergence of a state with a localization energy equal to ω. The maximum number of split-off states of all types cannot exceed the number of attraction centers, i.e., cannot exceed the quantity

$$N_A=N\sum_{s=1}^{\infty}n_s(c)s. \tag{22}$$

This estimate works in the strong scattering limit.

Sums of type (21) rapidly converge over the entire range of concentrations. Since the functions $n_s(c)$ are known for many lattices, at least for several small values of s, the resulting sum rule (21) can be used to estimate the number of states that have split off the bottom of the exciton band of the crystal. For instance, for an fcc lattice, de Gennes *et al.*[20,21] found that

$$\begin{gathered}n_1(c)=c(1-c)^{12},\quad n_2(c)=12c^2(1-c)^{18},\\ n_3(c)=c^3[24(1-c)^{23}+126(1-c)^{24}].\end{gathered} \tag{23}$$

Our estimates of the probability of the emergence of a cluster consisting of four atoms yield

$$n_4(c)\approx10^3c^4(1-c)^{30}. \tag{24}$$

These functions can be used along with the sum rule (21) to estimate the number of states that form below the bottom of the band of a virtual crystal. All further information can be obtained if we know, at least approximately, the dependence of the localization energy on the number of A atoms in the cluster. This makes it possible to evaluate the value and behavior of the total density of states and the nature of variation of the density of states.

2.2.2. Effect of fluctuations on cluster states. Perturbation theory

Up to this point we have assumed that the clusters are surrounded by a homogeneous medium, or a virtual crystal. To assess the role of fluctuations, we plug the solutions (18) into the expression (11) and find the first nonvanishing correction to the density of states for which the deviation from the virtual crystal approximation is responsible. As a result we obtain

$$\begin{aligned}\rho(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t\\ \times\exp\{-i(\omega-\omega_{\mathrm{loc}}^{st,\kappa})\tau\}\\ \times\left\langle\exp\left\{-i\sum_{\mathbf{n}\ni\mathcal{C}_{st,\kappa}}\Delta_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau\right\}\right\rangle,\end{aligned} \tag{25}$$

where the angle brackets indicate averaging over all possible realizations of occupation of the entire lattice except for a single specified cluster, and summation in the exponent of the exponential function is over the lattice sites that lie outside the cluster. When the occupation of the states is a random process, each is occupied by an A atom with probability c and a B atom with probability $1-c$. Bearing in mind that the exponential being averaged becomes a product of individual exponentials, each of which is averaged independently, we find that

$$\begin{aligned}&\left\langle\exp\left\{-i\sum_{\mathbf{n}\ni\mathcal{C}_{st,\kappa}}\Delta_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau\right\}\right\rangle\\ &\quad=\prod_{\mathbf{n}\ni\mathcal{C}_{st,\kappa}}^{N}\{c\exp(-i\Delta_A|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau)\\ &\quad+(1-c)\exp(-i\Delta_B|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau)\},\end{aligned} \tag{26}$$

where

$$\begin{aligned}\Delta_A&=E_A-E_G=-(1-c)\Delta,\\ \Delta_B&=E_B-E_G=c\Delta,\\ \Delta&=E_B-E_A>0.\end{aligned} \tag{27}$$

The averaged expression for the density of states can be written

$$\begin{aligned}\rho(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t\\ \times\exp\left\{i(\omega-\omega_{\mathrm{loc}}^{st,\kappa})\tau+\sum_{\mathbf{n}\in\mathcal{C}_{st,\kappa}}\ln[R_{\mathbf{n}}(\tau)]\right\},\end{aligned} \tag{28}$$

where

$$\begin{aligned}R_{\mathbf{n}}(\tau)=c\exp[i\Delta(1-c)|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau]\\ +(1-c)\exp[-i\Delta c|\phi_{\mathbf{n}}^{st,\kappa}|^2\tau].\end{aligned} \tag{29}$$

Confining ourselves to the first nonvanishing cumulant, we arrive at an expression for the density of states,

$$\begin{aligned}\rho(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t\\ \times\exp\left\{i(\omega-\omega_{\mathrm{loc}}^{st,\kappa})\tau-\frac{\gamma_{st,\kappa}^2\tau^2}{2}\right\},\end{aligned} \tag{30}$$

where

$$\gamma_{st,\kappa}^2=\sum_{\mathbf{n}\ni\mathcal{C}_{st,\kappa}}c(1-c)\Delta^2[|\phi_{\mathbf{n}}^{st,\kappa}|^2]^2,$$

with the sum in the last expression being taken over the lattice sites outside the given cluster. Evaluating the integral with respect to τ, we obtain

$$\begin{aligned}\rho(\omega)=\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t\frac{1}{\sqrt{2\pi\gamma_{st,\kappa}^2}}\\ \times\exp\left\{-\frac{(\omega-\omega_{\mathrm{loc}}^{st,\kappa})^2}{2\gamma_{st,\kappa}^2}\right\}.\end{aligned} \tag{31}$$

This expression differs from (20) in that the localization energy of each cluster is determined to within $\gamma_{st,\kappa}$. If the localization energies and $\gamma_{st,\kappa}$ prove to be of the same order, the total number of states that have split off the bottom of the band of the virtual crystal is smaller than when fluctuations were ignored [Eq. (21)], i.e.,

$$\mathscr{N}(0)=\int_0^{E_L}\rho(\omega)\,d\omega\leqslant\sum_s n_s(c). \tag{32}$$

This result is limited to the range of applicability of the perturbation theory, i.e., the lower the cluster concentration and the higher the localization energy of the cluster states that provide the leading contribution to the density of states, the better the result of using perturbation theory techniques.

2.2.3. Effect of fluctuations on cluster states. Variational approach

We now consider the problem of calculating the density of states by the variational method. A new element in this problem is the great diversity of clusters that split off states from the band bottom. This means we cannot select a variational wave function in a unique manner. For further calculations it is convenient to pass from the Fourier integral (30) to the density of states expressed in terns of the Laplace transform:

$$\rho(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\tau\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t \times\exp\left\{-i\left(\sum_{\mathbf{nm}}\phi_{\mathbf{n}}^{st,\kappa}[H_{\mathbf{nm}}^{vc}+\omega\delta_{\mathbf{nm}}]\phi_{\mathbf{m}}^{st,\kappa}\right)\times(\tau-i\theta_{st,\kappa})+\sum_{\mathbf{n}\ni\mathscr{C}_{st,\kappa}}\ln[R_{\mathbf{n}}(\tau-i\theta_{st,\kappa})]\right\}. \tag{33}$$

The variational problem in the independent-cluster approximation reduces to a set of independent nonlinear equations of motion for bound states of clusters, each of which has the form

$$\sum_{\mathbf{nm}}[H_{\mathbf{nm}}^{vc}+\omega\delta_{\mathbf{nm}}]\phi_{\mathbf{m}}^{st,\kappa}+U_{st,\kappa}(\mathbf{n})\phi_{\mathbf{n}}^{st,\kappa}=0. \tag{34}$$

For each cluster the potential energy within the cluster can be specified uniquely and does not vary. The most probable configuration corresponds to the case in which the potential energy $U_{st,\kappa}(\mathbf{n})$ of the cluster st,κ is taken in the form

$$U_{st,\kappa}(\mathbf{n})=\begin{cases}\Delta_{\mathbf{n}}, & \mathbf{n}\in\mathscr{C}_{st,\kappa},\\ -\dfrac{\partial\ln R_{\mathbf{n}}(-i\theta_{st,\kappa})}{\partial\theta_{st,\kappa}|\phi_{\mathbf{n}}^{st,\kappa}|^2}, & \mathbf{n}\ni\mathscr{C}_{st,\kappa},\end{cases} \tag{35}$$

where the matrix element $\Delta_{\mathbf{n}}$ has the value $[-(1-c)\Delta]$ at sites s and the value $c\Delta$ at sites t, and

$$-\frac{\partial\ln R_{\mathbf{n}}(-i\theta_{st,\kappa})}{\partial\theta_{st,\kappa}|\phi_{\mathbf{n}}^{st,\kappa}|^2}=-(1-c)\Delta\left[1-\frac{\exp(-\Delta\theta_{st,\kappa}|\phi_{\mathbf{n}}^{st,\kappa}|^2)}{c+(1-c)\exp(-\Delta\theta_{st,\kappa}|\phi_{\mathbf{n}}^{st,\kappa}|^2)}\right]. \tag{36}$$

Thus, the variational procedure yields the most probable configuration of fluctuations outside the cluster. Note that the self-consistent solutions of the equations for the bound states yield the functions $\phi_{\mathbf{n}}^{st,\kappa}$ and, at the same time, fix the values of the free parameters of the Laplace transform, $\theta_{st,\kappa}$. By expanding $\ln[R_{\mathbf{n}}(\tau-i\theta_{st,\kappa})]$ in (33) in a power series in τ and keeping the terms up to the second order in τ we can reduce the integral in (33) to a Gaussian integral. We then integrate and transform the resulting expression via the equation of motion (34). What we obtain is an expression for the density of states,

$$\rho(\omega)=\sum_s\sum_{\kappa=1}^{g_{st}}\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2c^s(1-c)^t\frac{1}{\sqrt{2\pi\gamma_{st,\kappa}^2(\omega)}} \times\exp\left\{\sum_{\mathbf{n}\in\mathscr{C}_{st,\kappa}}\left[\ln\left(\frac{c}{c+p(\mathbf{n})}\right)^{c+p(\mathbf{n})}\times\left(\frac{1-c}{1-c-p(\mathbf{n})}\right)^{1-c-p(\mathbf{n})}\right]\right\}, \tag{37}$$

where $p(\mathbf{n})=-U_{st,\kappa}(\mathbf{n})/\Delta$,

$$\gamma_{st,\kappa}^2(\omega)=\frac{m_2}{\left[\sum_{\mathbf{n}}|\phi_{\mathbf{n}}^{st,\kappa}|^2\right]^2}, \tag{38}$$

$$m_2=\Delta^2\sum_{\mathbf{n}}[|\phi_{\mathbf{n}}^{st,\kappa}|^2]^2[c+p(\mathbf{n})][1-c-p(\mathbf{n})], \tag{39}$$

with $\Delta=|E_B-E_A|$. The local values of concentration of atoms that are centers of attraction and repulsion are described by the expressions $[c+p(\mathbf{n})]$ and $[1-c-p(\mathbf{n})]$, respectively.

For each value of s, the above formula for the density of states describes a set of g_{st} bands whose shape near the peak is close to Gaussian. As we move away from the peak, each band acquires the Urbach shape. In the limit of low attraction-center concentrations the expression (37) for the density of states describes structural features that are due to the minimum-size clusters.

2.2.4. Approximate description of the ρ vs. ω dependence

To describe the behavior of $\rho(\omega)$ near the mobility edge and below, we assume that in the zeroth approximation all states with localization energy ω_j exceeding a certain value ω_0 can be assumed localized and reasonably independent. The value ω_0 measured from E_G must be chosen in such a way that the total number of states whose energy exceeds this value,

$$\mathcal{N}(\omega_0)=\int_{\omega_0}^{E_L}\rho(\omega)\,d\omega, \tag{40}$$

with E_L the Lifshits limit for the solid solutions, meets the condition

$$\mathcal{N}(\omega_0)a^3\ll 1. \tag{41}$$

Here $a\approx\sqrt{\hbar^2/2M\omega_0}$ is the characteristic length of exponential decay of the wave function of an isolated well outside the well. In what follows, the value ω_0 acts as the limit of the applicability range of the theory and constitutes an important parameter that can be used in varying $\rho(\omega)$ near the mobility edge.

To calculate $\rho(\omega)$, we use the effective mass approximation and the Halperin–Lax proposal concerning the possibility of using an average wave function to describe localized states.[12,13] We examine fluctuation wells whose strength exceeds that of a compact cluster needed to form a localized state of depth ω_0. Using the approach developed in Refs. 16–18, we obtain

$$\rho(\omega)\sim\frac{1}{v_0\sqrt{2\pi\gamma^2(\omega)}}\,\exp\left\{\int\frac{d^3r}{v_0}\left[\ln\left(\frac{c}{c+p(r)}\right)^{c+p(r)}\right.\right.$$
$$\left.\left.\times\left(\frac{1-c}{1-c-p(r)}\right)^{1-c-p(r)}\right]\right\}, \tag{42}$$

where

$$\gamma^2(\omega)=\frac{m_2}{\left[\int\frac{d^3r}{v_0}\varphi_{\mathrm{tr}}^2(r)\right]^2}, \tag{43}$$

$$m_2=\Delta^2\int\frac{d^3r}{v_0}\varphi_{\mathrm{tr}}^4(r)[c+p(r)][1-c-p(r)], \tag{44}$$

and $v_0=V/N$ is the volume per atom in the lattice, with $\Delta=|E_B-E_A|$. The local values of the concentration of atoms that are attraction and repulsion centers are described by the expressions $[c+p(r)]$ and $[1-c-p(r)]$. Here we use the quantity $p(r)$, which is the measure of deviation of the attraction center concentration from the average value. In this case the self-consistent potential well $U_{\mathrm{tr}}(r)$ can be described by the formula

$$U_{\mathrm{tr}}(r)=-\Delta p(r), \tag{45}$$

where

$$p(r)=\begin{cases}1-c, & r\leqslant R_0,\\ (1-c)\left[1-\dfrac{\exp(-t\varphi_{\mathrm{tr}}^2\Delta)}{c+(1-c)\exp(-t\varphi_{\mathrm{tr}}^2\Delta)}\right], & r>R_0,\end{cases} \tag{46}$$

and

$$R_0=\sqrt{\frac{\hbar^2}{2M[(1-c)\Delta-\omega_0]}}\times\left[\pi-\arctan\sqrt{\frac{(1-c)\Delta-\omega_0}{\omega_0}}\right] \tag{47}$$

can be found by solving the equation

$$\left\{-\frac{\hbar^2}{2M}\nabla^2+\omega-U_{\mathrm{tr}}(r)\right\}\varphi_{\mathrm{tr}}(r)=0. \tag{48}$$

To estimate the absolute value of the density of states, we find the proportionality factor in (42) by calculating the total density of states and matching the value at $\omega\approx\omega_0$ with the total density of states found by employing the sum rule (32) and Eqs. (23) and (24). The resulting density of states is given in Sec. 5.

3. PHONONLESS ABSORPTION AND LUMINESCENCE BANDS

To describe a phononless absorption band we employ the method developed in Ref. 18. Optical absorption near the ground state of an exciton in a perfect crystal is described by the spectral density of the state for a zero wave vector:

$$\alpha_{1s}(\omega)=\frac{v_0}{\pi}|\psi_{1s}(0)|^2\operatorname{Im}G_{00}(\omega-E_{1s}-i\delta), \tag{49}$$

where

$$G_{\mathbf{kk}}(\omega-E_{1s})=\frac{1}{\omega-\hbar^2\mathbf{k}^2/2M-E_{1s}-i\delta}.$$

Here E_{1s} and $\psi_{1s}(r=0)$ are, respectively, the eigenvalue and eigenfunction of the exciton's ground state. As before, the origin of energy coincides with the bottom of the band of the $1s$ state of the exciton, E_G, so that $E_{\Lambda=1s}=0$.

The imaginary part of the one-particle Green's function is the diagonal matrix element obtained as a result of performing a double Fourier transform and averaging over the configurations of the imaginary part of the Green's function (4). Generally, the result can be written

$$\langle\operatorname{Im}G_{\mathbf{kk}}(\omega)\rangle=\frac{\tau_{\mathbf{kk}}(\omega)}{[\omega-\hbar^2\mathbf{k}^2/2M-\Delta_{\mathbf{kk}}(\omega)]^2+\tau_{\mathbf{kk}}^2(\omega)}, \tag{50}$$

where $\tau_{\mathbf{kk}}(\omega)$ and $\Delta_{\mathbf{kk}}(\omega)$ are related through a dispersion law, which ensures correct normalization of the expression for the spectral density. In the region of localized states, where condition (41) is met, these two functions are the imaginary and real parts of the configuration average of the scattering matrix.[41,42] If only one localized state is formed in each fluctuation well, the relationship between $\tau_{\mathbf{kk}}(\omega)$ and $\rho(\omega)$ is

$$\tau_{\mathbf{kk}}(\omega)=\frac{6v_0}{\pi^3}\Delta^2\sqrt{\omega/\Delta}\left(\frac{E_{\mathrm{cr}}}{\Delta}\right)^{3/2}|I_1(\mathbf{k})|^2\rho(\omega), \tag{51}$$

where $\rho(\omega)$ is given by (42) with allowance for the proportionality factor found by using the sum rule (32), and

$$I_1(\mathbf{k})=\left(\frac{2M\omega}{\hbar^2}\right)^{3/2}\frac{1}{(2\pi)^3}\int d^3r\,\exp(i\mathbf{kr})\varphi_{\mathrm{tr}}(\mathbf{r}), \tag{52}$$

where at $k=0$ we have

$$\int d^3r\;\varphi_{\rm tr}(\mathbf{r})=\frac{\Delta}{\omega}\int d^3r\;\varphi_{\rm tr}(\mathbf{r})p(\mathbf{r}), \tag{53}$$

since $p(\mathbf{r})$ describes the configuration of the fluctuation well and is related to $\varphi_{\rm tr}(\mathbf{r})$ through Eqs. (45) and (48). Equation (51) makes it possible to find the imaginary part of the scattering matrix $\tau_{\mathbf{kk}}(\omega)$ in the region of interest.

Finding the function $\Delta_{\mathbf{kk}}(\omega)$ is more complicated since we must know $\tau_{\mathbf{kk}}(\omega)$ over a broad section of the spectrum. Knowing $\tau_{\mathbf{kk}}(\omega)$, we can calculate $\Delta_{\mathbf{kk}}(\omega)$ via the dispersion relation. In contrast to $\tau_{\mathbf{kk}}(\omega)$, which for $\omega>\omega_0$ varies exponentially, the real part of the scattering matrix, $\Delta_{\mathbf{kk}}(\omega)$, decreases slowly (by a power law) with increasing ω, tending in the limit of large ω to zero as $1/\omega$. Hence, to approximately describe the spectral density within a relative narrow energy range near its peak, we can replace $\Delta_{\mathbf{kk}}(\omega)$ with the constant value $\Delta^0_{\mathbf{kk}}$. As a result, at $k\approx 0$ we have

$$\langle \mathrm{Im}\,G_{00}(\omega)\rangle\approx\frac{\tau_{00}(\omega)}{[\omega-\Delta^0_{00}]^2+\tau^2_{00}(\omega)}, \tag{54}$$

where Δ^0_{00} is chosen is such a way that the calculated position of the peak in the $1s$-stateband in the absorption spectrum coincides with the experimental value.

3.1. Basic characteristics of the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ and estimates of the parameters of the problem

To compare our results with the experimental data we selected the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$. The perturbation introduced by Te atoms is strong enough for a localized state to form on a cluster consisting of two or more Te atoms.[40]

The experimentally measured concentration dependence of the band gap of the solid solution,

$$\mathscr{E}_G(c)=c\mathscr{E}^A_G+(1-c)\mathscr{E}^B_G-\Delta\mathscr{E}_G(c), \tag{55}$$

can be characterized by the bowing $\Delta\mathscr{E}_G(c)$, i.e., the deflection from a simple linear interpolation between the limits $\mathscr{E}^A_G$ and $\mathscr{E}^B_G$. The development of the electronic theory of solid solutions is closely related to studying the nonlinear dependence of the band gap on concentration (the bowing effect).[43–48] While in the earlier work devoted to explaining the bowing the observed effect was related to the effective averaging of the parameters of the electron Hamiltonian and/or the scattering off single-site fluctuations, recent studies in this field[46–48] show that the nature of the perturbation of the electronic states introduced by isoelectronic substitution is more complicated, and that to explain the observed effect one must allow for structural changes in the lattice that are due to the variation in the length of bonds between the substituting anion and the neighboring cations. According to the data in Refs. 46–49, band gap variations result from a self-consistent transformation of the electronic states initiated by the chemical and structural perturbations introduced by isoelectronic substitution.

A detailed analysis of this dependence[40,50] for the $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ system shows that the bowing $\Delta\mathscr{E}_G(c)$ consists of two terms,

$$\Delta\mathscr{E}_G(c)=\Delta\mathscr{E}_1(c)+\Delta\mathscr{E}_2(c), \tag{56}$$

and that the two terms display different concentration-dependent behavior. The larger of the two is the symmetric part of the bowing, which is described by the formula $\Delta\mathscr{E}_1(c)=b_1c(1-c)$. The peak value of this term is about 0.25 eV at $c=0.5$, which is comparable to the difference of the band gaps of ZnSe and ZnTe crystals (2.82 and 2.39 eV, respectively[51]). The component $\Delta\mathscr{E}_2(c)$, which is asymmetric with respect to $c=0.5$, reaches its peak value of approximately 0.1 eV at $c\approx 0.15$.

Both the absorption and luminescence of the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ (see Refs. 37, 40, 50, and 52–55) point to the formation of fluctuation-induced tails of localized states of excitons, tails that decrease exponentially with increasing depth of the band gap, $\propto\exp(-\omega/\varepsilon_U)$. The Urbach parameter ε_U reaches its peak value of approximately 0.025 eV at $c\approx 0.15$, and its variation with concentration almost coincides, over a broad range, with the asymmetric part of the bowing function.[50] This points to $\Delta\mathscr{E}_2(c)$ and the fluctuation-induced tail beingn of similar origin and makes it possible to link these two characteristics with the composition fluctuation responsible for the splitting off of localized states.

The large value of the bowing and the substantial smearing of the band edge by the fluctuation-induced tails suggest that the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ is a system in which the perturbation of the electronic states of the valence band where Te is substituted for Se is strong at least in the range of Te concentrations $c\leqslant 0.30$. This is corroborated by the fact that the tail of the fluctuation states in the low-concentration region $c\approx 0.01$ has structural features[37,40,50,52–55] that are due to the localized states of small-sized clusters consisting of two or three Te atoms. As the Te concentration increases and the average level of the potential decreases, the size of the perturbation introduced by each Te atom also decreases systematically.

Another feature of the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$, which sets it apart from other compounds of the given type, is the strong exciton–phonon interaction, which in the concentration range $c\leqslant 0.30$ leads in the luminescence spectra to structureless bands[40,50,52,53] with a halfwidth of approximately 0.1 eV. The larger value of the constant of Fröhlich interaction with LO phonons in the solid solution as compared to the values of that constant in the original ZnSe and ZnTe crystals corroborates the fact that in the given concentration range electron localization is fairly strong, as a result of which electron–vibrational transitions involving phonons whose wave vectors occupy a sizable portion of the Brillouin zone become possible.

Both Δ and $\Delta/E_{\rm cr}$ become the parameters whose values determine the Urhach energy ε_U. To find the values of these parameters we used the experimental values of the Urbach energies obtained from absorption coefficient measurements.[40,52] The values were $\Delta\approx 1$ eV and $\Delta/E_{\rm cr}\approx 0.8$.

The values of these parameters of the Hamiltonian can also be estimated by examining the concentration dependence of the shift of the peak of the exciton ground state (Eq. (55)). In the one-band model, the nonlinear contribution $\Delta E_1(c)$ is the first correction to the position of the bottom of the valence band due to scattering of single-site fluctuations, and can be expressed in terms of the same parameters as the Urbach energy ε_U. We know that in general such an estimate does not guarantee good agreement between theoretical and experimental data. However, if the parameters Δ and Δ/E_{cr} of the model Hamiltonian are selected to achieve the best agreement with the observable values of Urbach energies, the internal consistency of the model suggests that the fluctuation shift of the band edge, equal to

$$\Delta E_1(c)=\Delta^2 G^{vc}_{\mathbf{nn}}(0)c(1-c), \tag{57}$$

satisfies the inequality $\Delta E_1 \leqslant \Delta\mathscr{E}_1$, where $\Delta\mathscr{E}_1(c)$ is the experimental value of the band-gap bowing. The validity of this inequality follows from the fact that the calculated shift ΔE_1 pertains to the valence band, due to which the tail emerges, while the observed value of $\Delta\mathscr{E}_1$ also contains a conduction-band contribution, which usually has the opposite sign and increases the bowing. In the lattice variant of the model, the critical value of the single-site perturbation potential is linked to the electron Green's function through the relationship

$$G^{vc}_{\mathbf{nn}}(0)=\frac{1}{E_{\mathrm{cr}}}. \tag{58}$$

Thus, we see that the symmetric part of the shift can be expressed in terms of the same parameters as the Urbach energy.

The quantity $\Delta E_1(c)$ is not directly related to the formation of localized states, which means that in all calculations it is convenient to carry out an additional renormalization of the energy origin, eliminating this quantity from calculations of localization energies. In the effective mass approximation the shift of the band bottom, which is due to single-site fluctuations, also needs to be renormalized.[14]

With allowance for renormalization, the asymmetric part of the bowing, $\Delta\mathscr{E}_2(c)$, can be found from the experimental data to be the position of the peak of the observed absorption band with respect to the bottom of the band of the virtual crystal (Eq. (56)) after the symmetric part of the bowing, $\Delta\mathscr{E}_1(c)$, has been subtracted. Separation of the concentration shift of the band-gap edge, $\Delta\mathscr{E}_G$, into $\Delta\mathscr{E}_1(c)$ and $\Delta\mathscr{E}_2(c)$ makes it possible to accurately establish only the maximum value of $\Delta\mathscr{E}_1(c)$. On the other hand, when only one band is responsible for the formation of a density-of-states tail (in our case we assume that this is the valence band), the value of $\Delta\mathscr{E}_2(c)$ obtained from experiments and the value of $\Delta E_2(c)$ found from calculations in the one-band model coincide to high accuracy:

$$\Delta E_2(c)=\Delta\mathscr{E}_2(c). \tag{59}$$

Since $\Delta E_2(c)$ is related to the same fluctuations that lead to the formation of localized states, the concentration dependence $\Delta E_2(c)$ has the form

$$\Delta E_2(c)\sim\varepsilon_U\sim\frac{\Delta(1-c)}{\ln(1/c)}\left(\frac{\Delta(1-c)}{E_{\mathrm{cr}}}\right)^{3/2}. \tag{60}$$

3.2. Phononless luminescence band

The above reasoning assumes that fluctuation-induced potential wells, which lead to exciton localization, are clusters of atoms of the narrow-band component of the solid solution, with diagonal disorder being responsible for producing fluctuation states in the given model. When the atoms of two species are distributed at random, the fluctuation-induced potential wells are also distributed at random over the crystal volume. This, however, is true only for states whose localization energy is high enough. As we approach the mobility edge, an ever greater fraction of states belongs not to a separate cluster but to complexes of clusters, or superclusters. In this section we develop an approach to the luminescence problem that is based on the analogy between percolation along fluctuation states and the classical theory of percolation along overlapping spheres.

3.2.1. Classification of localized states

Enclosing each potential well in which there is a state with an energy $\omega>\omega_0$ in a sphere of radius $R_{\mathrm{int}} > \sqrt{\hbar^2/2M\omega_0}$ whose value is a parameter of the problem, we use the theory of percolation along overlapping spheres to calculate such characteristics of tail states as the average number of complexes (superclusters), $\langle n_s\rangle$, consisting of s overlapping spheres. We assume that the state belonging to a given potential well is isolated in space if the volume $4\pi R^3_{\mathrm{int}}/3$ of the sphere contains no other potential wells with a localization energy higher that the energy of localization in the given well. Concerning the problem in question, we can say that $\langle n_1\rangle$ is the number of isolated localizing potential wells, $\langle n_2\rangle$ is the number of clusters consisting of pair wells, and so on.

If we use the well-known expression for $\langle n_1\rangle$ (see Refs. 30–35), we arrive at

$$\langle n_1(\omega)\rangle\approx\exp(-[R_{\mathrm{int}}/\overline{r(\omega)}]^3)\equiv\exp\{-2\mathscr{P}(\omega)\}, \tag{61}$$

where

$$\overline{r(\omega)}=\left(\frac{3}{4\pi}\frac{1}{\mathscr{N}(\omega)}\right)^{1/3} \tag{62}$$

and $\langle n_1(\omega)\rangle\equiv\langle n_1(\mathscr{N}(\omega))\rangle$. The concentration $\mathscr{N}(\omega)$ of potential wells with a localization energy limited on the one side by ω and on the other by the Lifshits limit is determined by a formula of type (40). The function $\mathscr{P}(\omega)$ is the density of localized excitons in units of the first virial coefficient. The data of various researchers on the critical value of density for percolation along spheres are collected in Ref. 34. Such values lie in the range $1.17\geqslant\mathscr{P}^{\mathrm{cl}}_{\mathrm{cr}}\leqslant 1.40$. Equations (61) and (62) show that in our case the critical value $\mathscr{P}(\omega_{\mathrm{ME}})$ depends on the parameter R_{int} introduced earlier and the total density at the percolation threshold, $\mathscr{N}(\omega_{\mathrm{ME}})$.

The dependence of $\langle n_s\rangle$ on the concentration of spheres for $s=2-5$ was obtained in Ref. 34 as a power series in the concentration. To be able to use these functions near the percolation threshold, we derived the following extrapolations of the series obtained in Ref. 34 for $s=2,3,4$:

$$\langle n_2(\omega)\rangle=\mathscr{P}(\omega)\exp\{-3.073\mathscr{P}(\omega)\},$$

$$\langle n_3(\omega)\rangle = 1.375\mathcal{P}^2(\omega)\exp\{-4.09\mathcal{P}(\omega)\}, \tag{63}$$
$$\langle n_4(\omega)\rangle = 2.1842\mathcal{P}^3(\omega)\exp\{-5.084\mathcal{P}(\omega)\}.$$

A series expansion of these expressions in powers of $\mathcal{P}(\omega)$ leads to results that to high accuracy coincide with the series found in Ref. 34. Moreover, it is possible to estimate the accuracy of the extrapolation for the function $\langle n_2(\omega)\rangle$ if one uses the results of the numerical calculations of the function $P_3(p_{\mathrm{norm}})$ done in Ref. 30. The function $P_3(p_{\mathrm{norm}})$ is the probability of a potential well selected at random belonging to a supercluster whose size $s>3$, with $P_3(p_{\mathrm{norm}})$ linked to $\langle n_2(\omega)\rangle$ by the relationship $P_3(p_{\mathrm{norm}})=1-\langle n_1\rangle-2\langle n_2\rangle$, and to p_{norm} linked to $\mathcal{P}(\omega)$ by the relationship $p_{\mathrm{norm}}=\mathcal{P}(\omega)/8$. A comparison of $P_3(p_{\mathrm{norm}})$ with calculations done in Ref. 30 showed good agreement in the most important region, $0.3>p_{\mathrm{norm}}>0.1$, i.e., near the mobility edge.

3.2.2. Lifetimes of localized states

Let us examine the simplest situation in which a localized state can either recombine and emit a phonon or go to another state with a lower energy, transferring several phonons to the lattice. Here we ignore the interaction between two or more excited excitons, assuming all along that the occupation of the states is close to zero due to the weakness of excitation. Using the classification of localized states according to their membership in superclusters of different sizes, we claim that states belonging to single wells can only recombine. The states belonging to pair superclusters ($s=2$) are of two types: one (the lower) at absolute zero can also only recombine, while the second (the upper) can go to to the lower and emit phonons in the process. Obviously, the total concentration of states that can only recombine is

$$\mu_0(\omega)\equiv\mu_{\mathrm{rad}}(\omega)=\left\langle\sum_{s=1}^{\infty} n_s(\omega)\right\rangle, \tag{64}$$

where $n_s(\omega)$ is the number of superclusters consisting of s potential wells with a localization energy $\geqslant\omega$ per well, and the angle brackets indicate averaging over all possible realizations of a supercluster of the given size.

In the adopted model, $\mu_{\mathrm{rad}}(\omega)$ is the fraction of states for which the total lifetime is equal to the radiative time:

$$\tau^{(0)}\equiv\tau_{\mathrm{rad}}; \tag{65}$$

we assume that this quantity is independent of the localization energy. The other supercluster states have lifetimes limited by decay processes involving phonon emission. For instance, the upper states of pair clusters and the second states of clusters of larger size (supercluster states are numbered according to their decreasing localization energy) have only one feasible decay channel. The number of decay channels involving phonon emission increases with the number of the state in a supercluster. To simplify matters, we assume that the lifetime of a state depends only on the number of decay channels for the given state and is independent of the supercluster configuration. Then to a state with number $s\geqslant 1$ the corresponding total lifetime is

$$\tau^{(s-1)}=\frac{\tau_{\mathrm{dec}}^{(s-1)}\tau_{\mathrm{rad}}}{\tau_{\mathrm{dec}}^{(s-1)}+\tau_{\mathrm{rad}}}, \tag{66}$$

where the superscript $(s-1)$ indicates the number of decay channels and $\tau_{\mathrm{dec}}^{(s-1)}$ denotes the lifetime against transitions to the lower states of the supercluster. By analogy with (64), the number of states with s decay channels is

$$\mu_s(\omega)=\left\langle\sum_{k=s+1}^{\infty} n_k(\omega)\right\rangle. \tag{67}$$

To a first approximation, μ_{rad} can be written

$$\mu_{\mathrm{rad}}\approx\langle n_1(\omega)+n_2(\omega)+\cdots\rangle=\exp\{-2\mathcal{P}(\omega)\}$$
$$+\mathcal{P}(\omega)\exp\{-3.073\mathcal{P}(\omega)\}+\cdots, \tag{68}$$

and for the fraction of states with only one relaxation channel, μ_1, we have

$$\mu_1\approx\langle n_2(\omega)+n_3(\omega)+\cdots\rangle=\mathcal{P}(\omega)\exp\{-3.073\mathcal{P}(\omega)\}$$
$$+1.375\mathcal{P}^2(\omega)\exp\{-4.09\mathcal{P}(\omega)\}+\cdots. \tag{69}$$

To high accuracy, the sums (64) and (67) are determined by their lower limit,[36] i.e., by a small number of leading terms, which are given in (68) and (69). They are well-defined at low concentrations, in the critical region $|\delta\mathcal{N}(\omega)|\equiv|1-\mathcal{N}(\omega)/\mathcal{N}(\omega_{\mathrm{ME}})|\ll 1$, and at concentrations exceeding the critical, although in the latter two regions they are exponentially small. The approach described in Ref. 36 makes it possible, in the critical region, to represent the singular parts of the sums (64) and (67) as $|\delta\mathcal{N}(\omega)|^{\nu D+1}$, where D is the dimensionality of the system, and ν is the critical index of the order parameter, with ν depending on D. At $D=3$ we have $\nu\approx 0.87$, which leads to negligible singular parts of these sums in the critical region.

3.3. Contour of the phononless luminescence band

The steady-state concentration of occupied states for a continuous and relatively weak band-to-band excitation is proportional to the density of states for a given energy and lifetime of the states. Allowing for the fact that the contribution of each occupied state to the spectral density of radiation coincides with the contribution of the $1s$ state to the absorption coefficient, we can represent the phononless luminescence band in the form

$$I_{1s}^0(\omega)\sim\alpha_{1s}^0(\omega)P(\omega)\tau_{\mathrm{rad}}, \tag{70}$$

where

$$P(\omega)=\sum_{s=1}^{\infty}\frac{\mu_{(s-1)}(\omega)\tau^{(s-1)}}{\tau_{\mathrm{rad}}}\equiv\sum_{s=1}^{\infty}P^{(s-1)}(\omega) \tag{71}$$

yields the total relative occupancy of the states (with the localization energy ω) belonging to superclusters of different sizes. Allowance for the first two terms in $P(\omega)$, which means that only isolated clusters and ground states of superclusters with $s=2$ are taken into account, yields

$$I^0(\omega)\sim\alpha_{1s}^0(\omega)\tau_{\mathrm{rad}}[\exp\{-2\mathcal{P}(\omega)\}$$
$$+\mathcal{P}(\omega)\exp\{-3.073\mathcal{P}(\omega)\}], \tag{72}$$

which is independent of nonradiative relaxation times. Higher-order terms that are also independent of the nonradiative relaxation rate emerge due to the ground states of superclusters of larger sizes. Moreover, all superclusters beginning with $s=2$ provide contributions that depend on nonradiative relaxation times. For instance, the first correction for which the second state of the supercluster with $s=2$ is responsible can be written

$$\begin{aligned}\alpha^0_{1s}(\omega)P^{(1)}(\omega)\tau_{\text{rad}}&\\ \approx \alpha^0_{1s}(\omega)\langle n_2(\omega)\rangle\tau^{(1)}&=\alpha^0_{1s}(\omega)\mathscr{P}(\omega)\\ \times\exp\{-3.073\mathscr{P}(\omega)\}&\frac{\tau_{\text{dec}}\tau_{\text{rad}}}{\tau_{\text{dec}}+\tau_{\text{rad}}}. \qquad (73)\end{aligned}$$

This expression shows that the role played by the correction depends on the ratio of τ_{dec} and τ_{rad}.

In all cases the minimum-sized superclusters provide the leading contributions, while the higher-order terms provide only small corrections both in the low concentration region $\mathscr{N}(\omega)/\mathscr{N}(\omega_{\text{ME}})\ll 1$, because the number of such corrections is proportional to the degree of concentration $\mathscr{P}(\omega)$, which increases with supercluster size,[34] and near the mobility edge, because of the more rapid exponential decrease in higher-order terms.

4. ELECTRON–PHONON INTERACTION

The electron–phonon interaction has an additional effect on the shape of the absorption and luminescence spectra because absorption and luminescence is accompanied by emission of both acoustic and optical phonons.

To describe the interaction between excitons and LO and LA phonons, we can use our previous results.[18,19] The absorption coefficient and the luminescence intensity at absolute zero with allowance for interaction with phonons can be written

$$\alpha_{1s}(\omega)=\int_0^\infty dz\;\alpha^0_{1s}(\omega+z)F(z) \qquad (74)$$

and

$$I_{1s}(\omega)=\int_{-\infty}^0 dz\,I^0_{1s}(\omega-z)F(z), \qquad (75)$$

respectively. Here

$$\begin{aligned}F(\omega)=\frac{1}{2\pi}\int_{-\infty}^\infty dt\exp\Bigg[&i\omega t+\sum_q\frac{|H_{\Phi\Phi}(\mathbf{q})|^2}{\Omega_q^2}\\ &\times(\exp(-i\Omega_q t)-1)\Bigg] \qquad (76)\end{aligned}$$

yields the density of electron–vibrational states of the wing.

When the scattering is strong, a localized exciton consists of a hole localized in the random potential of the solid solution and an electron bound to the hole by Coulomb interaction. Corresponding to this model is the wave function

$$\Phi=\phi_{\text{tr}}(\mathbf{R})\psi_{1s}(\xi), \qquad (77)$$

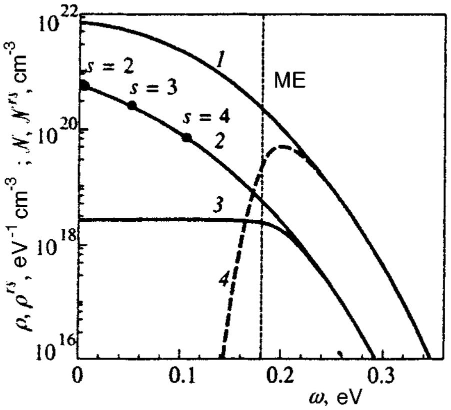

FIG. 1. Results of calculations: curve *1*, the density of states $\rho(\omega)$; curve *2*, the total density of states $\mathscr{N}(\omega)$; curve *3*, the density of "radiating" states $\mathscr{N}^{rs}(\omega)$; and curve *4*, the total density of radiating states $\rho^{rs}(\omega)$ of the fluctuation tail of the solid solution $\text{ZnSe}_{1-c}\text{Te}_c$ at $c=0.15$. The ● denotes estimates of the total density of states by the sum rule. The zero in energy corresponds to the position of the bottom of the band of the virtual crystal with allowance for the symmetric part $\Delta E_1(c)$ of the bowing. The vertical dotted line indicates the position of the mobility edge ω_{ME} (see the text).

where $\phi_{\text{tr}}(\mathbf{R})$ describes the state of a particle localized by the random potential of the solid solution, and $\psi_{1s}(\xi)$ is the wave function of the ground state in the Coulomb potential. The argument $\mathbf{R}$ of ϕ_{tr} represents in this case the radius vector of the hole, $\mathbf{R}=\mathbf{r}_h$, while the argument ξ of ψ_{1s} represents the radius vector of the electron, $\xi=\mathbf{x}_e$.

We write the exciton–phonon interaction Hamiltonian as $H(\mathbf{q})=H_{\text{LO}}(\mathbf{q})+H_{\text{LA}}(\mathbf{q})$, where $\mathbf{q}$ labels the wave vectors of the phonons, and each term is the sum of the electron and hole Hamiltonians: $H_{\text{LO,LA}}(\mathbf{q})=H^e_{\text{LO,LA}}(\mathbf{q})+H^h_{\text{LO,LA}}(\mathbf{q})$. The matrix elements of the exciton–phonon Hamiltonian can be expressed as follows:[56,57]

$$H^\alpha_{\Phi\Phi}(\mathbf{q})=[\mathscr{F}^\alpha_{e,q}\exp i(\mathbf{q}\cdot\mathbf{r_e})+\mathscr{F}^\alpha_{h,q}\exp i(\mathbf{q}\cdot\mathbf{r_h})]_{\Phi\Phi}, \qquad (78)$$

where the label α indicates the interaction mechanism. Since each exponential function in (78) depends only on one argument, we have

$$\begin{aligned}H^\alpha_{\Phi\Phi}(\mathbf{q})=&\mathscr{F}^\alpha_{e,q}[\exp i(\mathbf{q}\cdot\mathbf{r}_e)]_{1s1s}\\ &+\mathscr{F}^\alpha_{h,q}[\exp i(\mathbf{q}\cdot\mathbf{r}_h)]_{\phi_{\text{tr}}\phi_{\text{tr}}}. \qquad (79)\end{aligned}$$

The functions $\mathscr{F}^\alpha_{e,q}$ and $\mathscr{F}^\alpha_{h,q}$ can be found, for example, in Ref. 56.

In calculating the spectra $\alpha_{1s}(\omega)$ and $I_{1s}(\omega)$ we allowed for the Fröhlich interaction with LO phonons and for the deformation and piezoelectric interactions with LA phonons. The main parameters of the solid solution needed in calculations (the effective electron and hole masses and the constants of the deformation potential and of the piezoelectric and Fröhlich interactions) were obtained via linear interpolation between their values for ZnSe and ZnTe (see Ref. 58).

5. RESULTS OF CALCULATIONS AND DISCUSSION

Calculated values of the density of fluctuation states, $\rho(\omega)$, and the total density of states, $\mathscr{N}(\omega)$, are depicted in Fig. 1. The total density of states is the total number of states

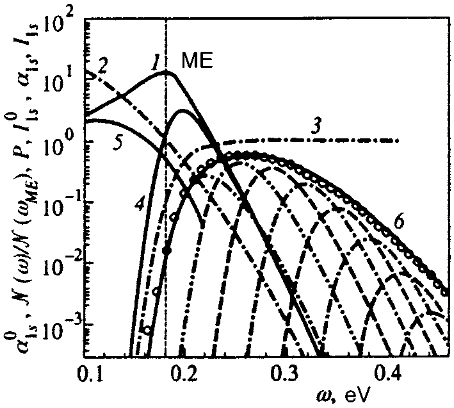

FIG. 2. Results of calculations: curve *1*, the contour of the phononless absorption band $\alpha^0_{1s}(\omega)$; curve *2*, the relative total density of states $\mathcal{N}(\omega)/\mathcal{N}(\omega_{\mathrm{ME}})$; curve *3*, the fraction $P(\omega)$ of radiative states; curve *4*, the phononless luminescence band $I^0_{1s}(\omega)$; and curves *5* and *6*, the absorption band $\alpha_{1s}(\omega)$ and the luminescence band $I^0_{1s}(\omega)$ of the 1*s* state with allowance for the interaction with phonons. The dot–dash curves represent the components of the contour of the luminescence band responsible for the interaction with LO phonons, and the ○ represents the luminescence spectrum of $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ at $c=0.15$ and $T=2$ K.

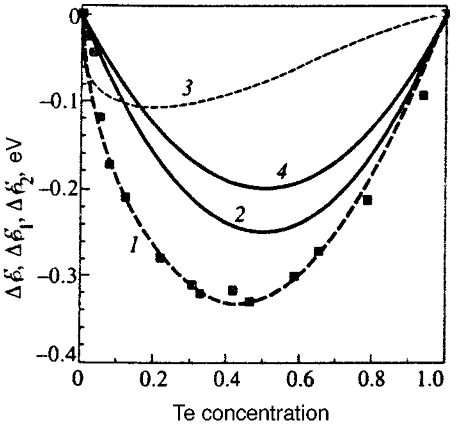

FIG. 3. Experimental data on concentration bowing of the band gap, $\Delta\mathscr{E}_G(c)$ (■). Curve *1*, the approximation of $\Delta\mathscr{E}_G(c)$ by Eq. (80), which is a combination of the functions (57) and (60), at $b_1=1.0$ eV and $b_2=0.3$ eV; curve *2*, the maximum possible value of the symmetric part of the bowing, $\Delta\mathscr{E}_1(c)$; curve *3*, the asymmetric part of the bowing, $\Delta\mathscr{E}_2(c)$; and curve *4*, the estimate of $\Delta\mathscr{E}_1(c)$ made by Eq. (57).

with energies exceeding ω, calculated on the basis of $\rho(\omega)$ via Eqs. (42)–(48). The proportionality factor in (42) and the parameter ω_0 must be selected in such a way that the total density determined via the sum rule (32) coincides with the value calculated from $\rho(\omega)$ at $\omega=\omega_0$. The value ω_0 was found to be 0.098 ± 0.005 eV, which places it in the range between the values of the localization energy for clusters with the number of Te atoms equaling $s=3$ and $s=4$. The localization energy for the clusters were estimated in the approximation of effective mass and spherical potential well of appropriate sizes. The total density of states for clusters, denoted by ● in Fig. 1, was calculated by the sum rule (32). The procedure yields values of the density of states, depicted in Fig. 1, in the region where $\omega\geqslant\omega_0$ and makes it possible to extrapolate this function into the region where $\omega\leqslant\omega_0$. Figure 1 also depicts the density of "radiating" states, $\rho^{rs}(\omega)=\rho(\omega)P(\omega)$, with $P(\omega)$ defined in (71), and the total density of states $\mathcal{N}^{rs}(\omega)$ as functions of the localization energy ω.

Figure 2 depicts the results of calculations of the contour of the phononless absorption and luminescence bands of the excitonic ground state for the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ at $c=0.15$. The position of the mobility edge (the dotted vertical line in Fig. 2) and the theoretical parameter R_{int} (introduced in Sec. 3.B.1) were selected in such a way that the shift of the peak in the luminescence band in relation to the peak in the absorption band corresponds to the observed quantity. In our calculations we used $\mathscr{P}(\omega_{\mathrm{ME}})=1.4$, which corresponds to the value obtained by Haan and Zwanzig[34] for the critical concentration in the classical problem of overlapping spheres.

The contour of the phononless luminescence band in Fig. 2 shows that only a relatively small number of fluctuation states, lying in the region of small absorption-coefficient values, determines the luminescence process. The peak in the distribution of these radiating states is shifted in relation to the absorption peak toward the red part of the spectrum even when the interaction with lattice vibrations is ignored, which is reasonably consistent with the existing experimental data on disordered systems. The interaction with phonons leads to additional Stokes shifts of the absorption and emission bands in opposite directions. Figure 2 depicts the results of allowing for the effect of the exciton–phonon interaction on the position and shape of the luminescence band and the experimental spectrum of the solid solution $\mathrm{ZnSe}_{1-c}\mathrm{Te}_c$ at $c=0.15$. In our calculations we assumed the localized-exciton wave function to be the product of the wave function of a localized hole and the electron Coulomb wave function. Since in this concentration range the excitonic binding energy $E_{\mathrm{ex}}(c)$ is much lower than the localization energy for the radiating states, the electron has not enough time to adiabatically follow the motion of the hole and, therefore, interacts with the hole distribution averaged over the fluctuation-induced well. Satisfactory agreement with the experimental curve is achieved when one allows for ten LO repetitions; the decomposition of the contour of the luminescence band into components is also shown in Fig. 2.

Figure 3 presents the experimental data on the concentration dependence of the band-gap bowing, $\Delta\mathscr{E}_G(c)$, the separation of the experimentally measured shift of the peak, $\Delta\mathscr{E}_G(c)$, into $\Delta\mathscr{E}_1(c)$ and $\Delta\mathscr{E}_2(c)$ according to the formula

$$\Delta\mathscr{E}_G(c)=b_1c(1-c)+b_2\frac{(1-c)^{5/2}}{\ln(1/c)}, \qquad (80)$$

and the estimate of $\Delta E_1(c)$ by (57) with the values of the parameters listed in this paper. Clearly, the resulting value of $\Delta E_1(c)$ obeys the earlier discussed inequality $\Delta E_1\leqslant\Delta\mathscr{E}_1$. Here the position of the peak in the absorption band, calculated at $c=0.15$ with allowance for the interaction with phonons, is approximately 0.1 eV and is also in good agreement with the observed value of $\Delta\mathscr{E}_2(c)$.

In conclusion we note that the proposed approach to describing the spectrum of fluctuation states of a disordered solid solution with diagonal disorder and a narrow-gap component concentration below the percolation threshold for a disordered sublattice can be generalized to two-dimensional systems and to other types of percolation.

The classification of localized states according to their membership in superclusters may also prove useful for disordered systems in which the density of fluctuation states is three-dimensional while the statistics of the superclusters is two-dimensional. Such a situation may occur, say, in the case of quantum wells formed by a solid solution if the wells are wide enough to accommodate three-dimensional fluctuation clusters but not wide enough to accommodate three-dimensional superclusters.

The present work was made possible by grants from INTAS (Grant No. 94-324) and the Russian Fund for Fundamental Research (Projects 96-02-16933 and 97-02-18138).

[1] I. M. Lifshits, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 549 (1965)].

[2] D. J. Thouless, Phys. Rep. C **13**, 95 (1974).

[3] Y. Nagaoka (Ed.), Prog. Theor. Phys. Suppl. No. 84, *Anderson Localization* (1985).

[4] T. Ando and H. Fukuyama (Eds.), *Anderson Localization*, Springer-Verlag, Berlin (1988).

[5] G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954).

[6] M. Lannoo and P. Lenglart, J. Phys. Chem. Solids **30**, 2409 (1969).

[7] J. Bernholc and Sokratas Pantelides, Phys. Rev. B **18**, 1780 (1978).

[8] I. M. Lifshitz, Nuovo Cimento, Suppl. Vol. 3, ser. X, 716 (1956).

[9] R. J. Elliott, J. A. Krumhansl, and P. L. Leath, Rev. Mod. Phys. **46**, 465 (1974).

[10] M. A. Ivanov and Yu. G. Pogorelov, Zh. Éksp. Teor. Fiz. **72**, 2198 (1977) [Sov. Phys. JETP **45**, 1155 (1977)].

[11] M. A. Ivanov and Yu. G. Pogorelov, Zh. Éksp. Teor. Fiz. **76**, 1010 (1979) [Sov. Phys. JETP **49**, 510 (1979)].

[12] B. I. Halperin and M. Lax, Phys. Rev. **148**, 722 (1966).

[13] B. I. Halperin and M. Lax, Phys. Rev. **149**, 802 (1966).

[14] E. Brezin and G. Parizi, Phys. Rev. B **48**, 3100 (1993).

[15] A. A. Klochikhin and S. G. Oglobin, Zh. Éksp. Teor. Fiz. **100**, 2026 (1991) [Sov. Phys. JETP **73**, 1122 (1991)].

[16] A. A. Klochikhin and S. G. Oglobin, Phys. Rev. B **48**, 3100 (1993).

[17] A. A. Klochikhin, Phys. Rev. B **53**, 10 979 (1995).

[18] A. A. Klochikhin, S. A. Permogorov, and A. N. Reznitskiĭ, JETP Lett. **65**, 289 (1997).

[19] A. A. Klochikhin, S. A. Permogorov, and A. N. Reznitskiĭ, Fiz. Tverd. Tela (St. Petersburg) **39**, 1170 (1997) [Phys. Solid State **39**, 1035 (1997)].

[20] P. G. de Gennes, P. Lafore, and J. P. Millot, J. Phys. Chem. Solids **11**, 105 (1966).

[21] P. G. de Gennes, P. Lafore, and J. P. Millot, J. Phys. Radium **20**, 624 (1959).

[22] Y. Cho and Y. Toyozawa, J. Phys. Soc. Jpn. **26**, 71 (1969).

[23] S. Kirkpatrick and T. P. Eggarter, Phys. Rev. B **6**, 3598 (1972).

[24] A. B. Harris and T. C. Lubensky, Phys. Rev. B **23**, 2640 (1981).

[25] A. B. Harris and T. C. Lubensky, Phys. Rev. Lett. **49**, 296 (1982).

[26] Y. Shapir, A. Aharony, and A. B. Harris, Phys. Rev. Lett. **49**, 486 (1982).

[27] W. Y. Ching and D. L. Huber, Phys. Rev. B **25**, 1096 (1982).

[28] Y. Meir, A. Aharony, and A. B. Harris, Europhys. Lett. **10**(3), 275 (1989).

[29] I. Chang, Zvi Lev, A. B. Harris, and A. Aharony, Phys. Rev. Lett. **74**, 2094 (1995).

[30] D. F. Holcomb and J. J. Rehr Jr., Phys. Rev. **183**, 733 (1969).

[31] V. K. S. Shante and S. Kirkpatrick, Adv. Phys. **20**, 325 (1971).

[32] J. W. Essam, Rep. Prog. Phys. **43**, 833 (1980).

[33] G. E. Pike and C. H. Seager, Phys. Rev. B **10**, 1421 (1974).

[34] S. W. Haan and R. Zwanzig, J. Phys. A **10**, 1547 (1977).

[35] W. T. Elam, A. R. Kerstein, and J. J. Rehr Jr., Phys. Rev. Lett. **52**, 1516 (1984).

[36] A. Bunde and S. Havlin, in *Fractals and Disordered Systems*, A. Bunde and S. Havlin (Eds.), Springer-Verlag, Berlin (1991), p. 97.

[37] A. Klochikhin, S. Permogorov, and A. Reznitsky, J. Cryst. Growth **159**, 848 (1996).

[38] Nguen Xuan Xinh, A. A. Maradudin, and R. A. Coldwell-Horsfall, J. Phys. (France) **26**, 717 (1965).

[39] J. W. Essam, in *Phase Transitions and Critical Phenomena*, C. Domb and M. S. Green (Eds.), Vol. 2, Academic Press, New York (1972), p. 197.

[40] S. Permogorov and A. Reznitsky, J. Lumin. **52**, 201 (1992).

[41] R. G. Newton, *Scattering Theory of Waves and Particles*, McGraw-Hill, New York (1966).

[42] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions, and Decay in Non-relativistic Quantum Mechanics* [in Russian], 2nd edn., Nauka, Moscow (1971) [English translation of the 1st edn.: NASA Techn. Transl. F-510 (1969)].

[43] J. A. Van Vechten and T. K. Bergstresser, Phys. Rev. B **1**, 3351 (1970).

[44] D. Richardson, J. Phys. C **5**, L27 (1972); **5**, L127 (1972).

[45] A. Baldereshi and K. Mashke, Solid State Commun. **16**, 99 (1975).

[46] J. E. Bernard and A. Zunger, Phys. Rev. B **36**, 3199 (1987).

[47] S.-H. Wei, L. G. Ferreira, J. E. Bernard and A. Zunger, Phys. Rev. B **42**, 9622 (1990).

[48] S.-H. Wei and A. Zunger, Phys. Rev. B **43**, 1662 (1991); **43**, 14 272 (1991).

[49] S.-H. Wei and A. Zunger, Phys. Rev. B **39**, 3279 (1989).

[50] A. Naumov, S. Permogorov, A. Reznitsky, S. Verbin, A. Klochikhin, J. Cryst. Growth **101**, 713 (1990).

[51] B. Segal and D. T. F. Marple, in *Physics and Chemistry of II–VI Semiconducting Compounds*, M. Aven and J. S. Prener (Eds.), North-Holland, Amsterdam (1967), p. 319.

[52] A. Yu. Naumov, S. A. Permogorov, T. B. Popova, A. N. Reznitskiĭ, V. Ya. Zhulaĭ, V. A. Novozhilov, and N. N. Spendiarov, Fiz. Tekh. Poluprovodn. **21**, 350 (1987) [Sov. Phys. Semicond. **21**, 213 (1987)].

[53] A. Yu. Naumov, S. A. Permogorov, A. N. Reznitskiĭ, V. Ya. Zhulaĭ, V. A. Novozhilov, and G. T. Petrovskiĭ, Fiz. Tverd. Tela (Leningrad) **29**, 377 (1987) [Sov. Phys. Solid State **29**, 215 (1987)].

[54] A. G. Abdukadyrov, S. D. Baranovskiĭ, S. Yu. Verbin, E. L. Ivchenko, A. Yu. Naumov, and A. Reznitskiĭ, Zh. Éksp. Teor. Fiz. **98**, 2056 (1990) [Sov. Phys. JETP **71**, 1155 (1990)].

[55] A. Reznitsky, S. D. Baranovskii, A. Tsekoun, and C. Klingshirn, Phys. Status Solidi B **184**, 159 (1994).

[56] E. Cohen and M. Sturge, Phys. Rev. B **25**, 3828 (1982).

[57] A. I. Ansel'm and Yu. A. Firsov, Zh. Éksp. Teor. Fiz. **28**, 151 (1955) [Sov. Phys. JETP **1**, 139 (1955)].

[58] *Physics and Chemistry of II–VISemiconducting Compounds*, M. Aven and J. S. Prener (Eds.), North-Holland, Amsterdam (1967).

Translated by Eugene Yankovsky

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Vacancy mobility in polymer crystals

E. A. Zubova and L. I. Manevich

N. N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia

N. K. Balabaev

Institute of Mathematical Problems in Biology, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

(Submitted 14 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1063–1069 (March 1999)

A molecular-dynamics model of the behavior of a vacancy in the chain of an equilibrium polymer crystal (the ''collective atom'' approximation for polyethylene) is developed for the first time. It is shown that a defect of this type in a polymer crystal has a soliton mobility, as opposed to vacancies in crystals of low-molecular substances. © *1999 American Institute of Physics.* [S1063-7761(99)02103-4]

1. INTRODUCTION

In order to predict the relaxation, plastic, and durability characteristics and the melting temperature of crystals it is necessary to consider localized mobile defects or deviations from the ideal structure. Point structural defects (such as vacancies or inclusions) are responsible for the relaxation properties of solids. Their inability to undergo directed motion is why relaxation processes are slow in low-molecular crystals. Because the properties of polymer crystals are highly anisotropic (the atoms in a polymer chain are bonded to one another by chemically covalent bonds, while the intermolecular interaction is through weak van der Waals forces), vacancies with breaking of interchain covalent bonds are essentially immobile. However, these crystals can contain other, specifically polymer, point defects, which are not caused by breaking of intrachain bonds, but by a deformation of a chain which is localized in a small portion of the chain.

The idea of such defects arose in polymer crystal physics after the observation of anomalously rapid dielectric relaxation of oxidized polyethylene.[1] An analysis of a number of possible molecular mechanisms for this process[2] showed that the most probable mechanism was the propagation of localized regions of twisting (by 180°) along the chains with stretching (over a half period of the chain) extending on the order of a few tens of periods in the absence of conformational changes. A quasi-one dimensional approximation for immobile neighboring chains (see Ginzburg *et al.*[3] and the references cited there) can be used to describe this kind of defect as a soliton-type topological excitation,[4] i.e., a localized nonlinear wave propagating at a constant near-sonic speed along the chain which changes its state after it passes and can, thereby, cause rapid relaxation in the crystal.

An approximate analytic description of static point defects in polyethylene has been proposed[5] and they have been studied numerically[6] using the techniques of molecular mechanics. However, these papers do not touch on the question of the mobility of the defects and, therefore, their contribution to relaxation in these crystals. Recently a numerical solution has been obtained for the problem of the propagation of near-sonic solitons in a polyethylene chain surrounded by immobile neighboring chains.[7] Only one paper[8] contains a molecular-dynamics study of the behavior of point twist defects in polyethylene crystals (with mobile molecules in the first coordination sphere) for a ''collective atom'' model (the CH_2-groups are replaced by point particles). Unfortunately, the use of an initially nonequilibrium sample crystal structure led, as will be shown below, to unjustified conclusions regarding the nonsoliton nature of the mobility of this defect in the crystal.

The purpose of this paper is to study the type of mobility of point structural defects in the framework of a molecular-dynamics model for a polymer crystal with chains of the simplest type (plane trans-zigzag in the collective atom approximation, Fig. 1a). The equilibrium configuration of a crystal of this type is found in advance (Figs. 1b and 1c). Here we limit ourselves to analyzing vacancies (Fig. 2a), the simplest defects which can exist in a polymer crystal chain. We also study the dependence of the defect behavior on the crystal structure.

2. NUMERICAL MODEL OF POLYMER CRYSTALS

We have taken the following model for a polymer crystal[9] (polyethylene with collective atoms; see Fig. 1a): the chains consist of plane trans-zigzags; the bonds between the atoms (point particles of mass m) are absolutely rigid of length l_0; the deformation energies of the valence (θ_n) and conformational (ϕ_n) angles are given by

$$U_3(\theta_n)=\tfrac{1}{2}K_\theta(\theta_n-\theta_0)^2,$$

$$U_4(\phi_n)=\alpha+\beta\cos\phi_n+\gamma\cos 3\phi_n;$$

and, atoms separated by more than two neighbors or belonging to different chains interact through the potential

$$U(r)=\begin{cases}U_{LJ}(r)-U_{LJ}(R), & r\leqslant R,\\ 0, & r>R,\end{cases}$$

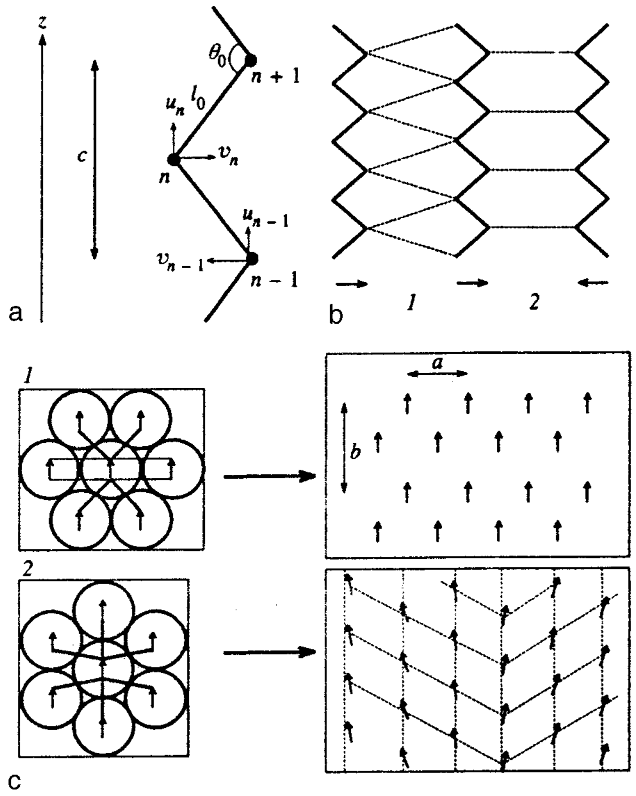

FIG. 1. Model of a polymer crystal (polyethylene with collective atoms): (a) chain parameters and local atomic coordinates; (b) two types of mutual locations of neighboring chains in a crystal (the lengths of the dashed lines are slightly less than the equilibrium distance for a van der Waals potential), and (c) the possible equilibrium configurations for packing of plane zigzags: (*1*) stable and (*2*) unstable (shown here is the plane of a transverse cross section of the molecules; the arrows denote the direction from the nearest atoms of the molecule under the plane to the nearest above the plane; the thick lines represent the bonds between molecules of type *1* in b and the thin lines, those of type *2* in b).

where $U_{LJ}(r)=4\varepsilon[(\sigma/r)^{12}-(\sigma/r)^6]$ is the Lennard-Jones potential with a minimum at $r_0=2^{1/6}\sigma$. The numerical values of the constants employed here are listed in Table I.

We have taken periodic boundary conditions for the crystal in all three directions. The cells of the computational grid were chosen to have the shape of a rectangular parallelepiped. The corresponding classical first order Lagrange equations were solved numerically using a Berle leap-frog algorithm[12] taking note of the limitations imposed by the rigid bonds.[13]

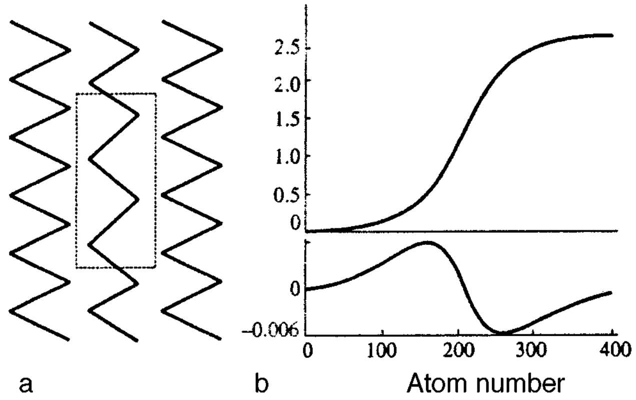

FIG. 2. A vacancy in a polymer crystal chain: (a) the shape of the defect (the dotted box represents the defect region) and (b) the longitudinal displacements of atoms from the equilibrium position (Å) in molecules with a defect (upper curve) and in molecules in the first coordination sphere (lower curve).

TABLE I. Parameters of the model crystal.

Parameter	Value	Ref.	Parameter	Value	Ref.
m	14 a.m.u.	–	β	1.675 kJ/mole	10
l_0	1.53 Å	10	γ	6.695 kJ/mole	10
θ_0	113°	10	ε	0.4937 kJ/mole	11
K_θ	331.37 kJ/mole	10	σ	3.8 Å	11
α	8.370 kJ/mole	10	$R=2r_0$	8.531 Å	–

The periodic boundary conditions make it possible to follow the soliton dynamics in the direction along the axis of the molecules for an unlimited time and to avoid introducing unphysical boundary conditions such as a rigidly attached second coordination sphere in the plane of the transverse cross section. In order to keep a soliton from interacting with itself, the number of molecules in the grid cell was chosen so that the image of each molecule lies no closer than its fourth coordination sphere, while the length of the molecule for a soliton extending on the order of 35 chain periods (c) was assumed to be $200c$ (two CH_2 groups fall within a single period).

3. EQUILIBRIUM CRYSTAL CONFIGURATION OF A POLYETHYLENE CRYSTAL IN THE COLLECTIVE ATOM MODEL

When the chains are packed in a (mechanically) equilibrium configuration, they can have two kinds of mutual positions (Fig. 1b). The first is stable (*1*) and is roughly a factor of two more favorable energetically than the second (*2*), which is unstable (a saddle point). Since the projected length of a molecule on its perpendicular cross sectional plane is $l_\perp=0.843$ Å and the van der Waals radius $r_0=4.265$Å $\approx 5l_\perp$, the packing of the molecules in the crystal will be close to cylindrical. Two different (mechanically) equilibrium configurations are conceivable (Fig. 1c). Both have a monoclinic cell and similar energies (each atom has six van der Waals bonds with atoms of other chains). The second, however, is unstable and stratifies into two domains, both of which correspond to the first configuration. The parameters a and b of the stable structure depend on the cutoff radius R. Table II lists some theoretical estimates of these constants (the length of the dotted lines in Fig. 1b equals the van der Waals radius) and data relating to the relaxation of samples

TABLE II. The parameters a and b (Å) for the equilibrium crystalline configuration with different cutoff radii R.

Parameter	$R=r_0$ (est.)	$R=1.8r_0$	$R=2r_0$
a	4.265	3.998	3.980
b	8.618	7.994	7.966

with different R. The period along the molecular axis is always equal to $c=2.554$ Å. The density of the sample is $\rho=1.155$ g/cm^3.

Molecular-dynamics modelling of a polyethylene crystal in the collective atom approximation has shown that there is no local minimum in the potential energy for an orthorhombic structure for any such cell parameters. The numerical simulation yields a minimum of this sort only for a model of polyethylene in which the CH_2 group is modelled by three spatially separated force centers. Thus, an orthorhombic structure in polyethylene can exist only because of the presence of lateral groups, and not of the shape of the chain skeleton.

4. THEORETICALLY EXPECTED CHARACTER OF THE VACANCY DYNAMICS IN A POLYMER CRYSTAL CHAIN

We shall show that the simplest description of the dynamics of a vacancy in a chain of a polymer crystal in the approximation of immobile neighbors can be reduced to a sine-Gordon equation for the transverse displacement u of the atoms in the chain. (See Fig. 1a.)

In fact, the effective potential of the matrix is obtained by calculating the crystal energy with all the molecules fixed but one, which moves along the axis. In the stable equilibrium configuration of the crystal it can be approximated to within two percent by two harmonics:

$$V(u)=A\left[1-\cos\left(\frac{2\pi}{c}u\right)\right] \tag{1}$$

with $A=0.274$ kJ/mole.

The condition of rigid bonds determines the relationship among the longitudinal and transverse displacements, u and v, of the atoms. (See Fig. 1a.) Going to the continuum approximation in the system lagrangian (justified by the relative weakness of the intermolecular interaction) and neglecting the dispersion and nonlinearity owing to the intramolecular interactions (a more accurate continuum description can be found elsewhere[14]), we obtain a sine-Gordon equation with $K=K_\theta(2\tan(\theta_0/2))^2$ for the longitudinal displacement $u(z,t)$:

$$mu_{tt}-Ku_{zz}+A\left(\frac{2\pi}{c}\right)^2\sin\left(\frac{2\pi}{c}u\right)=0.$$

Here K is the rigidity parameter of the chain, so that the sound speed $v_s=\sqrt{K/m}$, while A characterizes the height of the barrier between two neighboring positions of the atoms in the matrix potential (1).

The sine-Gordon equation has well known soliton solutions corresponding to vacancies:

$$u(z,t)=\frac{c}{2\pi}4\arctan\exp\left(\frac{z-vt}{L}\right), \tag{2}$$

where v is the soliton velocity ($v<v_s$), $L=L_0\sqrt{1-(v/v_s)^2}$, and $L_0=(c/2\pi)\sqrt{K/A}$ is the half width of a motionless defect, which is larger the more rigid the chain is relative to the matrix. For our values of the constants (Table I), we have $v_s\approx14.76$ km/s and $L_0\approx35(c/2)$.

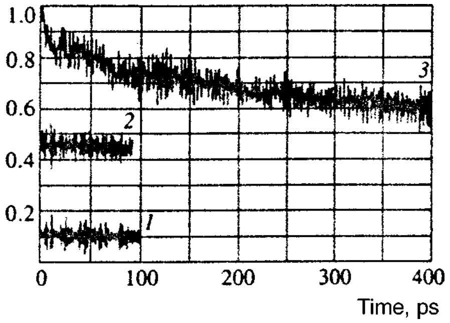

FIG. 3. Vacancy dynamics in an equilibrium crystal: curves *1* and *2* illustrate the conservation of low and medium (0.1 and 0.45 times the sound speed) velocities of defects in a chain surrounded by mobile neighbors; curve *3*, the slowing down of fast (0.9 times the sound speed) defects with a transition to stationary soliton motion at a lower velocity ($\sim$0.6 times the sound speed).

Therefore, a vacancy (without breaking of covalent bonds, Fig. 2a) can move along the chain with a near sonic velocity, maintaining its localization and without disrupting the crystal structure outside the region of the defect. This means that at velocities not too close to that of sound, the vacancy dynamics should be those of a soliton. (When $v\rightarrow v_s$ it is no longer possible to neglect either the discreteness or the intramolecular nonlinearity; solitons are narrow.) This conclusion, however, is based on a quasi-one dimensional approximation for the immobile neighboring chains. Now our goal is to study the vacancy dynamics in a crystal where all the chains are mobile.

5. RESULTS OF A MOLECULAR-DYNAMIC SIMULATION OF THE BEHAVIOR OF VACANCIES IN A POLYMER CRYSTAL

This system is characterized by the following time scales: the transit time for sound over one chain period $\approx1.7\cdot10^{-2}$ ps, the soliton width ≈1.22 ps, and the period of the oscillations of the atoms in a chain in the potential of neighboring chains ≈1.39 ps.

In the molecular-dynamics simulation, for one of the molecules in a crystal that had relaxed and cooled to a temperature of 0.1 K we specified atomic displacements and velocities in accordance with the approximate analytic formula (2) and followed the evolution of the defect for a long time (on the order of hundreds of picoseconds). Almost instantly the soliton acquired a shape exactly consistent with the crystalline environment. In a sample with mobile neighbors, the presence of a vacancy in one chain causes nonuniformities to appear in all the chains within the first coordination sphere, i.e., ''shadows'' (see Fig. 2b) which accompany the defect even when it moves. In the numerical simulation we tracked the velocity v_m of the center of mass of a chain with a defect, which rescales to the vacancy velocity $v_{\text{vac}}=-(N/2)v_{cm}$ (where N is the number of atoms in the chain). ''Imbedding'' a defect in a crystal excites (for both mobile and immobile neighbors) thermal vibrations of the atoms (up to a sample temperature of several K), so rapid oscillations are superimposed on the true vacancy velocity which have nothing to do with it (Fig. 3).

We now describe the results of simulating the evolution of vacancies with initial velocities of 0.9, 0.45, and 0.1 times that of sound (13.2, 6.6, and 1.5 km/s) in a stably equilibrium crystal (Fig. 1c, frame *1*). If the neighboring molecules are fixed, defects with all three of these velocities move smoothly along the chain with their velocities essentially unchanged for at least 100 ps. Over this time, they can cover 5186, 2593, and 576 chain periods, respectively. When all the molecules are mobile, the dynamics of the two slowest defects is the same (Fig. 3). Only for the fastest defect does the velocity fall off slowly, decreasing to ≈ 0.6 times the sound speed (9 km/s) over 400 ps. Therefore, the mobility of the surrounding chains does not affect the behavior of vacancies moving at low and medium velocities, but slows down the faster defects to medium velocities without changing their soliton-type dynamics.

This slowing down is not caused by the radiation of energy by the defect into the chain along which it moves owing to effects of the discreteness which show up at higher defect velocities:[15] for $v=0.9v_s$ the half width L of the solitons is still quite large $L\approx 15(c/2)$ [Eq. (2)]. Evidently, the reason for the slowing down of fast defects in a crystal is their more intense interaction with the mobile molecules of neighboring chains. This effect has not been observed before and requires separate theoretical and numerical study.

Note that the configuration of the crystalline environment has no effect on the character of the dynamics of a defect in a chain if the neighboring chains are immobile, but changes it fundamentally when the neighboring chains are mobile. Specifically, in nonequilibrium or unstable structures (e.g., a nonequilibrium orthorhombic structure in the collective atom model of polyethylene[8] or an unstable monoclinic structure; Fig. 1c, frame *2*) which require rotation of chains for relaxation, a vacancy will be slowed down rapidly. For example, in the latter case a vacancy with an initial speed of 4.6 km/s is stopped over a time on the order of 10 ps, having covered about 100 chain periods. The onset of this kind of slowing down was apparently observed in Ref. 8 and served as the basis for the unjustified claim that the twisting defect dynamics did not exhibit soliton behavior. In a molecule at the boundary between two domains in a relaxed stratifying sample (Fig. 1c, frame *2*), point defects are also slowed down and change the structure of the boundary between the domains.

Therefore, in an equilibrium polymer crystal, vacancies caused by localized tensile deformation of the chain have a soliton-type mobility.

This work was supported by the Russian Fund for Fundamental Research (Grants 98-03-33366a and 97-02-17825a).

[1] R. H. Boyd, Polymer **26**, 323 (1985).

[2] R. H. Boyd, Polymer **26**, 1123 (1985).

[3] V. V. Ginzburg, L. I. Manevich, and N. G. Ryvkina, Mekanika kompozitnykh materialov, No. 2, 249 (1991).

[4] K. Lonngren and E. Scott, Eds., *Solitons in Action* Academic Press, New York (1978).

[5] V. V. Ginzburg and L. I. Manevich, Fiz. Tverd. Tela **32**, 2414 (1990) [Sov. Phys. Solid State **32**, 1401 (1990)].

[6] D. H. Reneker and J. Mazur, Polymer **29**, 3 (1988).

[7] L. I. Manevich and A. V. Savin, Vysokomolekulyarnye soedineniya A **140**, 788 (1998).

[8] D. W. Noid, B. G. Sumpter, and B. Wunderlich, Macromolecules **24**, 4148 (1991).

[9] N. K. Balabaev, O. V. Gendel'man, M. A. Mazo, and L. I. Manevich, Zh. Fiz. Khim. **69**, 24 (1995).

[10] D. W. Noid, B. G. Sumpter, and B. Wunderlich, Macromolecules **23**, 664 (1990).

[11] D. Rigby and R. J. Roe, Macromolecules **22**, 2259 (1989).

[12] M. P. Allen and P. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford (1987).

[13] P. G. Khalatur, N. K. Balabaev, and A. S. Pavlov, Mol. Phys. **59**, 753 (1986).

[14] L. I. Manevitch and A. V. Savin, Phys. Rev. E **55**, 4713 (1997).

[15] M. Peyrard and M. D. Kruskal, Physica D **14**, 88 (1984).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Dependence of exchange interactions on chemical bond angle in a structural series: cubic perovskite–rhombic orthoferrite–rhombohedral hematite

I. S. Lyubutin,*) T. V. Dmitrieva, and A. S. Stepin

Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russia

(Submitted 28 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1070–1084 (March 1999)

Mössbauer spectroscopy is used to study the hyperfine magnetic fields at tin ^{119}Sn ions introduced as an isomorphic impurity in the lattices of the orthoferrites $RFeO_3$. The large reduction in the field H_{hf}^{Sn} (4.2 K) observed when R is changed from La to Lu correlates with the drop in the Néel point and indicates that the exchange interactions are decreasing over this series. A crystal chemical analysis of the structural series with the general formula ABO_3 shows that the ideal structure of cubic perovskite can be converted to a rhombohedral hematite-corundum structure by simple rotation of the $[BO_6]$ octahedra if the B–O interionic distances remain unchanged. The rhombic distortions are associated with a reduction in the B–O–B bond angle from $\theta=180°$ in perovskite to $\sim 132°$ in hematite. The rare earth orthoferrites $RFeO_3$ follow the same mechanism for structural transformations and the $LaFeO_3-LuFeO_3$ series occupies an intermediate position $(157° > \theta > 142°)$ between the extreme members of the series mentioned above. A reduction in the bond angle leads to weakening of the Fe–O–Fe exchange interaction, which shows up as a drop in the Néel temperature and in the hyperfine magnetic field at the nucleus. An analysis of theoretical models shows that for a suitable choice of the exchange and transfer parameters, the angular variation in the parameters of the exchange interaction is described fairly well by the Moskvin theory over a rather wide range of angles θ. The contributions to the fields H_{hf}^{Sn} and H_{hf}^{Fe} from the t_{2g}- and e_g-orbitals of neighboring paramagnetic ions in the orthoferrites and orthochromites are examined.

[S1063-7761(99)02203-9]

1. INTRODUCTION

The rare earth orthoferrites $RFeO_3$ (R is a rare earth element) have a distorted perovskite crystal structure.[1] All the Fe^{3+} ions are crystallographically equivalent and lie in an octahedral oxygen environment. Each iron ion is surrounded by six Fe^{3+} ions and the Fe–O–Fe bond angle is close to 180°.[2] This results in a strong negative indirect Fe^{3+}–O–Fe^{3+} exchange interaction, which causes the formation of two magnetic sublattices whose moments are almost antiparallel. The small angle between the moments of these sublattices produces a small ferromagnetic moment.[3,4] The rare earth elements become magnetically ordered only at very low temperatures, and for $T > 10$ K they have no effect on the basic magnetic properties of the orthoferrites.

Unusual behavior of the Néel temperature T_N has been observed in the $RFeO_3$ series from R=La to R=Lu:[5] T_N falls by almost 120 K, even though the dimensions of the unit cell decrease so the Fe–O–Fe exchange bond length should also decrease. Usually this leads to a strengthening of the exchange interaction and, therefore, to a rise in T_N. Note, for example, that in the series of rare earth ferrites with a garnet $R_3Fe_5O_{12}$ structure, T_N is independent of the atomic number of the rare earth element and is almost constant for all the rare earth elements.[6-8] A diminution in the exchange interactions as R is varied from La to Lu has also been observed in Mössbauer studies of the rare earth orthoferrite series $RFeO_3$ doped with the tin isotope ^{119}Sn.[9]

In this paper we perform a crystal chemical analysis of the structural series ABO_3, which includes cubic perovskite, the rhombic orthoferrites, and rhombohedral hematite (corundum). It is shown that the structural transformations in this series can be described in terms of a unified mechanism and, for fixed B–O interionic distances, are determined exclusively by the change in the angle θ of the B–O–B chemical bond. Mössbauer spectroscopy at $T=4.2$ K is used to measure the hyperfine magnetic field at tin Sn^{4+} ions (H_{hf}^{Sn}) implanted as measurement probes in the iron sublattice of the $RFeO_3$ orthoferrites. These observations are used to explain the unusual magnetic properties of the rare earth orthoferrites. Theoretical models for the angular dependence of the exchange interaction in the $RFeO_3$ orthoferrites and $RCrO_3$ orthochromites are examined. An attempt is made to apply the theory to a wider range of angles θ based on the experimental data.

2. CRYSTAL CHEMICAL ANALYSIS OF THE STRUCTURAL SERIES CUBIC PEROVSKITE-RHOMBIC ORTHOFERRITES-RHOMBOHEDRAL HEMATITE (CORUNDUM)

The rhombic orthoferrites $RFeO_3$ (space group[1] $Pbnm$) are usually attributed to the perovskite-like structures obtained as a result of all the possible distortions of the cubic $CaTiO_3$ perovskite lattice (space group[10] $Pm3m$) on replacing the Ca cations by R and Ti by Fe. These distortions can change the crystal symmetry. A transition from the cubic perovskite structure to a rhombic structure in the rare earth orthoferrites has been established experimentally.[10,11] RMO_3 compounds with M=Cr, Sc, V, Ga, and Al also have this type of structure.[12-16] It is interesting to note that the deformation which develops for the rhombic distortions has almost no effect on the oxygen octahedron surrounding the iron ions, while the coordination polyhedron of the R^{3+} cations is significantly distorted. The average interatomic Fe–O and O–O distances are essentially constant for the entire rare earth series $RFeO_3$, at 2.011 and 2.844 Å, respectively.[10] As R is changed, only the mutual positions of the oxygen octahedra change and this leads to a change in the angle of the Fe–O–Fe valence bond. The deviation from cubic symmetry increases as the R^{3+} ionic radius decreases; it is smallest in $LaFeO_3$ and greatest in $LuFeO_3$.

Based on the available crystal chemical data, we may assume that there is a single mechanism for the structural changes in the series cubic perovskite-rhombic orthoferrites-rhombohedral hematite-corundum (general formula ABO_3). Here we shall show by purely geometric arguments, that there is an analogy in the structural models for cubic perovskite and rhombohedral hematite (corundum) and one phase can be transformed into the other only through a certain rotation of the $[BO_6]$ octahedra about one another. During these rotations, the most important distortions take place in the $\{111\}$ planes for the cubic model and in the basis plane for the hexagonal (rhombohedral) model. Figure 1 shows structural diagrams of the extreme phases in this series, perovskite (Fig. 1a) and hematite (Fig. 1b), constructed for ideal $[BO_6]$ octahedra. A atoms, shown here in only one place, occupy all the empty spaces. The patterns of the structure which are formed by the upper facets of the $[BO_6]$ octahedra of the lower layer and the lower facets of the $[BO_6]$ octahedra of the upper layer are distinguished in Figs. 1a and 1b. Each of the two ''stages'' of the octahedral layer is represented by three isolated octahedra, while both ''stages'' form a six-member ring in the infinite layer of octahedra. The spheres shown in Figs. 1a and 1b (to the left) illustrate the dense packing of the large A ions in the plane grids.

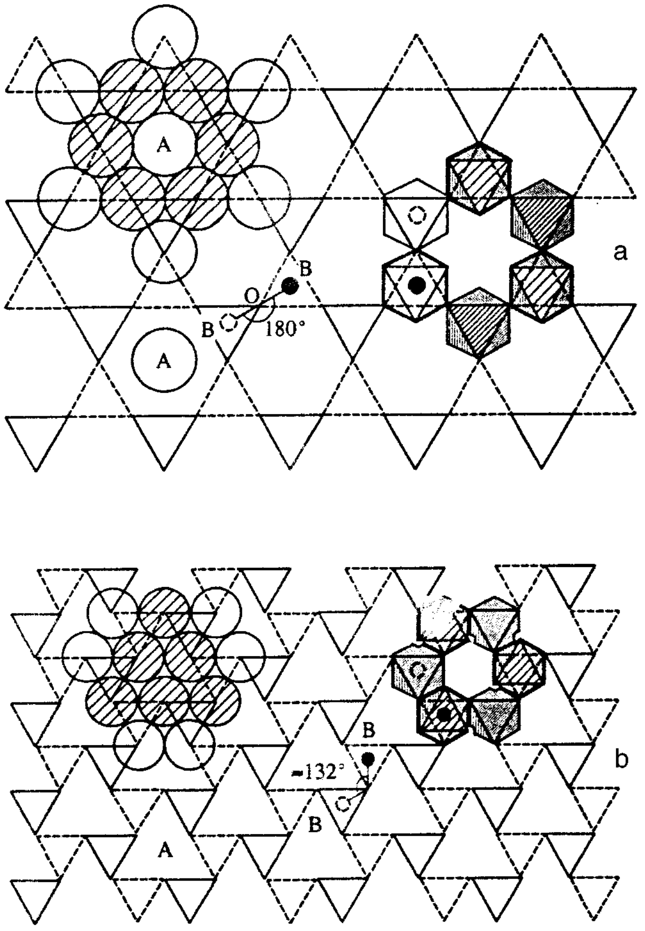

FIG. 1. Two dimensional lattices of the densest oxygen packing in a cubic perovskite structure (a) and in a rhombohedral hematite-corundum structure (b). The structural schemes of the final phases constructed from ideal $[BO_6]$ octahedra are shown here. The octahedra illustrate the pattern of the structure and lie above and below the lattice layer. (See text.)

Ideal perovskite ABO_3 can be represented as the densest packing of O and A atoms. (In the ideal case O=A with the general formula BO_4.) In the case O=A, shown on the left of Fig. 1a, dense packing occurs with a packing index of 0.7405 (packing or filling coefficient): $q=[\Sigma(V_{at})_i]/\Omega$, where $(V_{at})_i=(4/3)\pi r_i^3$ is the volume of a single atom and Ω is the cell volume. For the densest packing, $q=74.05\%$. B atoms occupy 1/4 of the octahedral voids (on the right of Fig. 1a), while the remaining octa- and tetrahedral voids remain unoccupied. The $[BO_6]$ octahedra with common vertices form the simplest octahedral skeleton with A atoms in its voids.

Figure 1b shows the other extreme end of the series, rhombohedral hematite α-Fe_2O_3 (corundum Al_2O_3). Here the customary densest packing of oxygen atoms is realized, with 1/3 of the octahedral voids occupies by iron atoms. There are two possible ways to fill the remaining octahedral voids with A atoms (in this case iron atoms, since A=B=Fe for hematite). In the first, that third of the octahedral voids which form common boundaries with $[FeO_6]$ are filled, while the second is enantiomorphic (a mirror reflection) to the first.

The transition of perovskite into hematite and the resulting distortions can be imagined by comparing Figs. 1a and 1b. The perovskite cubic skeleton can be regarded as the result of a symmetrizing of the skeleton of the $[BO_6]$ octahedra during densest oxygen packing of rhombohedral hematite. Since the rhombic distortions of the unit cells of perovskite-like structures are assumed to be negligible and the B–O distances hardly change, it seems, at first glance, that the changes in the B–O–B bond angle θ should also be small. However, an elementary geometrical analysis (see Figs. 1a and 1b) that this is not so. $\theta=180°$ for ideal perovskite (i.e., when the ratio of the ionic radii $r_A/r_O=1$). At the other extreme of the series (hematite), in the ideal case of dense packing (i.e., for $r_A/r_O=0.414$ and $r_B/r_A=1$) this angle is $2\arcsin\sqrt{5/6}=13141'$. This estimate coincides with the experimental value of $\theta\approx132°$ in hematite.[17] Note that the transition from a hematite-corundum structural type to a perovskite structure has been observed experimentally in $InGaO_3$ under high pressure.[18]

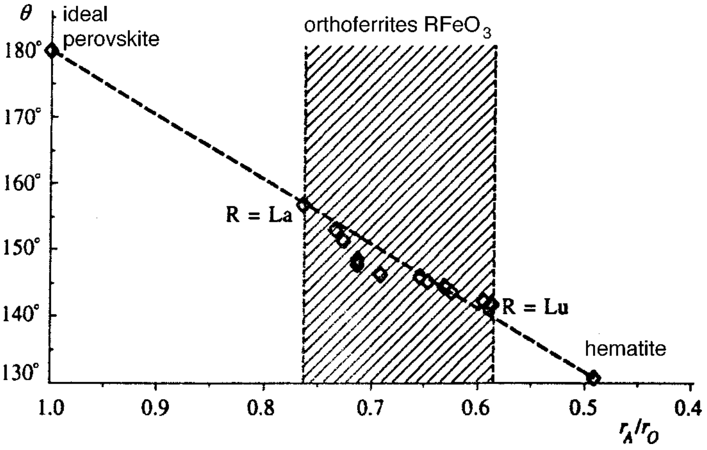

FIG. 2. The B–O–B valence bond angle in the structural series ABO_3 as function of the ratio r_A/r_O of the ion radii. The points are experimental values. The dashed line lies between the extreme members of the series, perovskite and hematite. The intermediate region which applies to the rare earth orthoferrite series $RFeO_3$ is shaded. The ion radii are taken from Belov and Bokiya.

According to the proposed mechanism, for the intermediate members of the rare earth orthoferrite series, θ should lie somewhere between 180° and 132°. (See Fig. 2.) Unfortunately, a geometric analysis cannot be used to determine how θ varies with the ionic radius r_A. However, assuming a linear variation $\theta = f(r_A/r_O)$, it is easy to make an approximate estimate of the ranges of variation in θ for the rare earth orthoferrites. (See Fig. 2). We found that from lanthanum orthoferrite ($r_{La}/r_O \approx 0.765$) to lutecium orthoferrite ($r_{Lu}/r_O \approx 0.588$) the angle θ varies from ~158° to ~140°. This estimate for the extreme members of the series is in good agreement with the experimental values of these angles.[10]

3. EXPERIMENT

We prepared a series of semicrystalline samples of the rare earth orthoferrites $(R_{1-z}Ca_z)[Fe_{1-z}Sn_z]O_3$ with R=La, Pr, Nd, Sm, Gd, Tb, Ho, Er, Yb, Lu, and Y, in which the Fe^{3+} iron ions were partially ($z=0.05$) replaced by Sn^{4+} ions. Here in order to maintain the neutrality of the molecules, rare earth R^{3+} ions were replaced by the doubly valent Ca^{2+} ion. In order to increase the accuracy of the Mössbauer studies, the tin was enriched in the isotope ^{119}Sn to 95%. X-ray analysis showed that all the samples are in a single phase, have a distorted perovskite-type structure (space group *Pbnm*), and are isostructural with the orthoferrite $GdFeO_3$.[1]

The Mössbauer absorption spectra of the ^{119}Sn nuclei were taken on an electrodynamic system in a constant acceleration regime. The $Ca^{119m}SnO_3$ gamma-ray source was kept at room temperature. The measurements were done at temperatures ranging from 4.2 K to the Néel point. Several control measurements of the Móssbauer spectra of ^{57}Fe were also made.

At low temperatures magnetic hyperfine splitting is observed in the Mössbauer spectra of ^{119}Sn owing to the Zeeman interaction of the tin nuclei with the effective magnetic field H_{hf}^{Sn}. This field develops because of the transfer of spin density from the paramagnetic Fe^{3+} ions to the diamagnetic Sn^{4+} ion.[19-22] The transfer occurs along the Fe–O–Sn chain with the participation of an intermediate oxygen ion. In terms of its physical nature, this interaction is analogous to the indirect Fe–O–Fe exchange interaction and the field H_{hf}^{Sn} is referred to as the supertransferred hyperfine magnetic field[23-25] or the indirect hyperfine interaction field H_{hf}^{Sn}.[26]

A calculation of the hyperfine interaction showed that over all the compositions there are no quadrupole shifts to within the limits of error. The isomeric chemical shifts relative to the source are zero (for equal temperatures of the source and absorber). No special studies of the temperature shifts were made. The magnitude of the magnetic field H_{hf}^{Sn} at a tin nucleus in the $RFeO_3$ lattice depends strongly on the rare earth element. At $T=4.2$ K it has a maximum of 257 ± 1 kOe for R=La and falls to 161 ± 1 kOe for R=Lu. The Mössbauer spectra remain well resolved over the entire temperature range up to the Néel point T_N. T_N was defined by the magnetic field's going to zero at the tin nucleus. Figure 3 shows the variation in T_N and H_{hf}^{Sn} in the series from $LaFeO_3$ to $LuFeO_3$. The clear analogy between these curves is confirmation of a direct relationship between the exchange interaction and spin density transfer from a magnetic to a diamagnetic atom and indicates a correlation of the quantities T_N

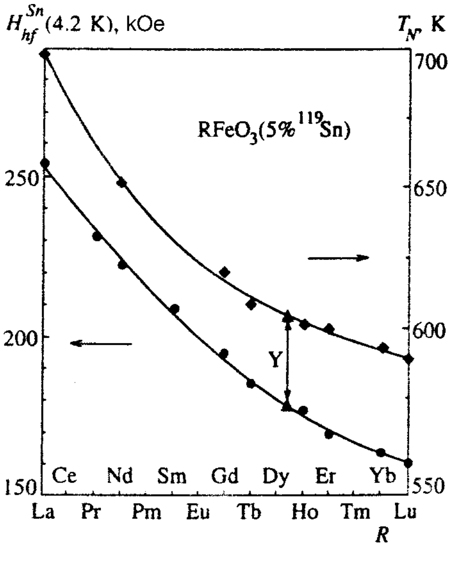

FIG. 3. The Néel temperature T_N and the hyperfine magnetic field H_{hf}^{Sn} at the tin nuclei at $T=4.2$ K as functions of the atomic number of the rare earth elements in the orthoferrites $(R_{0.95}Ca_{0.05})[Fe_{0.95}Sn_{0.05}]O_3$.

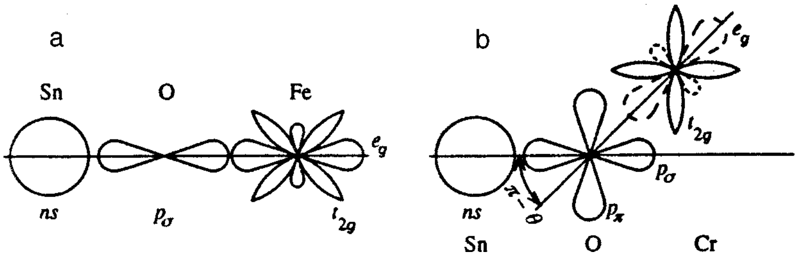

FIG. 4. Illustrating the overlapping of the atomic orbitals of a diamagnetic cation, anion, and paramagnetic cation: (a) for a bond angle $\theta=180°$, so the interaction through the e_g-p_σ-orbitals predominates; (b) for a bond angle $\theta<180°$; besides the σ-bonds, the π-bonds also participate in the exchange interaction. The dotted curves represent the unfilled e_g-orbital of the Cr^{3+} ion.

and H_{hf}^{Sn} with the Fe–O–Fe(Sn) exchange bond angle.

4. ANGULAR DEPENDENCE OF H_{hf} AND T_N IN THE RARE EARTH ORTHOFERRITES

A number of authors have attempted to obtain an analytic expression for the angular dependence of the supertransferred field H_{hf} in crystals with various structures. Zavadskiĭ, *et al.*[25,27] were among the first to obtain this dependence for a spinel and orthoferrite structure based on the molecular orbital technique including only bonding orbits. It was assumed that in a pure (undoped by diamagnetic ions) crystal, the field H_{hf} develops as a result of the transfer of spin density to a central Fe^{3+} (↑) ion from surrounding Fe^{3+} (↓) ions. This takes place through the following mechanism: $2p$-orbitals of oxygen are polarized as a result of charge transfer to unoccupied $3d$-orbitals of Fe^{3+} (↓) and overlap with occupied $3d$-orbitals of these ions (Fig. 4). This, in turn, leads to the transfer of $2p$-electrons into the empty $4s$-shell of Fe^{3+} (↑) ions and overlap of the polarized $2p$-orbits with the inner s-shells of the core of this ion. The following expression for H_{hf}^{Fe} at a central iron ion was obtained:[27]

$$H_{hf}^{\mathrm{Fe}}=525N^4k\left[-\sum_{n=1}^{3}S_{ns}\varphi_{ns}(0)+a_4\varphi_{4s}(0)\right]^2 \times\{(A_\sigma^2-A_\pi^2)\cos^2\theta+A_\pi^2\}. \tag{1}$$

Here S_{ns} are the overlap parameters of the oxygen orbitals with the inner s-shells of iron for the $O^{2-}-Fe^{3+}(\uparrow)$ pair, a_4 is the transfer integral for a $2p$-electron of oxygen into the outer $4s$-shell of iron, $\varphi_{ns}(0)$ are the wave functions of the Fe^{3+} ion, A_σ^2 and A_π^2 are the covalence parameters of the Fe–O bond, which characterize the number of unpaired spins (the fraction of unpaired spin density) in the $2p$-shell of O^{2-} formed as a result of transfer of $2p$-electrons of O^{2-} into empty $3d$-orbitals of $Fe^{3+}(\downarrow)$ ions ($A_{\sigma,\pi}=(B_{\sigma,\pi}+S_{\sigma,\pi})$, where $B_{\sigma,\pi}$ and $S_{\sigma,\pi}$ are, respectively, the $O^{2-}-Fe^{3+}(\downarrow)$) transfer and overlap parameters, k is the number of iron ions surrounding a central cation, and N is a normalization constant. Therefore, the cofactor in curly brackets characterizes the transfer of $2p$-electrons of the ligand into empty $3d$-shells of the surrounding ions, while the cofactor in square brackets characterizes the transfer of unpaired spin density into the outer s-shell of the central ion and the overlap of unpaired $2p$-electrons with inner ns-electrons of the central cation.

The above mechanism for the interaction of two paramagnetic Fe–O–Fe ions can also be applied to the interaction between paramagnetic and diamagnetic Fe–O–Sn ions (Fig. 4a). Extending Eq. (1) to the case of an impurity tin ion with an outer $5s$-shell in the orthoferrite matrix, we can, by analogy with Ref. 27, write

$$H_{hf}^{\mathrm{Sn}}=525N^4k\left[-\sum_{n=1}^{4}S_{ns}\varphi_{ns}(0)+a_5\varphi_{5s}(0)\right]^2 \times\{(A_\sigma^2-A_\pi^2)\cos^2\theta+A_\pi^2\}. \tag{2}$$

Here $\varphi_{ns}(0)$ are the s-wave functions of tin, and S_{ns} and a_5 are the overlap and transfer parameters, respectively, for $O^{2-}-Sn^{4+}$. The covalence parameters A_σ^2 and A_π^2 are generally independent of the $Fe^{3+}-O^{2-}$ interionic separation. In the $RFeO_3$ orthoferrites, however, this distance is essentially unchanged when R goes from La to Lu. This follows from structural data[10] and is also confirmed by the constancy of the isomeric shift in this series, as found by Mössbauer studies on ^{57}Fe nuclei.[27] It may also be assumed that the $Fe^{3+}-O^{2-}$ distances do not change for small amounts of substitution of iron by tin. (This is confirmed by our data on the isomeric shifts for the ^{119}Sn and ^{57}Fe nuclei.) Thus, we shall assume that A_σ^2 and A_π^2 are constant and the same as the corresponding quantities for the pure orthoferrites.

Unfortunately, the difficulty of theoretically estimating the parameters of the $O^{2-}-Sn^{4+}$ interaction makes it impossible to perform any sort of accurate calculation of the absolute magnitudes of the field H_{hf}^{Sn}. However, based on experimental data, we can attempt to estimate the ratio of A_σ^2 and A_π^2. Since the first cofactor in Eq. (2) can be regarded as constant for fixed Sn–O distances, Eq. (2) can be written in the form

$$H_{hf}^{\mathrm{Sn}}=K\{(A_\sigma^2-A_\pi^2)\cos^2\theta+A_\pi^2\}. \tag{3}$$

Then the ratio A_π^2/A_σ^2 can be calculated by plotting the experimental $H_{hf}^{\mathrm{Sn}}=f(\cos^2\theta)$ curve (Fig. 5). If we write Eq. (3) in the form

$$H_{hf}^{\mathrm{Sn}}=a_0+a_1x,\quad \text{where}\quad x=\cos^2\theta,$$

then it is easy to see that $A_\pi^2/A_\sigma^2=a_0/(a_0+a_1)$. A computer analysis of the $H_{hf}^{\mathrm{Sn}}=f(\cos^2\theta)$ curve for the orthoferrites (see Fig. 5, curve *1*) yields the values $a_0=-102$ and $a_1=421$. Thus, $A_\pi^2/A_\sigma^2\simeq-0.32$.

On the other hand, the Néel temperature T_N is determined by the same set of covalence parameters and[27]

$$T_N\sim A_\sigma^2\{2A_\pi^2+(A_\sigma^2-2A_\pi^2)\cos^2\theta\}. \tag{4}$$

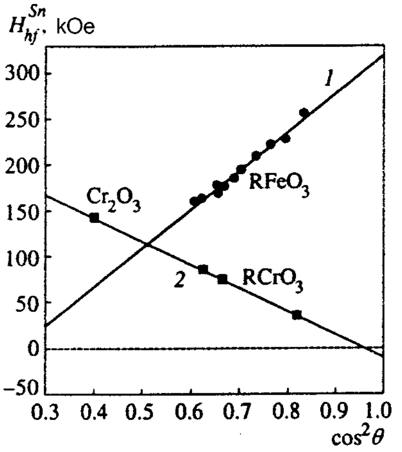

FIG. 5. Plots of $H_{hf}^{Sn}=f(\cos^2\theta)$ for the orthoferrites (data from this paper) and orthochromites (field values from Ref. 33). Also shown is the field for Cr$_2$O$_3$ from Ref. 30.

Thus the experimental $T_N=f(\cos^2\theta)$ curve can be used to make an independent estimate of the ratio A_π^2/A_σ^2 if we write Eq. (4) in the form $T_N=b_0+b_1x$, where $x=\cos^2\theta$ and $A_\pi^2/A_\sigma^2=b_0/2(b_0+b_1)$.

Figure 6 shows a plot of $T_N=f(\cos^2\theta)$ which we have constructed for values of T_N obtained from the temperature dependence of the fields H_{hf}^{Sn} for our tin-doped orthoferrite samples (curve *2*). As a comparison, the same dependence is shown for pure rare earth orthoferrites[5] (curve *1*). An analysis of these data shows that $A_\pi^2/A_\sigma^2=0.19\pm0.01$ for the pure orthoferrites, in accordance with the value obtained in Ref. 27, and $A_\pi^2/A_\sigma^2=0.20\pm0.01$ for the doped samples. Therefore, to within the experimental error the values of A_π^2/A_σ^2 are the same for the pure and doped samples. This confirms our assumption that for low doping A_σ^2 and A_π^2 are constant, although T_N changes significantly.

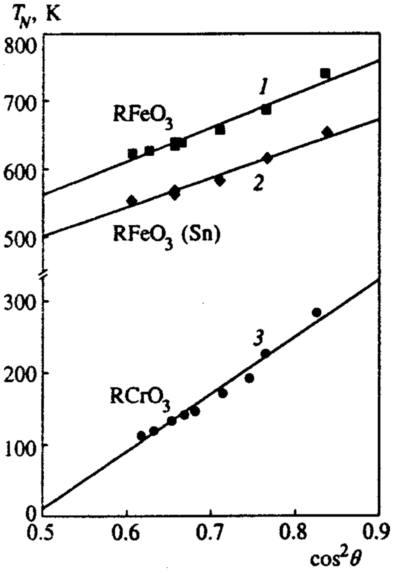

FIG. 6. Plots of $T_N=f(\cos^2\theta)$ for the orthoferrites and orthochromites. Curve *1* is T_N for the pure orthoferrites.[5] Curve *2* is our data for tin-doped orthoferrites. Curve *3* is T_N for the orthochromites.[13]

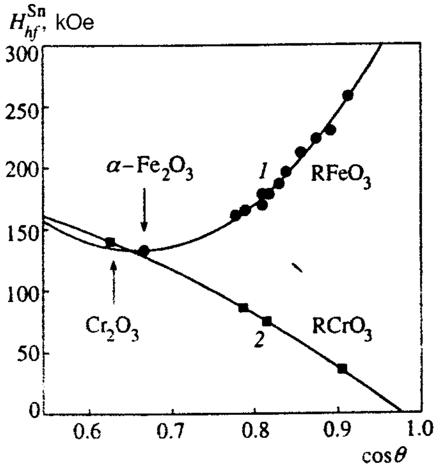

FIG. 7. Calculated (curves) and experimental (points) variation in the field H_{hf}^{Sn} (at $T=4.2$ K) with the cosine of the exchange bond angle. The calculated curves are (1) our data for the orthoferrites, including hematite, and (2) our calculation for the orthochromites, including Cr$_2$O$_3$.

The values of A_π^2/A_σ^2 for the pure orthoferrites obtained by various authors from experimental data differ by several times, ranging from $A_\pi^2/A_\sigma^2\approx0.16$[28] to $A_\pi^2/A_\sigma^2\approx0.38$.[29] Therefore, our result from data on T_N lies well within the range of published data. However, our value of $A_\pi^2/A_\sigma^2\simeq-0.32$ obtained from data on the fields H_{hf}^{Sn} at tin nuclei is negative, which conflicts with reality, since $A_\sigma^2>0$ and $A_\pi^2>0$. In addition, the field H_{hf}^{Sn} in the orthoferrites should be positive over the entire range of angles θ examined here, since $A_\sigma^2>A_\pi^2$. In addition, in describing the experimental dependence of the field with Eq. (3), we obtain $H_{hf}^{Sn}=-102$ kOe for $\theta=90°$.

Moskvin, *et al.*[26] have explained this conflict. They showed that including only the binding p-orbits of oxygen, as done in Ref. 27, may lead to an inaccurate result. In general, the contributions from the $2s$-orbits of oxygen must be included along with the $2p$-orbitals. Here the contribution of the $2s$-electrons, alone, to the field H_{hf}^{Sn} is independent of the exchange coupling angle θ, while the contribution of the $2p$-electrons remains proportional to $\cos^2\theta$ for a σ-bond and to $\sin^2\theta$ for a π-bond. However, there are overlapping sp-terms which make a contribution proportional to $\cos\theta$, only in a σ-bond. Then the angular dependence of the field H_{hf}^{Sn} is more precisely given by[26]

$$H_{hf}^{Sn}=\alpha+\beta\cos\theta+\gamma\cos^2\theta. \qquad (5)$$

Here the sign of the field should be positive over the entire range of angles θ, since the contributions to H_{hf}^{Sn} from $s-d$-exchange and from overlapping sp-terms are positive and increase as $\theta\to180°$. However, the resulting field H_{hf}^{Sn} should pass through a minimum within the interval $90°<\theta<180°$ when $\cos\theta=-\beta/2\gamma$.

Our experimental $H_{hf}^{Sn}=f(\cos\theta)$ curve for the orthoferrites is shown in Fig. 7, along with the curves calculated using Eq. (5). The experimental data on the orthoferrites are very well fit by the theoretical curve *1* for $140°<\theta<160°$ with the following parameters:

$$\alpha=+949\,\text{kOe},\quad \beta=+2487\,\text{kOe},\quad\text{and}\quad \gamma=+1889\,\text{kOe}.$$

This curve passes through a minimum at $\cos\theta = -0.658$ ($\theta \approx 132°$). The field at the minimum is $H_{hf}^{\mathrm{Sn}} \approx +130$ kOe, i.e., throughout the entire range of angles H_{hf}^{Sn} is positive. The curve with parameters $\alpha = +8$ kOe, $\beta = +241.5$ kOe, and $\gamma = +551$ kOe obtained in Ref. 26 provides an inferior approximation to the experimental data. Our analysis has shown that this curve should also have a minimum at $\cos\theta = -0.2191$, shifted toward smaller angles. However, at the minimum point, the field H_{hf}^{Sn} becomes negative, at $H_{hf}^{\mathrm{Sn}} = -18.45$ kOe. This contradicts the condition $H_{hf}^{\mathrm{Sn}} > 0$ and appears to indicate that the fit to the experimental data in Ref. 26 is not sufficiently accurate.

5. EXTENDING THE RANGE OF ANGLES TO θ=132° AND θ=180°

The range of angles θ corresponding to the series of rare earth orthoferrites can be extended by invoking data on the rhombic hematite α-Fe_2O_3. As shown in Sec. 2, the character of the distortions in hematite and the orthoferrites is identical and the Fe–O distances are essentially the same (1.98 and 2.01 Å, respectively); hence, the transfer and overlap parameters can be assumed to be close. For the theoretical analysis we have used the values of H_{hf}^{Sn} for hematite obtained by Fabrichnyĭ.[30] It turns out that the experimental point for α-Fe_2O_3, with astonishing accuracy, lies at the minimum of the parabola corresponding to the curve $H_{hf}^{\mathrm{Sn}} = f(\cos\theta)$ according to Eq. (5) for the orthoferrites with the parameters α, β, and γ given above (Fig. 7, curve *1*). Therefore, the theory of Moskvin, *et al.*[26] give a fully satisfactory description of the experimental data over a wide range of angles θ when the parameters α, β, and γ are properly selected.

An attempt can be made to extend the range of angles to $\theta \rightarrow 180°$ by invoking data on the fields H_{hf}^{Sn} in the compound MnO:^{119}Sn.[31] The antiferromagnetic material MnO has a face centered cubic NaCl structure where the Mn–O–Mn angles are 180 °. In a doped sample the magnetic moments of the 12 Mn^{2+} cations nearest to the Sn^{4+} ion (with 90-°ree Mn–O–Sn bonds) are completely compensated. Thus the total contribution to the field H_{hf}^{Sn} from the straight Mn-Sn, indirect 90-°ree Mn–O–Sn, and dipole interactions is equal to zero here. The field H_{hf}^{Sn} is created by six Mn^{2+} ions with identically directed spins through the 180-° Mn–O–Sn exchange interaction. However, although the electron shell of Mn^{2+} ($3d^5 - t_{2g}^3 e_g^2$) is the same as that of the Fe^{3+} ion, and the angle $\theta = 180°$ is more favorable for transfer of the spin density through $e_g - p_\sigma$-bonds, the field H_{hf}^{Sn} in MnO equals[30] 230 kOe and does not exceed that in $LaFeO_3$. This is evidently related to an increase in the interionic distances in manganese oxide (Mn–O=2.222 Å) compared to those in the orthoferrites (Fe–O=2.010 Å), which correlates with an increase in the radius of the paramagnetic ion $[r(Mn^{2+}) = 0.80\,\text{Å},\ r(Fe^{3+}) = 0.64\,\text{Å}]$. The °ree of covalency of the bond and the exchange interactions are reduced as the cation-anion distance is increased, and this also reduces spin density transfer to the diamagnetic cation. The field H_{hf}^{Sn} extrapolated to $\theta = 180°$ according to the data on the orthoferrites (Fig. 5) should be ~350 kOe, which is roughly 30% greater than that in MnO. This indicates that increasing the cation-anion distance by about 10% reduces, by ~30%, the covalence effects responsible for inducing the field H_{hf}^{Sn}. These estimates are in good agreement with similar estimates for a spinel structure[32] and indicate that the above theoretical ideas are applicable to a wide range of angles.

6. THE ANGULAR DEPENDENCE OF H_{hf} AND T_N IN THE RARE EARTH ORTHOCHROMITES

A similar analysis of the angular dependence of the field H_{hf}^{Sn} can be carried out for the rare earth orthochromites $RCrO_3$ by drawing on the experimental data of Moskvin, *et al.* [33] When R is changed from La to Lu in the orthochromites, rhombic distortions analogous to those in the orthoferrites take place, with a change in the Cr–O–Cr bond angle. Since $A_\pi^2 > A_\sigma^2$, it is to be expected that the field H_{hf}^{Sn} should have a maximum for $\theta = 90°$ and decrease with increasing θ. Transfer of oxygen $2p$-electrons into unfilled e_g-orbitals of chromium can occur for both spin orientations $(\downarrow)$ and $(\uparrow)$, so that here A_σ^2 represents the difference $(A_\sigma\downarrow)^2 - (A_\sigma\uparrow)^2$.[33] However, transfer of p-electrons with $(\uparrow)$ spin predominates because of the intraatomic polarization by the ''Hund'' interaction with the t_{2g}-electrons of the Cr^{3+} ion, which have a $(\uparrow)$ spin. This leads to a negative (relative to the direction of the magnetic moment of the Cr^{3+} ion) spin density on the $2p_\sigma$-orbitals of oxygen. Thus, the resulting value of A_σ^2 should be negative and for angles close to 180°, where the σ-bond plays a fundamental role, we should expect the sign of H_{hf}^{Sn} to change.

Besides the experimental data for the rare earth orthochromites, Fig. 7 includes the field H_{hf}^{Sn} for chromium oxide Cr_2O_3.[30] As for α-Fe_2O_3, the oxide Cr_2O_3 has a corundum-hematite crystal structure and, according to our ideology, can be regarded as an extreme member of the series $RCrO_3 - Cr_2O_3$. A computer analysis shows that all the experimental data can be fit satisfactorily to Eq. (5) (Curve *2* of Fig. 7) with the parameters

$$\alpha = +250\,\text{kOe}, \quad \beta = -17\,\text{kOe}, \quad \gamma = -244\,\text{kOe}.$$

Therefore, in the orthochromites the contribution proportional to $\cos\theta$ is negligibly small, since here the π-bonds are dominant. Thus, the experimental data give a satisfactory fit to the straight line $H_{hf}^{\mathrm{Sn}} = a_0 + a_1 x$, where $x = \cos^2\theta$ (Fig. 5, curve *2*). These data yielded $a_0 = 240 \pm 4$ kOe and $a_1 = -254 \pm 4$ kOe. Therefore, for the chromites we have $A_\pi^2 / A_\sigma^2 = -22 \pm 10$. Since A_π^2 is positive, $A_\sigma^2 = (A_\sigma\downarrow)^2 - (A_\sigma\uparrow)^2$ is negative, as is to be expected. Extrapolation to $\theta = 180$ indicates a sign shift in the field H_{hf}^{Sn} at $\theta \sim 167°$ (see Fig. 5) and gives $H_{hf}^{\mathrm{Sn}}(\theta = 180°) = -10 \pm 4$ kOe.

Using the values of the fields $H_{hf}^{\mathrm{Sn}}(180°) = -10 \pm 4$ kOe and $H_{hf}^{\mathrm{Sn}}(90°) = 240 \pm 5$ kOe obtained from this analysis, we have tried to estimate the absolute magnitudes of A_π^2 and A_σ^2. Equation (2) gives

$$H_{hf}^{\mathrm{Sn}}(180^\circ)=525N^4k\left[-\sum_{n=1}^{4}S_{ns}\varphi_{ns}(0)+a_5\varphi_{5s}(0)\right]^2A_\sigma^2,$$

$$H_{hf}^{\mathrm{Sn}}(90^\circ)=525N^4k\left[-\sum_{n=1}^{4}S_{ns}\varphi_{ns}(0)+a_5\varphi_{5s}(0)\right]^2A_\pi^2.$$

Setting $k=6$ and using the values of a_5, S_{ns}, and φ_{ns} given in Refs. 28 and 34, on comparing the computational and experimental data, we found $A_\pi^2=+6.5\pm0.5\%$ and $A_\sigma^2=-0.3\pm0.1\%$.

The published absolute values of A_π^2 and A_σ^2 for the orthochromites, as well as their ratio, differ greatly among the different papers. For example, $A_\pi^2=2.7\pm0.6\%$ and $A_\sigma^2=-2.2\pm0.6\%$ $(A_\pi^2/A_\sigma^2\approx-1)$ in Ref. 28, while $A_\pi^2=4\%$ and $A_\sigma^2=-0.7\%$ $(A^2\pi/A_\sigma^2\approx-6)$ according to Ref. 34. For the orthochromites with an iron impurity it was found that $A_\pi^2=9.6\%$ and $A_\sigma^2=-3.8\%$ $(A_\pi^2/A_\sigma^2=-2.5)$.[33] Therefore, our value of A_σ^2 is somewhat high and that of A_π^2 is low compared to the published data. This may be because in Eq. (2) we have neglected the contribution of the *s*-electrons, as well as that of the exchange $s-d$-interaction.

We have also tried to determine the ratio A_π^2/A_σ^2 for the orthochromites using the $T_N=f(\cos^2\theta)$ curve (see Fig. 6). By analogy with Ref. 27, for the orthochromites we can write $T_N\sim A_\pi^2\{2A_\sigma^2+(A_\pi^2-2A_\sigma^2)\cos^2\theta\}$. Rewriting this expression in the form $T_N=b_0+b_1x$, where $x=\cos^2\theta$, we obtain $A_\pi^2/A_\sigma^2=2(b_0+b_1)/b_0$. An analysis of the $T_N=f(\cos^2\theta)$ curve constructed from the experimental data of Ref. 35 (see Fig. 6, curve *3*) yields $A_\pi^2/A_\sigma^2=-2.4\pm0.5$. This value agrees better with the published data[29,34] than the value obtained from H_{hf}^{Sn}. This evidently means that the simplified theory of Ref. 27 does not describe correctly the mechanism by which the field H_{hf} is induced, a point that shows up especially for diamagnetic ions, at whose nuclei the magnetic field is entirely determined by covalence effects and is not ''screened'' by a strong field from the moments of its own electron shell (as happens, for example, in the iron ion).

7. PARTIAL CONTRIBUTIONS TO THE FIELD H_{hf}^{Sn} FROM THE T_{2g}- AND e_g–ORBITALS

Based on the experimental data and our earlier studies,[36] we have calculated the field H_{hf}^{Sn} induced by a single nearest paramagnetic cation (H_{red}^{Sn}) and constructed the dependence of this field on the exchange angle θ. In a crude approximation for the orthoferrites and orthochromites, this curve is linear for angles $140^\circ<\theta<160^\circ$. Thus, we can try to estimate the angular variation in the partial contributions to the field H_{red}^{Sn} from the t_{2g}- and e_g-orbitals of the paramagnetic cation in the following way.

For the orthoferrites $\mathrm{RFeO_3(Sn)}$:

$$\Delta\theta=\theta(\mathrm{R=La})-\theta(\mathrm{R=Lu})\approx+15^\circ,$$

$$\Delta H_{red}^{\mathrm{Sn}}=H_{red}^{\mathrm{Sn}}(\mathrm{La})-H_{red}^{\mathrm{Sn}}(\mathrm{Lu})=+16\,\mathrm{kOe}.$$

The Fe^{3+} ion has a $t_{2g}^3e_g^2$ electron configuration. A change in θ by one degree, produces an increment in the field H_{red}^{Sn} created by the combined t_{2g}- and e_g-orbitals of the paramagnetic cation with the participation of the oxygen p_σ- and p_π-orbitals of

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=+1.07\,\mathrm{kOe\cdot deg^{-1}}. \tag{6}$$

For the orthochromites $\mathrm{RCrO_3}$, we have

$$\Delta\theta=\theta(\mathrm{La})-\theta(\mathrm{Lu})\approx+13^\circ,$$

$$\Delta H_{red}^{\mathrm{Sn}}=-8.33\,\mathrm{kOe}.$$

The Cr^{3+} ion has a $t_{2g}^3e_g^0$ electron configuration. Thus, the angular variation in the field created by a single paramagnetic cation owing to the t_{2g}- and p_π-orbitals is given by

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=-0.64\ \ \mathrm{kOe\cdot deg^{-1}}. \tag{7}$$

Then, assuming that the Fe–O and Cr–O) interionic distances are close, from Eqs. (6) and (7) we obtain the angular variation in the field owing to the e_g- and p_σ-orbitals,

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=+1.71\ \ \mathrm{kOe\cdot deg^{-1}}. \tag{8}$$

As noted above, a supertransferred magnetic field H_{hf} is present at the nuclei of the iron ions (H_{hf}^{Fe}), as well as at the nuclei of the diamagnetic atoms, and in the series of rare earth orthoferrites and orthochromites its magnitude varies in manner analogous to the field H_{hf}^{Sn}. Thus, in $\mathrm{RFeO_3}$ the absolute magnitude of the field H_{hf}^{Fe} at the iron nuclei decreases from R=La to R=Lu by roughly 15 kOe (from 565.9 kOe in $\mathrm{LaFeO_3}$ to 550.5 kOe in $\mathrm{LuFeO_3}$[5]), while in iron-doped $\mathrm{RCrO_3}$, it increases by 9 kOe (from 510 to 519 kOe[33]). This behavior of the field at the nucleus of the paramagnetic Fe^{3+} ion is also explained by a change in the contribution to the field from the indirect hyperfine interaction, H_{hf}^{Fe}, as the Fe–O–Fe exchange bond angle is reduced.[27,33] The angular dependence of the field H_{hf}^{Fe} is adequately described by Eqs. (1). An estimate of the angular variation of the partial contributions from the t_{2g}- and e_g-orbitals to the field H_{hf}^{Fe} in the same range of angles as for H_{hf}^{Sn} yields

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=+0.33\,\mathrm{kOe\cdot deg^{-1}}\ \ \text{for the}\ \ e_g\text{-orbitals},$$

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=-0.115\,\mathrm{kOe\cdot deg^{-1}}\ \ \text{for the}\ \ t_{2g}\text{-orbitals},$$

$$(\Delta H_{red}^{\mathrm{Sn}}/\Delta\theta)=+0.21\,\mathrm{kOe\cdot deg^{-1}}\ \ \text{for the}\ \ (t_{2g}+e_g)\text{-orbitals}.$$

These estimates may be useful, for example, in planning experiments on exchange interactions at high pressures.

8. CONCLUSION

A crystal chemical analysis of the structural series with the general formula $\mathrm{ABO_3}$ has shown that the ideal cubic perovskite structure can undergo a transformation into a rhombohedral hematite-corundum structure by simple rotation of the $[\mathrm{BO_6}]$ octahedra if the B–O distances are fixed. The rhombic distortions are associated with a reduction in the B–O–B bond angle from $\theta=180^\circ$ in perovskite to $\sim132^\circ$ in hematite. The rare earth orthoferrites $\mathrm{RFeO_3}$ are subject to the same structural transformation mechanism and the series $\mathrm{LaFeO_3-LuFeO_3}$ occupies an intermediate position $(157>\theta$

>142°) between the extreme members of the series. A reduction in the bond angle θ leads to a weakening of the Fe–O–Fe exchange interaction, which shows up as a drop in the Néel temperature and in the magnitude of the hyperfine magnetic field at the nucleus.

Paramagnetic ions have an intrinsic magnetic moment which usually induces a strong field at their nuclei. Thus, subtle effects involving spin density transfer to a given ion from its neighbors are ''screened'' by the field from the intrinsic electron cloud. This difficulty is absent in the case of diamagnetic ions. Probe nuclei of diamagnetic tin ions have a greater sensitivity to the geometry of the exchange bonds and serve as a useful instrument for studying spin density delocalization and transfer. In addition, it turns out that in the oxide systems the absolute magnitudes of the supertransferred fields H_{hf} for tin ions are several times those for iron ions. This large difference in the fields H_{hf} at the nuclei of iron and tin appears to be related to the greater covalency of the Sn–O bonds compared to the Fe–O bonds.

An analysis of the theoretical models has shown that, with a suitable choice of the exchange and transfer parameters, Moskvin's model[26] provides a fairly good description of the variation in the exchange interaction parameters over a wide range of angles θ.

This work was supported by the Russian Fund for Fundamental Research (Project 98-02-17197a). The authors thank the American Physical Society (the Sloan Foundation Grant of the American Physical Society) for partial support of this work.

*)E-mail: root@magnet.crystal.msk.su

---

[1] S. Geller, J. Chem. Phys. **24**, 1236 (1956).
[2] W. C. Koehler and E. O. Wollan, J. Phys. Chem. Solids **2**, 100 (1957).
[3] V. E. Naĭsh and E. A. Turov, Fiz. Met. Metalloved. **11**, 161, 321 (1961).
[4] V. E. Naĭsh and E. A. Turov, Fiz. Met. Metalloved. **9**, 10 (1960).
[5] M. Eibschutz, S. Shtrikman, and D. Treves, Phys. Rev. **156**, 562 (1967).
[6] S. Geller, J. P. Remeika, R. C. Sherwood, H. G. Williams, and G. P. Espinoza, Phys. Rev. A **137**, 1034 (1965).
[7] S. Geller, H. J. Williams, and R. C. Sherwood, Phys. Rev. **123**, 1692 (1961).
[8] E. F. Bertaut and R. Pauthenet, Proc. IEE, Suppl. B **104**, 261 (1957).
[9] I. S. Lyubutin and Yu. S. Vishnhakov, Zh. Éksp. Teor. Fiz. **61**, 1962 (1971) [Sov. Phys. JETP **34**, 1045 (1971)].
[10] M. Marezio, J. P. Remeika, and P. D. Dernier, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **26**, 2008 (1970).
[11] P. Coppens and V. Eibschutz, Acta Crystallogr. **19**, 524 (1965).
[12] S. Geller and E. A. Wood, Acta Crystallogr. **9**, 563 (1956).
[13] S. Geller and V. B. Bala, Acta Crystallogr. **9**, 1019 (1956).
[14] S. Geller, Acta Crystallogr. **10**, 243 (1957).
[15] S. Geller, Acta Crystallogr. **10**, 248 (1957).
[16] M. A. Gilleo, Acta Crystallogr. **10**, 161 (1957).
[17] Y. Ishikawa and S. Akimoto, J. Phys. Soc. Jpn. **12**, 1083 (1957).
[18] M. Marezio, Trans. Am. Crystallogr. Assoc. No. 5 (1969), p. 29.
[19] K. P. Belov and I. S. Lyubutin, Zh. Éksp. Teor. Fiz. **49**, 747 (1965) [Sov. Phys. JETP **22**, 518 (1965)].
[20] S. L. Ruby, B. E. Evans, and S. S. Hafner, Solid State Commun. **6**, 277 (1968).
[21] B. J. Evans and L. J. Swartzendruber, Phys. Rev. B **6**, 233 (1972).
[22] P. B. Fabrichnyĭ, ZhVKhO **30**, 143 (1985).
[23] N. L. Huang, R. Orbach, E. Simanec, J. Owen, and D. R. Taylor, Phys. Rev. **156**, 383 (1967).
[24] G. A. Savatzky and F. Van der Woude, J. Phys. **35**, N C6, 47 (1974).
[25] F. Van der Woude and G. A. Sawatsky, Phys. Rev. B **4**, 3159 (1971).
[26] A. S. Moskvin, N. S. Ovanesyan, and V. A. Trukhtanov, Hyperfine Interact. **3**, 429; **5**, 13 (1977).
[27] C. Boekema, F. Van der Woude, and G. A. Sawatzky, Int. J. Magn. **3**, 341 (1972).
[28] P. Freund, J. Owen, and B. F. Hahn, J. Phys. C: Solid State Physics **6**, L139 (1973).
[29] D. C. Tofield and B. E. F. Fender, J. Phys. Chem. Solids **31**, 2741 (1970).
[30] P. B. Fabrichnyĭ, Izv. Akad. Nauk SSSR Ser. Fiz. **50**, 2310 (1986).
[31] P. B. Fabrichnyĭ, E. V. Lamykin, A. M. Babeshkin, and A. N. Nesmeyanov, Fiz. Tverd. Tela **13**, 3417 (1971) [Sov. Phys. Solid State **13**, 2874 (1971)].
[32] Y. Miyahara and S. Iida, J. Phys. Soc. Jpn. **37**, 1248 (1974).
[33] A. S. Moskvin, N. S. Ovanesyan, and V. A. Trukhtanov, Hyperfine Interact. **1**, 265 (1975).
[34] J. K. Lees and P. A. Flinn, J. Chem. Phys. **48**, 882 (1968).
[35] K. Motida and S. Miyahara, J. Phys. Soc. Jpn. **28**, 1188 (1970).
[36] I. S. Lyubutin and Yu. S. Visnyakov, Kristallografiya **17**, 960 (1972) [Sov. Phys. Crystall. **17**, 847 (1972)].

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Normal-metal hot-electron bolometer with Andreev reflection from superconductor boundaries

A. N. Vystavkin, D. V. Shuvaev, L. S. Kuz'min, and M. A. Tarasov*)

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

E. Aderstedt, M. Willander, and T. Claeson

Department of Microelectronics and Nanoscience, Chalmers University of Technology, Gothenburg, Sweden

(Submitted 24 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1085–1092 (March 1999)

This paper describes the design and experimental testing of a high-sensitivity hot-electron bolometer based a film of normal metal, exploiting the Andreev reflection from superconductor boundaries, and cooled with the help of a superconductor–insulator–normal metal junction. At the measured thermal conductivity, $G \approx 6 \times 10^{-12}$ W/K, and a time constant of $\tau = 0.2\,\mu$s, and a temperature of 300 mK, the estimated noise-equivalent power NEP$= 5 \times 10^{-18}$ W/Hz$^{1/2}$, assuming that temperature fluctuations are the major source of noise. At a temperature of 100 mK, the thermal conductivity drops to $G \approx 7 \times 10^{-14}$ W/K, which yields NEP$= 2 \times 10^{-19}$ W/Hz$^{1/2}$ at a time constant of $\tau = 5\,\mu$s. The microbolometer has been designed to serve as a detector of millimeter and FIR waves in space-based radio telescopes. © *1999 American Institute of Physics.* [S1063-7761(99)02303-3]

1. INTRODUCTION

A normal-metal hot-electron bolometer (NHEB) using the Andreev reflection from superconductor boundaries was first described by Nahum *et al.*[1,2] It demonstrated a very high voltage responsivity at operating temperatures of about 100 mK. The bolometer was based on a normal-metal film connected to two superconducting electrodes. A current induced by a microwave signal was transmitted via the electrodes to the film and increased the electron gas temperature in the normal metal. An important point is that electrons cannot impart their thermal energy to the electrodes owing to the Andreev reflection at the superconductor–normal metal interface. Electrons can give up their energy to the lattice, but, at temperatures below 1 K, electron–phonon coupling is inefficient, so the heat transfer from heated electrons to the lattice is very low. This weak coupling is characterized by thermal conductivity G and results in a considerable increase in energy engendered by low incident power, i.e., the electron heating effect is fairly large. Changes in the electron temperature are detected using an additional superconductor–insulator–normal metal (SIN) junction, where the normal metal is the heated normal-metal film (Fig. 1). The shape of the current–voltage characteristic of a SIN junction depends on the electron temperature in the normal-metal film. A dc bias current is fed to the junction, and changes in the voltage are measured, i.e., an output signal $\Delta V(T)$, which is linear over a wide range of incident power. One important feature of such a bolometer is the equality between the thermal time constant τ and the time $\tau_{\text{e-ph}}$ of electron–phonon relaxation. The typical values $\tau = 10\,\mu$s at 100 mK and $\tau = 0.4\,\mu$s at 300 mK are much shorter than the time constant required for most practical applications.[1,3]

Since the main application domain of high-sensitivity FIR detectors is space-based radio astronomy, NHEB detectors operated at a temperature of 300 mK, which can be obtained in relatively simple and small ^{3}He cryostats, show much promise. Conditions imposed by the European Space Agency on bolometric detectors for future IR and FIR space-based astronomical facilities were listed in the tender announcement (Ref. 3). The required detector should operate at a temperature of 300 mK and have a noise-equivalent power NEP$< 1 \times 10^{-17}$ W/Hz$^{1/2}$ at a time constant $\tau < 1$ ms.

There are three sources of noise contributing to the bolometer NEP:

$$\text{NEP} = \left[4k_B T_e^2 G + \frac{V_j^2}{S^2} + \frac{V_n^2}{S^2} \right]^{1/2}, \tag{1}$$

where T_e is the electron temperature in the absorber, $G = dP/dT$ is the heat sink thermal conductivity under operating conditions, $S = dV/dP = (dV/dT)\,G^{-1}$ is the detector response, V_j characteriz es voltage fluctuations in the SIN junction, and V_n is the measure of voltage noise in the amplifier. The first term on the right describes temperature fluctuations in the absorbing material and determines the fundamental noise minimum in the device at a given temperature. This equation clearly shows that G leads to a low value of NEP in such a bolometer.

We have suggested a technique for decreasing NEP using a well-known method[4] of decreasing the electron temperature T_e in the bolometer absorbing material with essentially no change in the actual lattice temperature of 300 mK.[5] In this device, the balance between electron heating by phonons and cooling due to the tunnel junction is controlled by tuning the bias voltage to the energy gap, which allows one to get rid of electrons with higher energies (Fig. 2).

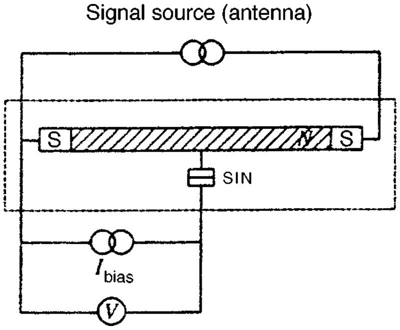

FIG. 1. Microbolometer structure. SIN is the junction biased by a low dc voltage. The voltage across the junction depends on how the current–voltage characteristic is flattened by thermal noise, which is a measure of the temperature in the normal-metal absorber (shaded region).

According to preliminary estimates,[5] the full thermal conductivity G cannot be reduced by electron cooling, because another channel of heat supply is added. The electron cooling can upgrade the NHEB parameters by decreasing T_e in the first term on the right of Eq. (1) and increasing the temperature response dV/dT, which yields higher S in the second and third terms of NEP.[5]

2. POWER DETECTION

We have developed and manufactured several versions of NHEB, including those with additional SIN junctions for cooling. First we built a microbolometer with a single tunnel junction for measuring the electron temperature in the absorbing material.[6] The absorber and superconducting electrodes were fabricated using direct electron-beam lithography and the shadow evaporation technique at different angles. Superconducting aluminum electrodes 40 nm thick were fabricated first and oxidized in an atmosphere of oxygen at a pressure of 4×10^{-2} mbar for two minutes to form a tunnel barrier. Then we deposited 3 nm of chromium and 35 nm of silver to fabricate an absorbing film 6 μm long and 0.25 μm wide. Then two superconducting lead electrodes with a thickness of 120 nm were fabricated using additional lithography and subsequent etching. The *in situ* ion etching prior to film deposition was used to remove oxides and contaminants on the interfaces between normal-metal and superconducting films.

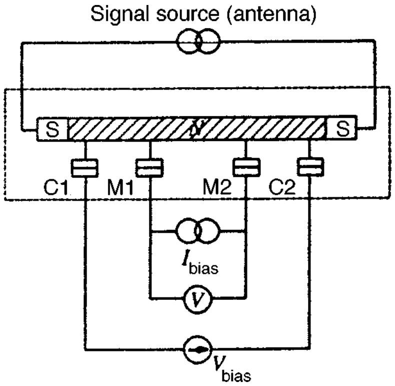

FIG. 2. Configuration of a microbolometer with electron cooling. The two low-resistance SIN junctions (C1 and C2) are biased by a dc voltage and reduce the effective electron temperature in the absorber normal-metal film (shaded area). The two high-resistance SIN junctions (M1 and M2) are biased by a low dc current and used in measurements of the electron temperature.

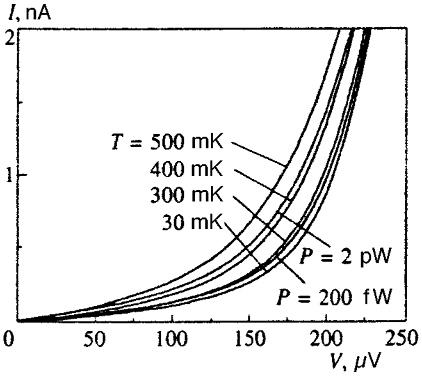

FIG. 3. Current–voltage characteristics of the SIN junction measured at various temperatures and zero signal current, and at two signal currents and an ambient temperature of 30 mK.

The current–voltage characteristics of the detector SIN junction were measured at various temperatures and incident powers (Fig. 3). The resistance of the shorter device ($l=6\,\mu$m) was 12 Ω. The voltage V as a function of signal current I_{abs} in the absorber at a constant bias current was measured at a constant temperature in two devices of lengths differing by a factor of two. The curves of $V(I)$ for the two devices are very similar (Fig. 4). The derivative dV/dI_{abs} is directly associated with the curve shape and can be expressed in terms of the temperature response, reciprocal thermal conductivity, and derivative dP/dI_{abs}:

$$\frac{dV}{dI_{abs}}=\frac{dV}{dP}\,\frac{dP}{dI_{abs}}=\frac{dV}{dT}\left(\frac{dP}{dT}\right)^{-1}\frac{dP}{dI_{abs}}.$$

One can derive from Joule's formula

$$P=P_{Joule}=I_{abs}^2R\rightarrow\frac{dP}{dI_{abs}}=2RI_{abs}.$$

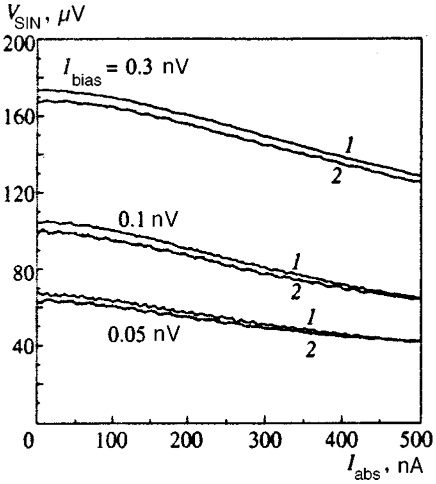

FIG. 4. Voltage across a SIN junction at a constant bias current I_{bias} as a function of the current across the absorber for two devices with lengths of 6 μm (curves *1*) and 12 μm (curves *2*) at an ambient temperature of 30 mK.

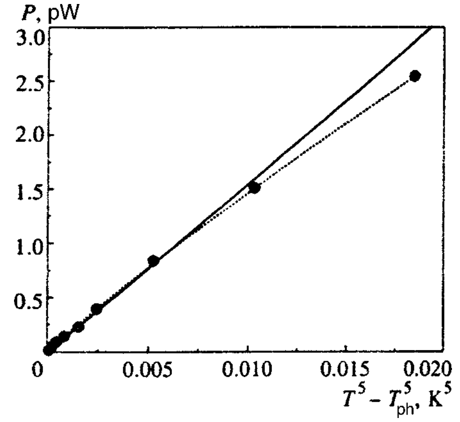

FIG. 5. Absorbed power as a function of $T^5 - T_{ph}^5$, where T is the electron temperature derived from measurements of $V(I_{bias}, T)$ and $T_{ph} = 30$ mK is the substrate temperature, assumed to equal that of the substrate holder. The straight line described by the equation $P = 1.55\times10^{-10}(T^5 - T_{ph}^5)$ corresponds to Eq. (2).

The reciprocal thermal conductivity can be derived from the expression for the heat exchange in the case of electron heating:

$$P_{e\to ph} = \Sigma U(T_e^5 - T_{ph}^5) \to \frac{dP}{dT} = 5\Sigma U T^4, \tag{2}$$

where Σ is the characteristic parameter of the material, and, supposing that the equilibrium condition holds, we have $P_{Joule} = P_{e-ph}$. After substituting the latter expression, we have

$$\frac{dV}{dI_{abs}} = \frac{dV}{dT}\left(\frac{2I_{abs}R}{5\Sigma T^4 U}\right) \propto \frac{R}{U}. \tag{3}$$

The coincidence of the IV curves of the different devices indicates that the increase in the dissipated power P owing to the higher resistance R is almost completely offset by the increase in the thermal conductivity due to the larger volume U, i.e., there is no significant thermal flux through the superconductor–normal metal interface. This provides direct evidence in favor of the Andreev reflection. Then the function $V(I_{abs})$ was used in calculating $V(P)$. The highest slope of the power–voltage characteristic at the optimal bias current $I_{bias} = 0.3$ nA is $S_{max} = |dV/dP| = 3\times10^7$ V/W. By combining the characteristics $V(I_{bias}, T)$ and $V(I_{bias}, P)$, one can calculate $P(T_e)$ (Fig. 5).

By comparing experimental data with calculations by Eq. (2), one can determine the material parameter $\Sigma \approx 3\times10^{-9}$ nW·K^{-5}·μm^{-3}, whence follows the thermal conductivity $G \approx 6\times10^{12}$ W/K at 300 mK. This parameter is half that derived from the data reported by Nahum and Martinis.[2] This difference is due to the smaller absorber volume in our case. The temperature fluctuation component in NEP [Eq. (1)] derived from our measurements of G is about 5×10^{-18} W/Hz$^{1/2}$, which is considerably lower than the parameter of future bolometers for space-based telescopes required by the European Space Agency.[3] At 100 mK the thermal conductivity decreases appreciably, to $G \approx 7\times10^{-14}$ W/K. This value of G contributes a temperature fluctuation component to NEP of 2×10^{-19} W/Hz$^{1/2}$.

The thermal time constant can be calculated by the formula $\tau = C/G$, where C is the electron specific heat. At the value of Σ measured in our experiments, we obtain $\tau \approx 5T^{-3}$ ns. At $T = 300$ mK the time constant $\tau \approx 0.2$ μs, and at $T = 100$ mK it is $\tau \approx 5$ μs, which is much better than usual requirements.

At ambient temperatures of 100 mK and below, the NHEB sensitivity did not improve significantly. The reason is the quality of the detector tunnel junction, whose current–voltage characteristic shape changes with temperature above 300 mK. This property might be due to the thermal processing during the second lithographic process.

3. ELECTRON COOLING

Using the same technology, we have fabricated an NHEB with four tunnel junctions (see Fig. 2, and Figs. 6 and 7). In this case, the absorber is a copper film 40 nm thick, 0.25 μm wide, and 7 μm long. In order to cool electrons efficiently, the resistance of each cooling junction should be

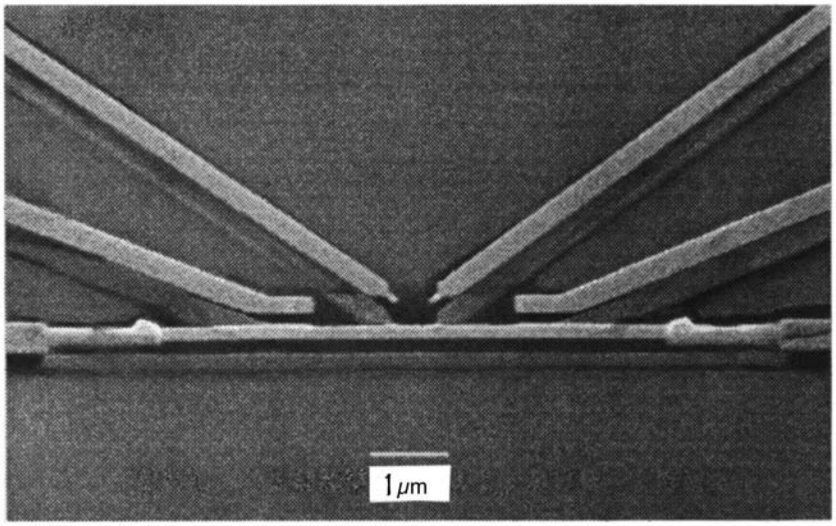

FIG. 6. Electron micrograph of a microbolometer with an electron microrefrigerator fabricated by the shadow evaporation technique at different angles. The upper horizontal strip is made of copper and acts as an absorber. The darker layer is aluminum, the two larger and two smaller junctions are fabricated where the absorber strip overlaps oxidized aluminum electrodes fabricated in a separate lithographic process.

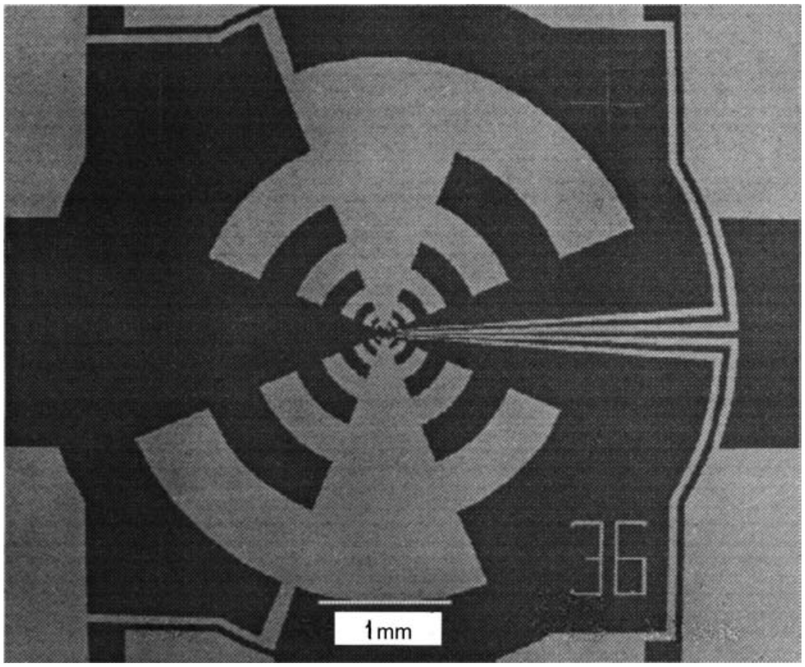

FIG. 7. View of a device with a planar log-periodic antenna designed for a frequency band of 100 to 1000 GHz.

of order 1 kΩ. At the same time, the resistances of detector junctions should be much higher than 10 kΩ in order to minimize the bias current. These two conditions were satisfied by fabricating junctions of very different areas (0.2 μm^2 for the large and 0.01 μm^2 for the small junctions). As a result, the ratio between the normal resistances of the junctions is more than 20, although the latter parameter can be affected by edge oxidation, which has more impact on small junctions than on large.

In order to demonstrate the effect of electron cooling, we have measured voltage variation across the detector junction at a constant bias current as a function of the voltage across the large junctions. Then the resulting values were calibrated for different temperatures in the cryostat, where the large junctions were not connected to a dc bias source.

The resulting curves measured at various initial temperatures $T_{\rm ph}=T_e$ at $V=0$ are plotted in Fig. 8. Both the cooling effect and its unexpected suppression at temperatures below 400 mK and 250 mK, respectively, have been detected. One reason for this behavior might be the trivial heating by the current conducted between the tunnel junctions. In order to check out this assumption, we fabricated large junctions separated by only 0.5 μm, but this modification had little effect on device characteristics.

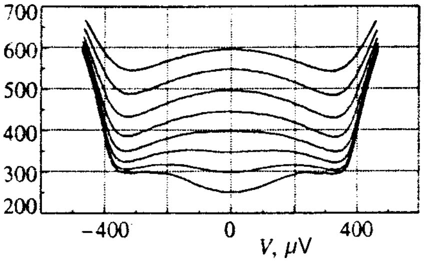

FIG. 8. Electron temperature $T_{\rm e}$ in the absorber as a function of the voltage applied to the two large SIN junctions ($R_n^{\rm C1}+R_n^{\rm C2}=625\,\Omega+645\,\Omega$) at various initial temperatures. Parameter T_e is derived from measurements of the voltage variations across the two small junctions ($R_n^{\rm M1}+R_n^{\rm M2}=62\,{\rm k}\Omega+116\,{\rm k}\Omega$) biased by a small dc current using a calibration curve obtained in a dedicated experiment.

4. CONCLUSIONS

We have developed, manufactured, and tested bolometers using hot electrons in small normal-metal films, exploiting the Andreev reflection from superconductor boundaries. Experiments with absorbers of different lengths confirmed that the thermal conductivity is largely determined by electron–phonon coupling, and there is no thermal conductivity through the interface between superconductors and normal metals. The measured thermal conductivity in the microbolometer with the short absorber is 6×10^{-12} W/K and 7×10^{-14} W/K at 300 mK and 100 mK, respectively. These values correspond to time constants $\tau=0.2\,\mu$s and $\tau=5\,\mu$s and noise-equivalent powers ${\rm NEP}=5\times10^{-18}$ W/Hz$^{1/2}$ and ${\rm NEP}=2\times10^{-19}$ W/Hz$^{1/2}$, if temperature fluctuations are considered to be the major source of noise. These characteristics, especially the time constants, are considerably better than those of conventional superconducting bolometers.[7] Our experiments support the concept of electron cooling in such bolometers using additional SIN junctions, which enables one to reduce the noise-equivalent power and use a less complicated cooling system.

This work was supported by the Ministry of Science and Technology of the Russian Federation, the Russian Fund for Fundamental Research, INTAS, and the Swedish Academy of Sciences. We express our gratitude to members of the

Russian Academy of Sciences A. F. Andreev and N. S. Kardashev for stimulating this research.

*)E-mail: tarasov@hitech.cplire.ru

---

[1] M. Nahum, P. L. Richards, and C. A. Mears, IEEE Trans. Appl. Supercond. **3**, 2124 (1993).
[2] M. Nahum and J. M. Martinis, Appl. Phys. Lett. **63**, 3075 (1993).
[3] ESA tender A03288: Hot-electron Microbolometers, opened Oct. 2, 1997.
[4] M. M. Leivo, J. P. Pekola, and D. V. Averin, Appl. Phys. Lett. **68**, 1996 (1996).
[5] L. Kuzmin, I. Devyatov, D. Golubev, in *Proceedings of the 4th Int. Conf. on Millimeter and Submillimeter Waves and Applications*, SPIE Int. Society Publ. (1998), p. 63.
[6] D. Shuvaev (Chouvaev), L. Kuzmin, M. Tarasov *et al.*, in *Proceedings of the 9th Int. Symp. on Space THz Technology*, JPL, California Institute of Technology, NASA (1998), p. 331.
[7] A. T. Lee, S-F. Lee, J. M. Gildemeister, and P. L. Richards, in *Proceedings of the 7th Int. Workshop on Low Temperature Detectors*, Munich (1997), p. 123.

Translation provided by the Russian Editorial office.

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Conductivity in a two-dimensional disordered model with anisotropic long-range hopping

E. A. Dorofeev and S. I. Matveenko

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia

(Submitted 6 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1093–1106 (March 1999)

We consider a two-dimensional system of particles localized on randomly distributed sites of a square lattice with anisotropic transition matrix elements between localized sites. The diagram and replica methods are used. The conductivity of a system in different limits of local sites and particles densities is calculated. The model is relevant to the problem of strong nonmagnetic impurities in superconductors with $d_{x^2-y^2}$ symmetry of the order parameter. © *1999 American Institute of Physics.* [S1063-7761(99)02403-8]

1. INTRODUCTION

We examine a system of randomly distributed impurities at a sites of a two-dimensional square lattice. An impurity potential generates a localized state with a strongly anisotropic (cross-shaped) wave function. The conductivity is produced by due to hopping of particles between local states on the same vertical or horizontal lines. This picture can be realized in two-dimensional $d_{x^2-y^2}$-wave superconductors, where local bound quasiparticle states can arise in the presence of unitary impurities.[1] The wave functions of the local states are strongly anisotropic, with exponential decay in all directions except $\varphi_n=(2n+1)\pi/4$, where the wave function is proportional to r^{-1}.

A similar anisotropy has a wave function of bound states in the vortex core in d-wave superconductors.[2] The wave function in the vicinity of gap nodes at large distances has the form

$$|\Psi|^2 \propto |\varphi-\varphi_n| \exp(-2|\varphi-\varphi_n| r/\xi),$$

with a maximum value $|\Psi|^2 \propto \xi/2r$ in the directions $\varphi-\varphi_n \simeq \xi/2r \to 0$.

We will consider the following tight-binding Hamiltonian:

$$H=\sum_{i,j} t(\mathbf{r}_j-\mathbf{r}_i)\psi^+(\mathbf{r}_i)\psi(\mathbf{r}_j)\rho(\mathbf{r}_i)\rho(\mathbf{r}_j), \qquad (1)$$

where $\psi^+(\mathbf{r}_i), \psi^+(\mathbf{r}_j)$ are creation and annihilation operators, $\rho(\mathbf{r}_i)$ is the density of impurities, equal to 1 at the impurity sites, and to 0 otherwise. The transition matrix element has a cross-shaped configuration

$$t(\mathbf{r})=(\delta_{x,0}+\delta_{y,0})f(r), \qquad (2)$$

with

$$f(r)=J\left(\frac{a}{r}\right)^{\gamma}\exp(-\kappa r),$$

and a is the lattice constant.

The plan of this article is as follows. In Sec. 2 we consider the case of low impurity density. In Sec. 3 we calculate the conductivity in the case of high impurity density. Results are discussed in the Conclusion.

2. LOW DENSITY

We consider now the limit of low impurity concentration $(c \ll 1)$. In an external electromagnetic field we substitute in (1)

$$t(\mathbf{r}_i-\mathbf{r}_j) \rightarrow t(\mathbf{r}_i-\mathbf{r}_j)\exp\left[ie\int_{\mathbf{r}_i}^{\mathbf{r}_j}\mathbf{A}(\mathbf{r},t)d\mathbf{r}\right].$$

The electric current is defined as usual with the Hamiltonian (1) by varying over a gauge-invariant vector potential $\mathbf{A}$

$$j_\alpha(t)=-ie\sum_{i,j}(\mathbf{r}_i-\mathbf{r}_j)_\alpha t_{ij}\Psi_i^+(t)\Psi_j(t)\rho_i\rho_j \times\exp(ie\mathbf{A}(t)(\mathbf{r}_i-\mathbf{r}_j)), \qquad (3)$$

where $t_{ij}=t(\mathbf{r}_i-\mathbf{r}_j)$. Since we will calculate $j(\omega)$, we assume that the potential $\mathbf{A}$ depends only on time t. Using the equation for the Green function $G(t_1,\mathbf{r}_1,t_2,\mathbf{r}_2)$

$$\frac{\partial}{\partial t_1}G(t_1,\mathbf{r}_1,t_2,\mathbf{r}_2)=-i\left\langle T\frac{\partial\Psi(1)}{\partial t_1}\Psi^+(2)\right\rangle -i\delta(\mathbf{r}_1-\mathbf{r}_2)\delta(t_1-t_2), \qquad (4)$$

we obtain after Fourier transformation in the linear $\mathbf{A}$ approximation

$$\begin{aligned} j_\alpha(\omega)=&\frac{e^2}{c}\sum_{i,j} t_{i,j}(\mathbf{r}_i-\mathbf{r}_j)_\alpha(\mathbf{r}_i-\mathbf{r}_j)_\beta A_\beta(\omega)\\ &\times\int\frac{d\Omega}{2\pi}G(\Omega,\mathbf{r}_j,\mathbf{r}_i)e^{i\Omega\alpha}-\frac{e^2}{c}\\ &\times\sum_{i,j,k,l} t_{i,j}t_{k,l}(\mathbf{r}_i-\mathbf{r}_j)_\alpha(\mathbf{r}_k-\mathbf{r}_l)_\beta A_\beta(\omega)\\ &\times\int\frac{d\Omega}{2\pi}G(\omega+\Omega,\mathbf{r}_j,\mathbf{r}_k)G(\Omega,\mathbf{r}_l,\mathbf{r}_i)e^{i\Omega\alpha}. \end{aligned} \qquad (5)$$

The summation in Eq. (5) is taken over impurity sites, and $\alpha \rightarrow +0$. In order to evaluate (5) in the lowest order with respect to the concentration, we examine the case of two randomly located sites. The Green function is found easily:

$$G(\omega,\mathbf{r}_i,\mathbf{r}_i)=\frac{\omega+\mu}{(\omega+\mu)^2-t_{i,j}^2},$$

$$G(\omega,\mathbf{r}_i,\mathbf{r}_j)=\frac{t_{1,2}}{(\omega+\mu)^2-t_{i,j}^2}, \tag{6}$$

where μ is the chemical potential. Substituting (6) into (5), we obtain in the case when the sites are located on the same horizontal chain

$$j_x(\omega)=\frac{e^2}{2c}(x_1-x_2)^2 t_{1,2}\frac{\omega^2}{(\omega+i\alpha)^2-4t_{1,2}^2}A_x(\omega)$$
$$=Q(\omega)A_x(\omega). \tag{7}$$

A similar equation can be derived for the case of nonzero temperature. The result differs only by the Fermi filling factor. After averaging over impurity sites we obtain for the conductivity $\sigma(\omega)=iQ(\omega)/\omega$

$$\sigma(\omega)=\frac{\pi e^2}{4}c^2L\int x^2 t(x)[n_F(\omega-\mu)$$
$$-n_F(-\omega-\mu)]\delta(\omega-2t(x))dx$$
$$=\frac{\pi e^2}{8}c^2L\frac{x_0^2 t(x_0)[n_F(\omega-\mu)-n_F(-\omega-\mu)]}{|t'(x_0)|}, \tag{8}$$

where n_F is the Fermi distribution function, and $2t(x_0)=\omega$.

Substituting $t(x)$ from (2) we get

$$\frac{\omega}{2J}=\frac{\exp(-\kappa x_0)}{x_0^\gamma}, \tag{9}$$

$$\sigma(\omega)=\frac{\pi e^2}{8}c^2L\frac{x_0^3[n_F(\omega-\mu)-n_F(-\omega-\mu)]}{\gamma+\kappa x_0}. \tag{10}$$

In the limit of low frequency we have the following asymptotic behavior:

1. $\kappa=0$, $x_0=(2J/\omega)^{1/\gamma}$:

$$\omega \gg T, \quad \sigma(\omega)\propto\omega^{-3/\gamma}, \tag{11}$$

$$\omega \ll T, \quad \sigma(\omega)\propto\omega^{-3/\gamma+1}; \tag{12}$$

2. $\kappa\neq 0$, $\kappa x_0 \sim \log(2J/\omega)$:

$$\omega \gg T, \quad \sigma(\omega)\propto\log^2(2J/\omega), \tag{13}$$

$$\omega \ll T, \quad \sigma(\omega)\propto\omega\log^2(2J/\omega). \tag{14}$$

3. HIGH DENSITY

3.1. Green function

In case of high density of impurities we assume that the distribution function of impurities is a Gaussian with variance g:

$$\rho(\mathbf{r}_i)=c+\delta\rho(\mathbf{r}_i), \quad \langle\delta\rho(\mathbf{r}_i)\delta\rho(\mathbf{r}_j)\rangle_\rho=g^2\delta_{ij}, \tag{15}$$

where $\langle\dots\rangle_\rho$ denotes the average over possible impurity configurations. We assume that the concentration $c\leqslant 1$.

The one-particle Green function for the arbitrary impurity distribution is usually defined in terms of a functional integral as

$$G(E,\mathbf{r},\mathbf{r}')=[E-t_{ij}\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)]^{-1}_{\mathbf{r},\mathbf{r}'}$$
$$=i\frac{\int D\bar{\psi}D\psi\,\psi(\mathbf{r})\bar{\psi}(\mathbf{r}')\exp(iS)}{\int D\bar{\psi}D\psi\exp(iS)}, \tag{16}$$

where

$$S=S_0+S_1, \tag{17}$$

$$iS_0=i\sum_{\mathbf{r}}\bar{\psi}(\mathbf{r})E\psi(\mathbf{r}), \tag{18}$$

$$iS_1=-i\sum_{\mathbf{r}_1\mathbf{r}_2}\bar{\psi}(\mathbf{r}_1)t(\mathbf{r}_1-\mathbf{r}_2)\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\psi(\mathbf{r}_2). \tag{19}$$

Introducing an additional integration over new fermion fields $\chi,\bar{\chi}$ in order to eliminate the second order terms $\rho\rho$, we get

$$\exp(iS_1)=\int D\bar{\chi}D\chi\exp\Bigg\{i\sum\bar{\chi}(\mathbf{r}_1)t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\chi(\mathbf{r}_2)$$
$$-ic\sum_{\mathbf{r}}[\bar{\chi}(\mathbf{r})\psi(\mathbf{r})+\bar{\psi}(\mathbf{r})\chi(\mathbf{r})]$$
$$-i\sum_{\mathbf{r}}\delta\rho(\mathbf{r})[\bar{\chi}(\mathbf{r})\psi(\mathbf{r})+\bar{\psi}(\mathbf{r})\chi(\mathbf{r})]\Bigg\}\frac{1}{Z}, \tag{20}$$

where

$$Z=\int D\bar{\chi}D\chi\exp\left\{i\sum_{\mathbf{r}_1,\mathbf{r}_2}\bar{\chi}(\mathbf{r}_1)t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\chi(\mathbf{r}_2)\right\}, \tag{21}$$

$$t^{-1}(\mathbf{r})=\frac{1}{V}\sum_{\mathbf{k}}\varepsilon^{-1}(\mathbf{k})e^{i\mathbf{kr}}, \tag{22}$$

$$\varepsilon(\mathbf{k})=\sum_{\mathbf{r}}t(\mathbf{r})e^{i\mathbf{kr}}=-J\ln\left[\left(\kappa^2+4\sin^2\frac{k_xa}{2}\right)\right.$$
$$\left.\times\left(\kappa^2+4\sin^2\frac{k_ya}{2}\right)\right]. \tag{23}$$

The Green function in terms of the new two-component field φ,

$$\varphi_1(\mathbf{r})=\psi(\mathbf{r}), \quad \varphi_2(\mathbf{r})=\chi(\mathbf{r}),$$
$$\bar{\varphi}_1(\mathbf{r})=\bar{\psi}(\mathbf{r}), \quad \bar{\varphi}_2(\mathbf{r})=\bar{\chi}(\mathbf{r}), \tag{24}$$

reads

$$\hat{G}(\mathbf{r}_1,\mathbf{r}_2)=-i\langle\hat{\varphi}(\mathbf{r}_1)\otimes\hat{\bar{\varphi}}(\mathbf{r}_2)\rangle, \tag{25}$$

where $\hat{\bar{\varphi}}(\mathbf{r})=[\bar{\varphi}_1(\mathbf{r}),\bar{\varphi}_2(\mathbf{r})]$, and angle brackets are defined as

$$\langle\dots\rangle=\frac{\int D\bar{\psi}D\psi D\bar{\chi}D\chi(\dots)\exp(iS_{\text{eff}})}{\int D\bar{\psi}D\psi D\bar{\chi}D\chi\exp(iS_{\text{eff}})} \tag{26}$$

with

$$iS_{\mathrm{eff}}=i\sum_{\mathbf{r}}\ \hat{\bar{\varphi}}(\mathbf{r})\begin{pmatrix} E & -c-\delta\rho(\mathbf{r}) \\ -c-\delta\rho(\mathbf{r}) & t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\end{pmatrix}\hat{\varphi}(\mathbf{r}). \tag{27}$$

The equation for the Green function after averaging over impurities in the Born approximation reads

$$\hat{G}(\mathbf{k})=\hat{G}^0(\mathbf{k})+\hat{G}^0(\mathbf{k})\hat{\Sigma}(\mathbf{k})\hat{G}(\mathbf{k}), \tag{28}$$

with the bare Green function

$$[\hat{G}^0(\mathbf{k})]^{-1}=\begin{bmatrix} E & -c \\ -c & \varepsilon^{-1}(\mathbf{k})\end{bmatrix} \tag{29}$$

and the self-energy $\hat{\Sigma}$ obtained by summing diagrams without intersections

$$\hat{\Sigma}(\mathbf{k})=g^2a^2\int\frac{d\mathbf{k}_1}{(2\pi)^2}\sigma^x\hat{G}(\mathbf{k}_1)\sigma^x. \tag{30}$$

The solution of Eqs. (28) and (30) is

$$\Sigma(\mathbf{k})=\begin{pmatrix} A & C \\ C & B\end{pmatrix}, \tag{31}$$

$$\hat{G}(\mathbf{k})=\frac{1}{[1+\varepsilon(\mathbf{k})R](E-A)-(c+iC)^2\varepsilon(\mathbf{k})}\times\begin{bmatrix} 1+\varepsilon(\mathbf{k})B & -i(c+iC)\varepsilon(\mathbf{k}) \\ -i(c+iC)\varepsilon(\mathbf{k}) & \varepsilon(\mathbf{k})(E-A)\end{bmatrix}, \tag{32}$$

where

$$A=g^2a^2\int\frac{d\mathbf{k}}{(2\pi)^2}\frac{(E-A)\varepsilon(\mathbf{k})}{[1+\varepsilon(\mathbf{k})B](E-A)-(c+iC)^2\varepsilon(\mathbf{k})},$$

$$B=-g^2a^2\int\frac{d\mathbf{k}}{(2\pi)^2}\frac{1-\varepsilon(\mathbf{k})B}{[1+\varepsilon(\mathbf{k})B](E-A)-(c+iC)^2\varepsilon(\mathbf{k})},$$

$$iC=g^2a^2\int\frac{d\mathbf{k}}{(2\pi)^2}\frac{(c+iC)\varepsilon(\mathbf{k})}{[1+\varepsilon(\mathbf{k})B](E-A)-(c+iC)^2\varepsilon(\mathbf{k})}.$$

In the limit $g^2\ll 1$, we obtain

$$A^{R,A}=\pm i\frac{\gamma}{2},\quad B^{R,A}=\mp i\frac{c^2}{E^2}\frac{\gamma}{2},\quad C^{R,A}=\pm i\frac{c}{E}\frac{\gamma}{2}, \tag{33}$$

$$\gamma=2\pi g^2a^2\frac{E^2}{c^4}\nu\left(\frac{E}{c^2}\right), \tag{34}$$

where $\nu(\varepsilon)$ is the density of states of the pure model ($c=1$, $g=0$):

$$\nu(\varepsilon)=\int\frac{d\mathbf{k}}{(2\pi)^2}\delta[\varepsilon-\varepsilon(\mathbf{k})]. \tag{35}$$

Taking into account that $A\ll E$, $BE\ll 1$, $C\ll c$, we find for the Green function in the limit $g^2\ll 1$

$$\hat{G}^{R,A}(\mathbf{k})=\frac{1}{E-c^2\varepsilon(\mathbf{k})\pm i\gamma[1+3(c^2/E)\varepsilon(\mathbf{k})]/2}\times\begin{pmatrix} 1 & c\varepsilon(\mathbf{k}) \\ c\varepsilon(\mathbf{k}) & \varepsilon(\mathbf{k})E\end{pmatrix}, \tag{36}$$

where G^R and G^A are the retarded and advanced Green functions.

3.2. Drude formula

The conductivity in our case is defined as in Sec. 1 in terms of the four-particle correlation function

$$\sigma_E(\omega)=\frac{e^2}{2\pi}\int\frac{d\mathbf{k}_1}{(2\pi)^2}\frac{d\mathbf{k}_2}{(2\pi)^2}v_\alpha(\mathbf{k}_1)v_\alpha(\mathbf{k}_2)K_{E\omega}\times(\mathbf{k}_1,\mathbf{k}_2;\mathbf{k}_2,\mathbf{k}_1), \tag{37}$$

where E is taken at the Fermi level,

$$K_{E\omega}(\mathbf{k}_1,\mathbf{k}_2;\mathbf{k}_2,\mathbf{k}_1)=\frac{1}{V}\sum_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{t}}\exp(i\mathbf{k}_1(\mathbf{x}-\mathbf{y}))\exp(i\mathbf{k}_2(\mathbf{z}-\mathbf{t}))\times\left\langle\rho_x\rho_y\rho_z\rho_t G^R_{11}\left(\mathbf{y},\mathbf{z},E+\frac{\omega}{2}\right)G^A_{11}\left(\mathbf{t},\mathbf{x},E-\frac{\omega}{2}\right)\right\rangle_\rho \tag{38}$$

and

$$v_\alpha(\mathbf{k})=\frac{\partial\varepsilon(\mathbf{k})}{\partial k_\alpha}. \tag{39}$$

Inserting the solution (36) into (37), we find in the lowest approximation

$$\sigma_E(\omega)=\frac{e^2}{2\pi}\int\frac{d\mathbf{k}}{(2\pi)^2}v_\alpha^2(\mathbf{k})G^R_{11}\left(\mathbf{k},E+\frac{\omega}{2}\right)\times G^A_{11}(\mathbf{k},E-\omega/2)=\frac{e^2c^6}{2\pi}\left(\frac{J}{B}\right)^2\frac{A(E)}{B(E)}, \tag{40}$$

where

$$A(E)=\int_{\varepsilon(\mathbf{k})=E/c^2}\frac{dl_{\mathbf{k}}}{v(\mathbf{k})}v_x^2(\mathbf{k}),\quad B(E)=\int_{\varepsilon(\mathbf{k})=E/c^2}\frac{dl_{\mathbf{k}}}{v(\mathbf{k})}, \tag{41}$$

$v(\mathbf{k})=\sqrt{v_x^2(\mathbf{k})+v_y^2(\mathbf{k})}$, and $dl_{\mathbf{k}}$ is the element of length of the Fermi surface.

The conductivity can be expressed in terms of a particle density defined by

$$n_0(E)=a^2\int_{-\pi/a}^{\pi/a}dk_x\int_{-\pi/a}^{\pi/a}dk_y\,\theta\left(\frac{E}{c^2}-\varepsilon(\mathbf{k})\right). \tag{42}$$

We obtain in the low and high density limits

$$\sigma=\begin{cases}\dfrac{e^2c^6}{g^2}\dfrac{n_0}{32\pi\log^2 2}, & \text{for } n_0\ll 1,\\[2ex] \dfrac{e^2c^6}{g^2}\dfrac{1-n_0}{4\kappa^4\log^2\kappa}, & \text{for } 1-n_0\ll 1.\end{cases} \tag{43}$$

The asymptotic behavior in the intermediate range $0<n_0<1$ is

$$\sigma=2\pi^4\frac{e^2c^6}{g^2}\frac{1}{(1-n_0)^2\log[1/(1-n_0)]}, \tag{44}$$

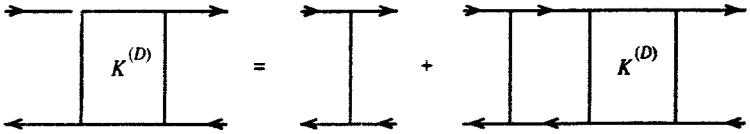

FIG. 1. Diffusion vertex.

with maximum value

$$\sigma_{\max}\sim\frac{1}{\kappa^{8/3}\log\kappa} \tag{45}$$

reached at $1-n_0\sim\kappa^{4/3}$.

3.3. Perturbation corrections

To go beyond the quasiclassical approximation we include contributions to the conductivity from ‘‘diffusion-ladder’’ and ‘‘crossed-ladder’’ or ‘‘cooperon’’ vertices.[3,4] The additional term is

$$\begin{aligned}\delta\sigma_E(\omega)=\frac{e^2c^4}{2\pi}\int\frac{d\mathbf{k}_1}{(2\pi)^2}\frac{d\mathbf{k}_2}{(2\pi)^2}v_\alpha(\mathbf{k}_1)v_\alpha(\mathbf{k}_2)\\ \times G^R_{1a}(\mathbf{k}_1)G^R_{c1}(\mathbf{k}_2)G^A_{1b}(\mathbf{k}_2)G^A_{d1}(\mathbf{k}_1)\\ \times K_{ac;bd}(\mathbf{k}_1,\mathbf{k}_2),\end{aligned} \tag{46}$$

where for the ‘‘diffusion’’ vertex contribution we obtain (see Fig. 1)

$$\begin{aligned}K^{(D)}_{ac;bd}&=g^2\sigma^x_{ac}\sigma^x_{bd}\\ &+g^2\int\frac{d\mathbf{l}}{(2\pi)^2}\sigma^x_{aa_1}G^R_{a_1c_1}(\mathbf{l})K^{(D)}_{c_1c;bb_1}G^A_{b_1d_1}(\mathbf{l})\sigma^x_{d_1d}.\end{aligned} \tag{47}$$

The solution of this equation does not depend on $\mathbf{k}_1$ and $\mathbf{k}_2$. Therefore the contribution of the ‘‘diffusion’’ vertex to the conductivity is equal to zero because

$$\int\frac{d\mathbf{k}_1}{(2\pi)^2}v_\alpha(\mathbf{k}_1)G^R_{1a}(\mathbf{k}_1)G^A_{d1}(\mathbf{k}_1)=0. \tag{48}$$

Now we consider the ‘‘cooperon’’ vertex contribution. The vertex $K^{(C)}$ obeys the equation (see Fig. 2)

$$\begin{aligned}K^{(C)}_{ac;bd}(\mathbf{q})=g^2\sigma^x_{ac}\sigma^x_{bd}+g^2\int\frac{d\mathbf{l}}{(2\pi)^2}\sigma^x_{ae}\sigma^x_{bg}G^R_{ea_1}(\mathbf{l})G^A_{gb_1}\\ \times(\mathbf{q}-\mathbf{l})K^{(C)}_{a_1c;b_1d}(\mathbf{q}),\end{aligned} \tag{49}$$

where $\mathbf{q}=\mathbf{k}_1+\mathbf{k}_2$.

Using the Green function from (36) we can rewrite this equation:

$$K^{(C)}_{ac;bd}-\frac{1}{4}b(\omega,k)\sigma^x_{ae}\sigma^x_{bg}\tau_{ea_1}\tau_{gb_1}K^{(C)}_{a_1c;b_1d}=g^2\sigma^x_{ac}\sigma^x_{bd}, \tag{50}$$

where

$$b=1+\frac{1}{4\gamma}(i\omega-Dk^2),\quad D=\frac{c^4\overline{v_x^2}}{4\gamma}, \tag{51}$$

and

$$\hat{\tau}=\begin{pmatrix}\exp(\theta) & 1\\ 1 & \exp(-\theta)\end{pmatrix}, \tag{52}$$

where $\exp(\theta)=c/E$.

We seek a solution of (50) in the form

$$K^{(C)}_{ac;bd}=g^2\sum_{\mu\nu}\sigma^\mu_{ab}\sigma^\nu_{cd}K^{\mu\nu}, \tag{53}$$

where $\mu,\nu\in\{0,x,y,z\}$, and $K^{\mu\nu}$ satisfies

$$K^{\mu\nu}-b(\omega,k)\sum_{\lambda=0,x,y,z}\Lambda_{\mu\lambda}K^{\lambda\nu}=\frac{1}{2}g_c^{\mu\nu}, \tag{54}$$

with

$$\hat{\Lambda}=\frac{1}{2}\begin{pmatrix}\cosh^2\theta & \cosh\theta & 0 & \sinh\theta\cosh\theta\\ \cosh\theta & 1 & 0 & \sinh\theta\\ 0 & 0 & 0 & 0\\ -\sinh\theta\cosh\theta & -\sinh\theta & 0 & -\sinh^2\theta\end{pmatrix} \tag{55}$$

and

$$\hat{g}_c=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}. \tag{56}$$

A solution of (54) can be found in terms of matrices *U*,*B*:

$$K^{\mu\nu}=U_{\mu M}B_{LN}(U^{-1})_{N\nu}, \tag{57}$$

where

$$(U^{-1})_{M\mu}\Lambda_{\mu\nu}U_{\nu N}=z_M\delta_{MN}, \tag{58}$$

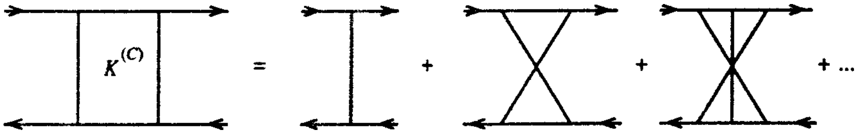

FIG. 2. Crossed-ladder vertex.

$$B_{MN}=\frac{1}{1-b(\omega,k)z_M}(U^{-1})_{M\mu}g^{\mu\nu}U_{\nu N}. \tag{59}$$

The eigenvalues z_M and matrix U are

$$z_0=1,\quad z_1=0,\quad z_2=0,\quad z_3=0, \tag{60}$$

$$\hat{U}=\begin{pmatrix} \frac{1}{\sqrt{2}}\cosh\theta & \frac{1}{\sqrt{2}}\cosh\theta & -i\sinh\theta & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}}\sinh\theta & -\frac{1}{\sqrt{2}}\sinh\theta & i\cosh\theta & 0 \end{pmatrix}. \tag{61}$$

We see from (51) and (60) that the eigenmode with $N=1$ only has singular behavior for $\omega,k\to 0$. Substituting solutions (60), (59), (57), (54), and (53) into (46) we obtain a logarithmically divergent correction to the conductivity:

$$\delta\sigma=-\frac{e^2}{2\pi^2}\int\frac{dq}{q}. \tag{62}$$

We consider the problem in detail in the next section using field-theoretic methods for diffusion mode interactions.

3.4. Field-theoretic description

Quantum corrections to the conductivity can be described in terms of a diffusion modes interactions. To derive an effective Lagrangian we make use of a replica method. Conductivity properties are determined by the density-density correlation function

$$K(\omega)=\langle G^R_{E+\omega/2}(\mathbf{r}_1,\mathbf{r}_2)G^A_{E-\omega/2}(\mathbf{r}_1,\mathbf{r}_2)\rangle, \tag{63}$$

where

$$\hat{G}^{(R,A)}(E)=[E\pm i\delta-t_{ij}\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)]^{-1}, \tag{64}$$

and angle brackets denote impurity averaging.

Integrating over anticommuting Grassmann variables κ^*,κ we can write

$$G^R_{E+\omega/2}(\mathbf{r}_1,\mathbf{r}_2)G^A_{E-\omega/2}(\mathbf{r}_1,\mathbf{r}_2)$$
$$=-\frac{\int D\kappa^* D\kappa\,\kappa_1(\mathbf{r}_1)\kappa_1^*(\mathbf{r}_2)\kappa_{N+1}(\mathbf{r}_1)\kappa^*_{N+1}(\mathbf{r}_2)\exp(iS)}{\int D\kappa^* D\kappa\exp(iS)}, \tag{65}$$

where

$$iS=i\sum_{\mathbf{r}_1,\mathbf{r}_2}\sum_{n=1}^{2N}\kappa_n^*(\mathbf{r}_1)\{[E+(\omega+i\delta)\lambda_n]I_{\mathbf{r}_1,\mathbf{r}_2}$$
$$-t_{\mathbf{r}_1,\mathbf{r}_2}\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\}\kappa_n(\mathbf{r}_2), \tag{66}$$

$$\lambda_n=\begin{cases}1, & n\leqslant N,\\ -1, & n>N,\end{cases} \tag{67}$$

and N is the number of replicas.

The quadratic term $\rho\rho$ can be transformed with the help of the additional Grassmann fields ν^*,ν

$$\exp\left\{-i\sum_{\mathbf{r}_1,\mathbf{r}_2}\sum_{n=1}^{2N}t(\mathbf{r}_1-\mathbf{r}_2)\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\kappa_n^*(\mathbf{r}_1)\kappa_n(\mathbf{r}_2)\right\}$$
$$=\int D\nu^* D\nu\exp\left\{i\sum_{\mathbf{r}_1,\mathbf{r}_2}\sum_{n=1}^{2N}t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\nu^*(\mathbf{r}_1)\right.$$
$$\times\nu(\mathbf{r}_2)-i\sum_{\mathbf{r}}\sum_{n=1}^{2N}\rho(\mathbf{r})$$
$$\left.\times[\nu_n^*(\mathbf{r})\kappa_n(\mathbf{r})+\kappa_n^*(\mathbf{r})\nu_n(\mathbf{r})]\right\}\left[\int D\nu^* D\nu\right.$$
$$\left.\times\exp\left\{i\sum_{\mathbf{r}_1,\mathbf{r}_2}\sum_{n=1}^{2N}t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\nu^*(\mathbf{r}_1)\nu(\mathbf{r}_2)\right\}\right]^{-1}. \tag{68}$$

It is convenient to define spinors Ψ, χ:

$$\sqrt{2}\Psi_n=\begin{pmatrix}\Psi_{n1}\\ \Psi_{n2}\end{pmatrix}=\begin{pmatrix}\kappa_n^*\\ \kappa_n\end{pmatrix},\quad \sqrt{2}\overline{\Psi}_n=\begin{pmatrix}-\kappa_n\\ \kappa_n^*\end{pmatrix}, \tag{69}$$

$$\sqrt{2}\chi_n=\begin{pmatrix}\chi_{n1}\\ \chi_{n2}\end{pmatrix}=\begin{pmatrix}\nu_n^*\\ \nu_n\end{pmatrix},\quad \sqrt{2}\overline{\chi}_n=\begin{pmatrix}-\nu_n\\ \nu_n^*\end{pmatrix}, \tag{70}$$

or

$$\overline{\Psi}=(C\Psi)^T,\quad \overline{\chi}=(C\chi)^T,\quad C=\delta_{mn}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}, \tag{71}$$

where C is the charge conjugation matrix. The same relations hold between the original and Fourier components

$$\overline{\Psi}_{ni}(p)=C_{ij}\Psi_{ni}(-p),\quad \Psi_{ni}(p)=C_{ji}\overline{\Psi}_{ni}(-p), \tag{72}$$

$$\overline{\chi}_{ni}(p)=C_{ij}\chi_{ni}(-p),\quad \chi_{ni}(p)=C_{ji}\overline{\chi}_{ni}(-p), \tag{73}$$

where

$$\Psi_{ni}(\mathbf{r})=\sum_{\mathbf{p}}\frac{1}{\sqrt{N}}\Psi_{ni}(\mathbf{p})e^{i\mathbf{pr}},\quad \chi_{ni}(\mathbf{r})=\sum_{\mathbf{p}}\frac{1}{\sqrt{N}}\chi_{ni}(\mathbf{p})e^{i\mathbf{pr}}. \tag{74}$$

The action (66) takes the form

$$iS=i\left\{\sum_{\mathbf{r}}\overline{\Psi}(\mathbf{r})\left[E+\left(\frac{\omega}{2}+i\delta\right)\Lambda\right]\Psi(\mathbf{r})\right.$$
$$+\sum_{\mathbf{r}_1,\mathbf{r}_2}\overline{\chi}(\mathbf{r}_1)t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\chi(\mathbf{r}_2)$$
$$\left.-\sum_{\mathbf{r}}[\overline{\chi}(\mathbf{r})\Psi(\mathbf{r})+\overline{\Psi}(\mathbf{r})\chi(\mathbf{r})]\right\}, \tag{75}$$

where Λ is a diagonal matrix consisting of elements λ_n. Introducing new bispinors

$$\varphi(\mathbf{r})=\begin{pmatrix}\Psi(\mathbf{r})\\ \chi(\mathbf{r})\end{pmatrix},\quad \overline{\varphi}(\mathbf{r})=(\overline{\Psi}(\mathbf{r})\overline{\chi}(\mathbf{r}))$$

we can rewrite the action as

$$iS=i\sum_{\mathbf{r}_1,\mathbf{r}_2}\overline{\varphi}(\mathbf{r}_1)G_0^{-1}(\mathbf{r}_1,\mathbf{r}_2)\varphi(\mathbf{r}_2)$$

$$-i\sum_{\mathbf{r}}\delta\rho(\mathbf{r})\overline{\varphi}(\mathbf{r})\sigma^x\varphi(\mathbf{r}),\tag{76}$$

where

$$G_0^{-1}(\mathbf{r}_1,\mathbf{r}_2)=\begin{pmatrix}E+\left(\dfrac{\omega}{2}+i\delta\right)\Lambda & -c\\ -c & t^{-1}(\mathbf{r}_1-\mathbf{r}_2)\end{pmatrix}\tag{77}$$

is the bare Green function.

Performing the average over a Gaussian distribution for the randomness we obtain

$$\langle\exp(iS_{\text{int}})\rangle=\left\langle\exp\left[-i\sum_{\mathbf{r}}\delta\rho(\mathbf{r})\overline{\varphi}(\mathbf{r})\sigma^x\varphi(\mathbf{r})\right]\right\rangle$$

$$=\exp\left\{-\frac{g^2}{2}\sum_{\mathbf{r}}[\overline{\varphi}(\mathbf{r})\sigma^x\varphi(\mathbf{r})]^2\right\}.\tag{78}$$

The order parameter in the localization theory is a traceless tensor proportional to $\langle\Psi\overline{\Psi}\rangle$. Taking into account only long-wavelength fluctuations we can rewrite (78) as

$$\langle\exp(iS_{\text{int}})\rangle=\exp\left[g^2\sum_{\mathbf{r}}\sigma^x_{ab}\sigma^x_{cd}P^{\nu\mu}_{da}(\mathbf{r})P^{\mu\nu}_{bc}(\mathbf{r})\right],\tag{79}$$

where

$$P^{\mu\nu}_{bc}(\mathbf{r})=\frac{1}{L}\sum_{\mathbf{p},\mathbf{q}}\varphi^{\mu}_{b}(\mathbf{p})\overline{\varphi}^{\nu}_{c}(\mathbf{p}+\mathbf{q})\exp(-i\mathbf{q}\mathbf{r}),\quad \nu=\{n,i\}.$$

Spinors φ and $\overline{\varphi}$ are related by charge conjugation (71). This imposes symmetry conditions

$$[P^{\mu\nu}_{bc}(\mathbf{r})]^*=P^{\nu\mu}_{cb}(\mathbf{r}),\quad\text{or}\quad P^{+}=P,\tag{80}$$

$$P=CP^TC^T.\tag{81}$$

Introducing a Gaussian integration over the c-number matrix field Q, we obtain

$$\exp(iS_{\text{eff}})=\int DQ\,\exp\left\{\sum_{\mathbf{r}_1,\mathbf{r}_2}\overline{\varphi}(\mathbf{r}_1)\right.$$

$$\times\left[iG_0^{-1}(\mathbf{r}_1,\mathbf{r}_2)+I_{\mathbf{r}_1\mathbf{r}_2}\frac{\beta}{4}Q(\mathbf{r}_1)\right]\varphi(\mathbf{r}_2)$$

$$\left.-\sum_{\mathbf{r}}\frac{\alpha}{4}\operatorname{Tr}(Q\sigma^xQ\sigma^x)\right\}\left\{\int DQ\right.$$

$$\left.\times\exp\left[-\sum_{\mathbf{r}}\frac{\alpha}{4}\operatorname{Tr}(Q\sigma^xQ\sigma^x)\right]\right\}^{-1}.\tag{82}$$

After integration over the bispinor field φ we obtain

$$iS_{\text{eff}}=\operatorname{Tr}\ln\left[iG_0^{-1}+\frac{\beta}{4}Q\right]-\sum_{\mathbf{r}}\frac{\alpha}{4}\operatorname{Tr}(Q\sigma^xQ\sigma^x).\tag{83}$$

In the saddle-point approximation we use the equation

$$\delta S/\delta Q=0,$$

or

$$\left(iG_0^{-1}+\frac{\gamma}{2}Q\right)^{-1}-\frac{\gamma}{2g^2}\sigma^xQ\sigma^x=0.\tag{84}$$

The solution of this equation to lowest order in g^2 is

$$Q_{sp}=\Lambda(\cosh\theta-\sigma^z\sinh\theta+\sigma^x)\exp\theta,\tag{85}$$

$$E/c=\exp\theta,\tag{86}$$

with the Green function

$$G=\frac{1}{E+\Lambda\omega/2-c^2\varepsilon(\mathbf{p})-2i\gamma\Lambda}\begin{pmatrix}1 & E/c\\ E/c & E^2/c^2\end{pmatrix}.\tag{87}$$

Expanding near the saddle point and using the symmetry of the tensor Q, we find

$$iS_{\text{eff}}=\operatorname{Tr}\ln\left(iG_0^{-1}+\frac{\gamma}{2}\delta Q\right)-\frac{\gamma^2}{16g^2}$$

$$\times\sum_{\mathbf{r}}\operatorname{Tr}\{[Q+\delta Q(\mathbf{r})]\sigma^x[Q+\delta Q(\mathbf{r})]\sigma^x\}.\tag{88}$$

Taking Fourier transforms, Eq. (88) reads

$$i\,\delta S_{\text{eff}}=-\frac{\gamma^2}{16g^2}\frac{1}{L}\sum_{\mathbf{q}}\left\{\operatorname{Tr}[\delta Q(\mathbf{q})\sigma^x\delta Q(-\mathbf{q})\sigma^x]\right.$$

$$-g^2\int\frac{d\mathbf{p}}{(2\pi)^2}\operatorname{Tr}[G(\mathbf{p})\delta Q(\mathbf{q})G(\mathbf{p}+\mathbf{q})$$

$$\left.\times\delta Q(-\mathbf{q})]+\operatorname{Tr}\left[\frac{i\omega\gamma}{4}\begin{pmatrix}1&0\\0&0\end{pmatrix}\delta Q(\mathbf{q})\right]\right\}.\tag{89}$$

Due to the symmetry of the Q we can write the variation δq as

$$\delta Q_{bc}=Q_{\alpha}\sigma^{\alpha}_{bc_1}\sigma^{x}_{c_1c},\tag{90}$$

where all matrices Q_α are real. Inserting the relation (90) into (89), we find, taking into account only low-energy transverse $(Q\,\delta Q+\delta Q Q=0)$ modes,

$$\delta S_{\text{eff}}=-\frac{\gamma^2}{8g^2}\sum_{\alpha,\beta}\left[\delta_{\alpha\beta}-\frac{b_0}{8}\operatorname{Tr}(\sigma^{\beta}\sigma^{x}\tau\sigma^{\alpha}\sigma^{x}\tau)\right]$$

$$\times Q_\alpha Q_\beta+Q_\alpha T_\alpha,\tag{91}$$

where

$$b_0=1-\frac{Dk^2}{4\gamma},\quad D=\frac{c^4\overline{v_x^2}}{4\gamma},\quad \tau=\cosh\theta+\sigma^z\sinh\theta+\sigma^x,$$

$$T_\alpha=\operatorname{Tr}\left[\frac{i\omega}{2}\Lambda\begin{pmatrix}1&0\\0&0\end{pmatrix}\sigma^{\alpha}\sigma^{x}\right].\tag{92}$$

After diagonalization we obtain

$$iS_{\text{eff}}=\frac{1}{L}\sum_{\mathbf{k}}\operatorname{Tr}\left\{-\frac{\gamma Dk^2}{32g^2}q_1(\mathbf{k})q_1(-\mathbf{k})\right.$$

$$-\frac{\gamma}{8g^2}\sum_{l=2,3,4}\operatorname{Tr}[q_l(\mathbf{k})q_l(-\mathbf{k})]$$

$$\left.+\frac{i\omega\gamma}{16g^2}\Lambda U_{\alpha l}q_l(\mathbf{k})\right\},\tag{93}$$

where

$$Q_\alpha = U_{\alpha l} q_l , \tag{94}$$

$$U_{\alpha l} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}}\cosh\theta & \frac{1}{\sqrt{2}}\cosh\theta & -i\sinh\theta & 0 \\ -\frac{i}{\sqrt{2}}\sinh\theta & -\frac{i}{\sqrt{2}}\sinh\theta & -\cosh\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{95}$$

We see that the action (93) consists of only one diffusion type mode $(l=1)$. Taking into account only these modes and neglecting interactions with other higher-energy modes, we obtain the well-known action of the nonlinear σ model,

$$S = \frac{1}{t}\int d\mathbf{r}[\mathrm{Tr}(\nabla Q)^2 - \tilde{\omega}\,\mathrm{Tr}(\Lambda Q)], \tag{96}$$

where

$$\tilde{\omega} = \frac{\gamma}{16g^2}\frac{\omega}{D}, \quad \frac{1}{t} = \frac{\gamma}{16g^2}D.$$

The renormalization group equations for this case were studied in Ref. 5. In the limit $N \to 0$, we see that in lowest order

$$\frac{dt}{d\ln k} = \frac{t^2}{8}, \tag{97}$$

$$\frac{d\ln\tilde{\omega}}{d\ln k} = 0. \tag{98}$$

Since the conductivity σ is proportional to the diffusion constant D, equations (97) and (98) determine the frequency and system length dependence of the conductivity. In particular, we have from (98) that $\sigma(\omega) \propto D \propto \omega$ for small ω.

4. CONCLUSIONS

We have investigated a two-dimensional model with a new type of disorder due to a random distribution of local states with strongly anisotropic overlaps of wave functions. This type of disorder, described by a quadratic impurity density Hamiltonian (1), was not considered previously. The conductivity of the system was calculated in the limits of low (Sec. 2) and high densities (Sec. 3) of local states. Since perturbation theory leads to divergent terms (''Cooperon'' vertices) we used the field theoretic description in terms of a diffusion mode interaction. Introducing an additional integration over fermion field and performing the average over impurities with the help of a replica trick, we obtained the action of the nonlinear σ model. Renormalization group equations for this model determine the behavior of the diffusion constant and the conductivity. We have shown that this type of disorder leads to weak localization phenomena in the high density limit, as in the usual two-dimensional case.[6]

As mentioned in the Introduction, a similar picture can be realized in nontrivial superconductors. A strong scattering impurity potential produces a resonant or marginally bound state inside the gap in a *d*-wave superconductor. The wave function of the impurity bound state is highly anisotropic, with $1/r$ decay along the nodes of the gap, and, an exponential, angle-dependent decay range otherwise. A finite concentration of impurities leads to an formation of the narrow quasiparticle band. If we simplify a picture and take into account scattering processes of quasiparticles inside this band only, we obtain our model. We note that our consideration is applicable only in the case of strong unitary impurities producing local bound states. Opposite cases of weak impurity scattering with different types of disorder were studied in Refs. 7 and 8.

This work was supported by the Russian Fund for Fundamental Research under Grant No. 960217791.

[1] A. V. Balatsky and M. T. Salkola, Phys. Rev. Lett. **76**, 2386 (1996), e-print archive cond-mat/9702087. D. N. Aristov and A. G. Yashenkin, e-print archive cond-mat/9602087.

[2] N. B. Kopnin and G. E. Volovik, JETP Lett. **64**, 690 (1996).

[3] L. P. Gor'kov, A. I. Larkin, and D. E. Khmelnitskii, JETP Lett. **30**, 228 (1979).

[4] E. Abrahams, P. W. Anderson, and T. V. Ramakrishnan, Philos. Mag. B **42**, 827 (1980).

[5] K. B. Efetov, A. I. Larkin, and D. E. Khmelnitskii, Zh. Éksp. Teor. Fiz. **79**, 1120 (1980) [Sov. Phys. JETP **52**, 548 (1980)].

[6] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).

[7] A. A. Nersesyan, A. M. Tsvelik, and F. Wenger, Phys. Rev. Lett. **72**, 2628 (1994).

[8] K. Ziegler, M. H. Hettler, and P. J. Hirschfeld, Phys. Rev. Lett. **77**, 3013 (1996).

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

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Magnetization of $FeBO_3$ by microwave pumping

L. E. Svistov

Institut für Festkörperphysik, TH-Darmstadt, D-64289, Darmstadt, Germany;
A. V. Shubnikov Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russia

H. Benner

Institut für Festkörperphysik, TH-Darmstadt, D-64289, Darmstadt, Germany

(Submitted 19 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1107-1115 (March 1999)

Changes in the magnetic moment of crystalline $FeBO_3$ when an antiferromagnetic resonance is excited in it are studied. This done using a SQUID magnetometer in combination with a microwave spectrometer. At temperatures $T>15$ K a reduction in the total magnetic moment of the sample is observed when an antiferromagnetic resonance is excited in it. At liquid helium temperatures, an induced rise in the sample magnetic moment was observed. This type of magnetization of the sample can be explained qualitatively if it is assumed that under antiferromagnetic resonance excitation conditions, magnetoelastic modes with high wave numbers are excited along with magnons. Efficient excitation of magnetoelastic modes under antiferromagnetic resonance conditions is confirmed by the experimental observation of a size effect in thin, high quality single crystal slabs of $FeBO_3$. © *1999 American Institute of Physics.* [S1063-7761(99)02503-2]

1. INTRODUCTION

The electron spin resonance of antiferromagnetic $FeBO_3$ is studied in this paper. In these experiments simultaneous measurements were made of the microwave power absorbed by the sample and the accompanying change in the total magnetic moment of the sample.

The change in the sample magnetic moment is a result of all the nonequilibrium perturbations of the sample caused directly by the microwave pumping, as well as by ''secondary'' quasiparticles created through relaxation of the primaries. A simultaneous analysis of the microwave power absorbed by the sample and the corresponding change δM in the sample magnetic moment yields information on the nature of the quasiparticles that are excited and on their relaxation mechanism.

Since the measured δM is an integral quantity, this analysis is effective for materials with a known excitation spectrum. For magnetically ordered substances measurements have been made for compounds with well studied excitation spectra, such as nickel ferrite[1] and iron yttrium garnet,[2] as well as antiferromagnetic materials with an easy-plane magnetic anisotropy ($CoCO_3$, $FeBO_3$, and $MnCO_3$), in which magnons, nuclear magnons, and magnetoelastic oscillations are excited.[3–7]

The properties of $FeBO_3$ ($T_N=348$ K), an antiferromagnetic material with a easy-plane magnetic anisotropy, have been well studied and are, to a great extent, determined by a magnetoelastic interaction. The magnetoelastic interaction causes coupling of magnetic and elastic excitations. Thus, the normal modes (quasimagnons and quasiphonons) contain elastic, as well as magnetic components. The quasimagnon spectrum includes an additional gap, while the quasiphonon branch acquires a dependence on the magnitude of the static magnetic field. The calculated spectra of the low frequency branch of the quasimagnons (ω_k) and of the acoustic branch of the quasiphonons (Ω_k) for a infinite crystal (which makes it possible to restrict the analysis to the magnetoelastic coupling of magnons and phonons with equal wave vectors) are given by[8–10]

$$\omega_k=\gamma[H(H+H_D)+H_\Delta^2+(\alpha k)^2]^{1/2}, \tag{1}$$

$$\Omega_k=c_e k\cdot[1-(\gamma H_\Delta \xi/\omega_k)^2]^{1/2}, \tag{2}$$

where H is a static magnetic field directed in the easy plane of the crystal; $\gamma=g\mu_B/\hbar=17.8\times10^9\ \mathrm{s}^{-1}\mathrm{kOe}^{-1}$ is the gyromagnetic ratio; $H_D\simeq100$ kOe is the Dzyaloshinskii field; γH_Δ is the magnetoacoustic gap in the spin wave spectrum ($H_\Delta\simeq1.9$ kOe); and, $\alpha\simeq0.8\times10^{-5}$ kOe·cm is a phenomenological exchange constant that is proportional to the exchange field $H_E=2.6\times10^6$ Oe. The sound speed c_e and the coefficient ξ, which describes the efficiency of the linear interaction between the magnons and phonons, depend on the direction of the wave vector and the polarization of the quasiphonons. The values of the constants are specified for low temperatures, far from T_N.

Figure 1a shows spectra of quasimagnons and quasiphonons with wave vectors directed along the C_3 axis of the crystal and with the polarization of the quasiphonons parallel to **H**. For such quasiphonons, $c_e\simeq4.8\times10^5\ \mathrm{cm\,s}^{-1}$ and $H_\Delta\xi\simeq2$ kOe. (The values of all elastic and magnetoelastic constants of $FeBO_3$ can be found, for example, in Ref. 11). In Fig. 1b we have plotted, as a function of the wave vector, the amount by which the magnetic

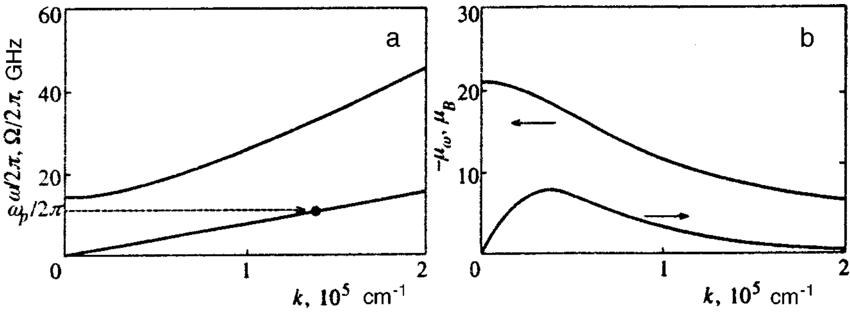

FIG. 1. (a) Spectra of quasimagnons and quasiphonons with wave vectors directed along the C_3 axis of a crystal and with the quasiphonons polarized parallel to **H**. (b) The magnitudes of the changes in the sample magnetic moment during excitation of a single quasimagnon and quasiphonon as functions of the wave vector. The curves were obtained using Eqs. (1)–(3) for a magnetic field $H=200$ Oe.

moment of the sample is reduced by excitation of a single quasimagnon ω_k or quasiphonon Ω_k, as obtained from (1) and (2) using the formula

$$\mu_k=-\partial\varepsilon_k/\partial H, \tag{3}$$

where $\varepsilon_k=\hbar\omega_k/2\pi$ for the quasimagnons or $\varepsilon_k=\hbar\Omega_k/2\pi$ for the quasiphonons. ($\hbar$ is Planck's constant.)

2. SAMPLES AND TECHNIQUE

The excitation spectrum of $FeBO_3$ is sensitive to the internal stresses of the crystal.[12] Thus, the samples, grown in the form of thin slabs (20-150 μm thick), were examined by x-ray diffraction topography (the Lang method) at a wavelength corresponding to the Mo $K_{\alpha 1}$ line. The samples which yielded a uniform image without block segments and dislocations were selected. These samples were placed in envelopes of cigarette paper and attached to them at two ''points.'' The samples prepared in this fashion were fastened by the end of the envelope to the loop of a coaxial line or in a cavity. This method of attaching the samples was used in order to avoid the deformations caused by gluing them directly. The thick samples were glued at the end of the crystal.

An apparatus consisting of a combination of a SQUID magnetometer and a microwave spectrometer[6,7] was constructed for measuring the change (δM) in the magnetic moment of the sample when it was acted on by the microwave magnetic field. Figure 2 is a sketch of the apparatus. An outer vessel filled with liquid helium contained the SQUID magnetometer cell, a differential flux transformer, and a superconducting solenoid operated in short circuit. The sample and a system for delivering the microwave power were located inside an inner Dewar flask. The inner Dewar flask was metallic in order to prevent the microwaves' acting directly on the SQUID magnetometer cell. The inner Dewar flask was filled with helium as a heat exchanger gas or with liquid nitrogen or helium. The working range of fields in the apparatus was 0-450 Oe. The apparatus made it possible to measure changes in the magnetic moment of a sample with an accuracy of 5×10^{-7} Oe$\cdot$cm^3.

A microwave magnetic field was created at the sample either by a loop which served as the load of a coaxial line or in coaxial (at frequencies of 7-10 GHz) or helical (0.5-1.2 GHz) cavities.

Using a coaxial line loaded with a loop makes it possible to excite microwaves over a wide range of frequencies in a single experiment. The amplitude of the field at the sample, however, depends on the frequency and it is difficult to determine its magnitude in this experimental setup.

IN order to obtain quantitative results, a set of resonators was used. By measuring the power passing through the cavity and reflected from it, as well as its parameters, one can determine the microwave power absorbed by the sample and estimate the magnitude of the microwave magnetic field at the sample.

In all the experiments the microwave field **h** and the static field **H** were mutually perpendicular and lay in the easy magnetization plane of the crystal.

3. EXPERIMENTAL RESULTS

Figure 3 shows the results of measuring δM when a microwave magnetic field produced by a loop load on a coaxial line is applied to a sample. The abscissa is a voltage proportional to the frequency of the microwave generator operating with frequency wobble. It is clear that the magnetic moment of the sample is sensitive to the microwave interaction at low frequencies and in the range 7-10 GHz. The existence of a large number of peaks in this curve is related to the excitation of resonances in the waveguide line. It is natural to relate the low frequency response (Fig. 3a) to the excitation of magnetoelastic modes. The high frequency peak (Fig. 3b) is close to the antiferromagnetic resonance

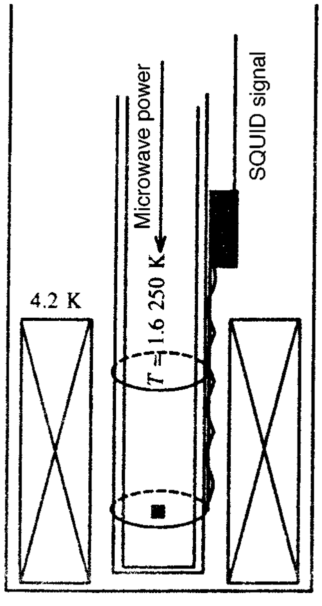

FIG. 2. A sketch of the apparatus.

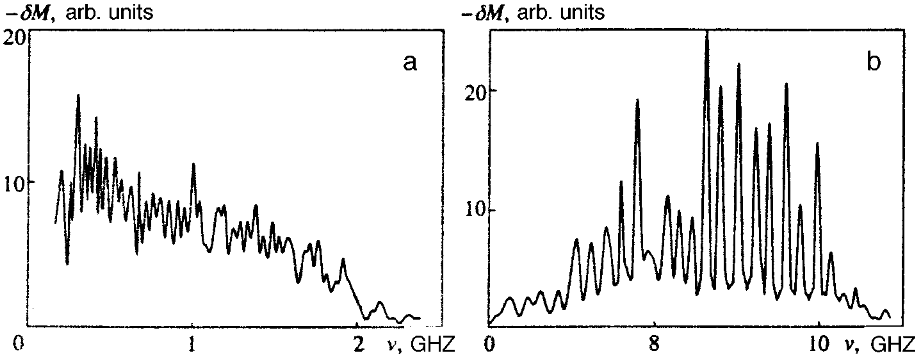

FIG. 3. The change δM in the sample magnetic moment when a microwave magnetic field produced by a loop attached to a coaxial line is applied to it. The abscissa is a voltage proportional to the frequency of the microwave generator. The sample magnetic moment is sensitive to microwave pumping at low frequencies (a) and at frequencies of 7-10 GHz (b). The oscillatory character of these curves is caused by the excitation of resonances in the waveguide line. $H=200$ Oe and $T=77$ K.

frequency given by Eq. (1). Curves of this sort were measured for different values of the magnetic field H and temperature T.

Figure 4 shows plots of δM as a function of the static magnetic field for various frequencies at temperatures of 135 and 4.2 K.

At "high" temperatures the magnetic moment of the sample decreases when it is acted on by a microwave magnetic field (Fig. 4a). The plot of δM as a function of the magnetic field has a resonant feature, which shifts to higher fields as the frequency is raised. This behavior is consistent with the spectrum of the antiferromagnetic resonance calculated using Eq. (1). When the temperature is reduced, the resonance shifts to lower fields.

An unexpected result was obtained near the temperature of liquid helium (Fig. 4b). The magnetic moment of the sample increased when a microwave magnetic field was applied to it, both at the low frequencies corresponding to effective excitation of magnetoelastic modes and at frequencies near the antiferromagnetic resonance. The $\delta M(H)$ curve was independent of frequency at low temperatures (to within the accuracy of the experiment) (Fig. 4b).

We should point out, at once, that the temperature dependence of the sample magnetic moment was investigated experimentally. Over the entire range of fields we examined, the sample magnetic moment decreased as the temperature was raised. Thus, trivial overheating of the crystal cannot explain the observed effect.

The fractional change in the total magnetic moment of the sample in all these experiments was less than 5×10^{-4}. With such weak interactions, δM was proportional to the microwave power applied to the sample for the fields and temperatures studied here.

Figure 5 shows δM as a function of temperature when microwave fields with frequencies of 730 MHz and 8.5 GHz act on the sample. δM changes sign at $T\simeq 8-10$ K.

Figure 6 shows the field dependence of the microwave power transmitted through a coaxial cavity tuned to 7.8 GHz. A sample in the form of a thin slab of thickness 114 μm was placed at an antinode of the microwave magnetic field. A number of resonance features can be seen in the plot of the microwave power transmitted through the cavity.

4. DISCUSSION

The variation in the magnetic moment of a sample when quasielastic vibrations with frequencies of 0.1-2 GHz are excited in it was studied in an previous paper.[7] Quasielastic excitations were created in the sample by three different methods: (1) by contact, using a piezoelectric transducer, (2) a variable microwave field at the magnetoelastic resonance frequencies of the crystal, and (3) parametrically, by parallel pumping. All three methods yielded the same field dependences for the variation in the magnetic moment of the sample when quasielastic modes were excited in it. Near the temperature of liquid nitrogen, the total magnetic moment of the sample decreased when nonequilibrium quasielastic vibrations were excited in it. The measured magnitude of δM was the same as the estimated value obtained assuming that the secondary quasiparticles do not make a significant contribution to the change in the total magnetic moment or, in other words, that relaxation of the nonequilibrium packet to a thermal level is a one-step process. Near the temperature of liquid helium, magnetization of the sample was observed when quasielastic waves were excited in it. This sort of stimulation of magnetization in a crystal by phonon pumping

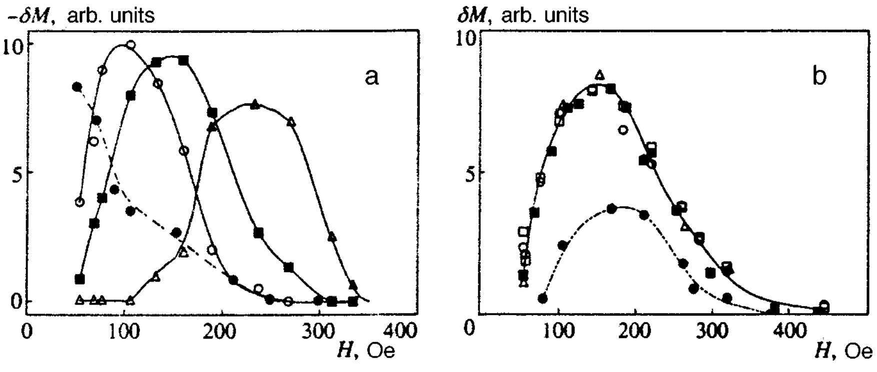

FIG. 4. The change δM in the magnetic moment of a sample when a microwave magnetic field is applied to it as a function of the magnitude H of the static field for different frequencies and temperatures: (a) $T=135$ K, $\nu=8.00$ GHz (hollow circles), 9.23 GHz (solid squares), 10.28 GHz (triangles); $T=32$ K, $\nu=8.00$ GHz (solid circles); (b) $T=4.2$ K, $\nu=8.00$ GHz (hollow circles), 9.00 GHz (triangles), 11.00 GHz (hollow squares), 12.24 GHz (solid squares), and 1.00 GHz (solid circles). At "low" temperatures δM is positive, at "high" temperatures it is negative.

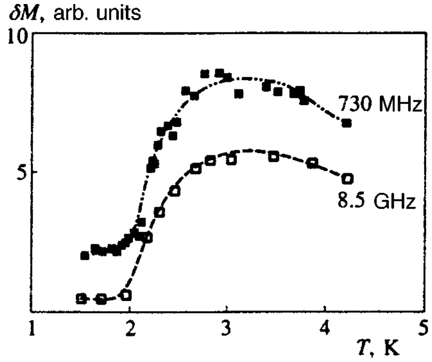

FIG. 5. δM as a function of temperature when microwave magnetic fields with frequencies of 730 MHz (solid squares) and 8.5 GHz (hollow squares) are applied to the sample.

in antiferromagnetic materials with a strong magnetoelastic interaction had been predicted theoretically before.[13] This effect is related to the fact that at low temperatures the relaxation of a nonequilibrium packet of phonons is a multistep process.

It has been shown experimentally[14] that three-body magnon-phonon interactions (magnon+phonon=magnon) in $FeBO_3$ are the most effective at low temperatures. The characteristic feature of this process is that it does not change the total number of magnons. In the equilibrium state the populations have a Bose-Einstein distribution. When phonon pumping and efficient three-body magnon-phonon interactions are present, the populations of the excitations should relax at different rates. Thus, phonon pumping will cause thermal magnons to be transferred from the bottom of the band to higher frequency regions. As can be seen in Fig. 1, the magnetic moment corresponding to excitation of a single magnon decreases with its frequency, so this kind of redistribution leads to an increase in the total magnetic moment of a sample. Of course, this increase should be less than the reduction in the magnetic moment of the sample owing to the presence of thermal magnons.

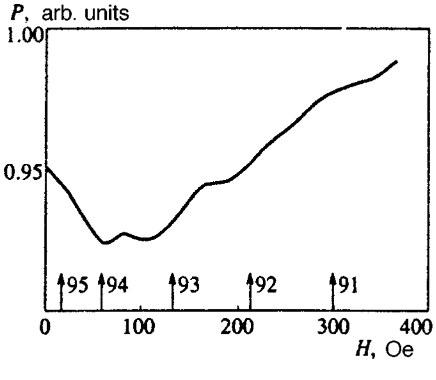

FIG. 6. The field dependence of the microwave power transmitted through a coaxial cavity. The sample is in the shape of a thin slab with dimensions $1\times1.5\times0.114\ \mathrm{mm}^3$. The arrows denote the static fields for which a pump frequency of $\nu=7.8$ GHz coincides with the frequency of a standing magnetoelastic wave calculated using Eq. (2). The numbers next to the arrows are the numbers of half waves across the sample thickness under the resonance conditions.

In our experiments the calculated reduction in the sample magnetic moment associated with thermally equilibrium excitations was at least 20 times the observed increase in the sample magnetic moment when microwave power was applied to it. The increased efficiency with which the sample was magnetized at low temperatures $T<3$ K (Fig. 5) is apparently a consequence of a drop in the number of thermal magnons (i.e., of the fact that the efficiency of the three-particle magnon-phonon interaction decreases with temperature[14]).

Although the increase in the sample magnetic moment when magnetoelastic modes are excited in it can be explained, it is not possible to explain the increase in the sample magnetic moment when magnetic oscillations are excited in it.

Apparently, for low fields (where the magnetoelastic interaction is most efficient) near the antiferromagnetic resonance frequency it is not the magnon branch which is most efficiently excited, but magnetoelastic modes with large wave vectors. In Fig. 1 we have indicated with an arrow the quasiphonons which we believe are efficiently excited when a microwave magnetic field acts on the sample at a frequency ω_p close to the antiferromagnetic resonance frequency.

The efficiency with which magnetoelastic modes are linearly excited in the sample should, on one hand, decrease with increasing wave number and increase resonantly when the frequencies of the magnon and elastic modes coincide. Thus, we may expect efficient linear excitation of magnetoelastic modes both at low frequencies and near the antiferromagnetic resonance frequency.

The linear excitation of magnetoelastic modes in ferromagnetic materials with finite dimensions under ferromagnetic resonance conditions has been studied theoretically.[15] It is a complicated matter to distinguish experimentally the absorption of microwave power owing to the excitation of quasimagnon and quasielastic modes, so this effect has apparently not been studied experimentally before.

An experimental confirmation of this hypothesis is the presence of resonance features in the absorption line measured for samples in the form of a thin slab (Fig. 6). For a sample of this shape, the eigenfrequencies of the magnetoelastic vibrations are nonuniformly distributed in frequency. Their density is greatest near the frequencies corresponding to standing magnetoelastic waves in an infinite slab. The calculated values of the static fields for which the pump frequency coincides with the frequency of a standing magnetoelastic wave in an infinite slab of this thickness are indicated by arrows in Fig. 6. The calculations were based on Eq. (2) for the magnetoelastic waves most strongly coupled to the microwave magnetic field. Note that the situation is much more complicated for a finite crystal and the excitation of magnetoelastic modes with other polarizations may be no less efficient.

Nevertheless, the calculated and measured separations between the resonances are in satisfactory agreement. The resonance fields shown in Fig. 6 were obtained through fitting of the slab thickness within a range of one percent. Here the calculated mode number for each of the resonances is determined with an accuracy of $N\pm3$.

The hypothesis proposed here can be used to make a lower bound estimate of the time for relaxation of the nonequilibrium magnon distribution to thermodynamic equilibrium. To do this we have measured the absolute magnitudes of the microwave power absorbed by the sample and the corresponding change in the magnetic moment. Thus, at a temperature of 4.2 K, when the sample absorbed a microwave power of $P_a=30\mu$W at a frequency of 8 GHz, the magnetic moment of the sample was observed to increase by 7.5×10^{-7} G·cm3.

The characteristic relaxation time for the nonequilibrium magnon distribution to a thermodynamic equilibrium state can be estimated using the formula

$$\tau=\frac{\delta M}{\mu}\frac{h\nu}{P_a},$$

where μ is the characteristic increase in the magnetic moment of the sample when a magnon with frequency ν is transferred to higher frequencies: $\nu+8$ GHz. As can be seen from Fig. 1, each act of transferring a magnon with frequency ν to higher frequencies $\nu+8$ GHz leads to an increase in the sample magnetic moment by up to five Bohr magnetons. The power P_a absorbed by the sample divided by the energy $h\nu$ of a single excitation is an estimate of the number of transfers per unit time. Substituting the experimental values of δM and P_a, we obtain $\tau\simeq3\mu$s. This is a lower bound estimate and is only valid when the three-body magnon-phonon process is most efficient.

It is possible that this redistribution of thermal magnons is the reason for the observed ''rigidity'' of parametric magnon excitation in $FeBO_3$.[16] This is supported by the fact that the rigidity phenomenon can undergo saturation with further application of microwave pumping to the sample ($\mathbf{h}\perp\mathbf{H}$) near the antiferromagentic resonance frequency.[16] It is interesting that the ''rigid'' character of the parametric excitation of magnons is recovered roughly 100 μs after the additional pumping ceases.[16] This time is apparently determined by the relaxation time of the magnon system discussed earlier[7] and in this paper.

We thank L. A. Prozorova, A. I. Smirnov, I. L. Smol'skiĭ, and Yu. S. Kivshar for useful discussions. This work was supported by the Russian Fund for Fundamental Research (Grant 96-02-16575) and an Alexander von Humboldt stipend.

[1] N. Bloembergen and S. Wang, Phys. Rev. **93**, 72 (1954).
[2] H. Le Gall and J. P. Jamet, in *Proc. of the International Conference on Magnetism*, I (II), Moscow (1973), p. 20.
[3] A. S. Borovik-Romanov, V. G. Zhotikov, N. M. Kreĭnes, and A. A. Pankov, Zh. Éksp. Teor. Fiz. **70**, 1924 (1976) [Sov. Phys. JETP **43**, 1002 (1976)].
[4] B. Ya. Kotyuzhanskiĭ and L. A. Prozorova, Zh. Éksp. Teor. Fiz. **85**, 1461 (1983) [Sov. Phys. JETP **58**, 846 (1983)].
[5] L. I. Svistov, Zh. Éksp. Teor. Fiz. **99**, 1612 (1991) [Sov. Phys. JETP **72**, 900 (1991)].
[6] L. E. Svistov, J. Löw, and H. Benner, J. Phys.: Condens. Matter **5**, 4215 (1993).
[7] L. E. Svistov, V. L. Safonov, J. Löw, and H. Benner, J. Phys.: Condens. Matter **6**, 8051 (1994).
[8] A. S. Borovik-Romanov and E. G. Rudashevskiĭ, Zh. Éksp. Teor. Fiz. **47**, 2095 (1964) [Sov. Phys. JETP **20**, 1407 (1964)].
[9] E. A. Turov and V. G. Shavrov, Fiz. Tverd. Tela **7**, 217 (1965) [Sov. Phys. Solid State **7**, 166 (1965)].
[10] I. E. Dikshteĭn, V. V. Tarasenko, and V. G. Shavrov, Zh. Éksp. Teor. Fiz. **67**, 816 (1974) [Sov. Phys. JETP **40**, 404 (1974)].
[11] A. V. Andrienko, L. V. Podd'yakov, and V. L. Safonov, Zh. Éksp. Teor. Fiz. **101**, 1083 (1992) [Sov. Phys. JETP **74**, 579 (1992)].
[12] L. E. Svistov, V. L. Safonov, and K. R. Khachevatskaya, Zh. Éksp. Teor. Fiz. **112**, 564 (1997) [JETP **85**, 307 (1997)].
[13] V. P. Seminozhenko and V. L. Sobolev, Fiz. Nizk. Temp. **8**, 830 (1982) [Sov. J. Low Temp. Phys. **8**, 417 (1982)].
[14] B. Ya. Kotyuzhanskiĭ, L. A. Prozorova, and L. E. Svistov, Zh. Éksp. Teor. Fiz. **92**, 238 (1987) [Sov. Phys. JETP **65**, 134 (1987)].
[15] A. A. Lugovoĭ and E. A. Turov, Zh. Éksp. Teor. Fiz. **94** (10), 358 (1988) [Sov. Phys. JETP **67**, 2161 (1988)].
[16] B. Ya. Kotyuzhanskiĭ, L. A. Prozorova, and L. E. Svistov, Zh. Éksp. Teor. Fiz. **93**, 692 (1986) [Sov. Phys. JETP **66**, 391 (1986)].

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Hydrodynamics of modulated finite-amplitude waves in dispersive media

A. V. Tyurina and G. A. Él'*)

Institute of Terrestrial Magnetism, the Ionosphere, and Radio Wave Propagation, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

(Submitted 16 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1116–1136 (March 1999)

Analytic approaches are developed for integrating the nondiagonalizable Whitham equations for the generation and propagation of nonlinear modulated finite-amplitude waves in dissipationless dispersive media. Natural matching conditions for these equations are stated in a general form analogous to the Gurevich-Pitaevskii conditions for the averaged Korteweg-de Vries equations. Exact relationships between the hydrodynamic quantities on different sides of a dissipationless shock wave, an analog of the shock adiabat in ordinary dissipative hydrodynamics and first proposed on the basis of physical considerations by Gurevich and Meshcherkin,[4] are obtained. The boundaries of a self similar, dissipationless shock wave are determined analytically as a function of the density jump. Some specific examples are considered. © *1999 American Institute of Physics.* [S1063-7761(99)02603-7]

1. INTRODUCTION

It is well known that the applicability of exact analytic methods in the theory of nonlinear dispersive waves is limited to fully integrable small amplitude approximations to the general, ''unintegrable'' systems of equations.[1] These ''unintegrable'' equations arise naturally in plasma physics, the hydrodynamics of waves on liquid surfaces, magnetohydrodynamics, etc. Many of them have a number of general properties which are independent of the physical nature of the phenomena being described and have an highly characteristic structure. Here are some examples:

(a) Gravitational waves on the surface of a shallow liquid:[2]

$$\partial_t h+\partial_x(hu)=0,$$

$$\partial_t u+u\partial_x u+\partial_x h+\varepsilon^2\partial^3_{xxx}h=0. \tag{1}$$

Here h is the depth of the liquid, u is the horizontal component of the velocity, and ε is a small dispersion parameter (equilibrium depth).

(b) Ion-acoustic waves in a two temperature ($T_e \gg T_i$), collisionless plasma:[3–5]

$$\partial_t \rho+\partial_x(\rho u)=0,$$

$$\partial_t u+u\partial_x u+\partial_x \varphi=0,$$

$$\varepsilon^2\partial^2_{xx}\varphi=e^{\varphi}-\rho. \tag{2}$$

Here ρ and u are the ion density and velocity, φ is the electric potential, and ε is the Debye radius.

(c) Magnetosonic waves in a cold plasma moving transverse to a magnetic field:[6–8]

$$\partial_t \rho+\partial_x(\rho u)=0,$$

$$\partial_t u+u\partial_x u+\frac{B}{\rho}\partial_x B=0,$$

$$\varepsilon^2\partial_x\left(\frac{1}{\rho}\partial_x B\right)=B-\rho. \tag{3}$$

Here u and ρ are the plasma velocity and density, B is the magnetic field strength, and $\varepsilon=c/\omega_p$, where ω_p is the electron plasma frequency.

All these systems of equations are of fourth order with respect to the spatial derivatives and have at least four independent conservation laws.[5] When the dispersion parameter ε goes to zero, the systems (1)–(3) transform to the Euler equations of ideal hydrodynamics with different Poisson adiabatic indices γ: $\gamma=2$ in the case of Eqs. (1) and (3) and $\gamma=1$ in the case of Eqs. (2). Systems of equations of this type are usually referred to as dispersion-hydrodynamic.

Another common feature of these systems of equations (1)–(3) is that they lack complete integrability properties. As a consequence, one of the few exact results in this area up to now has been just the existence of travelling single phase solutions $f(kx-\omega t)$ of bounded amplitude. If the wave amplitude exceeds a certain critical value a_{cr}, a solution in the form of a travelling wave ceases to exist;[9] then the very description by a single-stream, dispersion hydrodynamics of the type (1)–(3) also ceases to be valid. In this paper we only study single-stream flows with $a<a_{cr}$.

Physically interesting nonlinear solutions of dispersion-hydrodynamic equations arise in problems concerning the evolution of smooth perturbations (Fig. 1a). In the initial stage of the evolution of such perturbations, nonlinearity plays a dominant role, leading by some time $t=t_c$ to a inversion (breaking) singularity with an infinite derivative (gradient catastrophe).[10] Because of dispersion, nonlinear oscillations are generated near the inversion point for $t>t_c$.[5] The oscillating region expands with time, occupying an ever larger region of space (Fig. 1b). Since the oscillating zone arises in situations such that a shock wave would develop in ordinary dissipative hydrodynamics, the resulting wave structure is usually called a dissipationless (collisionless)

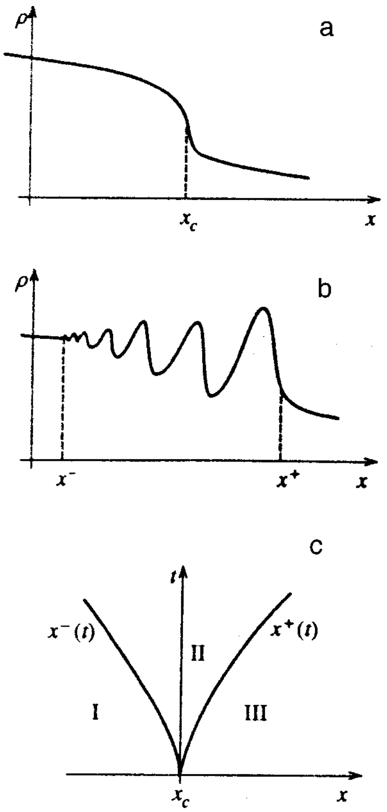

FIG. 1. Inversion (breaking) of a smooth profile and the formation of a dissipationless shock wave: (a) initial data at the inversion time $t=t_c$, (b) oscillatory structure of the shock for $t>t_c$, (c) the xt plane for the dissipationless shock problem.

shock wave.[9] Exact analytic results[11–13] for completely integrable systems, as well as numerical calculations for nonintegrable systems,[4] suggest some plausible general ideas regarding the structure of the solutions describing dissipationless shock waves.

1. The solution of the dispersion-hydrodynamic equations in the neighborhood of a dissipationless shock wave can be represented in the form of a modulated single phase travelling wave.

2. In one of the fronts of a dissipationless shock wave, the wave amplitude goes to zero [the linear front $x^-(t)$] and in the opposite, the wave number [the soliton front $x^+(t)$].

This last condition allows us, in Section 3 of this paper, to formulate natural matching conditions for the solution of the modulation equations in the neighborhood of a dissipationless shock with the solutions of the Euler equations of ideal hydrodynamics on the outside. The assumption that the oscillations are quasistationary in the region of the dissipative shock makes it possible to construct a solution using Whitham’s method,[6,14] a generalization of the WKB method to the theory of nonlinear partial differential equations. This method was first applied to dissipationless shocks by Gurevich and Pitaevskii,[15] who constructed an exact solution of Whitham’s equation that ensured continuous matching of the central flow in the oscillatory region (region II in Fig. 1b) to the smooth external flow outside the dissipationless shock wave (regions I and III in Fig. 1b). They made an analytic study[15] of the case of small initial step perturbations, where the initial disperison-hydrodynamic equations reduce to the Korteweg-de Vries (KdV) equation. More general solutions of the KdV-Whitham equation have been constructed.[16–19] A key point in these papers is the construction of exact analytic solutions is the diagonalizability of the Whitham equations for the case of the KdV equations, i.e., the existence of Riemann invariants.[6,20,22] Similar studies have been done on the nonlinear Schrödinger equation with defocussing.[23–27]

The Gurevich-Pitaevskii stability regimes for the KdV equations have been analyzed numerically.[28] The effect of small dissipative corrections on the structure of the solution to the KdV-Whitham equation has been studied analytically.[29,30]

The existence of a Riemann structure in the Whitham equations is a consequence of the exact integrability of the initial system of equations. We note that, although small dissipative corrections do not destroy the Riemann structure of the modulation system,[29,30] including them leads to new terms on the right which make the system unintegrable (in the sense of the existence of a general local solution). In this paper we study nonintegrable systems without dissipation. For nonintegrable systems of the type (1)–(3) the modulation equations can be written by formal use of Whitham’s recipe, averaging the required number of conservation laws over the period of the single-phase travelling solution $f(kx-\omega t)$. The desired equations describe the slow variation in the local integrals of motion. (The number of such equations equals the differential order of the system with respect to the spatial derivatives.) Slow variation means that the characteristic length of the oscillations is small compared to the size of the oscillatory region. Whitham’s recipe has been applied to the systems of Eqs. (2) and (3) in a study[5] of the asymptotic regimes of a self-similar dissipationless shock wave near its boundaries, the linear $x^-(t)$ and soliton $x^+(t)$ fronts (see Fig. 1c), whose positions are assumed known.

In this paper we develop analytic approaches to integrating the nondiagonalizable Whitham modulation equations. The principal method for these studies is the method of characteristics, whose applicability does not depend on the diagonalizability properties. Since the problem under study is nonlinear, the behavior of the characteristics is determined by the specific solution being studied, as well as by the structure of the coefficients in the system of equations. In Section 4 we shall show that in the family of solutions for the problem of generating a dissipationless shock wave (we shall call it the inversion problem) along some curves in the xt plane (the fronts of this shock), the general Whitham equations become degenerate, so that the modulation system acquires a local Riemann structure. In addition, it turns out that here the local Riemann invariants transform to solutions of the inversion problem along the characteristics through the ‘‘Whitham’’ zone, just as happens with the ordinary Riemann variables in the diagonalizable case. This property makes it possible to:

(1) obtain a relationship between the hydrodynamic variables on both sides of a dissipationless shock, the analog of the shock adiabat in ordinary dissipative hydrodynamics,

(2) formulate and prove a general theorem about a contact flow in dissipationless dispersive hydrodynamics, and

(3) obtain exact analytic expressions for the boundaries of a self-similar dissipationless shock wave in the form of a function of the density jump at its boundaries (Section 5).

The general construction of a study of the nondiagonalizable Whitham equations on a family of solutions to the inversion problem is illustrated with concrete examples, including: a KdV modulation system in hydrodynamic form, which does not use the diagonalizability of the KdV-Whitham equations (Section 6), and a modulation system for nonlinear ion-acoustic waves corresponding to Eqs. (2) (Section 7). Exact analytic expressions are found for the boundaries of a self-similar ion-acoustic dissipationless shock. These formulas are in complete agreement with the results of direct numerical calculations.[4]

2. HYDRODYNAMIC FORM OF THE MODULATION EQUATIONS

It has been shown[5] that it is convenient to present the averaged equations which do not have Riemann invariants in the form of a hydrodynamic system of equations for the parameters of the system motion and excited waves: the average density $\overline{\rho}\equiv n$, the average flow velocity $\overline{u}\equiv v$, the energy density in the oscillations $\overline{\rho u}-\overline{\rho}\,\overline{u}=A^2$, and the density of the waves (of the wave number) k. Averaging is done on a family of single-phase travelling solutions of the form $f(x-Ut)$, where U is the phase velocity:

$$\begin{gathered}\frac{\partial n}{\partial t}+\frac{\partial(nv)}{\partial x}+\frac{\partial A^2}{\partial x}=0,\\ \frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}+\frac{1}{n}\frac{\partial P(n)}{\partial x}+\frac{\partial E^2}{\partial x}=0,\\ \frac{\partial A^2}{\partial t}+\frac{\partial(VA^2)}{\partial x}+A^2\frac{\partial v}{\partial x}+E^2\frac{\partial n}{\partial x}=0,\\ \frac{\partial k}{\partial t}+\frac{\partial(kU)}{\partial x}=0.\end{gathered}\tag{4}$$

Here

$$\begin{gathered}E^2=E^2(n,k,A^2),\quad V=v+W(n,k,A^2),\\ U=v+U_1(n,k,A^2)\end{gathered}\tag{5}$$

is a function whose specific form is determined by averaging the conservation laws for the initial equations of dispersive hydrodynamics over the fast oscillations. (For details, see Ref. 5.) In the following, only a few general properties of these functions, which determine the asymptotic properties of the system of Eqs. (4), will be important.

The nonstationary flow described by Eqs. (4) can be divided arbitrarily into two parts: a slow hydrodynamic motion characterized by the variables n and v ($P(n)$ is the ordinary hydrodynamic pressure), and quasistationary nonlinear oscillations with an energy density A^2 and wave number k. These equations, of course, are not independent. The terms $\partial A^2/\partial x$ and $\partial E^2/\partial x$ in the first two equations describe the effect of the oscillations on the density and velocity of the hydrodynamic flow. On the other hand, the terms $A^2\partial v/\partial x$ and $E^2\partial n/\partial x$ in the energy transport equation for the nonlinear waves characterize the effect of changes in the hydrodynamic flow parameters on the oscillations.

We now consider some formal limiting cases.

2.1. Zero amplitude waves, $A^2=0$

The analysis given in Ref. 5 shows that for $A\to 0$,

$$E^2\propto A^2,\quad V\to\partial\omega_0/\partial k,\quad U\to\omega_0/k.\tag{6}$$

Here $\omega=\omega_0(k,n,v)$ is the usual linear dispersion relation, in which, however, the constants ρ_0 and u_0, relative to which the linearization

$$\rho=\rho_0+\rho_1e^{i(kx-\omega t)},\quad u=u_0+u_1e^{i(kx-\omega t)}$$

is carried out, are replaced by the slowly varying parameters n,v of the average flow. The quantities V and U, therefore, are converted into the ordinary linear group and phase velocities, which now depend on the hydrodynamic variables n and v, as well as on k. Equations (4)–(6) imply that the system of Eqs. (4) has the exact reduction

$$A^2=0,\quad \frac{\partial n}{\partial t}+\frac{\partial(nv)}{\partial x}=0,$$

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}+\frac{1}{n}\frac{\partial P(n)}{\partial x}=0,\quad \frac{\partial k}{\partial t}+\frac{\partial\omega_0(k,n,v)}{\partial x}=0.\tag{7}$$

The first two equations of the system of Eqs. (7) are the equations of Euler hydrodynamics and can be obtained from the initial system, which contains high order spatial derivatives, by letting the dispersion parameter approach zero directly. The third equation is the conservation equation for the number of waves; in the present case a zero amplitude is purely formal. Nevertheless, this equation contains important information on the asymptotic behavior of the characteristics of the system of Eqs. (4).

2.2. Solitons, $k=0$

We stipulate at once that in this paper by ''solitons'' we mean the formal solutions of the equations of dispersive hydrodynamics in the form of solitary waves. Of course, in the general case these solutions do not have many properties in common with the real solitons in integrable systems, in particular, a collision of ''nonintegrable'' solitons is not elastic and is accompanied by radiation.[31]. As $k\to 0$ we have the following asymptotes:[5]

$$E^2\propto A^2\propto k,\quad V-U\propto k.$$

$\zeta=A^2/k=O(1)$ has the significance of the average energy of a single vibration and remains finite as k approaches zero. The equation for ζ, which follows from Eq. (4), has the form

$$\frac{\partial\zeta}{\partial t}+U\frac{\partial\zeta}{\partial x}+\zeta\left[\frac{\partial v}{\partial x}+f(k,n,\zeta)\frac{\partial n}{\partial x}\right]+\frac{1}{n}\frac{\partial}{\partial x}[k\zeta(V-U)]=0, \tag{8}$$

where

$$f(k,n,\zeta)=E^2/A^2=O(1).$$

Thus, the system of Eqs. (4) has yet another exact reduction:

$$k=0,\quad \frac{\partial n}{\partial t}+\frac{\partial}{\partial x}(nv)=0,$$

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}+\frac{1}{n}\frac{\partial P(n)}{\partial x}=0,$$

$$\frac{\partial\zeta}{\partial t}+U_s(n,v,\zeta)\frac{\partial\zeta}{\partial x}+\zeta\left[\frac{\partial v}{\partial x}+f(0,n,\zeta)\frac{\partial n}{\partial x}\right]=0, \tag{9}$$

where

$$U_s(n,v,\zeta)=\lim_{k\to 0}U. \tag{10}$$

It should be noted that, as opposed to the complete system of Eqs. (4), which is of fourth order, the reduced system of Eqs. (7) and (9) consists of three equations. This is evidently a consequence of the merger of two families of characteristics in the linear and soliton limits. In the first case, a multiple velocity coincides with the group velocity $\partial\omega_0/\partial k$ and in the second, to the soliton speed U_s. We emphasize that the system of Eqs. (7) and (9) has at least two Riemann invariants; these are the invariants of ideal Euler hydrodynamics,[10]

$$\partial_t J_\pm+V_\pm\partial_x J_\pm=0, \tag{11}$$

where

$$J_\pm(n,v)=v\pm\int\frac{c_s(n)}{n}dn \tag{12}$$

are the Riemann invariants,

$$c_s^2(n)=dP(n)/dn, \tag{13}$$

where c_s is the sound speed, and

$$V_\pm(n,v)=v\pm c_s \tag{14}$$

are the characteristic velocities. Setting J_+ or J_- constant [$J_\pm=\text{const}$ is the exact solution of the system of Eqs. (11)], it is always possible to diagonalize the remaining system of two equations. It is important that these properties do not depend on the diagonalizability of the complete modulation system. The fact that the Euler equations ''split off'' in the system of Eqs. (4) for $A^2=0$ and $k=0$ means that it can be described using a dissipationless shock wave which joins the two Euler regimes.

3. NATURAL MATCHING CONDITIONS FOR THE CHARACTERISTICS IN THE INVERSION PROBLEM

Let the solution of the Euler equations with inverted monotonic initial conditions on the right (Fig. 1a),

$$\rho(x,0)=\rho_0(x),\quad u(x,0)=u_0(x), \tag{15}$$

have the form

$$\rho=\rho_e(x,t),\quad u=u_e(x,t). \tag{16}$$

Then, for $t>t_0$, where t_0 is the time of inversion (without limiting the generality, we assume $t_0=0$), in the xt plane of the solutions of the complete system including the dispersion terms, the oscillatory region is divided by a dissipationless shock (Fig. 1b and 1c) and bounded by the curves $x^-(t)$, its trailing edge, and $x^+(t)$ its leading edge. The behavior of the averaged quantities in this region is governed by the Whitham equations (4). Outside the region of the dissipationless shock the Euler equations for the variables ρ and u are valid, as before, and have the same Riemann form as Eq. (11):

$$\partial_t r_\pm+W_\pm\partial_x r_\pm=0, \tag{17}$$

where

$$r_\pm=J_\pm(\rho,u),\quad W_\pm=V_\pm(\rho,u). \tag{18}$$

We now formulate the boundary conditions for the modulation system (4) which will allow us to match the average flow in the Whitham region to the smooth flow (9) at the (previously unknown) boundaries $x^\pm(t)$. We note, first of all, that the number of characteristics in the outer and inner regions is different, so matching can take place only at points where pairs of families of the characteristics of the Whitham system merge, i.e., for $A^2=0$ [Eq. (7)] and for $k=0$ [Eq. (9)]. From the beginning we limit ourselves to the case of negative dispersion, where waves are generated ($A^2\to 0$) at the trailing edge and solitons are created ($k\to 0$) at the leading edge. (See Refs. 15, 4, and 5).

The $x^\pm(t)$ curves separating the two analytically different regimes are characteristics (in our case the envelops of multiple characteristics, i.e., caustics) on the family of solutions under study, in accordance with the properties of quasilinear hyperbolic systems.[32] Let us examine the behavior of the characteristics in the Whitham region in more detail. The characteristic equations of the modulation system (4) have the form

$$\frac{dx}{dt}=V_i(n,v,k,A^2),\quad i=1,2,3,4, \tag{19}$$

where the V_i are the characteristic (group) velocities or eigenvalues of the matrix of the coefficients of the modulation system (4). We restrict ourselves to the case of real and, generally speaking, different V_i, which is the hyperbolicity condition. Let $V_1\geqslant V_2\geqslant V_3\geqslant V_4$. Then the continuous matching conditions for the characteristics on the caustic surfaces take the following form (see Fig. 2):

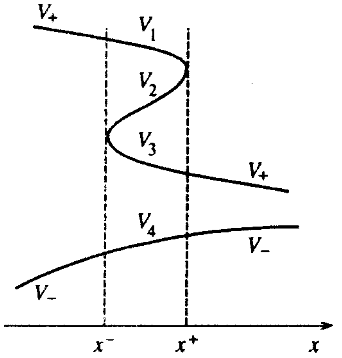

FIG. 2. The behavior of the characteristic velocities in a dissipationless shock.

$$V_2(n,v)|_{x^-}=V_3(n,v)|_{x^-}=\frac{dx^-}{dt},$$

$$V_1(n,v)|_{x^-}=W_+(\rho^-,u^-),$$

$$V_4(n,v)|_{x^-}=W_-(\rho^-,u^-) \qquad (20)$$

for $x=x^-(t)$; and

$$V_2(n,v)|_{x^+}=V_1(n,v)|_{x^+}=\frac{dx^+}{dt},$$

$$V_3(n,v)|_{x^+}=W_+(\rho^+,u^+),$$

$$V_4(n,v)|_{x^+}=W_-(\rho^+,u^+) \qquad (21)$$

for $x=x^+(t)$, where

$$\rho^\pm\equiv\rho_e(x^\pm,t), \quad u^\pm\equiv u_e(x^\pm,t).$$

We shall refer to the problem stated in Eqs. (4), (20), and (21) as the *inversion problem*. Note that the inversion problem is not posed for the initial system, but for its Whitham equations.

It is impossible not to notice the similarity between the qualitative behavior of the characteristic velocities in the inversion problem (Fig. 2) and the behavior of the Riemann invariants in the solution of the Gurevich–Pitaevskii problem for fully integrable systems.[4,25,27] The qualitative variation of the characteristics in the xt plane corresponding to the conditions (20) and (21) is shown in Fig. 3.

We now introduce yet another form of the matching conditions (20) and (21). The analysis of this section shows that for $A^2=0$ and for $k=0$ the modulation system is degenerate and the Riemann invariants $J_\pm(n,v)$ appear for it. For the problem under consideration, this structure has a local (in the (x,t) plane) character, since the corresponding couplings among the quantities are realized in the solution only along the boundaries of the dissipationless shock. Thus, we shall refer to the Riemann variables $J_\pm(n,v)$ as local Riemann invariants.

Conditions (20) and (21) can then be rewritten as matching conditions for the local Riemann invariants of a modulation system with the invariants of the outer Eulerian hydrodynamics $r_\pm\equiv J_\pm(\rho,u)$ at the fronts of the dissipationless shock.

For $x=x^-(t)$

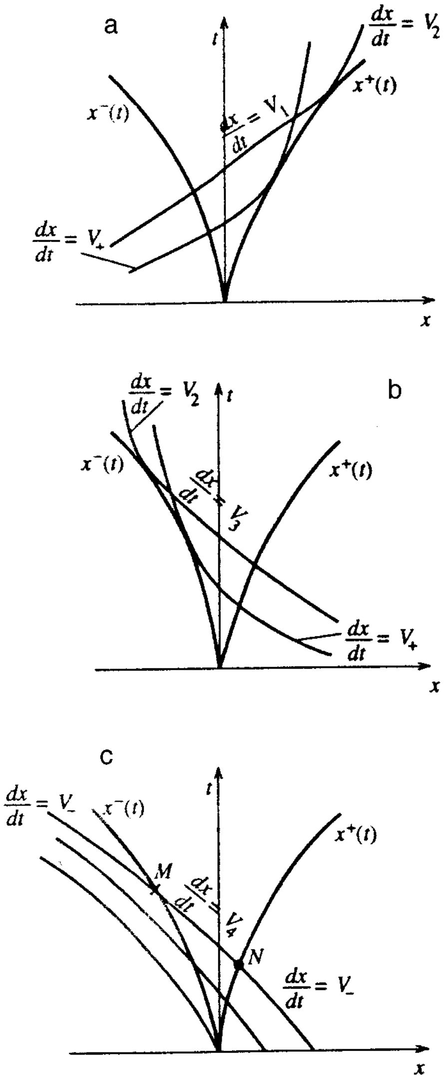

FIG. 3. The behavior of the characteristics in a dissipationless shock: (a) the families $dx/dt=V_1$ and $dx/dt=V_2$, (b) the families $dx/dt=V_3$ and $dx/dt=V_2$, and (c) the family $dx/dt=V_4$.

$$A^2=0, \quad J_+|_{x^-}=r_+|_{x^-}, \quad J_-|_{x^-}=r_-|_{x^-}. \qquad (22)$$

For $x=x^+(t)$

$$k=0, \quad J_+|_{x^+}=r_+|_{x^+}, \quad J_-|_{x^+}=r_-|_{x^+}. \qquad (23)$$

We emphasize once again that $J_\pm=J_\pm(n,v)$ are the ''inner'' (Whitham) variables, while $r_\pm=r_\pm(\rho,u)$ are the ''outer'' (Euler) variables.

Finally, using Eqs. (12) and (18), we introduce the concept of the conditions (22) and (23) in terms of a continuous matching of the average flow density and velocity.

For $x=x^-(t)$

$$A^2=0, \quad n(x^-,t)=\rho^-, \quad v(x^-,t)=u^-. \qquad (24)$$

For $x=x^+(t)$

$$k=0, \quad n(x^+,t)=\rho^+, \quad v(x^+,t)=u^+. \qquad (25)$$

In conclusion, we note that satisfying the matching conditions for the average hydrodynamic flow does not guarantee continuity of the shock wave parameters and their derivatives at the fronts. This latter point does not contradict the physical statement of the problem; mathematically the existence of such discontinuities was possible because the equations are hyperbolic.[32] Discontinuities develop at the boundaries $x^{\pm}$:[5,15] the derivatives $\partial A^2/\partial x$ and $\partial k/\partial x$ go to infinity at the soliton front (this is the so-called singular discontinuity[15]) while *k*, itself (at this point the conservation of the number of waves is violated), and, in general, the derivative $\partial A^2/\partial x$ have a discontinuity at the generation point. It should also be kept in mind that the existence of singularities in the solution for the modulation equations is a consequence of the asymptotic character of the Whitham method; in the exact solution (if it exists), there are no singularities.

4. TRANSPORT OF THE LOCAL RIEMANN INVARIANTS THROUGH THE WHITHAM ZONE. JUMP CONDITIONS ACROSS A DISSIPATIONLESS SHOCK WAVE AND A THEOREM ON CONTACT FLOW

Let us consider the family of characteristics corresponding to the velocity V_4 (Fig. 3c). We select a characteristic which intersects the boundary of the dissipationless shock at the points *M* and *N*. Then, as we shall show, the condition

$$J_-(M)=J_-(N) \tag{26}$$

is satisfied for this characteristic; that is, in the solutions to the inversion problem, the local Riemann invariant is carried through the dissipationless shock wave region along ''its own'' characteristic. For diagonal systems the condition (26) is an obvious consequence of the constancy of the corresponding Riemann invariant along the characteristics. In general (the nondiagonalizable case), this relationship is nontrivial and must be justified.

Let us consider some quantity *R* determined by an ordinary differential equation along an arbitrary smooth curve Γ, $x=x_\Gamma(t)$, joining the points *M* and *N*:

$$D_\Gamma R_4=f(\Gamma) \tag{27}$$

with the boundary conditions

$$R_4(M)=J_-(M). \tag{28}$$

The function

$$f(\Gamma)=\sum_{i=1}^{4} l_i^4(\mathbf{y})|_\Gamma D_\Gamma u_i \tag{29}$$

is specified along Γ in the solution of the inversion problem, $y_i(x,t)$, which we shall regard as solved here. The operator D_Γ denotes differentiation along Γ; $\mathbf{l}^{(4)}$ is the left eigenvector of the matrix of coefficients of the modulation equations (1) corresponding to the velocity V_4, $\mathbf{y}=(n,v,k,A^2)$. We emphasize that the characteristics $dx/dt=V_4$ on both fronts are matched with the Euler characteristics from one and the same family $dx/dt=W_-$ [See Eqs. (20) and (21)].

It is easy to see that for $A^2=0$, Eq. (27) has a general integral that does not depend on the curve Γ and this integral is nothing other than the Riemann invariant $J_-(n,v)$.[10,32] In fact, for $A^2=0$, because of the reduction (7), the vector $\mathbf{l}^{(4)}$ transforms into the left eigenvector of the matrix of coefficients of the Euler system corresponding to the velocity W_- (14). Thus, the function $f(\Gamma)$ in this limit coincides with the characteristic form for Euler hydrodynamics and is the total derivative D_Γ, independently of the curve Γ.[32,33] A similar statement holds for $k=0$. Thus,

$$R_4|_{A^2=0}=R(n,v)=J_-(n,v)+C^1, \tag{30}$$

$$R_4|_{k=0}=R(n,v)=J_-(n,v)+C^2 \tag{31}$$

are the integrals of Eq. (27) on the family of solutions of interest to us. It is important that the integrals (30) and (31) are, themselves, local and their values are independent of the global properties of the solution. Thus, the relationship between the constants of integration C^1 and C^2 is easily established using the natural requirement that the local integrals (30) and (31) coincide as the width of the dissipationless shock approaches zero. In this case, it is evident that $C^1=C^2\equiv C$ and Eq. (28) implies that $C=0$. Note that a situation will arise below in which the relationship between the constants of integration is not so trivial. Thus, $R=R(\Gamma)$ is a continuous function of the curve Γ and takes the values

$$R_4(M)=J_-(M) \text{ and } R_4(N)=J_-(N) \tag{32}$$

at its ends.

Now let the curve Γ be the characteristic determined by the equation $dx/dt=V_4$ on the solution of this problem. Then the expression on the right hand side of Eq. (27) is the same as the characteristic form[14,32,33] of the modulation system (4) corresponding to the velocity V_4, so that

$$D_4R_4=0, \tag{33}$$

where

$$D_4=\frac{\partial}{\partial t}+V_4(x,t)\frac{\partial}{\partial x}.$$

Now Eq. (26) follows immediately from Eq. (32). We emphasize that, despite the external similarity of Eq. (33) to the Riemann form of the equations for diagonalizable quasilinear systems, $R_4(x,t)$ is not, in general, a ''real'' Riemann invariant, since D_4 is a linear differential operator specified on the solution of the inversion problem. Of course, for diagonalizable systems the dependence on the solution vanishes and R_4 becomes a local function of the dependent variables u_i, i.e., is transformed into an ordinary Riemann invariant. The consequence of Eq. (26) and the matching conditions (22) and (23) is a substantial restriction imposed on the invariants of the outer hydrodynamic flow adjacent to the dissipationless shock, specifically,

$$r_-(M)=r_-(N). \tag{34}$$

Note that in the above discussion it has been assumed that the integrating factor ensuring the existence of the total derivative D_Γ on the right hand side of Eq. (28) is equal to unity. This is true for Euler hydrodynamics and, therefore, is satisfied near the boundaries of the dissipationless shock for the Whitham characteristic $dx/dt=V_4$ which joins the two Euler regimes. Furthermore, it is easy to show that this statement about the transfer of a local Riemann invariant along a characteristic (through the Whitham zone) can also be gen-

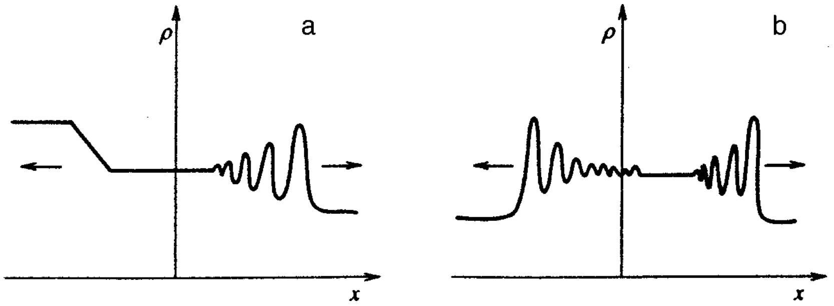

FIG. 4. Typical decay patterns for a discontinuity in dispersive hydrodynamics: (a) rarefaction wave on the left, dissipationless shock on the right, (b) two dissipationless shocks travelling in opposite directions.

eralized to the case of an integrating factor that differs from unity in one or several regions where a local invariant exists. As a result, the rule about the transfer of a local invariant is valid, not only for the characteristics $dx/dt=V_4$, but also for characteristics from any other family on which a local Riemann structure exists in several regions of xt. This is used in the next section, where the transfer of data along the characteristics $dx/dt=V_1$ and $dx/dt=V_3$ is exploited. Recall that the condition (34) applies to an arbitrarily chosen characteristic from the family that is being studied. If, however, the flow to the right or left of the dissipationless shock wave is a simple wave or constant, then the invariant r_- is constant throughout this entire region and, therefore, along the boundary. Then, Eq. (34) yields an equation relating the hydrodynamic variables along both sides of such a dissipationless shock:

$$r_-(\rho^-,u^-)=r_-(\rho^+,u^+). \tag{35}$$

Therefore, the value of the Riemann invariant for ideal hydrodynamics is transferred through the Whitham zone, even though the Whitham equations, themselves, have no Riemann invariants.

Condition (35) plays the same role in dispersive hydrodynamics as the shock adiabat in ordinary dissipative hydrodynamics.[10] Note that Eq. (35) applies to waves travelling to the right. Analogous results can be obtained for leftward propagating waves; in that case the invariant r_+ will be conserved on passing through the dissipationless shock. Equation (35) was first proposed[4] on the basis of physical considerations.

Equation (35) can also be used to formulate a general theorem on contact flows[10,32,34] in dispersive hydrodynamics. If the flow along one side of a dissipationless shock is constant ($\rho=\rho_0$, $u=u_0$), then the flow along the other side is a simple wave. In fact, one of the families of characteristics carries the value of ''its own'' Riemann invariant through the dissipationless shock, and this means that the flow on the other side is a simple wave (or as a special case, is constant).

5. DETERMINATION OF THE COORDINATES OF THE FRONTS OF THE DISSIPATIONLESS SHOCK IN THE SELF-SIMILAR SOLUTIONS OF THE WHITHAM EQUATIONS

We now turn to one of the most important nonlinear problems of hydrodynamics, the decay of an initial discontinuity. The decay of a discontinuity produces two waves on both sides of it, a dissipative shock and a rarefaction wave, in different combinations depending on the initial data.[4,26] A plateau region in which the flow is constant develops between these waves.1) Figure 4 shows typical pictures of the breakup of a discontinuity. The fact that an arbitrary discontinuity, parametrized by the four constants $\rho^\pm$ and $u^\pm$, cannot be ''matched'' in the case of a single dissipationless shock wave follows from the existence of the additional relation (35), which reduces the number of parameters to three. The desired matching is achieved by introducing two waves travelling in opposite directions. The hydrodynamic quantities on both sides of the wave travelling to the right are related by Eq. (35). For the leftward travelling wave, as noted above, we have the analogous relation

$$r_+(\rho^-,u^-)=r_+(\rho^+,u^+). \tag{36}$$

Note that one of Eqs. (35) or (36) is automatically satisfied even if one of the waves that develops is a rarefaction wave; in this case it is a consequence of the constancy of the corresponding Riemann invariant of the equations of ideal hydrodynamics. The time evolution of a dissipationless shock obeys the self-similar ($\tau=x/t$) solutions of the corresponding modulation system. We shall limit ourselves to considering dissipationless shocks travelling to the right. It is not necessary to be interested in the exact structure of the solution in the region of the dissipationless shock, limiting oneself to just the Euler equations, but introducing a discontinuity into them that occupies a finite region in the self-similar variables from τ^- to τ^+.[4] τ^- and τ^+ have been determined[4] as functions of the density jump at a discontinuity for an ion-acoustic dissipationless shock. For an analytic solution of this problem, it turns out that it is enough to know the local Riemann structure and the asymptotic behavior of the characteristics of the modulation system near the fronts of the dissipationless shock. Let, as before, the constant density and velocity in front of the shock be ρ^+ and u^+ and behind it, ρ^- and u^-. Without loss of generality, we take $u^+=0$, which corresponds to transforming to the moving coordinate system. First of all, we note that a self-similar dissipationless shock is described by solutions of the modulation equations in the form of centered simple waves. This means that the family of characteristics corresponding to the velocity V_2 forms a fan of straight lines with its center at the point (0,0) and the boundaries of the dissipationless shock, the straight lines $x=\tau^-t$ and $x=\tau^+t$, represent multiple characteristics (Fig. 5).

The rectilinear characteristics of the outer flow, which correspond to the velocity V_+, ''punch through'' the front of the dissipationless shock. Then they curve and at the op-

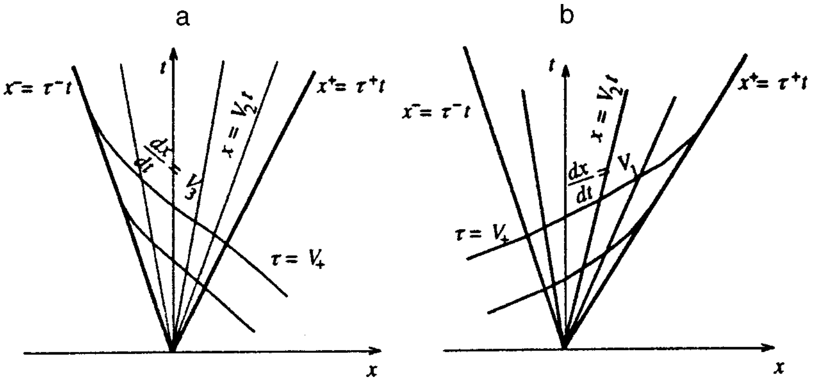

FIG. 5. Characteristics in a centered simple dissipationless shock: (a) the families $dx/dt = V_2$ and $dx/dt = V_3$, (b) the families $dx/dt = V_2$ and $dx/dt = V_1$.

posite front they merge with the characteristics from the centered family. An important difference between this special case of simple wave and the general situation described in Section 2, where the boundaries are not characteristics but are envelopes with a multiple characteristic direction at every point. (See Figs. 3a and 3b.) Figure 5 does not show the previously described family of characteristics $dx/dt = V_4$, since there are no fundamental changes from the general case (Fig. 3c). (Of course, the characteristics in the outer region become rectilinear.)

For concreteness, we now examine the trailing edge of the shock, $x^- = \tau^- t$, at which the amplitude goes to zero ($A^2 = 0$) and the matching conditions (24) are satisfied ($v^- = u^-$, $n^- = \rho^-$). As shown in Section 2, the multiple characteristic velocity in this case coincides with the linear group velocity of the generated wave, i.e.,

$$\omega_0'(k_0, u^-, \rho^-) = \tau^-, \tag{37}$$

where k_0 is the wave number from which nonlinear generation takes place at the trailing edge.[5] Therefore, the problem of determining the coordinate τ_- reduces to finding the function $k_0(u^-, \rho^-, \rho^+)$. In order to find k_0 it is necessary to determine the local Riemann invariant corresponding to the multiple characteristic velocity for $A^2 = 0$, and then equate it to the outer invariant r_+ arriving with the Euler characteristics at the leading edge of the dissipationless shock. We now describe this process in more detail.

The jump condition (35) takes the form $r_-(\rho^-, u^-) = r_+(\rho^+, 0)$ and makes it possible to determine u^- as a function of ρ^+ and ρ^-. As a result, we have only two nontrivial parameters ρ^+ and ρ^- characterizing the self-similar problem. We now consider the system of Eqs. (7), in which we set the hydrodynamic invariant $J_-(n, v)$ constant [this invariant is matched with $r_-(\rho^+, \rho^-)$ in the dissipationless shock) (see Eq. (16)], i.e.,

$$J_-(n, v) = r_-(\rho^+, \rho^-) = \text{const.} \tag{38}$$

The condition (38) automatically ensures matching at the trailing edge and determines the function $n(v)$ in the system of Eqs. (7). The system, itself, then takes the form

$$\partial_t J_+(n) + V_+(n) \partial_x J_+(n) = 0,$$

$$\partial_t k + \partial_x \omega_0(k, n) = 0. \tag{39}$$

Since Eq. (39) is a second order hyperbolic system, it can always be diagonalized.[32] One of its Riemann invariants, $J_+(n) \equiv J_+(n, v(n))$, has already been pointed out, while the second necessarily exists and can be found in the form $J_0 = J_0(k, n)$. This invariant, according to Section 4, is transported through the Whitham zone along the characteristic $dx/dt = V_3$ (see Fig. 5a), i.e.,

$$J_0(k_0, \rho^-) = r_+(u^+ = 0, \rho^+). \tag{40}$$

Equation (40) yields the desired function $k_0(\rho^+, \rho^-)$. Note that the Riemann invariant is an arbitrary function $f(J_0)$, so in order to determine a specific dependence $J_0(k, n)$ in Eq. (40), it is necessary to impose the natural condition

$$J_0(k = 0, n) = J_+(n) \tag{41}$$

on J_0, which ensures that these invariants are equal when the width of the dissipationless shock goes to zero. (Recall that $k = 0$ at the leading edge, while J_0 is calculated at the trailing edge, so examining the limit $J_0(k \to 0)$ is equivalent to calculating J_0 for a dissipationless shock of infinitely small intensity).

The coordinate τ^+ of the leading edge can be calculated in completely analogous fashion. To do this it is necessary to determine the Riemann invariants of the system of Eqs. (9) with the integral (38), and then to equate the local invariant $J_s(n, \zeta)$, corresponding to the velocity U_s, to the outer invariant $r_+(u^-, \rho^-)$. (Transport takes place along the characteristic $dx/dt = V_1$.) The condition analogous to Eq. (41) for a dissipationless shock of zero intensity has the form

$$J_s(n, \zeta = 0) = J_+(n). \tag{42}$$

As a result, we obtain the dependence

$$\zeta_s = \zeta_s(\rho^-, \rho^+), \tag{43}$$

where ζ_s is the value of ζ at the soliton front τ^+.

The self-similar coordinate of the soliton front [see Eq. (21)] is determined by the multiple characteristic of the velocity for $k = 0$, which coincides with the soliton velocity U_s [Eq. (10) (recall that $v(\tau^+ = u^+ = 0)$]:

$$\tau^+ = U_s(\rho^+, 0, \zeta_s) \equiv U_s(\rho^-, \rho^+). \tag{44}$$

We now proceed to consider some concrete examples.

6. THE KORTEWEG-DE VRIES EQUATION

We shall consider the KdV equation in the form

$$\partial_t u + u \partial_x u + \partial_{xxx}^3 u = 0. \tag{45}$$

The modulation system for the KdV equation has a Riemann invariant.[6,20] The corresponding problem of the decay of the discontinuity

$$u(0,x)=0, \quad x>0, \quad u(0,x)=1, \quad x\leqslant 0, \tag{46}$$

has been fully studied.[15] We shall calculate the coordinates of the fronts of the dissipationless shocks, using the general technique described in this section (i.e., without using the Riemann form for the complete modulation system). The matching conditions analogous to the conditions (24) and (25) for the average (averaged over the oscillations) quantity $\eta\equiv\overline{u}$ take the following form:

$$A^2=0, \quad \eta=1 \tag{47}$$

for $\tau=\tau^-$, and

$$k=0, \quad \eta=0 \tag{48}$$

for $\tau=\tau^+$. Equation (35) can be regarded as automatically satisfied, since the outer flow in the KdV hydrodynamics is necessarily a simple wave. Let us write down the dispersion relation for the nonlinear KdV waves against the evolving hydrodynamic background $\eta(x,t)$:

$$\omega_0(k,\eta)=k\eta-k^3. \tag{49}$$

Equation (49) makes it possible to obtain a reduction of the modulation system for the linear front that is analogous to Eq. (7) at once:

$$A^2=0, \quad \partial_t\eta+\eta\partial_x\eta=0, \quad \partial_t k+\partial_x(k\eta-k^3)=0. \tag{50}$$

It is easy to find the desired local Riemann invariant from Eq. (50):

$$J_0(k,\eta)=\eta-3k^2/2. \tag{51}$$

The role of the hydrodynamic invariant J_+ is played by the variable η, itself. [See Eq. (39)]. Note that the condition (41) is satisfied, since $J_0(0,\eta)=\eta$. Then,we have the equation $J_0(k_0,1)=0$ for k_0, which gives $k_0^2=2/3$. [See Eq. (40)]. From this we find at once that [see Eq. (37)]

$$\tau^-=\frac{\partial\omega_0}{\partial k}\left(\sqrt{\frac{2}{3}},1\right)=-1. \tag{52}$$

For the leading (soliton) front we have the following reduction of the KdV modulation system:[5]

$$k=0, \quad \partial_t\eta+\eta\partial_x\eta=0, \quad \partial_t\zeta+U_s(\eta,\zeta)\partial_x\zeta+\zeta\partial_x\eta=0, \tag{53}$$

where

$$U_s(\eta,\zeta)=\eta+2\zeta^{2/3}/3.$$

We find the local Riemann invariant that is constant along the soliton front from Eq. (53):

$$J_s(\eta,\zeta)=\eta+\zeta^{2/3}. \tag{54}$$

Equation (42) is satisfied, since $J_s(\eta,0)=\eta$. Transport of the local Riemann invariant along the family of characteristics $dx/dt=V_1$ yields the equation $J_s(0,\zeta_s)=1$, which implies that $\zeta_s=1$. Then [see Eq. (44)],

$$\tau^+=U_s(0,\zeta_s)=2/3. \tag{55}$$

The values of $\tau^\pm$ found here correspond to the results of Ref. 15.

7. NONLINEAR ION-ACOUSTIC WAVES

Let us consider the system of Eqs. (2) describing a nonlinear flow in a nonisothermal plasma. The modulation system of Eqs. (2) has been found in Ref. 5. As noted above, the absence of Riemann invariants for this system means that it is impossible to construct an exact solution of the inversion problem by known methods. Nevertheless, the existence of a local Riemann structure, which ensures generation of a dissipationless shock wave, makes it possible to calculate analytically the coordinates of the fronts of the dissipationless shock.

7.1 Trailing edge of an ion-acoustic dissipationless shock

In order to determine the trailing edge of a dissipationless shock wave it is sufficient to know only the linear dispersion relation, which in this case has the form[5]

$$\omega_0(k,v,n)=k(v+\gamma^{1/2}), \tag{56}$$

where

$$\gamma=\frac{1}{1+k^2/n}. \tag{57}$$

Then we have a system of three equations when $A^2=0$ for determining the local Riemann invariant:

$$\partial_t J_\pm+V_\pm\partial_x J_\pm=0,$$
$$\partial_t k+\partial_x(k(v+\gamma^{1/2}))=0, \tag{58}$$

where

$$J_\pm(n,v)=v\pm\ln n, \quad V_\pm=v\pm 1. \tag{59}$$

The conditions (22) and (35) fix the invariant J_-:

$$v-\ln n=u^- -\ln\rho^- =-\ln\rho^+. \tag{60}$$

Equation (60) implies that

$$u^-=\ln\Delta, \quad v=\ln(n/\rho^+), \tag{61}$$

where

$$\Delta=\rho^-/\rho^+ \tag{62}$$

is the density jump, which is now the only parameter of the problem.

Given Eq. (61), the system of Eqs. (58) now takes the form

$$\partial_t J_+ +[\ln(n/\rho^+)+1]\partial_x J_+=0,$$
$$\partial_t k+\partial_x[k(\ln(n/\rho^+)+\gamma^{1/2})]=0, \tag{63}$$

where

$$J_+(n)=2\ln n-\ln\rho^+.$$

Diagonalizing the system (63) leads to the following Riemann invariants:

$$J_1=J_+(n), \quad J_2=\ln n+\ln\gamma+\frac{2}{1+\gamma^{1/2}}. \tag{64}$$

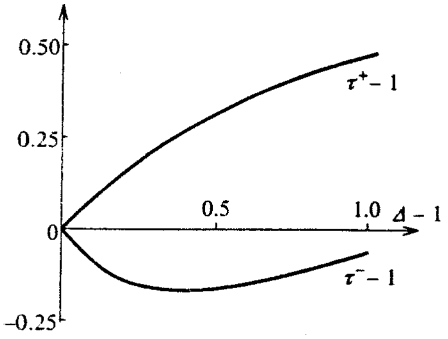

FIG. 6. Self-similar boundaries of an ion-acoustic dissipationless shock, τ^- and τ^+, as functions of the density jump $\Delta=\rho^-/\rho^+$.

The local Riemann invariant $J_0(k,n)=f[J_2(k,n)]$ is found using the condition (41), which in this case takes the form

$$f(J_2(0,n))=J_+(n). \tag{65}$$

Equations (64) and (65) imply that

$$f(x)=2(x-1)-\ln\rho^+. \tag{66}$$

Then

$$J_0(k,n)=2\left(\ln n+\ln\gamma+\frac{2}{1+\gamma^{1/2}}-1\right)-\ln\rho^+. \tag{67}$$

The quantity $\gamma_0=1/(1+k_0^2/\rho^-)$ is determined from condition (40) for the transport of the local Riemann invariant J_0 along the characteristics $dx/dt=V_3$:

$$2\left(\ln\rho^-+\ln\gamma_0+\frac{2}{1+\gamma_0^{1/2}}-1\right)-\ln\rho^+=\ln\rho^+. \tag{68}$$

As a result, for the coordinate of the trailing edge,

$$\tau^-=(\partial\omega_0/\partial k)|_{n=\rho_2,\gamma=\gamma_0}=\ln\Delta+\gamma_0^{3/2} \tag{69}$$

we have the equation

$$\ln[\Delta\cdot(\tau_*^-)^{2/3}]=\frac{(\tau_*^-)^{1/3}-1}{(\tau_*^-)^{1/3}+1}, \tag{70}$$

where $\tau_*^-=\tau^- -\ln\Delta$. The function $\tau^-(\Delta)$ is shown in Fig. 6 and is fully consistent with the results of a numerical calculation.[4]

7.2. Leading front of an ion-acoustic dissipationless shock

Finding the coordinates of the leading front of a self-similar ion-acoustic dissipationless shock is ideologically the same as the previous calculations for the trailing edge, but involves some cumbersome calculations. This is because of the complexity of the coefficients of the modulation system (4) for nonlinear ion-acoustic waves in the soliton limit. (See Ref. 5.) Here we present only the final result. The implicit dependence of the coordinate τ^+ of the leading front on the density jump Δ at the boundaries of the dissipationless shock is given by

$$\ln\Delta=\int_1^{\tau_+}\frac{z-1}{z}\,\frac{3a_-z-a_+z^{-1}+2a'_-z^2+2a'_+}{a_-z^2+a_-z+a_+z^{-1}-2a_--a_+}\,dz, \tag{71}$$

where

$$a_+(z)=-z^{1/4}\int_{-z^2/2}^{\psi(z)}(z-y)^{1/2}\times\left\{-z+\frac{\exp[(z^2-y^2)/2]-1}{z-y}\right\}^{-1/2}y\,dy,$$

$$a_-(z)=z^{-1/4}\int_{-z^2/2}^{\psi(z)}(z-y)^{1/2}\times\left\{-z+\frac{\exp[(z^2-y^2)/2]-1}{z-y}\right\}^{-1/2}dy. \tag{72}$$

The upper limit of integration $\psi(z)$ is specified by the algebraic equation

$$(-2\psi)^{1/2}-z+\frac{\exp(\psi+z^2/2)-1}{z}=0. \tag{73}$$

A plot of $\tau^+(\Delta)$ constructed using Eqs. (71)–(73) is shown in Fig. 6; this is also in agreement with the numerical calculations.[4]

We thank A. A. Krylov, V. V. Khodorovskiĭ, and A. V. Gurevich for many useful comments. This work was supported by the CRDF foundation (grant RM1-145). G. A. É. also thanks the Russian Fund for Fundamental Research for partial financial support (grant 96-01-01453).

[*)] E-mail: el@center.izmiran.troitsk.ru

[1)] In some cases a vacuum zone develops instead of a plateau.[4,26]

---

[1] F. Calogero, in: V. G. Bar'yakhtar and V. E. Zakharov, eds., *Integrability and Kinetic Equations for Solitons* [in Russian], Naukova Dumka, Kiev (1990), p. 65.

[2] V. I. Karpman, *Nonlinear Waves in Dispersive Media* [in Russian], Nauka, Moscow (1973).

[3] B. B. Kadomtsev, *Collective Phenomena in Plasmas* [in Russian], Nauka, Moscow (1976).

[4] A. V. Gurevich and A. P. Meshcherkin, Zh. Éksp. Teor. Fiz. **87**, 1277 (1984) [Sov. Phys. JETP **60**, 732 (1984)].

[5] A. V. Gurevich, A. L. Krylov, and G. A. Él', Zh. Éksp. Teor. Fiz. **98**, 1605 (1990) [Sov. Phys. JETP **71**, 899 (1990)].

[6] G. B. Whitham, Proc. R. Soc. London, Ser. A **283**, 238 (1965).

[7] A. V. Gurevich and A. P. Meshcherkin, Trudy FIAN **165**, 207 (1986).

[8] G. A. Él', in: B. S. Moiseev, ed., *Wave Processes in the Near-earth Plasma* [in Russian], Nauka, Moscow (1992), p. 163.

[9] R. Z. Sagdeev, in: M. A. Leontovich, ed., Reviews of Plasma Physics, Consultants Bureau, New York (1968), Vol. 4, p. 20.

[10] L. D. Landau and E. M. Lifshitz, *Hydrodynamics* [in Russian], Nauka, Moscow (1986).

[11] P. D. Lax and C. D. Levermore I–III, Commun. Pure Appl. Math. **36**, 253, 571, 809 (1983).

[12] S. Venakides, Commun. Pure Appl. Math. **38**, 883 (1985).

[13] P. Deift, S. Venakides, and X. Zhou, IMR No. 6, 285 (1997).

[14] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York (1974).

[15] A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **65**, 590 (1973) [Sov. Phys. JETP **38**, 291 (1973)].

[16] F. R. Tian, Commun. Pure Appl. Math. **46**, 1093 (1993).

[17] A. V. Gurevich, A. L. Krylov, and G. A. Él', Zh. Éksp. Teor. Fiz. **101**, 1797 (1992) [Sov. Phys. JETP **74**, 957 (1992)].

[18] G. A. El, Phys. Lett. A **222**, 393 (1996).

[19] B. Dubrovin, Am. Math. Soc. Trans. **179**, 35 (1997).

[20] H. Flaschka, M. G. Forest, and D. W. McLaughlin, Commun. Pure Appl. Math. **33**, 739 (1980).

[21] B. A. Dubrovin and S. P. Novikov, Usp. Mat. Nauk **44**, 28 (1989).

[22] S. P. Tsarev, Dokl. Akad. Nauk SSSR **282**, 534 (1985).

[23] M. G. Forest and J. E. Lee, in C. Dafermos *et al.*, eds., *Oscillation Theory Computation and Methods of Compensated Compactness*, Springer-Verlag (1986), p. 35.

[24] M. V. Pavlov, Teor. Mat. Fiz. **71**, 351 (1987).

[25] A. V. Gurevich and A. L. Krylov, Zh. Éksp. Teor. Fiz. **92**, 1684 (1987) [Sov. Phys. JETP **65**, 944 (1987)].

[26] G. A. El, V. V. Geogjaev, A. V. Gurevich, and A. L. Krylov, Physica D **87**, 186 (1995).

[27] G. A. El and A. L. Krylov, Phys. Lett. A **203**, 77 (1995).

[28] V. V. Avilov and S. P. Novikov, Dokl. Akad. Nauk SSSR **294**, 325 (1987) [Sov. Phys. Dokl. **32**, 366 (1987)].

[29] V. V. Avilov, I. M. Krichever, and S. P. Novikov, Dokl. Akad. Nauk SSSR **295**, 345 (1987) [Sov. Phys. Dokl. **32**, 564 (1987)].

[30] A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **93**, 871 (1987) [Sov. Phys. JETP **66**, 490 (1987)].

[31] A. E. Kudryavtsev, JETP Lett. **22**, 82 (1975).

[32] B. L. Rozhdestvenskiĭ and N. N. Yanenko, *Systems of Quasilinear Equations and their Applications in Gas Dynamics* [in Russian], Nauka, Moscow (1978).

[33] R. Courant, *Partial Differential Equations* [Russian translation], Mir, Moscow (1964).

[34] L. V. Ovsyannikov, *Lectures on the Foundations of Gas Dynamics* [in Russian], Nauka, Moscow (1981).

Translated by D. H. McNeill

JOURNAL OF EXPERIMENTAL AND THEORETICAL PHYSICS VOLUME 88, NUMBER 3 MARCH 1999

Instability and chaos in nonlinear wave interaction: a simple model

I. M. Khalatnikov

School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 69978 Tel Aviv, Israel; L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia

M. Kroyter*)

School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, 69978 Tel Aviv, Israel

(Submitted 7 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1137–1146 (March 1999)

We analyze the stability of a system which contains a harmonic oscillator nonlinearly coupled to its second harmonic, in the presence of a driving force. It is found that there always exists a critical amplitude of the driving force above which a loss of stability appears. The dependence of the critical input power on the physical parameters is analyzed. For a driving force of greater amplitude, chaotic behavior is observed. The generalization to interactions which include higher modes is discussed. © *1999 American Institute of Physics.* [S1063-7761(99)02703-1]

1. INTRODUCTION

In a series of experiments, the motion of the surface of a superfluid liquid in a cylindrical vessel was studied. This motion was induced by standing waves of second sound propagating in the bulk of the liquid. Above a critical value of the input power, the motion lost stability.[5,6]

To account for this loss of stability we analyzed a model that explained this phenomenon,[8] and found it to be in good agreement with the experimental results. The model is general enough to account for the loss of stability in other wave systems.

2. MODEL

The model consists of two nonlinearly coupled harmonic oscillators, of which one is coupled to an external driving force. First, we justify the use of two oscillators, with frequencies close to ω and 2ω, for describing the physics of systems such as the one above (ω is the frequency of the driving force). We assume that in the linear approximation the free, nondissipative (classical) theory is given by the Hamiltonian

$$H=\sum_{n=1}^{\infty} \omega_n a_n^* a_n , \tag{1}$$

where a_n is the (complex) amplitude of the n^{th} mode and a_n^* is its complex conjugate. Dissipation and the driving force are added subsequently. The modes are the eigenfunctions of the wave equation with the appropriate Sturm–Liouville boundary conditions. We neglect terms higher than cubic in the Hamiltonian, as well as terms which are far from resonance and therefore have small coupling constants.[1] The Hamiltonian becomes

$$H=\sum_{n=1}^{\infty} \omega_n a_n^* a_n + \sum_{\substack{k+l-m\simeq 0\\ k,l,m=1}}^{\infty} (\lambda_{k,l;m} a_k a_l a_m^* + \text{c.c.}), \tag{2}$$

where c.c. stands for complex conjugate, and $\lambda_{k,l;m}=\lambda_{l,k;m}$.

We now couple an external driving force to one of the modes. Note that in order to describe a physical problem, attenuation must be added as well. Modes which are not strongly coupled to the excited mode will decay. Again, we assume that to describe the onset of instability, a minimal number of modes is needed. We therefore take the excited mode and the mode with frequency closest to twice the frequency of the first. With a harmonic driving force, the Hamiltonian takes the form

$$\begin{aligned} H={}&\omega_d a_d^* a_d + \omega_{2d} a_{2d}^* a_{2d} + (\lambda a_d^2 a_{2d}^* + \text{c.c.})\\ &+(f e^{i\omega t} a_d^* + \text{c.c.}), \end{aligned} \tag{3}$$

where ω is the frequency of the driving force, which should be close to ω_d in order to establish resonance.

We use

$$i\dot{a}_d=\frac{\partial H}{\partial a_d^*}, \tag{4}$$

Hamilton's equations in the amplitude formalism,[1] to derive the equations of motion:

$$i\dot{a}_d=\omega_d a_d + 2\lambda^* a_d^* a_{2d} + f e^{i\omega t}, \tag{5}$$

$$i\dot{a}_{2d}=\omega_{2d} a_{2d} + \lambda a_d^2. \tag{6}$$

The equations are invariant under the transformation

$$\begin{aligned} &a_d\to a_d e^{i(\phi+\theta)}, \quad a_{2d}\to a_{2d} e^{i(\phi-\theta)},\\ &\lambda\to\lambda e^{i(-\phi-3\theta)}, \quad f\to f e^{i(\phi+\theta)}. \end{aligned} \tag{7}$$

It is therefore possible to eliminate two independent phases from the equations, so we can choose λ and f to be real.

We now add dissipative terms to the equation in the usual manner.[1] The equations become

$$i\dot{a}_d=(\omega_d-i\gamma_d)a_d+2\lambda a_d^* a_{2d}+fe^{i\omega t}, \tag{8}$$

$$i\dot{a}_{2d}=(\omega_{2d}-i\gamma_{2d})a_{2d}+\lambda a_d^2, \tag{9}$$

where γ are the dissipation constants.

The final stage before analyzing the equations is to introduce "slow variables" to eliminate the time dependence. Under the transformation

$$a_d\rightarrow a_d e^{-i\omega t}, \quad a_{2d}\rightarrow a_{2d}e^{-2i\omega t}$$

the equations become

$$i\dot{a}_d=(\Delta_d-i\gamma_d)a_d+2\lambda a_d^* a_{2d}+f, \tag{10}$$

$$i\dot{a}_{2d}=(\Delta_{2d}-i\gamma_{2d})a_{2d}+\lambda a_d^2, \tag{11}$$

where $\Delta_d\equiv\omega_d-\omega$ and $\Delta_{2d}\equiv\omega_{2d}-2\omega$ are the frequency offsets with respect to the driving force.

We note that in deriving the Hamiltonian (3) we have neglected one nonlinear term of the same order with respect to a_d as the one that we have kept, namely $\kappa a_d^2 a_d^{*2}$. For systems in which $\kappa\ll\lambda^2/\omega$, this term is small, but it turns out that even for κ much larger the importance of this term is not crucial. Note that ω is absent from the equations. From dimensional considerations, κ can appear in the equations only as $\kappa\Delta$ or $\kappa\gamma$; this is the term that has to be of order λ^2.

Hereafter we analyze three aspects of the model: stationary solutions, stability, and numerical calculations. For stationary solutions it is easy to verify that the effect of κ is merely to renormalize Δ_{2d} and γ_{2d}. This is the well-known effect of shifting the resonance.[2] We have seen that κ is not of great importance in the stability analysis or in our numerical calculations, even for $\kappa>\lambda^2/\Delta_{2d},\lambda^2/\gamma_{2d}$. We will not include this term in what follows.

Although the model we use is a very simplified one, it still contains five parameters in addition to the driving force amplitude f. Not all the parameters are important. The amplitude f of the driving force is an effective expression which is in fact a function of Δ_d; moreover, the driving force couples to all other modes as well, and we may neglect all other couplings only when the one that we are left with is the dominant one. For this to be the case, we must have $\omega\simeq\omega_d$, that is, Δ_d must be small compared to all other parameters with dimensions of frequency. The value of Δ_{2d} will be dictated by geometry. Both our analytic and our numerical results depend on this assumption. In most physical systems there is a relation between γ_d and γ_{2d}. We shall assume that these two parameters are of the same order of magnitude.

3. STATIONARY SOLUTIONS

We begin our analysis by finding the fixed points of the equations, i.e., solving the equations

$$(\Delta_d-i\gamma_d)a_d+2\lambda a_d^* a_{2d}+f=0, \tag{12}$$

$$(\Delta_{2d}-i\gamma_{2d})a_{2d}+\lambda a_d^2=0. \tag{13}$$

We eliminate a_{2d} from the second equation, and substitute in the first one to obtain

$$(\Delta_d-i\gamma_d)(\Delta_{2d}-i\gamma_{2d})a_d-2\lambda^2 a_d|a_d|^2 = -(\Delta_{2d}-i\gamma_{2d})f. \tag{14}$$

The equation for $\zeta\equiv2\lambda^2|a_d|^2/|\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d}|$ then becomes

$$((\zeta\pm1)^2+\beta)\zeta=h, \tag{15}$$

where

$$h\equiv\frac{2(\gamma_{2d}^2+\Delta_{2d}^2)\lambda^2}{|\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d}|^3}f^2 \tag{16}$$

is the scaled force, and

$$\beta\equiv\left(\frac{\Delta_d\gamma_{2d}+\Delta_{2d}\gamma_d}{\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d}}\right)^2. \tag{17}$$

The sign in equation (15) coincides with the sign of $\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d}$.

This equation has either one or three solutions. For a given value of h, the equation will have three solutions if and only if

$$\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d}<0, \tag{18}$$

$$0\leqslant\beta<\frac{1}{3}, \tag{19}$$

$$\frac{2}{27}[1+9\beta-(1-3\beta)^{3/2}]\leqslant h\leqslant\frac{2}{27}[1+9\beta+(1-3\beta)^{3/2}]. \tag{20}$$

In the following it will be illustrated that when three solutions are present, the middle one is unstable, as expected.

We note that the situation of three solutions is, in a sense, nonphysical. Both γ_d and γ_{2d} are positive, so we can use (17)–(19) to deduce that

$$\left(\frac{\gamma_d}{\Delta_d}\right)^2<\left(\frac{\gamma_{2d}}{\Delta_{2d}}+\frac{\gamma_d}{\Delta_d}\right)^2<\frac{1}{3}. \tag{21}$$

But this suggests that $\gamma_d<\Delta_d$, which contradicts our assumptions. In this region of parameters our model is inappropriate.

4. STABILITY

To check whether the stationary solutions are stable we linearize the equations around these solutions, and check whether small perturbations grow or decay. To simplify the calculations, we recall the symmetry (7) and use it with $\phi+3\theta=0$ to redefine the stationary value of the first mode, $a_d^{(0)}$, to be real, without altering λ. The change in f is not important, since f will be absent from the linearized equations. We substitute in the linearized equations:

$$a_{2d}^{(0)}=-\frac{\lambda a_d^{(0)^2}}{\Delta_{2d}-i\gamma_{2d}}. \tag{22}$$

Stability is now determined by $a_d^{(0)}$. Also, to simplify the notation, we use a_d, a_{2d} rather than $\delta a_d, \delta a_{2d}$ for deviations from the stationary solution.

The linearized equations are

$$i\dot{a}_d=(\Delta_d-i\gamma_d)a_d+2\lambda\left(a_d^{(0)}a_{2d}-\frac{\lambda a_d^{(0)^2}}{\Delta_{2d}-i\gamma_{2d}}a_d^*\right), \tag{23}$$

$$i\dot{a}_{2d}=(\Delta_{2d}-i\gamma_{2d})a_{2d}+2\lambda a_d^{(0)}a_d; \tag{24}$$

multiplying by $-i$ and separating into real and imaginary parts, we obtain the differential equation:

$$\frac{d}{dt}\begin{pmatrix}\mathrm{Re}(a_d)\\ \mathrm{Im}(a_d)\\ \mathrm{Re}(a_{2d})\\ \mathrm{Im}(a_{2d})\end{pmatrix} = \begin{pmatrix} -\gamma_d-p\gamma_{2d} & \Delta_d+p\Delta_{2d} & 0 & 2a_d^{(0)}\lambda \\ -\Delta_d+p\Delta_{2d} & -\gamma_d+p\gamma_{2d} & -2a_d^{(0)}\lambda & 0 \\ 0 & 2a_d^{(0)}\lambda & -\gamma_{2d} & \Delta_{2d} \\ -2a_d^{(0)}\lambda & 0 & -\Delta_{2d} & -\gamma_{2d} \end{pmatrix} \times\begin{pmatrix}\mathrm{Re}(a_d)\\ \mathrm{Im}(a_d)\\ \mathrm{Re}(a_{2d})\\ \mathrm{Im}(a_{2d})\end{pmatrix}, \tag{25}$$

where $p=2a_d^{(0)^2}\lambda^2/(\gamma_{2d}^2+\Delta_{2d}^2)$.

To ensure stability, we require that the real part of all eigenvalues of this matrix be negative. We find the coefficients of the characteristic polynomial $u^4+au^3+bu^2+cu+d$ to be

$$a=2(\gamma_d+\gamma_{2d}), \tag{26}$$

$$b=-\frac{4\lambda^4}{\gamma_{2d}^2+\Delta_{2d}^2}a_d^{(0)^4}+8\lambda^2 a_d^{(0)^2} +(\gamma_d^2+\Delta_d^2+\gamma_{2d}^2+\Delta_{2d}^2+4\gamma_d\gamma_{2d}), \tag{27}$$

$$c=-\frac{8\lambda^4\gamma_{2d}}{\gamma_{2d}^2+\Delta_{2d}^2}a_d^{(0)^4}+8\lambda^2(\gamma_d+\gamma_{2d})a_d^{(0)^2} +2[(\gamma_d^2+\Delta_d^2)\gamma_{2d}+(\gamma_{2d}^2+\Delta_{2d}^2)\gamma_d], \tag{28}$$

$$d=12\lambda^4 a_d^{(0)^4}+8\lambda^2(\gamma_d\gamma_{2d}-\Delta_d\Delta_{2d})a_d^{(0)^2} +(\gamma_d^2+\Delta_d^2)(\gamma_{2d}^2+\Delta_{2d}^2). \tag{29}$$

To ensure that all roots of this polynomial have a negative real part, we use the Routh–Hurwitz criterion:[3,4]

$$a>0, \tag{30}$$

$$b>0, \tag{31}$$

$$d>0, \tag{32}$$

$$abc>c^2+a^2d. \tag{33}$$

The condition (30) is trivial for a physical problem. The condition (31) is a quadratic equation in $a_d^{(0)^2}$, and is easily solved to give

$$a_d^{(0)^2}<\frac{\gamma_{2d}^2+\Delta_{2d}^2}{4\lambda^2}\left(4+\sqrt{5+\frac{\gamma_d^2+\Delta_d^2+4\gamma_d\gamma_{2d}}{\gamma_{2d}^2+\Delta_{2d}^2}}\right). \tag{34}$$

The third condition, (32), is again a quadratic equation in $a_d^{(0)^2}$, but with a positive rather then negative coefficient of a_d^4.

It is easily seen that for a negative d to occur in the physical region $a_d^{(0)^2}>0$, we need to have

$$\gamma_d\gamma_{2d}>\Delta_d\Delta_{2d}. \tag{35}$$

When this condition is fulfilled, an unstable region appears when

$$\beta<\frac{1}{3}. \tag{36}$$

Direct solution of the quadratic equation then shows that the central region of solutions coincides exactly with this unstable region (20). As mentioned above, this region is not physically important.

We combine (26)–(29) and (33), and define

$$z=\lambda^2 a_d^{(0)^2} \tag{37}$$

to obtain the last inequality:

$$a_0z^4+a_1z^3+a_2z^2+a_3z+a_4>0, \tag{38}$$

where

$$a_0=\frac{64\gamma_d\gamma_{2d}}{(\gamma_{2d}^2+\Delta_{2d}^2)^2}, \tag{39}$$

$$a_1=-\frac{64(\gamma_d+\gamma_{2d})^2}{\gamma_{2d}^2+\Delta_{2d}^2}, \tag{40}$$

$$a_2=\frac{32\gamma_d\gamma_{2d}}{\gamma_{2d}^2+\Delta_{2d}^2}(\Delta_{2d}^2-\Delta_d^2-(\gamma_d+\gamma_{2d})^2), \tag{41}$$

$$a_3=16(\gamma_d+\gamma_{2d})^2[(\gamma_d+\gamma_{2d})^2+(\Delta_d+\Delta_{2d})^2)], \tag{42}$$

$$a_4=4\gamma_d\gamma_{2d}[(\gamma_d+\gamma_{2d})^2+(\Delta_d+\Delta_{2d})^2] \times[(\gamma_d+\gamma_{2d})^2+(\Delta_d-\Delta_{2d})^2)]. \tag{43}$$

It is seen that for any parameter value there exists an open neighborhood of zero in which the stationary solution is stable. It is very tedious to solve the inequality for the general case. We solve it for two special cases: one-dimensional geometry, and a cylindrical wave with a large Q-factor, both with reflecting boundary conditions.

Recall the assumption $\Delta_d \ll \gamma_d$. It is natural to assume that γ_d and γ_{2d} are of the same order of magnitude. In a wide class of cases $\gamma\propto\omega^2$, and therefore

$$\gamma_{2d}\simeq 4\gamma_d. \tag{44}$$

We shall consider this case for both geometries. The value of Δ_{2d} is dictated by geometry.

For one-dimensional geometry, the d^{th} mode is $\cos(d\pi x/L)$, where L is the length of the vessel. This dependence yields

$$\Delta_{2d}=\omega_{2d}-2\omega=2\Delta_d-(2\omega_d-\omega_{2d})=2\Delta_d-c(2k_d-k_{2d})=2\Delta_d-c(2\pi d/L-\pi 2d/L)=2\Delta_d. \tag{45}$$

We therefore have for the one-dimensional case

$$\Delta_d,\Delta_{2d}\ll\gamma_d,\quad \gamma_{2d}. \tag{46}$$

We define

$$x=z/\gamma_d^2, \tag{47}$$

$$s=\gamma_{2d}/\gamma_d. \tag{48}$$

We use (46) to derive the inequality

$$x^4-s(1+s)^2x^3-\frac{1}{2}s^2(1+s)^2x^2+\frac{1}{4}s^3(1+s)^4x+\frac{1}{16}s^4(1+s)^4>0. \tag{49}$$

The solution of this inequality combined with (34) yields the final result,

$$x<\frac{1}{2}(s+s^2), \tag{50}$$

from which one easily finds an expression for the critical input power:

$$f_c=\frac{3\gamma_d+\gamma_{2d}}{\lambda}\sqrt{\gamma_{2d}\frac{\gamma_d+\gamma_{2d}}{2}} \tag{51}$$

or, using (44),

$$f_c\simeq\frac{22}{\lambda}\gamma_d^2. \tag{52}$$

We now substitute $\gamma_d=\alpha\omega^2$ to obtain

$$f_c\simeq\frac{22\alpha^2}{\lambda}\omega^4. \tag{53}$$

A full description of the loss of stability for the specific problem can be obtained if we take into account the dependence of α and λ on the relevant physical parameters, for example, the temperature.

For a cylindrical vessel of radius R, the modes are given by $J_n(kr)\cos(n\theta)$, where J_n is the n^{th} Bessel function, and $k=\omega/c$, where c is the wave velocity. The boundary conditions impose the relation $k_{n,m}R=\chi_{n,m}$ where $\chi_{n,m}$ is the m^{th} zero of $J_n'(\chi)$. For simplicity we consider here only the J_0 modes.

The value of Δ_{2d} is dictated by the Bessel function asymptotic behavior,

$$\chi_m\equiv\chi_{0,m}\simeq n\pi+\pi/4, \tag{54}$$

whereupon

$$\begin{aligned}\Delta_{2d}&=\omega_{2d}-2\omega=2\Delta_d-(2\omega_d-\omega_{2d})\\&=2\Delta_d-c(2k_d-k_{2d})\simeq2\Delta_d-\frac{c}{R}(2\chi_d-\chi_2d)\\&=2\Delta_d-\frac{c}{R}\left(2\left(d\pi+\frac{\pi}{4}\right)-\left(2d\pi+\frac{\pi}{4}\right)\right)\\&=2\Delta_d-\frac{\pi c}{4R}\simeq2\Delta_d-\frac{\omega_d}{4d+1}.\end{aligned} \tag{55}$$

Since $\omega_d=2Q\gamma_d$, the higher is Q, the higher the values of d for which the inequality

$$\Delta_{2d}\gg\gamma_d \tag{56}$$

holds.

We solve now equation (38) for the case

$$\Delta_d\ll\gamma_d,\quad \gamma_{2d}\ll\Delta_{2d}. \tag{57}$$

We define s as before, but now

$$x=z/\Delta_{2d}^2, \tag{58}$$

and we obtain

$$x^4-\frac{(1+s)^2}{s}x^3+\frac{1}{2}x^2+\frac{(1+s)^2}{4s}x+\frac{1}{16}>0. \tag{59}$$

When this condition is combined with (34), we have

$$x<\frac{1}{4}(v+\sqrt{2uv}), \tag{60}$$

where $u=(1+s)^2/s$ and $v=u-\sqrt{u^2-4}$. We use (44) to obtain $x<0.59$. For other values of s there are only small changes in the result. In all cases the critical value is in the range $0.5<x_0<0.65$. The maximum is attained at $s=1$, and the minima are at $x=0$ and $x\to\infty$ (note that $x_0(s)=x_0(1/s)$). The critical input power f_c can now be calculated:

$$f_c=\frac{2\Delta_{2d}^2x_0^{1.5}}{\lambda}\simeq0.56\frac{c^2}{\lambda R^2}. \tag{61}$$

A full description of the loss of stability in this geometry can be obtained if we take into account the dependence of λ and c on the relevant physical parameters.

5. BEYOND NUMERICAL CALCULATIONS

Some questions arise. Does the system always reach a stationary solution in the stable region? What happens above the stable region? In what way would the theory be modified if we include the full Hamiltonian (2)?

We solved the equation numerically with parameters suitable to describe the cylindrical geometry:

$$\Delta_d=0,\quad \Delta_{2d}=1500,$$

$$\gamma_d=30,\quad \gamma_{2d}=120,$$

$$\lambda=5400, \tag{62}$$

with initial conditions

$$a_d(t=0)=0,\quad a_{2d}(t=0)=0. \tag{63}$$

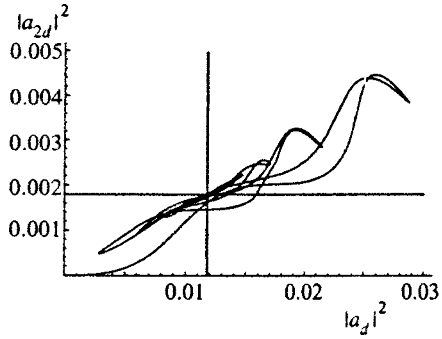

FIG. 1. The system approaches the fixed point for $f=50$; the position of the fixed point is indicated.

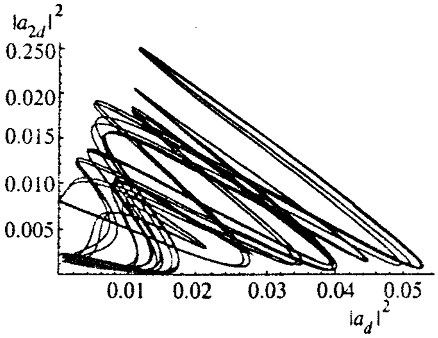

FIG. 3. For $f=500$. One of the limit cycles which bifurcates towards chaos.

Results for other parameter values may be very similar due to the scaling properties of the equation discussed above.

For small enough values of f, the system reaches the stationary solution after wandering in phase space (Fig. 1). For $f \simeq 0.3 f_c$ with the initial conditions above, the system escapes the basin of attraction of the fixed point, and instead approaches a limit cycle (Fig. 2). The basin of attraction of the stationary solution shrinks to zero as the instability is approached. This limit cycle is not unique. By choosing various initial conditions, other limit cycles can be approached. In the higher f regime, the behavior is harder to determine.

It is easy to prove that the motion of the system is bounded in phase space, and that the volume in phase space decays exponentially with decay factor $2(\gamma_d+\gamma_{2d})$.

A necessary condition for chaos to evolve is that the system be locally unstable. Our analysis shows that the phase of a_d, a_{2d} is irrelevant to this question. Given the parameters, the potentially chaotic regions are defined in the $(|a_d|^2, |a_{2d}|^2)$ plane. Our calculations show that the region

$$|a_{2d}|^2\lambda^2 \gg \gamma^2, \Delta^2 \tag{64}$$

is always locally unstable. The numerical calculations show that when f is increased the system enters this region, bifurcations appear, as in the usual route to chaos. For large enough f, chaos will evolve.

In Fig. 3 we see bifurcations at $f=500$. Chaos evolves at $f \simeq 506$, as we see in Fig. 4.

When more modes are added to the system, the behavior changes. The projection of, say, the 3-mode system in the (four-dimensional) phase space of two modes yields, in general, trajectories which are very different from the original ones. Yet we argue that the main conclusion does not change. Indeed, if we examine the original set of equations (10) and (11), we note that the transformation

$$a_d \to \alpha a_d, \quad a_{2d} \to \alpha a_{2d}, \quad f \to \alpha f, \quad \lambda \to \frac{1}{\alpha}\lambda, \tag{65}$$

which is a generalization of (7), leaves the equations invariant. We could then deduce that $f_c \propto 1/\lambda$. From dimensional considerations, f should be proportional to γ^2, Δ^2. It is seen that for the one-dimensional case, the leading behavior is $f_c \propto \gamma^2$, while for the high-Q case $f_c \propto \Delta_{2d}^2$.

All our calculations were in fact needed just to illustrate that there is only one transition from stability to instability (i.e., no unstable windows), to validate the assumption that the largest constant with frequency dimensions is not absent from the expression for f_c, and to calculate x_0. When we add new modes, new constants are added to the system. From (54), we find that for all $j\Delta_j \propto c/R$, so these constants do not cause a problem. The same is true for the one-dimensional case as for the new λ's if they scale in some way, e.g., if

$$\lambda_{k,l;m} = f(T,R,\ldots)\, h\left(\frac{l}{k}, \frac{m}{k}\right) k^u, \tag{66}$$

where $f(T,R,\ldots)$ is any function of all physical parameters but the wavelength, $h(l/k, m/k)$ are constants, and u is an exponent, then the symmetry still holds, and then given that the general picture remains the same, all that we need to change is the value of x_0. This necessary modification of x_0, plus the shrinking of basin of attraction, which effectively

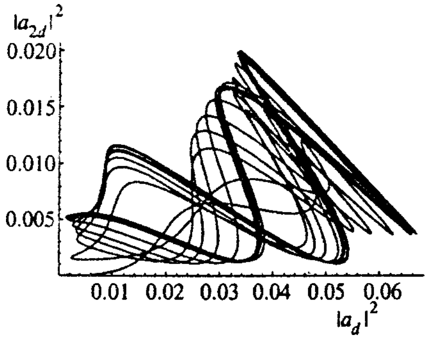

FIG. 2. The system approaches a limit cycle for $f=100$, which is below the critical value.

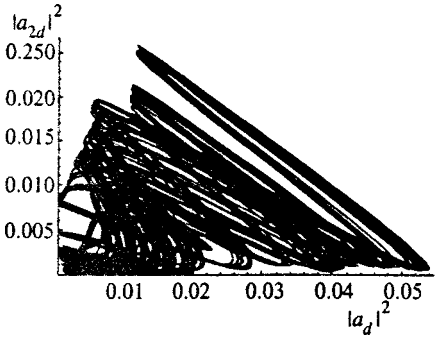

FIG. 4. For $f=507$, the system is chaotic.

lowers x_0, suggests that this part of our calculations is not accurate. Nevertheless, the dependence of the critical input power on all physical parameters remains the same even for the full Hamiltonian (2). There are values of $u,h(l/k,m/k)$ for which other predictions, such as the distribution of the chaotic regions of the 2-mode system, would not be dramatically changed as well. More extensive investigation of this system is therefore highly desirable.

The authors express their gratitude to Prof. Boris Chirikov for his kind interest in this work, and for his many important remarks. One of the authors (I. M. K.) would like to thank Prof. Valery Pokrovsky, with whom the problem analyzed in this work was first formulated. We wish to thank Prof. Naum Meiman for useful discussions.

*)E-mail: mikroyt@post.tau.ac.il

---

[1] V. S. L'vov, *Wave Turbulence Under Parametric Excitations*, Springer-Verlag, Berlin (1994).
[2] L. D. Landau and E. M. Lifshitz, *Mechanics*, Pergamon Press, New York (1976).
[3] E. A. Guillemin, *The Mathematics of Circuit Analysis*, John Wiley and Sons Inc., New York (1965).
[4] J. V. Uspensky, *Theory of Equations*, McGraw-Hill, New York (1948).
[5] J. L. Olsen, J. Low Temp. Phys. **61**, 17 (1985).
[6] P. W. Egolf, J. L. Olsen, B. Roehricht, and D. A. Weiss, Physica B **169**, 217 (1991).
[7] I. M. Khalatnikov, G. V. Kolmakov, and V. L. Pokrovsky, Zh. Éksp. Teor. Fiz. **107**, 1563 (1995) [JETP **80**, 873 (1995)].
[8] M. Kroyter, Masters Thesis, School of Physics and Astronomy, Tel-Aviv University (1998).

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