Propagation of axions in a strongly magnetized medium

A. V. Borisov*) and P. E. Sizin

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 5 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 3–11 (January 1999)

The polarization operator of an axion in a degenerate gas of electrons occupying the ground-state Landau level in superstrong magnetic fields $H \gg H_0 = m_e^2 c^3/e\hbar = 4.41 \times 10^{13}$ G is investigated in a model with a tree-level axion-electron coupling. It is shown that a dynamic axion mass, which can fall within the allowed range of values 10^{-5} eV $\lesssim m_a \lesssim 10^{-2}$ eV, is generated under the conditions of strongly magnetized neutron stars. As a result, the dispersion relation for axions is appreciably different from that in a vacuum. © *1999 American Institute of Physics.* [S1063-7761(99)00101-8]

1. The *a priori* strong nonconservation of CP parity in the standard model can be eliminated in a natural manner by introducing axions — pseudo-Goldstone bosons associated with the spontaneous breaking of the additional Peccei–Quinn global symmetry $U(1)_{PQ}$.[1,2] According to the experimental data,[3] the energy scale v_a for the $U(1)_{PQ}$ symmetry breaking is much greater than the electroweak scale — $v_a \gtrsim 10^{10}$ GeV, and the constants of the possible couplings of an axion to the standard particles ($\sim 1/v_a$) are very small (the ''invisible'' axion: see Ref. 4 for a review of various axion models).

Axion effects can be appreciable under the astrophysical conditions of high matter densities, high temperatures, and strong magnetic fields (for example, in neutron stars[5]). Axion production processes, which result in additional energy losses by stars, and the limits obtained by astrophysical methods on the parameters of axion models are examined in Ref. 4. In so doing, the influence of electromagnetic fields were neglected.

The investigation of axion processes in strong magnetic fields commenced comparatively recently. The Compton and Primakoff mechanisms of axion production on nonrelativistic electrons by thermal photons ($\gamma + e \to e + a$) in the presence of a magnetic field are studied in Ref. 6. The extension to relativistic electrons in a constant external electromagnetic field is given in Ref. 7 (Primakoff effect) and Refs. 8 and 9 (Compton effect), where[7,9] estimates were also obtained for the contributions of the indicated processes to the axion luminosity of a magnetized strongly degenerate relativistic electron gas under the conditions of the crust of a neutron star. A new axion production mechanism — synchrotron emission of axions ($e \to e + a$) by relativistic electrons — was proposed in Ref. 10 and its contribution to the energy losses by a neutron star was calculated. In Refs. 6–10 it was assumed that the external field intensity $F \ll H_0 = m_e^2 c^3/e\hbar \simeq 4.41 \times 10^{13}$ G. In Ref. 11, numerical methods were used to extend the results of Ref. 10 to superstrong magnetic fields $H \gtrsim H_0$. It was found that the basic equation derived in Ref. 10 for the axion synchrotron luminosity for the semiclassical case of high electron energies ($\varepsilon \gg m_e c^2$) and fields $H \ll H_0$ agrees with the numerical calculations up to $H/H_0 \lesssim 20$. The axion synchrotron luminosity of neutron stars and white dwarfs was also investigated in Ref. 11.

In Refs. 8–11 a model with a derivative axion-electron coupling eae, described by the interaction Lagrangian[4]

$$\mathscr{L}_{ae} = \frac{g_{ae}}{2m_e}(\bar{\psi}\gamma^\mu\gamma^5\psi)\partial_\mu a, \tag{1}$$

was used. Here m_e is the electron mass and $\gamma^5 = -i\gamma^0\gamma^1\gamma^2\gamma^3$; the system of units such that $\hbar = c = 1$ is used; the signature of the metric is $(+---)$; and

$$g_{ae} = c_e \frac{m_e}{v_a} \tag{2}$$

is a dimensionless coupling constant, where the numerical factor c_e depends on the choice of the specific axion model.[4]

In models where axions are coupled only with heavy fermions by a tree-level coupling there arises an effective direct low-energy axion–photon interaction of the type $\gamma a \gamma$.[4] This interaction is the basis of the Primakoff axion photoproduction mechanism employed in Refs. 6 and 7. The synchrotron process $e \to ea$ in the absence of a tree-level axion–electron coupling was considered recently in Ref. 12. This process is due to resonant conversion of a longitudinal plasmon (a photon in a medium), emitted by a relativistic electron in a magnetic field, into an axion.

Decay of an axion in a strong magnetic field into a fermion pair ($a \to f\bar{f}$)[13] and two photons ($a \to \gamma\gamma$)[14] are also of interest for astrophysics and cosmology.

In the present paper the model (1) is used to calculate the polarization operator of an axion moving in a strongly magnetized degenerate electron gas and the change in the dispersion relation of an axion in a medium is investigated using this operator.

2. Taking account of the contribution of the electrons only [see Eq. (1)] we obtain, using the real-time formalism of the finite- temperature quantum field theory (see, for example, Ref. 15), the following momentum representation for the one-loop polarization operator of an axion:

$$\Pi(k,k')=-iG_a^2\int d^4x\,d^4x'\exp(ikx-ik'x')$$
$$\times\mathrm{Tr}[\hat{k}\gamma^5 G(x,x')\hat{k}'\gamma^5 G(x',x)]. \quad (3)$$

Here k (k') is the final (initial) 4-momentum of an axion; $G(x,x')$ is the time-dependent single-particle Green's function of an ideal electron-positron gas in a constant magnetic field;[15] notations have also been introduced for the contraction $\hat{a}=\gamma^\mu a_\mu$ of a 4-vector a^μ with the Dirac γ matrices and for the dimensional coupling constant

$$G_a=\frac{g_{ae}}{2m_e}. \quad (4)$$

On account of the translational invariance (constant external field, homogeneous isotropic medium) the polarization operator (3) is diagonal in momentum space:

$$\Pi(k,k')=(2\pi)^4\delta^{(4)}(k-k')\Pi(k). \quad (5)$$

Here $\Pi(k)$ determines the axion propagator $D(k)$ in the momentum representation according to the Dyson equation

$$D(k)=[k^2-m_a^2-\Pi(k)]^{-1}, \quad (6)$$

where m_a is the free-axion mass (in the absence of a field and a medium), which is generated by the chiral anomaly of QCD:[2] $m_a\sim\Lambda_{\mathrm{QCD}}^2/v_a$. The renormalized value $\Pi_R(k)$ (see below) gives the dispersion relation

$$k^2=m_a^2+\Pi_R(k). \quad (7)$$

3. We give the constant uniform magnetic field $\mathbf{H}\|\mathbf{z}$ in terms of the 4-potential A^μ in the gauge

$$A^\mu=(0,0,xH,0). \quad (8)$$

Then the Green's function $G(x,x')$ can be represented in the following form after summing over the spin quantum number and the sign of the energy in the general expression for G in the form of a series in quadratic combinations of the eigenfunctions of the Dirac operator[15]:

$$G(x,x')=[\gamma^\mu(i\partial_\mu+eA_\mu)+m_e]K(x,x'),$$
$$K(x,x')=\frac{\sqrt{h}}{(2\pi)^3}\sum_{n=0}^{\infty}\int_{-\infty}^{\infty}dp_0dp_ydp_z$$
$$\times\exp[-ip_0(t-t')+ip_y(y-y')+ip_z(z-z')]$$
$$\times u_n(\eta)u_n(\eta')(R_{n+1}\Sigma_+ +R_n\Sigma_-), \quad (9)$$
$$R_n=[p_0^2-p_z^2-2hn-m_e^2+i0]^{-1}$$
$$+2\pi i\,\delta(p_0^2-p_z^2-2hn-m_e^2)N_F(p_0).$$

Here the electron charge $-e<0$, $h=eH$; $n=0,1,2,\ldots$ is the principal quantum number (the number of the Landau level); p_y and p_z are the eigenvalues of the projection operators of the canonical momentum — the constants of motion in the gauge (8); and $u_n(\eta)$ is a Hermite function of argument

$$\eta=\sqrt{h}(x+p_y/h),\quad \eta'=\eta(x\to x'),$$
$$\Sigma_\pm=(1\pm\Sigma_3)/2,\quad \Sigma_3=i\gamma^1\gamma^2.$$

The first term in R_n has poles at the points $p_0=\pm\varepsilon=\pm[m_e^2+2hn+p_z^2]^{1/2}$, determining the energy spectrum of an electron in a magnetic field. The second term ($\propto\delta(p_0^2-\varepsilon^2)$) describes the effect of the electron-positron medium, and

$$N_F(p_0)=\theta(p_0)[\exp[\beta(p_0-\mu)]+1]^{-1}+\theta(-p_0)$$
$$\times[\exp[\beta(-p_0+\mu)]+1]^{-1} \quad (10)$$

is expressed in terms of the Fermi distribution function of electrons and positrons in a medium with temperature $T=1/\beta$ and chemical potential μ, and $\theta(\pm p_0)$ is the Heaviside step function.

4. It is difficult to make a general analysis of the axion polarization operator for arbitrary values of the parameters H, T, and μ. In the present paper we confine our attention to superstrong magnetic fields and comparatively low temperatures

$$H\gg H_0,\quad T\ll\mu-m_e, \quad (11)$$

and we require the chemical potential to satisfy

$$\mu^2-m_e^2<2h. \quad (12)$$

It follows from Eqs. (11) and (12) that in this case the contribution of positrons in Eq. (10) can be neglected (it is suppressed by the factor $\exp[-\beta(\mu-m_e)]$) and the medium is a degenerate gas of electrons occupying the ground-state Landau level ($n=0$):

$$N_F(p_0)=\theta(p_0)\,\theta(\mu-p_0),\quad p_0=\sqrt{m_e^2+p_z^2}. \quad (13)$$

We also limit the range of the axion 4-momentum

$$|k_0^2-k_z^2|\ll h. \quad (14)$$

Then the main contribution of virtual (vacuum) electrons and positrons is likewise formed by states with $n=0$. As a result, retaining on the basis of Eqs. (11), (12), and (14) terms with $n=0$ in the sum (9), we obtain the following approximate expression for the Green's function in a superstrong magnetic field:

$$G(x,x')\simeq\left(\frac{h}{\pi}\right)^{1/2}\int_{-\infty}^{\infty}\frac{dp_y}{2\pi}\exp\Big[-\frac{1}{2}(\eta^2+\eta'^2)+ip_y(y-y')\Big]\int\frac{d^2p}{(2\pi)^2}\exp[-ip_0(t-t')$$
$$+ip_z(z-z')]G(p)\Sigma_-. \quad (15)$$

Here $p=(p_0,0,0,p_z)$ and

$$G(p)=(\hat{p}+m_e)[(p^2-m_e^2+i0)^{-1}$$
$$+2\pi i\,\delta(p^2-m_e^2)N_F(p_0)] \quad (16)$$

is the Fourier transform of the Green's function in the two-dimensional space (0, 3). For $N_F=0$ (no medium) the expression (15) is the well-known, effectively two-dimensional, electron propagator used in the theory of electrodynamic processes in superstrong magnetic fields and, specifically, for investigation of the photon polarization operator.[16]

5. Let us substitute the expression (15) into Eq. (3) and integrate over $t,t',y,y',z,$ and z'. This gives in the form of a product of delta functions

$$\prod_{n=0,y,z}\delta(k_n'-k_n)\,\delta(p_n'+k_n-p_n),$$

the laws of conservation of energy and of the corresponding projections of the momentum. The subsequent calculation of the Gaussian integrals over x, x' and the trivial integral over p_y gives $\delta(k_x'-k_x)$. As a result, as should be the case, we obtain a diagonal representation of the polarization operator (5), where

$$\Pi(k)=\frac{G_a^2}{\pi}h\exp\left(-\frac{k_\perp^2}{2h}\right)[F(l)+M(l)], \tag{17}$$

$$F(l)=-i\int\frac{d^2p}{(2\pi)^2}T(l,p)[p^2-m_e^2+i0]^{-1}\times[(p-l)^2-m_e^2+i0]^{-1}, \tag{18}$$

$$M(l)=2\pi\int\frac{d^2p}{(2\pi)^2}\delta(p^2-m_e^2)N_F(p_0)\times\left[\frac{T(l,p)}{(p-l)^2-m_e^2+i0}+(l\to-1)\right]. \tag{19}$$

Here $p=(p_0,0,0p_z)$ and $l=(k_0,0,0,k_z)$ are two-dimensional vectors, and

$$T(l,p)=\frac{1}{2}\mathrm{Tr}[\hat{k}\gamma^5(\hat{p}+m_e)\Sigma_-\hat{k}\gamma^5(\hat{p}-\hat{l}+m_e)\Sigma_-]. \tag{20}$$

In Eq. (17) the function F corresponds to the purely field contribution, and M describes the influence of the medium. We note that M does not contain a term $\sim N_F(p_0)N_F(p_0-l_0)$, since

$$\delta(p^2-m_e^2)\,\delta((p-l)^2-m_e^2)\,\theta(p_0)\,\theta(p_0-l_0)=0.$$

Using the relations

$$[\Sigma_-,\hat{p}]=0,\quad [\Sigma_-,\gamma^5]=0,$$

$$\gamma^n\Sigma_-=\Sigma_+\gamma^n\quad(n=1,2),\quad \Sigma_+\Sigma_-=0$$

the trace of Eq. (18) reduces to a two-dimensional form and can be easily calculated as

$$T(l,p)=\frac{1}{4}\mathrm{Tr}[\hat{l}(\hat{p}+m_e)\hat{l}(\hat{p}-\hat{l}-m_e)]=2(lp)^2-l^2(lp+p^2+m_e^2). \tag{21}$$

We calculate the Gaussian integrals over p_0 and p_z in Eq. (17) using the trace (21) and the well-known Fock–Schwinger proper-time representation for propagators of the form

$$(\Delta+i0)^{-1}=-i\int_0^\infty ds\exp[is(\Delta+i0)].$$

As a result, we find for the function $F(l)$ the integral representation

$$F_R(l^2)=-i\frac{m_e^2\tau}{4\pi}\int_0^1 dv\int_0^\infty dx\{[1+(1-v^2)\tau]\times\exp[-ix[1-(1-v^2)\tau]]-\exp(-ix)\}, \tag{22}$$

$$\tau=l^2/4m_e^2.$$

Here renormalization is performed according to the well-known rule[16]

$$F_R(l^2)=F(l^2)-F(m_a^2)-(l^2-m_a^2)F'(m_a^2).$$

In Eq. (22) the small mass parameter $\delta_a=m_a^2/4m_e^2$ is neglected. For $m_a\lesssim10^{-3}$ eV,[9,12] we have $\delta_a\lesssim10^{-18}$.

For $\tau<0$ we obtain from Eqs. (22) and (17) the field contribution

$$\Pi_R^{(F)}=-\frac{\alpha_a}{\pi}m_e^2\frac{H}{H_0}\exp\left(-\frac{k_\perp^2}{2h}\right)\left[\frac{(1-\xi)^2}{\xi}+\frac{1-\xi}{1+\xi}\ln\,\xi\right] \tag{23}$$

to the axion polarization operator. Here $\alpha_a=g_{ae}^2/4\pi$ (see Eq. (4)), and the standard variable[17] ξ was introduced as

$$\tau=-\frac{(1-\xi)^2}{4\xi}, \tag{24}$$

which is convenient for analytical continuation in $l^2=4m_e^2\tau$.

For $\tau>1$, a channel is open for axion decay into an electron- positron pair $(a\to e^-e^+)$ in a magnetic field. Its rate w for a real axion is related with the imaginary part of the polarization operator on the mass shell by the well-known relation

$$w=-\frac{1}{\omega}\mathrm{Im}\,\Pi_R^{(F)}=\alpha_a\frac{m_e^2}{\omega}\frac{H}{H_0}\times\exp\left(-\frac{k_\perp^2}{2h}\right)\theta(\tau-1)\left(1-\frac{1}{\tau}\right)^{-1/2}, \tag{25}$$

where ω is the axion energy.

This result, which follows from Eq. (23) with $\xi=|\xi|\exp(i\pi)$ (see Eq. (24)), is identical to the result obtained in Ref. 13 on the basis of a calculation of the elastic scattering amplitude of an axion in a magnetic field. It can also be found immediately, taking account of Eq. (17), from the representation (22):

$$\mathrm{Im}\,F_R=-\frac{l^2}{2}\int_0^1 dv\;\delta[1-(1-v^2)\tau]=-m_e^2\theta(\tau-1)\left(1-\frac{1}{\tau}\right)^{-1/2}. \tag{26}$$

Let us consider the contribution M (19) of the medium to the axion polarization operator. We note that it does not renormalize.[15] Integrating over p_z in Eq. (19), using the delta function and taking account of Eqs. (13) and (21), gives

$$M=-\frac{m_e^2}{2\pi}l^2\int_{m_e}^{\mu}\frac{d\varepsilon}{q}[D(l,p)+D(-l,p)$$

$$+D(l,\tilde{p})+D(-l,\tilde{p})], \tag{27}$$

$$D(l,p)=[l^2-2(lp)+i0]^{-1}.$$

Here ε is the energy of electrons in the medium, $q=\sqrt{\varepsilon^2-m_e^2}$, and the two-dimensional scalar products are $lp=k_0\varepsilon-k_zq$ and $l\tilde{p}=k_0\varepsilon+k_zq$.

The imaginary part of the expression (27) is determined using Sokhotskiĭ's formula

$$\frac{1}{x+i0}=\mathrm{P}\frac{1}{x}-i\pi\delta(x), \tag{28}$$

where P signifies a principal value. From Eqs. (27) and (28) we obtain on the mass shell

$$\mathrm{Im}\,M=\frac{m_e^2}{2}\theta(\tau-1)[\theta(\mu-\varepsilon_+)+\theta(\mu-\varepsilon_-)], \tag{29}$$

$$\varepsilon_\pm=\frac{\omega}{2}\pm\frac{k_z}{2}\left(1-\frac{1}{\tau}\right)^{1/2}.$$

Here $\varepsilon_\pm$ are the roots of the equations $l^2-2\omega\varepsilon\pm2k_zq=0$.

From Eqs. (17), (26), and (29) we find, taking account of the expression (25), the rate

$$w_M=\frac{1}{2}[\theta(\varepsilon_+-\mu)+\theta(\varepsilon_--\mu)]w, \tag{30}$$

where w is the decay rate (25) in the absence of a medium, for the axion decay into an e^-e^+ pair in the presence of a magnetized degenerate electron gas. We underscore that the imaginary part of the contribution (29) of the medium is positive, and summed with the negative field contribution (26) it gives a blocking Pauli factor $1-\theta(x)=\theta(-x)$ in Eq. (30). It forbids electron production inside a filled Fermi sphere (for $\varepsilon_\pm<\mu$).

Taking account of Eq. (28), we obtain for the real part of Eq. (27) on the mass shell the representation

$$\mathrm{Re}\,M=-\frac{m_e^2}{\pi}\tau\int_0^{\lambda}dx\left[\frac{1}{\tau-\cosh^2(x-\psi)}\right.$$

$$\left.+\frac{1}{\tau-\cosh^2(x+\psi)}\right]. \tag{31}$$

Here the substitution of the variable $\varepsilon\to x$ was used: $\varepsilon=m_e\cosh x$ and $q=m_e\sinh x$, and the parameters λ and ψ defined as

$$\cosh\lambda=\frac{\mu}{m_e},\quad \tanh\psi=\frac{k_z}{\omega}. \tag{32}$$

were introduced. The integral (31) can be expressed in terms of elementary functions.

We shall confine our attention below to the limiting cases that are of interest for astrophysical applications.

6. For an axion on the mass shell

$$l^2=4m_e^2\tau=\omega^2-k_z^2=m_a^2+k_\perp^2>0, \tag{33}$$

and the condition (14) gives $k_\perp^2\ll h$, so that $\exp(-k_\perp^2/2h)\simeq1$. We note that the imaginary part of the polarization operator is formed by the contribution of real electrons and positrons, and the expression for it holds under the weaker condition $k_\perp^2<2h$. Therefore the exponential factor can be retained in Eq. (25).

For $\tau\ll1$ (substantially below the threshold of the decay process $a\to e^-e^+$), we find from Eqs. (22), (31), and (17)

$$\Pi_R=\Pi_R^{(F)}+\Pi^{(M)}=\frac{\alpha_a}{\pi}m_e^2\frac{H}{H_0}\tau\left\{\nu_++\nu_-+\tau\right.$$

$$\left.\times\left[\nu_++\nu_--\frac{1}{3}(\nu_+^3+\nu_-^3)-\frac{4}{3}\right]\right\}. \tag{34}$$

Here

$$\nu_\pm=\tanh(\lambda\pm\psi)=\frac{\nu\omega\pm k_z}{\omega\pm\nu k_z},$$

$$\nu=\tanh\lambda=\left[1-\left(\frac{m_e}{\mu}\right)^2\right]^{1/2}.$$

In this case the medium makes the main contribution ($\sim\tau$). The purely field contribution is $\sim\tau^2$. We note that if the axion moves in the direction of the field $\mathbf{H}(k_\perp=0)$, then according to Eqs. (33) and (34) $\Pi_R\to0$ in the limit of a massless axion ($m_a\to0$).

At high energies ($\tau\gg1$) we obtain for the polarization operator the asymptotic representation

$$\Pi_R=\frac{\alpha_a}{\pi}m_e^2\frac{H}{H_0}\left[4\tau+\ln(4\tau)-4\,\mathrm{Arccosh}\frac{\mu}{m_e}-i\pi\right], \tag{35}$$

and the field contribution predominates.

Let us write the dispersion relation (7) in the form

$$\omega^2=k_\perp^2+k_z^2+m_a^2+\Pi_R(k). \tag{36}$$

It follows from Eqs. (34)–(36) that in a magnetized medium a radiation shift of the axion mass is generated — a dynamic mass, whose square, according to the definition in Ref. 15, is

$$\delta m_a^2=\mathrm{Re}\,\Pi_R.$$

For $\tau\gtrsim1$ we obtain the estimate

$$\delta m_a\sim g_{ae}m_e\left(\frac{H}{H_0}\tau\right)^{1/2}. \tag{37}$$

Assuming $k_\perp\gtrsim m_e$, we obtain

$$\delta m_a\gtrsim10^6g_{ae}\left(\frac{k_\perp}{1\,\mathrm{MeV}}\right)\left(\frac{H}{10^{13}\,\mathrm{G}}\right)^{1/2}\ \mathrm{eV}. \tag{38}$$

For $g_{ae}\sim10^{-13}$ (Refs. 4 and 10) and $H\gtrsim10^{17}$ G (such fields[18,19] and even $H\sim10^{18}-10^{20}$ G (Ref. 20) can exist in the interior regions of neutron stars), Eq. (38) gives $\delta m_a\gtrsim10^{-5}$ eV.

The chemical potential μ of a degenerate gas of electrons occupying the ground-state Landau level ($n=0$) in a magnetic field is related with the electron density n_e by the well-known relation

$$n_e=\frac{hp_F}{2\pi^2}, \tag{39}$$

where $p_F=\sqrt{\mu^2-m_e^2}$ is the Fermi momentum. Writing Eq. (12) in the form

$$\frac{H}{H_0}>\frac{1}{2}\left(\frac{p_F}{m_e}\right)^2, \tag{40}$$

we obtain, taking account of Eq. (39), an upper limit on the density

$$n_e<\frac{\lambda\!\!\!^{-}_e{}^3}{\sqrt{2}\,\pi^2}\left(\frac{H}{H_0}\right)^{3/2}, \tag{41}$$

where $\lambda\!\!\!^{-}_e=1/m_e$ is the electron Compton wavelength. For $H=2\times10^{17}$ G Eqs. (40) and (41) give $p_F<50$ MeV and $n_e<10^{36}$ cm^{-3}. Next, let $T\sim10^{10}$ K ~1 MeV and $k_\perp\gtrsim T$. Then the conditions (11), (12), and (14) can be satisfied and the estimate (38) can be justified.

In summary, under the conditions of strongly magnetized neutron stars a dynamic axion mass, which can fall within the existing limits on the axion mass[3,4,12] — 10^{-5} eV $\lesssim m_a\lesssim10^{-2}$ eV — is generated. Therefore $\delta m_a\sim m_a$ and the dispersion relation (36) differs appreciably from the vacuum relation ($k^2=m_a^2$). This must be taken into account, for example, when investigating the resonant conversion of a plasmon into an axion ($\gamma\to a$) in a magnetic field as a result of the crossing of the corresponding dispersion curves (as already noted above, this process in fields $H\ll H_0$ and in the absence of the direct coupling (1) was studied in Ref. 12). We also note that the rate (25) of the decay $a\to e^-e^+$ in a magnetic field has a square-root threshold singularity (as $\tau\to1+0$). This singularity can be removed by taking into account accurately the dispersion law of an axion near threshold, and the decay rate is found to be finite:[13] $w\sim m_e(\alpha_a H/H_0)^{2/3}$. A detailed analysis of the same threshold singularity (of cyclotron resonance) in a magnetic field and its elimination for the photon decay process ($\gamma\to e^-e^+$) was given earlier in Ref. 21, where, specifically, it is underscored that the indicated singularity can be explained by the quantization of the phase space of charged particles in a magnetic field.

We thank O. F. Dorofeev, P. A. Éminov, and V. Ch. Zhukovskiĭ for helpful discussions of the results.

*)E-mail: borisov@ave.phys.msu.su

[1] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[2] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978); F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[3] Particle Data Group: R. M. Barnett *et al.*, Phys. Rev. D **50**(1) (part I) 1 (1996).
[4] G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
[5] V. M. Lipunov, *Astrophysics of Neutron Stars* (Nauka, Moscow, 1987).
[6] A. V. Averin, A. V. Borisov, V. Ch. Zhukovskiĭ, and A. A. Él'sabbakh, Preprint No. 3/1993 (Department of Physics, Moscow State University, 1993).
[7] A. V. Borisov and K. V. Zhukovskiĭ, Yad. Fiz. **58**, 1298 (1995) [Phys. At. Nucl. **58**, 1218 (1995)].
[8] A. V. Borisov and V. Yu. Grishina, Vest. MGU, Fiz., astron., No. 4, 24 (1996).
[9] A. V. Borisov and V. Yu. Grishina, Zh. Éksp. Teor. Fiz. **110**, 1575 (1996) [JETP **83**, 868 (1996)].
[10] A. V. Borisov and V. Yu. Grishina, Zh. Éksp. Teor. Fiz. **106**, 1553 (1994) [JETP **79**, 837 (1994)].
[11] M. Kachelriess, C. Wilke, and G. Wunner, Phys. Rev. D **56**, 1313 (1997).
[12] N. V. Mikheev, G. Raffelt, and L. A. Vassilevskaya, E-print archive hep-ph/9803486.
[13] N. V. Mikheev and L. A. Vassilevskaya, Phys. Lett. B **410**, 203 (1997).
[14] N. V. Mikheev and L. A. Vassilevskaya, Phys. Lett. B **410**, 207 (1997); L. A. Vassilevskaya, N. V. Mikheev, and A. Ya. Parkhomenko, Yad. Fiz. **60**, 2224 (1997) [Phys. At. Nucl. **60**, 2041 (1997)].
[15] A. V. Borisov, A. S. Vshivtsev, V. Ch. Zhukovskiĭ, and P. A. Éminov, Usp. Fiz. Nauk **167**, 241 (1997).
[16] V. V. Skobelev, Izv. Vyssh. Uchebn. Zaved. Fiz. No. 10, 142 (1975); V. V. Skobelev, Doctoral Dissertation in Physicomathematical Sciences (MIIGAiK, Moscow, 1982).
[17] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Pergamon Press, New York) [Russian original, Nauka, Moscow, 1989, p. 563].
[18] I. M. Ternov and O. F. Dorofeev, Fiz. Élement. Chastits At. Yadra **25**, 5 (1994) [Phys. Part. Nuclei **25**, 1 (1994)].
[19] M. Bocquet, S. Bonazzola, E. Gourgoulhon, and J. Novak, Astron. Astrophys. **301**, 757 (1995).
[20] D. Bandyopadhyay, S. Chakrabarty, and S. Pal, Phys. Rev. Lett. **79**, 2176 (1997).
[21] A. E. Shabad, Tr. FIAN **192**, 5 (1988).

Translated by M. E. Alferieff

The inertial mass tensor of a polaron in an isotropic medium

A. É. Myasnikova*) and É. N. Myasnikov

Rostov Teachers' Training University, 344082 Rostov-on-Don, Russia
(Submitted 21 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 180–186 (January 1999)

Owing to a funadmentally erroneous approach to calculations of the effective polaron mass (calculations that use a model without spatial dispersion of the lattice polarizability), the polaron inertial mass has never before been distinguished from the mass as a measure of kinetic energy. In this paper we derive an expression for the tensor of the inertial mass of a large polaron. The tensor is found to be fully determined by two components: the longitudinal component, corresponding to the case where the force acting on the polaron is parallel to the polaron velocity, and the transverse component, corresponding to the case where the acceleration is perpendicular to the polaron velocity. The components of the polaron inertial mass tensor depend quasirelativistically on the polaron velocity due to the quasirelativistic compression of the polarization field in the direction of motion, which constitutes the effect of spatial dispersion of the lattice polarizability. We derive a formula that approximates the dependence of the components of the polaron mass tensor on all the parameters: the frequency and dispersion of the phonons, the polaron velocity, and the effective dielectric constant. [S1063-7761(99)01601-7]

1. INTRODUCTION

The investigation of the problem of the effective mass of an autolocalized charge carrier (a large polaron) began when it was first demonstrated that such carriers can exist[1] and still continues.[2–7]. The problem is interesting both from the standpoint of experimentally observing polaron-related effects and because of the potential of polaron electronics and in connection with the possibility of realizing high-T_c bipolaron superconductivity. Recently we derived expressions for the ''energy'' effective polaron mass (commonly known as the effective polaron mass)[6] and the polaron longitudinal inertial mass[7] in a model with spatial dispersion of the lattice polarizability. As shown in Ref. 8, only in such a model is the polaron mobile, i.e., not destroyed in its motion through the crystal, if its velocity v is less than the minimum phase velocity u of the phonons responsible for carrier autolocalization (there is no autolocalized state when $v>u$). The results of earlier studies that ignored the spatial dispersion of lattice polarizability[2–4] suggest that the polaron inertial and energy masses coincide and depend neither on the polaron velocity nor on the dispersion of the phonons participating in the formation of the polaron. Only in Ref. 5 was an approximate expression for the polaron energy mass derived for the case of weak dispersion of the phonon branch (small u).

The research done in Refs. 6 and 7 demonstrated that in contrast to ordinary ideas, the polaron inertial and energy masses differ when the polaron velocity v is finite (the difference is the greater the higher the values of v and u) and strongly depend on the polaron velocity v and the value of the minimum phase velocity u of the photons participating in polaron formation.

However, contrary to what was assumed in Ref. 7, the force acting on a polaron moving in a medium is not always longitudinal. An example is the Hall effect. To solve the problems of polaron electronics we must generally know the polaron inertial mass tensor. In other words, the results of Ref. 7 must be augmented by a ''transverse'' effective polaron mass. Indeed, even in Ref. 9, where the necessity of allowing for the spatial dispersion of the lattice polarizability in describing the motion of the cloud of polarization charge together with the charged particle generating this charge was demonstrated for the first time, it was shown that the polarization field of a moving point charge undergoes quasirelativistic compression and resembles a disk. When the polaron is in motion, such compression leads to effects similar to relativistic effects: the velocity dependence of the energy polaron mass[6] and the inertial polaron mass[7] due to a change in the degree of compression of the polarization force in the direction of motion. If there is also a nonlongitudinal field acting on the polaron, the disk rotates. This effect differs from the result of action of a longitudinal force, so that it is natural to expect that the inertial polaron mass differs from the longitudinal polaron mass. The present paper studies the dependence on the polaron velocity of the transverse inertial mass. On the basis of this study we construct the polaron inertial mass tensor in an isotropic medium.

2. THE POLARON TRANSVERSE INERTIAL MASS

The polaron inertial mass can be obtained from the time derivative of the polaron momentum P, assuming that the velocity is time-dependent, $v=v(t)$. If the z axis is directed parallel to the polaron velocity, the component m_{zz} of the effective mass tensor can be obtained by assuming that the force acting on the polaron is directed along the same z axis:

$$m_{zz}^{**}=\frac{dP}{dt}\left(\frac{dv}{dt}\right)^{-1}. \tag{1}$$

An expression for this (arbitrary) component of the polaron inertial mass tensor was derived in Ref. 7. In the local system of coordinates tied to the instantaneous directions of the velocity and acceleration the polaron inertial mass tensor is always diagonal. Its components $m_{xx}=m_{yy}$ (the purely transverse inertial mass) can be calculated by assuming that the force is directed along the x (or y) axis, with the z axis directed along the polaron velocity.

As noted in Ref. 7, the total polaron momentum can be written as the sum of the average value of the charge-carrier momentum, which can be assumed equal to m^*v (where m^* is the effective mass of the charge carrier in the crystal), and the average momentum of the phonon participating in the formation of the polaron. According to Ref. 7, the second term can be written as follows:

$$\mathbf{p}_{\rm ph}=\int \hbar\mathbf{k}\left[\frac{\Omega(k)\beta}{2\hbar}\mathbf{P}_k\cdot\mathbf{P}_{-k}+\frac{1}{2\hbar\Omega(k)\beta}\mathbf{T}_k\cdot\mathbf{T}_{-k}+\frac{i}{2\hbar}\mathbf{P}_k\cdot\mathbf{T}_{-k}-\frac{i}{2\hbar}\mathbf{P}_{-k}\cdot\mathbf{T}_k\right]\frac{d\mathbf{k}}{(2\pi)^3}, \tag{2}$$

where $\mathbf{P}_k$ is the Fourier transform of the polarization vector, $\mathbf{T}_k=\beta\mathbf{P}_k$, $\beta=4\pi\varepsilon^*/\Omega^2$, Ω is the frequency of longitudinal optical vibrations at the center of the Brillouin zone, $\Omega(k)$ is the law of dispersion of such vibrations, and $1/\varepsilon^*=1/\varepsilon_\infty-1/\varepsilon_0$ is the reciprocal effective dielectric constant.[2]

We assume that the force acting on the polaron is perpendicular to the polaron velocity and is such that the radius of the curvature of the polaron's path is much larger than the polaron radius, so that the polaron trajectory can be considered a straight line if we examine sections of the path of order the polaron radius. Then $\mathbf{T}_k=i(\mathbf{k}\cdot\mathbf{v})\beta\mathbf{P}_k$. If we also allow for the fact that $\mathbf{P}_{-k}=-\mathbf{P}_k$, we can write an expression for the average phonon momentum:

$$\mathbf{p}_{\rm ph}=\int\frac{d\mathbf{k}}{(2\pi)^3}\frac{\mathbf{k}\beta\Omega(k)}{2}P_k^2\left[-1-2\frac{\mathbf{k}\cdot\mathbf{v}}{\Omega(k)}-\left(\frac{\mathbf{k}\cdot\mathbf{v}}{\Omega(k)}\right)^2\right]. \tag{3}$$

The Fourier transform of the vector of the polarization generated by a charge carrier moving in a straight line with a velocity v smaller than the minimum phase velocity v of the phonons participating in the formation of the polaron (only in this case will the carrier wave function $\psi(\mathbf{r},t)$ be localized in space) has the form[7]

$$\mathbf{P}_k=\frac{e}{\varepsilon^*}\Omega^2\frac{i\mathbf{k}}{|k|^2}\frac{1}{(\mathbf{k}\cdot\mathbf{v})^2+\Omega^2(k)}\psi_k^2, \tag{4}$$

where ψ_k^2 is the Fourier transform of the square of the wave function of the charge carrier in the polaron. The wave function $\psi(\mathbf{r},t)$ of the charge carrier in the polaron can be obtained by minimizing the carrier energy functional, in which the polarization charge density $\rho(\mathbf{r},t)$ is expressed in terms of $\psi(\mathbf{r},t)$ as follows:[8]

$$\rho(\mathbf{r},t)=\frac{e}{\varepsilon^*}\Omega^2\int G(\mathbf{r}-\mathbf{r}',t)\psi^2(\mathbf{r}')\,d^3r', \tag{5}$$

where $G(\mathbf{r},t)$ is the Green's function of the equation of motion for the density of the polarization charge in the polariton. Equation (5) shows that the physical meaning of the function $G(\mathbf{r},t)$ is that this function is the density of the polarization charge generated by a charged point particle moving with a velocity v (this is how this density was obtained in Ref. 9). When the dispersion law of the phonon branch has the form $\Omega^2(k)=\Omega^2+u^2k^2$ in a cylindrical system of coordinates whose x axis is directed along the particle velocity, the function $G(\mathbf{r},t)$ has the form[9]

$$G(\mathbf{r},t)=\begin{cases}\dfrac{\exp\{-\Omega[(z-vt)^2/\beta_1^2+r^2]^{1/2}/u\}}{4\pi u^2\beta_1[(z-vt)^2/\beta_1^2+r^2]^{1/2}}, & v<u,\quad \beta_1^2=1-\dfrac{v^2}{u^2};\\[2ex] \dfrac{\cos\{\Omega[(z-vt)^2/\beta_2^2-r^2]^{1/2}/u\}}{2\pi u^2\beta_2[(z-vt)^2/\beta_2^2-r^2]^{1/2}}, & v>u,\quad \begin{matrix}z-vt<0,\\ r<|z-vt|/\beta_2;\end{matrix}\\[2ex] 0, & v>u,\quad \begin{matrix}z-vt<0,\\ r>|z-vt|/\beta_2,\end{matrix}\\[2ex] & z-vt>0,\quad \beta_2^2=\dfrac{v^2}{u^2}-1\end{cases}. \tag{6}$$

The physical meaning of u is that u is the minimum phase velocity of the phonons participating in polaron formation. Equation (6) shows that only when $v<u$ holds will the polarization charge generated by a moving charged point particle be localized, with the characteristic size of the localization region equal to $\sqrt{u^2-v^2}/\Omega$ in the direction of motion and u/Ω in the perpendicular direction. The typical values of u/Ω are much smaller than the polaron radius in the absence of spatial dispersion (e.g., at $\Omega=100\,\mathrm{cm}^{-1}$ and $u=2\times10^5\,\mathrm{cm\,s}^{-1}$ the ratio u/Ω is smaller than one ångström), i.e., the ''smearing'' of the carrier wave function in the polaron due to spatial dispersion is much smaller than the smearing due to the wave properties of the carrier. Hence, to simplify the calculations of the polaron mass, we found it appropriate to ignore the variation of the carrier wave function in the polaron due to spatial dispersion and allowed only for the effect of this factor on the polarization charge and to use in (4) the expression for the Fourier transform of the Pekar wave function given in Ref. 7. Substituting (4) in (3) and differentiating $\mathbf{p}_{\rm ph}$ with respect to time, we arrive at

$$\frac{d\mathbf{p}_{\rm ph}}{dt}=\frac{e^2\Omega^4\beta}{\varepsilon^{*2}(2\pi)^3}\int\frac{\mathbf{k}\,d\mathbf{k}}{|k|^2}\frac{\Omega(k)(\psi_k^2)^2}{[\Omega^2(k)-(\mathbf{k}\cdot\mathbf{v})^2]^3}\frac{d(\mathbf{k}\cdot\mathbf{v})}{dt}\times\left[\mathbf{k}\cdot\mathbf{v}\left(3+\frac{(\mathbf{k}\cdot\mathbf{v})^2}{\Omega^2(k)}\right)+\Omega(k)\left(1+3\frac{(\mathbf{k}\cdot\mathbf{v})^2}{\Omega^2(k)}\right)\right]. \tag{7}$$

In accordance with what has been said earlier, we select a system of coordinates in such a way that the z axis is directed along the polaron velocity and the x axis, along the force acting on the polaron. Then

$$\mathbf{k}\cdot\mathbf{v}=k_z v, \quad \frac{d(\mathbf{k}\cdot\mathbf{v})}{dt}=k_x v\frac{ds}{dt}, \tag{8}$$

where ds/dt is the derivative of the unit vector determining the velocity's direction. According to (1), the transverse effective mass can be obtained by the formula

$$m_{xx}=\frac{dP}{dt}\left(v\frac{ds}{dt}\right)^{-1}. \tag{9}$$

Plugging (9) into the expression (7) for $d\mathbf{p}_{\rm ph}/dt$, allowing for (8), and bearing in mind that the first term in (7) is odd in k_z and thus vanishes under integration, we arrive at an expression for the polaron transverse inertial mass:

$$m_{xx}^{**}=m^*+\frac{4e^2\Omega^2}{\varepsilon^*\pi^2u^4}\int_0^\infty\frac{k_x^2\,dk_x\,dk_y\,dk_z}{k_x^2+k_y^2+k_z^2}$$
$$\times\frac{k_z^2(1+3v^2/u^2)+k_x^2+k_y^2+\Omega^2/u^2}{(k_z^2(1-v^2/u^2)+k_x^2+k_y^2+\Omega^2/u^2)^3}(\psi_k^2)^2, \tag{10}$$

where the dispersion law of the phonon branch has been taken in the form $\Omega^2(k)=\Omega^2+u^2k^2$. The expression for the polaron's longitudinal inertial mass derived in Ref. 7 differs somewhat from (10):

$$m_{zz}^{**}=m^*+\frac{4e^2\Omega^2}{\varepsilon^*\pi^2u^4}\int_0^\infty\frac{dk_x\,dk_y\,k_z^2\,dk_z}{k_x^2+k_y^2+k_z^2}$$
$$\times\frac{k_z^2(1+3v^2/u^2)+k_x^2+k_y^2+\Omega^2/u^2}{(k_z^2(1-v^2/u^2)+k_x^2+k_y^2+\Omega^2/u^2)^3}(\psi_k^2)^2. \tag{11}$$

As noted earlier, $m_{xx}=m_{yy}$ and m_{zz} are the only finite components of the tensor of the effective polaron inertial mass in the local system of coordinates linked to the instantaneous directions of the velocity and acceleration, so that Eqs. (10) and (11) fully determine the tensor of the polaron effective inertial mass.

3. DISCUSSION

Figure 1 shows the dependence of the polaron longitudinal inertial mass m_{zz}^{**} (curves *1*, *1′*, and *1″*), the polaron energy mass[6] $m_{\rm en}^{**}$ (curves *2*, *2′*, and *2″*), and the polaron transverse inertial mass $m_{xx}^{**}=m_{yy}^{**}$ (curves *3*, *3′*, and *3″*) on the polaron velocity for three values of the minimum phase velocity u of the phonons participating in the formation of the polaron. Curves *1*, *2*, and *3* correspond to $u=5\times10^5\,\mathrm{cm\,s^{-1}}$, curves *1′*, *2′*, and *3′* to $u=10^6\,\mathrm{cm\,s^{-1}}$, and curves *1″*, *2″*, and *3″* to $u=2\times10^6\,\mathrm{cm\,s^{-1}}$. The values of the other parameters of the medium for the case shown in Fig. 1 are $1/\varepsilon^*=0.27$, $\Omega=6.78\times10^{13},\mathrm{s^{-1}}$, and $m^*=m_e$. We see that the longitudinal inertial mass increases with polaron velocity much faster than the energy mass and the transverse inertial mass, as also happens in the relativistic case. The transverse inertial mass coincides almost perfectly with the energy mass, but the masses are not equal, in contrast to the relativistic case, where they coincide perfectly. What also sets the polaron mass apart from the mass of a relativistic particle is that the polaron mass increases to a finite value as $v\to u$ rather than to infinity (this finite value depends on u).

The behavior of the polaron longitudinal, transverse, and energy masses can easily be understood if we examine formula (6) for the Green's function $G(\mathbf{r},t)$ of the equation of motion for the polarization charge density in the polaron. As noted earlier, $G(\mathbf{r},t)$ is the density of the polarization charge generated by a charged point particle moving with a velocity v. For $v<u$ the polarization charge, as Eq. (6) implies, is localized within a region whose size in the direction of motion is of order $\sqrt{u^2-v^2}/\Omega$ and perpendicular to that direction, of order u/Ω. Thus, as the polaron velocity approaches its critical value ($v=u$), the polarization charge generated by each point section of the charge carrier distribution in the polaron undergoes a quasirelativistic compression in the direction of motion. As a result all polaron masses increase in the limit $v\to u$ as,obviously, negative powers of the difference $1-v^2/u^2$, which can be interpreted as the ratio of the squares of the smearing of the polarization charge parallel and perpendicular to that direction due to spatial dispersion,:

$$1-\frac{v^2}{u^2}=\frac{u^2-v^2}{u^2}=\frac{u^2-v^2}{\Omega^2}\frac{\Omega^2}{u^2}. \tag{12}$$

The smearing parameter in the direction of motion tends to zero as $v\to u$. However, in the limit $v\to u$, the polaron mass tends to a finite value, since in addition to smearing related to spatial dispersion the polaron has smearing R related to the wave properties of the carrier. If we introduce this pa-

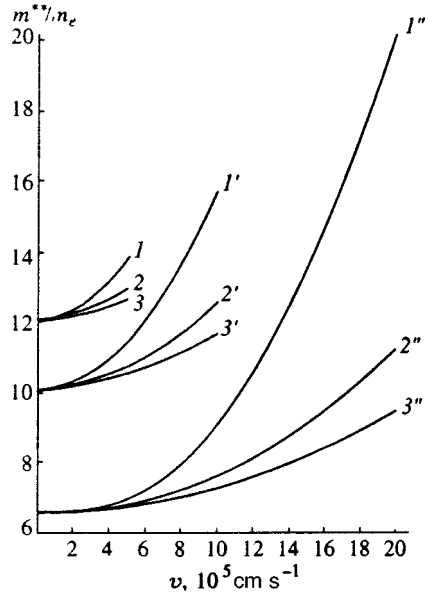


FIG. 1. Dependence of the polaron longitudinal inertial mass m_{zz}^{**} (curves *1*, *1′*, and *1″*), the polaron energy mass $m_{\rm en}^{**}$ (curves *2*, *2′*, and *2″*), and the polaron transverse inertial mass $m_{xx}^{**}=m_{yy}^{**}$ (curves *3*, *3′*, and *3″*) on the polaron velocity for three values of the minimum phase velocity u of the phonons participating in the formation of the polaron. Curves *1*, *2*, and *3* correspond to $u=5\times10^5\,\mathrm{cm\,s^{-1}}$, curves *1′*, *2′*, and *3′* to $u=10^6\,\mathrm{cm\,s^{-1}}$, and curves *1″*, *2″*, and *3″* to $u=2\times10^6\,\mathrm{cm\,s^{-1}}$. The values of the other parameters are $1/\varepsilon^*=0.27$, $\Omega=6.78\times10^{13}\,\mathrm{s^{-1}}$, and $m^*=m_e$.

rameter R, which characterizes the polaron size in the absence of spatial dispersion, into the numerator and denominator of (12), we arrive at the expression

$$\frac{u^2-v^2+R^2\Omega^2}{\Omega^2}\,\frac{\Omega^2}{u^2+R^2\Omega^2}=1-\frac{v^2}{u^2+R^2\Omega^2}, \tag{13}$$

which leads to a finite value for the polaron mass at $v=u$, while at $v=0$ the right-hand side of Eq. (13) is unity. Calculations have shown that the velocity dependence of the polaron mass can be approximated by the following formulas:

$$m_{zz}=m_0(u)\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-3/2},$$
$$m_{xx}=m_0(u)\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-1/2}, \tag{14}$$
$$m_{\rm en}=m_0(u)\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-1.3/2},$$

where R is a parameter dependent on the effective dielectric constant ε^*. By comparing the values of mass calculated by (14), by (10) and (11), and in Ref. 6 we can define R as $R=0.751/b$, where $b=m^*e^2/2\hbar^2\varepsilon^*$ is the value of the parameter from Ref. 2 at which the energy functional of a polaron with the trial wave function of the form

$$\psi(r)=\frac{b^{2/3}}{\sqrt{7\pi}}(1+br)e^{-br} \tag{15}$$

attains its minimum.

Obviously, the dependence of the polaron mass on the minimum phonon phase velocity u must contain the ratio of u to the one velocity parameter $R\Omega$ remaining at $v=0$, i.e., $m_0(u)$ must depend on $u/R\Omega$. Calculations yield

$$m_{zz}^{**}=\frac{m^*}{192\varepsilon^{*4}}\left(\frac{m^*c^2}{h\Omega}\right)^2\left(\frac{e^2}{\hbar c}\right)^4\left(1+\left[\frac{u}{R\Omega}\right]^2\right)^{-1}\times\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-3/2},$$

$$m_{xx}^{**}=\frac{m^*}{192\varepsilon^{*4}}\left(\frac{m^*c^2}{h\Omega}\right)^2\left(\frac{e^2}{\hbar c}\right)^4\left(1+\left[\frac{u}{R\Omega}\right]^2\right)^{!-1}\times\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-1/2},$$

$$m_{\rm en}^{**}=\frac{m^*}{192\varepsilon^{*4}}\left(\frac{m^*c^2}{h\Omega}\right)^2\left(\frac{e^2}{\hbar c}\right)^4\left(1+\left[\frac{u}{R\Omega}\right]^2\right)^{-1}\times\left(1-\frac{v^2}{u^2+R^2\Omega^2}\right)^{-1.3/2},$$

where the numerator is the Pekar polaron mass.[2,4]. These expressions approximate the values of the corresponding polaron masses to within 10%.

Thus, the velocity dependence of the polaron mass can be described by relativistic formulas if we allow for the fact that the polaron size in the direction of motion tends as $v\to u$, not to zero but to a finite value, which we characterize by the parameter R. The ratio of the smearing u/Ω due to spatial dispersion to the quantum smearing parameter R distinguishes between the polaron mass for $v\to 0$ and the Pekar mass: as this ratio increases the polaron mass decreases with $v\to 0$.

*)E-mail: rochal@phys.rnd.runnet.ru

[1] L. D. Landau and S. I. Pekar, Zh. Éksp. Teor. Fiz. **18**, 419 (1948).
[2] S. I. Pekar, *Studies in Electron Crystal Theory*, Rept. AEC-tr-5575, U.S. Atomic Energy Commission (1963).
[3] R. P. Feynman, Phys. Rev. **97**, 660 (1955).
[4] J. Appel, ''Polarons,'' in *Solid State Physics: Advances in Research and Applications*, H. Ehrenreich, F. Seitz, and D. Turnbull (Eds.), Vol. 21, Academic Press, New York (1968).
[5] A. S. Davydov and V. Z. Énol'skiĭ, Phys. Status Solidi B **143**, 167 (1987); Zh. Éksp. Teor. Fiz. **94**(2), 177 (1988) [Sov. Phys. JETP **67**, 313 (1988)].
[6] A. É. Myasnikova and E. N. Myasnikov, Zh. Eksp. Teor. Fiz. **112**, 278 (1997) [JETP **85**, 152 (1997)].
[7] A. E. Myasnikova and E. N. Myasnikov, Phys. Rev. B **56**, 5316 (1997).
[8] É. N. Myasnikov and A. P. Popov, Dokl. Akad. Nauk Ukr. SSSR A **5**, 73 (1980).
[9] A. E. Myasnikova, Phys. Rev. B **52**, 10 457 (1995).

Translated by Eugene Yankovsky

Magnetopolaron states involving confined phonons in a semiconductor quantum well

L. I. Korovin and I. G. Lang

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

S. T. Pavlov*)

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 22 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 187–204 (January 1999)

Energy splitting $\Delta E_{\rm res}$ in double magnetopolaron energy spectrum in rectangular quantum wells as functions of the well width d have been calculated. We have considered in the capacity of interaction leading to resonant coupling between electrons and phonons the interaction with confined phonons and (for comparison) with bulk LO phonons. We have obtained the conditions when the interaction with bulk phonons yields correct results. Calculations for AlAs/GaAs/AlAs and AlSb/InSb/AlSb structures have been performed. Alongside the parameter $\Delta E_{\rm res}$ for a polaron, whose resonant magnetic field is determined by the condition $\Omega=\omega_{L1}$, where Ω is the cyclotron frequency and ω_{L1} is the LO phonon frequency in the quantum well (A-polaron), we have calculated $\Delta E_{\rm res}$ for D- ($\Omega=2\omega_{L1}$) and F-polarons ($\Omega=3\omega_{L1}$), which is a factor of $\sqrt{2}$ and $\sqrt{3}$, respectively, smaller than $\Delta E_{\rm res}$ for the A-polaron. Since the splitting $\Delta E_{\rm res}$ for the A-polaron is very large (up to $0.2\hbar\omega_{L1}$), it is more convenient to study in experiments D- and F-polarons since their resonant magnetic fields are lower. We have predicted existence of ''weak'' magnetopolarons, in which the splitting is proportional to a higher power of Frölich's coupling constant α than $\alpha^{1/2}$. © *1999 American Institute of Physics.* [S1063-7761(99)01701-1]

1. INTRODUCTION

When a polaron state is formed in magnetic field, the contribution of electron–phonon coupling increases considerably when the resonant condition is satisfied, i.e., when the phonon frequency ω_{L1} and electron (hole) cyclotron frequency Ω are related by the formula

$$\omega_{L1}=j\Omega, \quad j=1,2,3,\ldots \tag{1}$$

In this case, resonant coupling between electron levels takes place. Condition (1) determines the crossing points between the energy levels of the electron–phonon system as functions of magnetic field (Fig. 1). Inclusion of magnetopolaron effects leads to anticrossing of energy levels. Anticrossing effects were detected in the interband magnetoabsorption spectra of InSb.[1–3]

Formation of polaron states takes place in both three-dimensional (3D) and quasi-two-dimensional (2D) systems. In both systems, these states have considerable effect on spectra of various magneto-optical processes, such as interband light absorption, cyclotron resonance, and Raman scattering (see, for example, the reviews[4–6]). The main difference between the two systems is in the electron (hole) spectrum. In 3D systems these are one-dimensional Landau bands, whereas in 2D structures these are discrete energy levels. As a result, the splitting between levels of the electron–phonon system due to anticrossing can vary: in the 3D case it is proportional to $\alpha^{2/3}$,[7] whereas in 2D structures the splitting is proportional to $\alpha^{1/2}$,[8–19] where α is Frölich's dimensionless electron–LO-phonon coupling constant.

Below we consider a single quantum well as a 2D system. Instead of bulk LO phonons, such systems have phonons of three types.[20–24] First, these are so-called half-space phonons, which are vibrations of the barrier material which do not penetrate into the quantum well. Second, there are interface phonons, which decay fast with the separation from the quantum well interfaces. Third, there are phonons confined within the quantum well. These vibrations do not penetrate into the barrier, and their amplitudes go to zero on the interfaces. Interaction between electrons and holes on one side, and three types of phonons on the other has been studied previously.[25–27] Das Sarma and Madhukar[10] calculated the spectrum of the A-magnetopolaron (Fig. 1) taking into account interaction of electrons with confined and interface phonons in an AlAs/GaAs/AlAs structure at three values of the well width (20 Å, 100 Å, and 200 Å). They came to a conclusion that the magnetopolaron spectrum in narrow quantum wells is controlled by interaction with interface phonons.

Many calculations of magnetopolaron spectra in a quantum well took into account only interaction with Frölich's bulk LO phonons. In other words, they took into account the effect of size-quantization in a quantum well on the electron (hole) spectrum, whereas the phonon spectrum was the same as in a homogeneous medium, as if the entire space were filled with the quantum well material. Strictly speaking, this approach is inconsistent and can yield only an approximate description of the magnetopolaron spectrum.

In the reported work we have investigated the conditions when Frölich's interaction with LO phonons can be used in

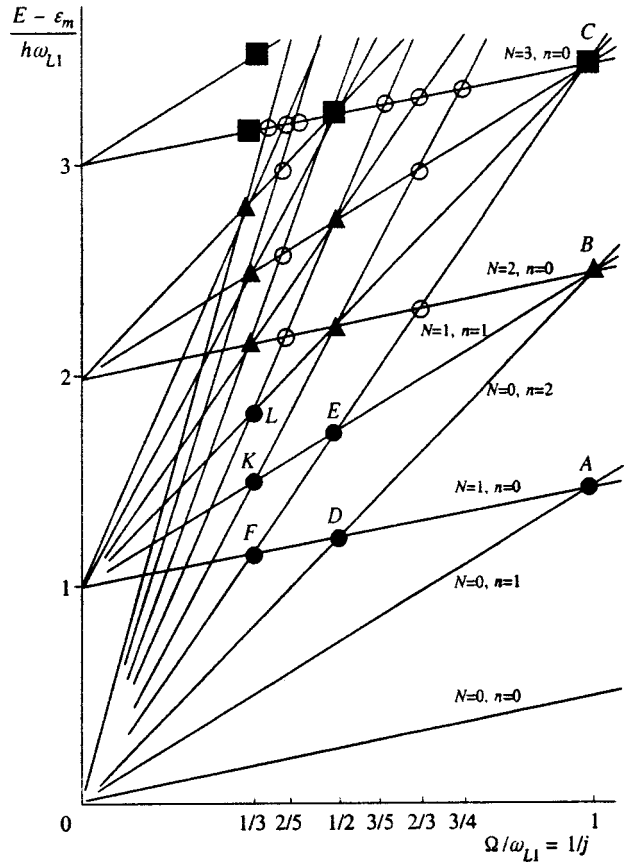


FIG. 1. Energy levels of an electron–phonon system in a quantum well in a strong magnetic field: Ω is the cyclotron frequency, ω_{L1} is the frequency of LO phonons in the well material, E is the electron energy, and ε_m is the size-quantized energy. Full circles mark double polarons, open circles denote weak polarons, triangles mark triple and squares four-fold polarons.

calculations of magnetopolaron spectra in quantum wells. With this end in view, we calculated the splitting of electron–phonon energy levels as functions of the well width d and number j. The j-dependence is interesting because at $j>1$ the resonant magnetic field is, in accordance with Eq. (1), a factor of j lower, which is easier to realize in an experiment. Since the electron and hole effective masses are different, the resonant coupling can involve either electrons or holes. Below we will consider for definiteness magnetopolarons including electrons, and spectra of hole magnetopolarons must be similar.

The paper is organized as follows. In Sec. 2 magnetopolarons are classified. Section 3 contains general expressions for the mass operator, which is included in the Dyson equation that determines the magnetopolaron spectrum. Interactions with both confined and bulk LO phonons will be considered. Section 4 is devoted to calculations of magnetopolaron spectra. Section 5 describes magnetopolaron spectra in the limiting case of wide quantum wells. In Sec. 6 we will discuss the results and draw conclusions.

2. CLASSIFICATION OF MAGNETOPOLARONS

Figure 1 shows energy levels of the electron–phonon system in a quantum well at a fixed size-quantization number m. The limiting frequencies of the confined and bulk LO phonons are assumed to be equal at ω_{L1}, and their dispersion is neglected. The abscissa is the ratio $\Omega/\omega_{L1}=1/j$, where

$$\Omega=|e|H/m_e c, \tag{2}$$

e is the electron charge, H is the magnetic field strength, c is the speed of light in vacuum, and m_e is the electron effective mass. The ordinate is the dimensionless energy E (in units of $\hbar\omega_{L1}$) of an electron measured with respect to ε_m, which is the mth level of size-quantization. The crossing points of these levels correspond to polaron states. Full circles mark double polarons, corresponding to crossing of only two levels. Let n be the number of the Landau level passing through a given crossing point when the number of phonons is $N=0$. Then the conditions for the existence of a double polaron are

$$2j>n, \quad n\geqslant j. \tag{3}$$

It is clear that $j=1$ corresponds to one double polaron denoted by A. The value $j=2$, i.e., $\Omega/\omega_{L1}=1/2$, corresponds to two double polarons, D and E, the number $j=3$, i.e., $\Omega\omega_{L1}=1/3$, to three double polarons, F, K, and L, and so on. The polarons that have the parameter Ω/ω_{L1} below 1/3 are not marked in Fig. 1. Above the double polarons, there are triple polarons corresponding to crossings of three energy levels, in the range of higher energies there are four-fold polarons, etc. The number of polarons of each sort at given j is j. The triple polarons in bulk crystals were first considered in Ref. 28 and in quantum wells in Ref. 19.

All the polarons mentioned above correspond to integer j. In addition, there are more crossing points in Fig. 1 marked by open circles. They correspond to fractional j. Since the levels crossing at these points have the difference between their numbers $\Delta N\geqslant 2$, one-phonon transitions between them are forbidden. Let us term these polarons weak. Since the levels cross at certain points, their splitting is inevitable, but the resulting splitting should be calculated taking into account virtual transitions via intermediate states. As a result, the energy splitting due to weak polarons is of higher order in α than $\alpha^{1/2}$.

If two or more values of the size-quantization number m are taken into account, the pattern of level crossing is complicated considerably. Previously we analyzed combined magnetopolarons in which two electron levels with different Landau numbers n and different m, or with different m and equal n are coupled by the electron–phonon interaction.[18] This paper is not concerned with combined polarons,[18] but considers only double polarons marked by full circles in Fig. 1. The polarons of the new types can be considered separately and, moreover, the pattern of energy levels in Fig. 1 can be used under the condition that the energy splitting between neighboring size-quantized levels are wider than the splitting $\Delta E_{\rm res}$ due to the polaron effect. Since the separations between the size-quantized energy levels are smaller at larger well widths d, there is an upper limit for d (see Sec. 6). Our theory applies to all numbers n and j that satisfy condition (3). Numerical calculations have been performed for A, D, and F polarons. For E, K, and L polarons, approximate values of the splitting are given. The letters B and C in Fig. 1 mark the three- and four-fold polarons, respectively, at $j=1$. The

pattern shown in Fig. 1 is based on the assumptions that the Landau levels are equidistant and excitonic effects can be neglected.

3. MODEL AND BASIC RELATIONS

We are considering a type I quantum well with a band gap E_g and a height ΔE_e of the electron barrier. The magnetic field $\mathbf{H}$ is perpendicular to the quantum well plane (aligned with the z-axis), and its vector potential $\mathbf{A}_0$ is expressed in the Landau gauge:

$$\mathbf{A}_0=(-yH,0,0). \tag{4}$$

The wave function of an electron not coupled to phonons has the form

$$\Psi_{m,n,k}(x,y,z)=\frac{1}{\sqrt{L_x}}e^{ikx}\varphi_n(y-y_k)\chi_m(z), \tag{5}$$

where

$$\varphi_n(y-y_k)=\frac{1}{\sqrt{2^n n!\sqrt{\pi}R_0}}\exp\left[-\frac{(y-y_k)^2}{2R_0^2}\right]H_n\left(\frac{y-y_k}{R_0}\right), \tag{6}$$

$$R_0^2=\frac{c\hbar}{|e|H},\quad y_k=-\frac{c\hbar k}{eH}, \tag{7}$$

k is the electron wave vector projection on the x-axis, $H_n(y)$ is a Hermite polynomial, L_x is the length to which the wave functions are normalized. In a rectangular quantum well of a finite depth, the function $\chi_m(z)$ has the form

$$\chi_m(z)=C_m(-1)^{(m-1)/2}\times\begin{cases}\cos(k_md/2)\exp(\kappa_m z), & z\leqslant 0,\\ \cos[k_m(z-d/2)], & 0\leqslant z\leqslant d,\\ \cos(k_md/2)\exp[-\kappa_m(z-d)], & z\geqslant d\end{cases} \tag{8}$$

for $m=1,3,5,\ldots$ and

$$\chi_m(z)=C_m(-1)^{m/2}\times\begin{cases}-\sin(k_md/2)\exp(\kappa_m z), & z\leqslant 0,\\ \sin[k_m(z-d/2)], & 0\leqslant z\leqslant d,\\ \sin(k_md/2)\exp[-\kappa_m(z-d)], & z\geqslant d\end{cases} \tag{9}$$

for $m=2,4,6,\ldots$, where d is the quantum well width, so that $0\leqslant z\leqslant d$,

$$C_m=\sqrt{\frac{2\kappa_m}{1+\kappa_m d\pm\cos(k_md)\pm(\kappa_m/k_m)\sin(k_md)}} \tag{10}$$

is the normalization constant, the upper signs are taken when m is odd and the lower signs when m is even. The parameter k_m, hence the energy ε_m^e of a level in the quantum well, is determined by the equations

$$\cot t=t/\sqrt{\beta_e^2-t^2},\quad m=1,3,5,\ldots, \tag{11}$$

$$\tan t=-t/\sqrt{\beta_e^2-t^2},\quad m=2,4,6,\ldots, \tag{12}$$

$$t=k_md/2,\quad \beta_e=Q_ed/2,\quad Q_e=\sqrt{2m_e\Delta E_e/\hbar^2},$$

$$\kappa_m=\sqrt{Q_e^2-k_m^2},\quad k_m=\sqrt{2m_e\varepsilon_m/\hbar^2}. \tag{13}$$

The electron energy measured with respect to the quantum well bottom is

$$E_{m,n}=\varepsilon_m+(n+1/2)\hbar\Omega. \tag{14}$$

In an infinitely deep quantum well ($\Delta E_e\to\infty$)

$$\varepsilon_m=\frac{\hbar^2\pi^2m^2}{2m_ed^2}. \tag{15}$$

The magnetopolaron energy spectrum is determined by poles of the one-particle electron Green's function[7]:

$$\varepsilon-E_{m,n}-\Sigma(m,n,\varepsilon)=0, \tag{16}$$

where $\Sigma(m,n,\varepsilon)$ is the mass operator.

The Hamiltonian of interaction between electrons and confined phonons has the form[26]

$$\mathscr{H}_C=\sum_{\mathbf{q}}\left\{\sum_{p=1,3,\ldots}C_{\mathbf{q},p}\cos\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right](a_{\mathbf{q},p}+a^+_{-\mathbf{q},p})+\sum_{p=2,4,\ldots}C_{\mathbf{q},p}\sin\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right](a_{\mathbf{q},p}+a^+_{-\mathbf{q},p})\right\}e^{i\mathbf{qr}} \tag{17}$$

for $0\leqslant z\leqslant d$ and $\mathscr{H}_C=0$ for $z<0$ and $z>d$. The factor $C_{\mathbf{q},p}$ is expressed as

$$C_{\mathbf{q},p}=-\hbar\omega_{L1}\sqrt{\frac{8\pi\alpha l}{S_0d[q^2+(p\pi/d)^2]}}, \tag{18}$$

where the dimensionless constant of coupling with confined phonons is

$$\alpha=\frac{e^2}{2\hbar\omega_{L1}l}\left(\frac{1}{\varepsilon_{\infty 1}}-\frac{1}{\varepsilon_{01}}\right),\quad l=\sqrt{\frac{\hbar}{2m_e\omega_{L1}}}, \tag{19}$$

ω_{L1} is the confined phonon frequency (its dispersion is neglected), ε_{01} ($\varepsilon_{\infty 1}$) is the static (high-frequency) permittivity of the quantum well material, $\mathbf{r}=(x,y)$ and $\mathbf{q}=(q_x,q_y)$ are the two-dimensional radius vector of the electron and the two-dimensional phonon wave vector, respectively, $a^+_{\mathbf{q},p}$ ($a_{\mathbf{q},p}$) is the creation (annihilation) operator of a confined phonon with wave vector $\mathbf{q}$ and quantum number p, which is a substitute for the wave vector projection q_z in the 3D case, and S_0 is the normalization area.

If two energy levels of the electron–phonon system cross, it suffices to take into account the simplest diagram (two vertices connected by electron and phonon lines) in the mass operator $\Sigma(m,n,\varepsilon)$. The contributions of diagrams with more vertices are smaller owing to the small coupling constant α. Using standard techniques for calculation of Feynman diagram contributions to the mass operator, we obtain the following expression:

$$\Sigma(m,n,\varepsilon)=\sum_{m_1,n_1}\sum_{\mathbf{q},p=1,3,\ldots}C^2_{\mathbf{q},p}\left[M^{(1)}_{m,m_1}\times(p)\frac{1+(-1)^{p+1}}{2}+M^{(2)}_{m,m_1}\times(p)\frac{1-(-1)^{p}}{2}\right]^2\frac{|I_{n,n_1}(\mathbf{q})|^2}{\varepsilon-E_{m_1,n_1}-\hbar\omega_{L1}+i\delta},$$

$$\delta\to 0. \tag{20}$$

The numerator of the last factor in the right-hand side of Eq. (20) is the squared absolute value of the matrix element of Hamiltonian (17) calculated using wave functions (5), and the denominator is determined by the zero-order Green's functions of the electron and phonon. The temperature is assumed to be low, so that no phonons are present, and the electron–phonon interaction leads only to emission of phonons. Equation (20) uses the following notation:

$$M^{(1)}_{m,m_1}(p)=\int_0^d dz\,\chi_m(z)\chi_{m_1}(z)\cos\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right],$$

$$p=1,3,5,\ldots, \tag{21}$$

$$M^{(2)}_{m,m_1}(p)=\int_0^d dz\,\chi_m(z)\chi_{m_1}(z)\sin\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right],$$

$$p=2,4,6,\ldots,$$

$$I_{n,n_1}(\mathbf{q})=\frac{1}{L_x}\int d^2r\,\exp\{i(k-k_1)x+i\mathbf{qr}\}\times\varphi_n(y-y_k)\varphi_{n_1}(y-y_{k_1}). \tag{22}$$

Although the functions $\chi_m(z)$ are nonvanishing outside the quantum well, integration in Eq. (21) is performed over the interval $0\leqslant z\leqslant d$, because $\mathscr{K}_C=0$ outside the quantum well. As follows from the expressions for $\chi_m(z)$, the integrands are products of sines and cosines. One can easily prove that the integral of the product of three sines or one sine and two cosines is zero, therefore, the numbers m and m_1 in $M^{(1)}_{m,m_1}(p)$ should be of the same parity, and in $M^{(2)}_{m,m_1}(p)$ they should have different parities. Since only the case $m=m_1$ will be discussed in what follows, only the terms with $M^{(1)}_{m,m}(p)$ are retained in Eq. (20). The following notation is introduced:

$$M^{(p)}_m=M^{(1)}_{m,m}(p),\quad p=1,3,5,\ldots \tag{23}$$

After performing integration over z in Eq. (21) in a quantum well of a finite depth, we have

$$M_m(p)=(-1)^{(p-1)/2}C^2_m d\,\pi p\left[\frac{1}{\pi^2p^2}\pm\frac{\cos(k_md)}{\pi^2p^2-4k_m^2d^2}\right], \tag{24}$$

where plus is taken in the case of odd m, and minus in the case of even m. After calculating the integral in Eq. (22), we obtain

$$|I_{n,n_1}(\mathbf{q})|^2=\frac{\min(n!,n_1!)}{\max(n!,n_1!)}u^{|n-n_1|}e^{-u}[L^{|n-n_1|}_{\min(n,n_1)}(u)]^2, \tag{25}$$

where

$$u=l_H^2q^2,\quad l_H^2=\frac{c\hbar}{2|e|H}=\frac{R_0^2}{2},$$

L^s_n is an associated Laguerre polynomial. After replacing summation over $\mathbf{q}$ with integration in Eq. (20), taking into account Eqs. (18), (19), (21), and (22), and neglecting the dispersion of confined phonons, we obtain the following expression for the mass operator:

$$\Sigma(m,n,\varepsilon)\equiv\Sigma_C(m,n,\varepsilon)=\sum_{n_1}\frac{w_C(m,n,n_1)}{\varepsilon-E_{m,n_1}-\hbar\omega_{L1}+i\delta}, \tag{26}$$

where

$$w_C(m,n,n_1)=(\hbar\omega_{L1})^2\frac{\alpha l}{2l_H}\frac{\min(n!,n_1!)}{\max(n!,n_1!)}\times\int_0^\infty du\,u^{|n-n_1|-1/2}e^{-u}\times[L^{|n-n_1|}_{\min(n,n_1)}(u)]^2\mathscr{F}^C_m(\beta_0\sqrt{u}). \tag{27}$$

The function $\mathscr{F}^C_n(x)$ for confined phonons is given by

$$\mathscr{F}^C_m(x)=4x\sum_{p=1,3,\ldots}\frac{\pi^2p^2}{x^2+\pi^2p^2}\times\left[\frac{1}{\pi^2p^2}\pm\frac{\cos(k_md)}{\pi^2p^2-4k_m^2d^2}\right]^2. \tag{28}$$

Equation (27) uses the notation

$$\beta_0=d/l_H=d\sqrt{2|e|H/c\hbar};\quad l/l_H=\sqrt{\Omega/\omega_{L1}}. \tag{29}$$

Alongside the interaction with confined phonons, let us consider Frölich's interaction, in which the phonon spectrum is determined for a homogeneous material. In calculating Frölich's interaction, one should select a limiting LO phonon frequency: either frequency ω_{L1} in the quantum well material, assuming that it occupies the entire space, or ω_{L2} of the barrier material. In the case under discussion, it is reasonable to choose ω_{L1}, since the electrons interact with LO phonons mostly within the quantum well. Frölich's interaction is described by the formula

$$\mathscr{H}_F=\sum_{\mathbf{q},q_z}C_{\mathbf{q},q_z}\exp\{i(\mathbf{qr}+q_zz)\}b_{\mathbf{q},q_z}+\text{H.c.}, \tag{30}$$

$$C_{\mathbf{q},q_z}=-i\hbar\omega_{\text{LO}}\left(\frac{4\pi\alpha l}{V_0}\right)^{1/2}\frac{1}{q}, \tag{31}$$

where $b^+_{\mathbf{q},q_z}$ ($b_{\mathbf{q},q_z}$) is the creation (annihilation) operator of a bulk optical phonon, and V_0 is the normalization volume. Using Eq. (30), we obtain the following expression for the mass operator:

$$\Sigma_F(m,n,\varepsilon)=\sum_{n_1}\sum_{\mathbf{q},q_z}\frac{|C_{\mathbf{q},q_z}|^2|R_m(q_z)|^2|I_{n,n_1}(\mathbf{q})|^2}{\varepsilon-E_{m,n_1}-\hbar\omega_{L1}+i\delta}, \quad (32)$$

where

$$R_m(q_z)=\int_{-\infty}^{\infty}dz[\chi_m(z)]^2\exp(iq_z z). \quad (33)$$

In contrast to Eq. (21), here integration is performed over the entire z-axis, since, at a finite quantum well depth, the electron wave function penetrates into the barrier, where the electron interacts with phonons. After integration over q_z, the mass operator $\Sigma_F(m,n,\varepsilon)$ is reduced to a form similar to Eqs. (26) and (27). The difference is in function $\mathscr{T}^F_m(x)$, which should be substituted for $\mathscr{T}^C_m(x)$ in Eq. (27) and can be expressed as

$$\mathscr{T}^F_m(x)=C_m^4 d^2[D_{QW}(x)+D_B(x)]. \quad (34)$$

The function $D_{QW}(x)$ is due to the electron–phonon interaction in the quantum well and has the form

$$\begin{aligned}D_{QW}(x)=&\frac{1}{2x}\left(1\pm\frac{\sin\xi_m}{\xi_m}\right)+\frac{x}{4(x^2+4\xi_m^2)}\\&\times\left(1\pm\frac{2\sin\xi_m}{\xi_m}+\frac{\sin 2\xi_m}{2\xi_m}\right)-e^{-x/2}\\&\times\left(\frac{1}{x}\pm\frac{x\cos\xi_m-2\xi_m\sin\xi_m}{x^2+4\xi_m^2}\right)\left[\frac{\sinh(x/2)}{x}\right.\\&\left.\pm\frac{1}{x^2+4\xi_m^2}\left(x\sinh\frac{x}{2}\cos\xi_m+2\xi_m\cosh\frac{x}{2}\sin\xi_m\right)\right].\end{aligned} \quad (35)$$

The electron–phonon interaction in the barriers is described by function $D_B(x)$:

$$\begin{aligned}D_B(x)=&\frac{P^4}{\zeta_m(2\zeta_m+x)}+\frac{2P^4e^{-x}}{(2\zeta_m+x)^2}+\frac{4P^2e^{-x/2}}{2\zeta_m+x}\left[\frac{1}{x}\sinh\frac{x}{2}\right.\\&\left.\pm\frac{2\xi_m\sin\xi_m\cosh(x/2)+x\cos\xi_m\sinh(x/2)}{x^2+4\xi_m^2}\right].\end{aligned} \quad (36)$$

Equations (35) and (36) uses the notation

$$\xi_m=k_m d,\quad \zeta_m=\kappa_m d,\quad x=qd=\beta_0\sqrt{u}. \quad (37)$$

The function $P=\cos(\xi_m/2)$ and plus are used when $m=1,3,5,\ldots$, whereas $P=\sin(\xi_m/2)$ and minus when $m=2,4,6,\ldots$. In the limiting case of an infinitely deep well, $\Delta E_e\to\infty$, we derive from Eqs. (8)–(12)

$$k_m d\to m\pi,\quad Q_e\to\infty,$$

$$\kappa_m\to\infty,\quad \varepsilon_m\to(\hbar\pi m)^2/2m_e d^2,$$

$$\chi_m\to\sqrt{2/d}\sin(m\pi z/d),\quad 0\leqslant z\leqslant d,$$

$$\chi\to 0,\quad z\leqslant 0,\quad z\geqslant d$$

for both odd and even m. The formulas for $\mathscr{T}^C_m$ and $\mathscr{T}^F_m$ are simplified, accordingly:

$$\mathscr{T}^C_m\to\left(\frac{16}{\pi}\right)^2\beta_0\sqrt{u}\times\sum_{p=1,3,\ldots}\frac{m^4}{p^2(\beta_0^2u+p^2\pi^2)(4m^2-p^2)^2}, \quad (38)$$

$$\mathscr{T}^F_m\to\frac{8}{\beta_0^2u+4\pi^2m^2}\left[\frac{3\beta_0\sqrt{u}}{8}+\frac{\pi^2m^2}{\beta_0\sqrt{u}}-\frac{4\pi^4m^4(1-\exp(-\beta_0\sqrt{u}))}{\beta_0^2u(\beta_0^2u+4\pi^2m^2)}\right]. \quad (39)$$

These equations show that the difference between interactions with confined and bulk LO phonons is in functions $\mathscr{T}^C_m$ and $\mathscr{T}^F_m$, which act as form factors in this case.

4. ENERGY SPECTRUM OF A DOUBLE MAGNETOPOLARON

The mass operator is a sum over Landau quantum numbers n_1. If there is resonant coupling between Landau levels n and n_1,

$$j=n-n_1 \quad (40)$$

and condition (3) is satisfied, the sum over n_1 contains one large term where the denominator is small, and the rest of the terms are small since they are proportional to $\alpha\ll 1$, so they all can be neglected. Suppose that the larger term corresponds to n_1 (this is the final state where an electron transfers by emitting a phonon). Then we derive from Eq. (16) a quadratic equation which determines the double magnetopolaron spectrum:

$$\varepsilon-E_{m,n}-\frac{w(m,n,n_1)}{\varepsilon-\varepsilon_{m,n}+\lambda}=0, \quad (41)$$

$$\lambda=(n-n_1)\hbar\Omega-\hbar\omega_{L1}. \quad (42)$$

Parameter λ is a deviation of magnetic field from the condition (1). Equation (41) determines two branches of the magnetopolaron spectrum with the separation between them

$$\Delta E(\lambda)=\sqrt{\lambda^2+4w_C(m,n,n_1)}. \quad (43)$$

At the resonance, where $\lambda=0$,

$$\Delta E_{\rm res}=2\sqrt{w(m,n,n_1)}. \quad (44)$$

It follows from Eq. (27) or the similar formula for Frölich's interaction with bulk LO phonons, where index C is replaced by F, that

$$\Delta E_{\rm res}\propto\alpha^{1/2}. \quad (45)$$

Equation (43) is valid if $\lambda\leqslant 2\sqrt{w}$. If $\lambda\gg 2\sqrt{w}$, terms with all Landau numbers should be included. Lang *et al.*[16] determined the energy splitting for polarons in a different manner. The result was presented as a sum over all types of phonons in a quantum well [Eqs. (30) and (20) in Ref. 16] with frequency ω_{L1} without dispersion. The form of electron–phonon interaction was not specified. After substituting expression (17) describing interaction with confined phonons in Eqs. (30) and (20) from Ref. 16, one obtains Eq. (44) for $n-1$, $n_1=0$, i.e., for the A polaron.

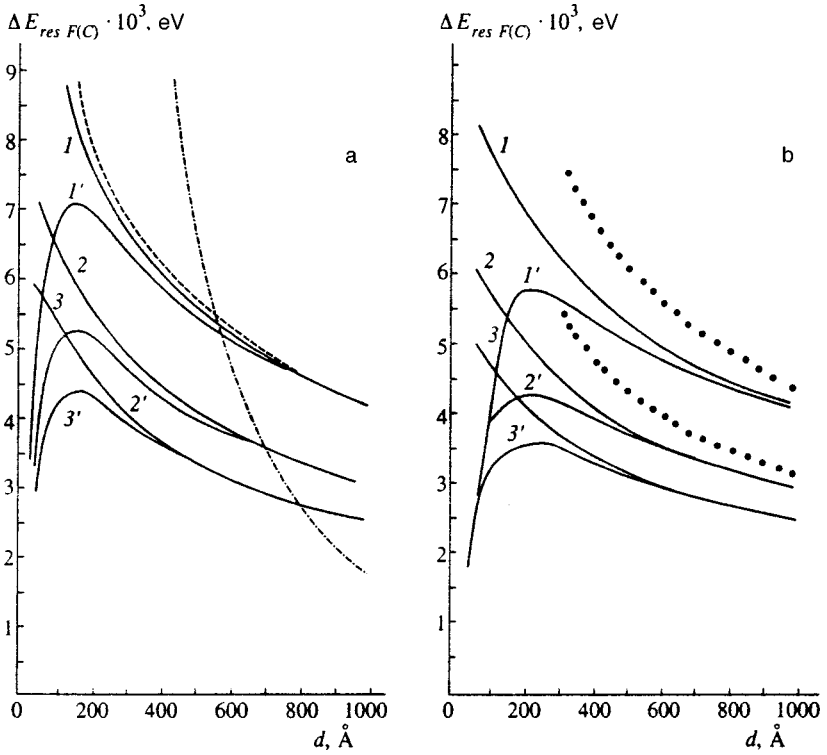


FIG. 2. Energy splitting between branches in magnetopolaron spectra $\Delta E_{\mathrm{res}\,F}$ (Frölich's interaction) and $\Delta E_{\mathrm{res}\,C}$ (interaction with confined phonons) as functions of the quantum well width in $Al_{0.32}Ga_{0.68}/GaAs/Al_{0.32}Ga_{0.68}As$ at (a) $m=1$ and (b) $m=2$. Curves *1–3* show $\Delta E_{\mathrm{res}\,F}$, curves *1′–3′* $\Delta E_{\mathrm{res}\,C}$; curves *1* and *1′* plot energy splitting due to the A-polaron, *2* and *2′* refer to the D-polaron, and *3*, *3′* to the F-polaron. The calculation parameters are listed in the text. The dashed curve in Fig. 2a plots $\Delta E_{\mathrm{res}\,F}$ for an infinitely deep well, the dash-dotted line shows the separation between $m=1$ and $m=2$ levels in an infinitely deep well in the case of Frölich's interaction. The dotted lines in Fig. 2b plot ΔE_{res} calculated by the asymptotic formula (56).

Solid lines in Figs. 2 and 3 show $\Delta E_{\mathrm{res}C}$ and $\Delta E_{\mathrm{res}F}$ versus the quantum well width d calculated by Eqs. (27), (28), (34)–(36), and (44). The dashed line in Fig. 2a shows $\Delta E_{\mathrm{res}\,F}$ for the *A*-polaron in an infinitely deep quantum well. The curves in Fig. 2 are plotted for the $Al_{0.32}Ga_{0.68}As/GaAs/Al_{0.32}Ga_{0.68}As$ structure with parameters $\Delta E_e=0.35$ eV,[29] $m_e/m_0=0.067$, $\hbar\omega_{L1}=0.036$ eV, $\hbar\omega_{L2}=0.047$ eV,[30] and $\alpha=0.06$. Figure 2a corresponds to the size-quantization quantum number $m=1$, curves *1* and *1′* show $\Delta E_{\mathrm{res}\,F}$ and $\Delta E_{\mathrm{res}\,C}$ for the *A*-polaron, curves labeled by *2* and *2′* show similar curves for the *D*-polaron, and curves *3* and *3′* for the *F*-polaron. Figure 2b is similar to 2a, but for $m=2$. Figure 3 corresponds to the case of $m=1$ in the AlSb/InAs/AlSb structure, where $\Delta E_e=1.8$ eV, $m_e/m_0=0.023$, $\hbar\omega_{L1}=0.030$ eV, $\hbar\omega_{L2}=0.042$ eV, and $\alpha=0.042$. The labels at curves in Fig. 3 are similar to those in Fig. 2. Figures 2 and 3 clearly show that the curves of $\Delta E_{\mathrm{res}\,F}(d)$ and $\Delta E_{\mathrm{res}\,C}(d)$ converge as d increases. This means that the effect of interaction with confined phonons is the same as with bulk LO phonons. Section 5 will give an explanation of this fact and determine the conditions when an interpretation based on interaction with bulk phonons is admissible.

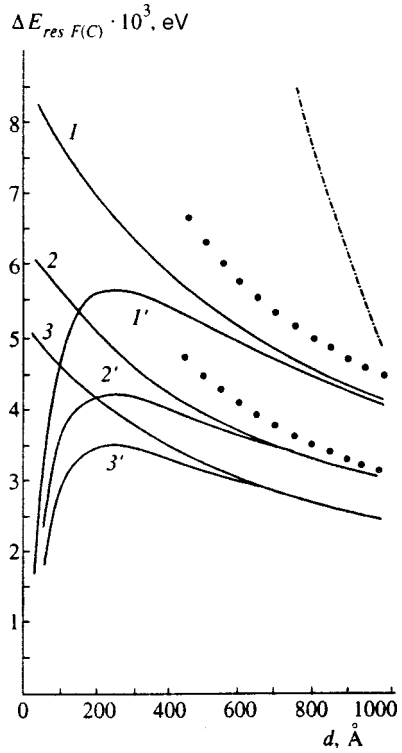


FIG. 3. The same as in Fig. 2a for the AlSb/InAs/AlSb structure.

5. ENERGY SPLITTING DUE TO MAGNETOPOLARONS IN WIDE QUANTUM WELLS

Let us investigate what approximate expression can be substituted for exact formulas given above at large d. To this end, let us reconsider Eq. (27) for function $w_C(m,n,n_1)$, which occurs in Eq. (44), and move one step back. Before integration over z we have

$$w_C(m,n,n_1)=\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\int_0^\infty du f_{n,n_1}(u)$$
$$\times\sum_{p=1,3,\ldots}\left\{\int_0^1 dy\cos\left[\pi p\left(y-\frac{1}{2}\right)\right]\psi_m^2(y)\right\}^2\frac{1}{u+(\pi p/\beta_0)^2}, \quad (46)$$

where $y=z/d$, function (22) for $n>n_1$ can be replaced by

$$f_{n,n_1}(u)=\frac{n_1!}{n!}e^{-u}u^{n-n_1}[L_{n_1}^{n-n_1}(u)]^2, \tag{47}$$

$$\psi_m(y)=\sqrt{d}\chi_m(yd). \tag{48}$$

Similarly, by introducing the integration variable $t=q_z d$ for interaction with bulk phonons, we obtain

$$w_F(m,n,n_1)=\frac{\hbar\omega_{L1}}{4\pi}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1}) \times\int_0^\infty du f_{n,n_1}(u)\int_{-\infty}^{\infty}dt \times\left|\int_{-\infty}^{\infty}dy e^{iyt}\psi_m^2(y)\right|^2\frac{1}{u+(t/\beta_0)^2}. \tag{49}$$

Let parameter β_0 in Eqs. (46) and (49) go to infinity and let us discard $(\pi p/\beta_0)^2$ and $(t/\beta_0)^2$ in the two respective denominators. By performing the sum over p in Eq. (46) and integration over t in Eq. (49), we obtain

$$w_C(\beta_0\to\infty)=\frac{\hbar\omega_{L1}}{2j}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\int_0^1 dy\,\psi_m^4(y), \tag{50}$$

$$w_F(\beta_0\to\infty)=\frac{\hbar\omega_{L1}}{2j}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\int_{-\infty}^{\infty}dy\,\psi_m^4(y). \tag{51}$$

In these calculations we have used the formula

$$\frac{n_1!}{n!}\int_0^\infty du e^{-u}u^{n-n_1-1}[L_{n_1}^{n-n_1}(u)]^2=\frac{1}{n-n_1}=\frac{1}{j}. \tag{52}$$

Thus, it turns out that in the limit $\beta\to\infty$, the difference between the formulas for $\Delta E_{\rm res}$ based on interaction with confined and bulk phonons is only in the limits of integration over $y=z/d$. This difference between integration limits is significant only in the case when the exact electron wave functions (8) and (9) differ from those in an infinitely deep quantum well, since in the latter case they equal zero outside the quantum well. But in wide quantum wells penetration of electrons into the barrier should be insignificant (the criterion will be specified below), therefore, we can use wave functions for quantum wells of infinite depth in the limit $d\to\infty$ and set

$$\psi_m(y)\simeq\psi_{m\infty}=\sqrt{2}\sin(m\pi y),\quad 0\leqslant y\leqslant 1, \qquad \psi_m(y)=0,\quad y\geqslant 1,\quad y\leqslant 0. \tag{53}$$

For all quantum numbers m

$$4\int_0^1 dy\sin^4(m\pi y)=\frac{3}{2}, \tag{54}$$

and we obtain very simple asymptotic expressions:

$$w_C(d\to\infty)=w_F(d\to\infty)=\frac{3}{4j}\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1}), \tag{55}$$

$$\Delta E_{\rm res}(d\to\infty)=\sqrt{\frac{3}{j}\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})}. \tag{56}$$

Note that the latter two equations could be also derived from Eqs. (27), (38), and (39) since in the limit $d\to\infty$

$$\mathscr{F}_m^{F(C)}(\beta_0\to\infty)=\frac{3}{\beta_0\sqrt{u}}. \tag{57}$$

Certainly, this description of the transition to limit $d\to\infty$ is far from accurate. At large d one should obtain results for a conventional bulk crystal. The concept of a magnetopolaron with fixed m makes no sense when the separation between size-quantization levels is comparable to or less than the splitting $\Delta E_{\rm res}$ (the criterion will be discussed in Sec. 6). The dotted lines in Fig. 2b show calculations of $\Delta E_{\rm res}(d)$ by Eq. (56). It is clear that the error of Eq. (56) is considerable even in the range of d where the calculations of $\Delta E_{\rm res}$ labeled by C and F are very close. Calculations indicate that Eq. (56) is fairly accurate when $d\geqslant 2000$ Å. The cause of the large error of Eq. (56) can be understood by analyzing Eq. (39). In order to obtain the limiting formula (57) from this equation, one must satisfy the condition

$$\beta_0^2 u_{n,n_1}\gg 4\pi^2 m^2, \tag{58}$$

where u_{n,n_1} is the characteristic value of variable u as a function of n and n_1. Condition (58) is a more stringent condition than equality of $\Delta E_{\rm res\,C}$ and $\Delta E_{\rm res\,F}$. In order to show this, let us reconsider Eqs. (46) and (49). Exact calculation of the sum in Eq. (46) yields

$$w_C(m,n,n_1)=\frac{1}{4}\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\beta_0\int_0^\infty du\frac{f_{n,n_1}(u)}{\sqrt{u}} \times\int_0^1 dy\int_0^1 dy'\psi_m^2(y)\psi_m^2(y') \times\frac{\exp(-|y-y'|\beta_0\sqrt{u})-\exp[-(|y-y'|-1)\beta_0\sqrt{u}]}{1+\exp(-\beta_0\sqrt{u})}. \tag{59}$$

Integration over t in Eq. (49) leads to the exact formula

$$w_C(m,n,n_1)=\frac{1}{4}\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\beta_0\int_0^\infty du\frac{f_{n,n_1}(u)}{\sqrt{u}} \times\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}dy'\psi_m^2(y)\psi_m^2(y') \times\exp(-|y-y'|\beta_0\sqrt{u}). \tag{60}$$

In the limit $\beta_0\to\infty$ Eqs. (59) and (60) are identical to (50) and (51), respectively.

Given that

$$\beta_0\sqrt{u_{n,n_1}}\gg 1, \tag{61}$$

we derive from Eq. (59) the following expression:

$$w_C(m,n,n_1)\simeq\frac{1}{4}\hbar\omega_{L1}\frac{e^2}{d}(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\beta_0\int_0^\infty du\,\frac{f_{n,n_1}(u)}{\sqrt{u}}$$
$$\times\int_0^1 dy\int_0^1 dy'\,\psi_{m\infty}^2(y)\psi_{m\infty}^2(y')$$
$$\times\exp(-|y-y'|\beta_0\sqrt{u}), \tag{62}$$

which differs from Eq. (60) in the limits of integration over y and y'. If the penetration of electrons into the barrier is negligible, functions $\psi_m(y)$ in Eqs. (60) and (62) can be replaced by functions (53), then we have

$$w_C(m,n,n_1)\simeq w_F(m,n,n_1)\simeq\frac{1}{4}\hbar\omega_{L1}\frac{e^2}{d}$$
$$\times(\varepsilon_\infty^{-1}-\varepsilon_0^{-1})\beta_0\int_0^\infty du\,\frac{f_{n,n_1}(u)}{\sqrt{u}}$$
$$\times\int_0^1 dy\int_0^1 dy'\,\psi_{m\infty}^2(y)\psi_{m\infty}^2(y')$$
$$\times\exp(-|y-y'|\beta_0\sqrt{u}). \tag{63}$$

By substituting Eq. (63) in (44), we obtain a formula describing curves of $\Delta E_{\rm res}(d)$ in Figs. 2 and 3 in the ranges of d where curves labeled by C and F are very close.

Functions $\psi_m(y)$ can be substituted by $\psi_{m\infty}(y)$ in the case of infinitely deep wells, provided that

$$d\gg d_p,\quad d_p=\frac{\hbar}{\sqrt{2m_e(\Delta E_e-E_m)}}, \tag{64}$$

where d_p is the electron penetration range inside the barrier. For deep levels, $\varepsilon_m\ll\Delta E_e$, d_p tends to the limit

$$d_{p_0}=\frac{\hbar}{\sqrt{2m_e\Delta E_e}}. \tag{65}$$

6. DISCUSSION

Thus, if conditions (61) and (64) are satisfied, interactions with confined and bulk LO phonons yield identical formulas (44) and (63) for the energy splitting due to magnetopolaron effects. Let us give a qualitative interpretation of this fact. The formulas obtained by Mori and Ando[26] for interaction between electrons and phonons of three types, namely, half-space, interface, and confined phonons, contain frequencies ω_{L1} and ω_{L2}, and permittivities $\varepsilon_{\infty1}$, $\varepsilon_{\infty2}$, ε_{01}, and ε_{02}. The subscript 1 (2) corresponds to the material of the quantum well (barrier). Assuming

$$\omega_{L1}=\omega_{L2},\quad \varepsilon_{\infty1}=\varepsilon_{\infty2},\quad \varepsilon_{01}=\varepsilon_{02} \tag{66}$$

and taking into account all three types of phonons in calculations of the energy splitting, one obtains for $\Delta E_{\rm res}(d)$ the same formula as with interaction (30) involving bulk LO phonons of frequency ω_{L1}, i.e., Eq. (44) after substitution of Eq. (27) and (34), or (60), which is equivalent. In other words, all these equations take into account all three types of phonons, but only if condition (66) holds. Since Eq. (66) does not hold, the quantities with index F calculated in the previous sections contain, in addition to the correct term due to confined phonons, incorrectly calculated contributions of half-space and interface phonons. On the other hand, the expressions with index C contain only the contribution of confined phonons. Should correct contributions of half-space and interface phonons be added? As a matter of fact,[26] the frequency ω_{L2} of half-space phonons is far from ω_{L1}, therefore these phonons do not take part in formation of magnetopolarons. The case of interface phonons is more complicated. Figure 2 in Ref. 26 shows various branches of the interface phonon frequency plotted against q. On the lower branch, their frequencies are close to ω_{L1}, on the upper to ω_{L2}, so the interface phonons, generally speaking, contribute to the magnetopolarons. In this paper the contribution of the interface phonons has not been taken into account, therefore the results are correct only for such quantum well widths d at which this contribution is small. Note that in calculating exactly the contribution of interface phonons, one should have taken into account their dispersion, which would complicate the problem considerably. The criterion for small contribution of interface phonons is identical to Eq. (61). Really, the range of interface phonons in the quantum well is of order of q^{-1} (see Table IV in Ref. 26), and since the mean value $\overline{q}l_H\sim\sqrt{u_{n,n_1}}$ [see the notation in Eq. (25)], the condition $\overline{q}^{-1}\ll d$ again yields Eq. (61).

Thus, if condition (61) holds, the contribution of interface phonons can be neglected, alongside the ''incorrect'' contribution of interface phonons to quantities with index F. If condition (61) is satisfied, we have Eq. (60) for quantities labeled by C and Eq. (62) with index F. As was noted above, the only difference between them is in the limits of integration over y and y'. The cause is that Eq. (62) contains the spurious contribution of half-space phonons. This spurious contribution, however, is negligible if condition (64) holds, and under this condition the expressions with indices C and F become identical, which leads to Eq. (61). If $d_p\gg l_H/\sqrt{u_{n,n_1}}$, condition (61) may be satisfied, whereas condition (64) may fail. This is possible, in principle, at large d_p, which, according to Eq. (64), are the larger, the higher the level m in the quantum well. In this case, one must use Eq. (62) [or Eq. (27) with substitution (28)], which yields the correct contribution of confined phonons. The latter case, however, is unrealistic in quantum wells discussed here at $m=1,2$ since condition (64) is satisfied at smaller d than those determined by condition (61) (see the dashed curve in Fig. 2a). Note that under resonant conditions

$$l_H=l\sqrt{j}, \tag{67}$$

so condition (61) transforms to

$$d\gg\frac{l\sqrt{j}}{\sqrt{u_{n,n_1}}}. \tag{68}$$

Since l is inversely proportional to $(\omega_{L1}m_e)^{1/2}$, in InAs l is larger than in GaAs, as a result, the curves of $\Delta E_{\text{res }C}(d)$ and $\Delta E_{\text{res }F}(d)$ in Fig. 3 converge at larger d than in Fig. 2a. Figures 2 and 3 also show that the larger j, the smaller d at which the curves of $\Delta E_{\text{res }C}(d)$ and $\Delta E_{\text{res }F}(d)$ converge. This probably means that the left-hand side of Eq. (68) drops

with j, i.e., $u_{n,n_1}^{1/2}$ grows with j faster than $j^{1/2}$. The parameters l and d_{p_0} of the quantum wells discussed above are as follows:

$$l=40\ \text{Å}, \quad d_{p_0}=12.3\ \text{Å} \quad (\text{GaAs}),$$

$$l=74.5\ \text{Å}, \quad d_{p_0}=9.6\ \text{Å} \quad (\text{InAs}).$$

The dash-dotted curves in Figs. 2a and 3 show the separation $\Delta\varepsilon=\varepsilon_2-\varepsilon_1$ between size-quantized levels $m=1$ and $m=2$ as functions of the quantum well width d. Our results apply only when the condition

$$\Delta E_{\text{res}}(d)<\Delta\varepsilon(d) \tag{69}$$

is satisfied. In calculations of the double magnetopolaron spectrum, we have assumed the conduction band to be parabolic and neglected excitonic effects. The nonparabolicity and excitonic effects, however, cannot radically change the calculated spectra because only two electronic levels are coupled in the magnetopolaron. These effects can only slightly change the resonant magnetic field defined by condition (1).

Now let us summarize our results. When conditions (61) and (64) are satisfied, i.e., the quantum well is sufficiently wide, one can determine the energy splitting due to magnetopolaron effect in terms of interaction with bulk phonons. If condition (61) does not hold, the interaction between electrons and interface phonons is significant. Figures 2 and 3 show that ΔE_{res} may be fairly large, about 1/5 of the phonon frequency ω_{L1} for A polarons. It follows from the asymptotic expression (56) that

$$\Delta E_{\text{res}}\sim\frac{1}{\sqrt{jd}},$$

where j is the factor in Eq. (1), which equals unity for the A-polaron, 2 for the D- and E-polarons, and so on. The limiting value is independent of the level number m. Since ΔE_{res} weakly depends on j, it will be more convenient to experiment with polarons at $j>1$, since the resonant fields corresponding to these polarons are a factor of j lower than for the A polaron. The parameter ΔE_{res} can be determined by measurements of light absorption, reflection, and Raman scattering in quantum wells.

This work was supported by the Russian Fund for Fundamental Research (Grant Nos. 96-02-17115a and 97-02-16495a) and the Physics of Solid-State Nanostructures program (Grant 97-1049).

*) E-mail: pavlov@sci.lpi.ac.ru

[1] D. M. Larsen and E. J. Johnson, in *Proc. of 8th Intern. Conf. on Physics of Semiconductors*, Kyoto, 1966 [J. Phys. Soc. Japan, Suppl. **21**, 443 (1966)].
[2] E. J. Johnson and D. M. Larsen, Phys. Rev. Lett. **16**, 655 (1966).
[3] D. M. Larsen, in *Proc. of X Intern. Conf. on the Physics of Semiconductors*, Cambridge, Mass., 1970, S. P. Keller, J. C. Hensel (Eds.), and F. Stern, U. S. AEC, Oak Ridge (1970).
[4] A. Petron and B. D. McComb, in *Landau Level Spectroscopy*, G. Landwer and E. I. Rashba (Eds.), *Modern Problems in Condensed Matter Sciences*, Vol. 27.2 (1988).
[5] R. J. Nicholas, D. J. Barnes, D. R. Seadley, C. J. Langerak, J. Singleton, P. J. van der Wel, J. A. A. J. Perenboom, J. J. Harris, and C. T. Foxon, in *Spectroscopy of Semiconductor Microstructures*, vol. 206 of NATO Advanced Study Institute, Series B: Physics, G. Fasol, A. Fasolino, and P. Lugli (Eds.), Plenum, New York (1980), p. 451.
[6] R. J. Nicholas, in *Handbook of Semiconductors*, M. Balkanski (Ed.), North Holland, Amsterdam (1994), Vol. 2.
[7] L. I. Korovin and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **53**, 1708 (1967) [Sov. Phys. JETP **26**, 979 (1967)]; JETP Lett. **6**, 50 (1967).
[8] L. I. Korovin, S. T. Pavlov, and B. É. Éshpulatov, Fiz. Tverd. Tela (Leningrad) **20**, 3594 (1978) [Sov. Phys. Solid State **20**, 2077 (1978)].
[9] Das Sarma and O. Madhukar, Phys. Rev. B **22**, 2823 (1980).
[10] Das Sarma and O. Madhukar, Phys. Rev. Lett. **52**, 859 (1984).
[11] G. O. Hai, F. M. Peeters, and J. T. Devreese, in *Phonons in Semiconductor Nanostructures*, *NATO ASI Series, Series E: Applied Sciences*, I. P. Leburston, I. Pascual, and C. Sotomayor Torres (Eds.), Kluver Academic Publishers, Dordrecht, Boston, London (1993), Vol. 236, p. 509.
[12] A. O. Govorov, Solid State Commun. **92**, 977 (1994).
[13] R. J. Nicholas, S. Sasaki, N. Niura, F. M. Peeters, J. M. Shi, C. O. Hai, J. T. Devreese, M. I. Lawless, D. E. Ashenlord, and B. Lunn, Phys. Rev. B **50**, 7596 (1994).
[14] J. M. Shi, F. M. Peeters, and J. T. Devreese, Phys. Rev. B **50**, 15182 (1994).
[15] L. I. Korovin, S. T. Pavlov, and B. É. Éshpulatov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1562 (1993) [Phys. Solid State **35**, 788 (1993)].
[16] I. G. Lang, V. I. Belitsky, A. Cantarero, L. I. Korovin, S. T. Pavlov, and M. Cardona, Phys. Rev. B **54**, 17768 (1996).
[17] L. I. Korovin, I. G. Lang, and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **111**, 2194 (1997) [JETP **84**, 1197 (1997)].
[18] L. I. Korovin, I. G. Lang, and S. T. Pavlov, JETP Lett. **65**, 528 (1997).
[19] I. G. Lang, V. I. Belitsky, A. Cantarero, L. I. Korovin, S. T. Pavlov, and M. Cardona, Phys. Rev. B **56**, 6880 (1997).
[20] R. Fuchs and K. L. Kliever, Phys. Rev. **140**, A2076 (1965).
[21] A. A. Lucas, E. Kartheuser, and R. G. Bardo, Phys. Rev. B **2**, 2448 (1970).
[22] J. J. Licari and R. Evrard, Phys. Rev. B **15**, 2254 (1977).
[23] E. P. Pokatilov and S. I. Beril, Phys. Status Solidi B **118**, 567 (1983).
[24] L. Wender, Phys. Status Solidi B **129**, 513 (1985).
[25] C. Trallero-Giner and F. Comas, Phys. Rev. B **37**, 4583 (1988).
[26] N. Mori and T. Ando, Phys. Rev. B **40**, 6175 (1989).
[27] D. L. Lin, R. Chen, and T. E. George, in *Quantum Well and Superlattice Physics III*, SPIE Proceedings Number 1283, G. H. Döhler, E. S. Koteles, and J. H. Schulman (Eds.), Bellingham, Washington (1990), p. 273.
[28] L. I. Korovin, Fiz. Tverd. Tela (Leningrad) **13**, 842 (1971) [Sov. Phys. Solid State **13**, 695 (1971)].
[29] R. L. Green and K. K. Bajaj, Phys. Rev. B **31**, 6488 (1985).
[30] S. Adachi, J. Appl. Phys. **58**, R1 (1985).

Translation provided by the Russian Editorial office.

de Haas–van Alphen oscillations and angular magnetoresistance oscillations in the organic metal κ-$(BETS)_2GaCl_4$

S. I. Pesotskiĭ*)

Institute of Chemical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Russia; International Laboratory of Strong Magnetic Fields and Low Temperatures, 53-529 Wrocław, Poland; Walther Meissner Institut, Bayerische Akademie der Wissenschaften, D-85748 Garching, Germany

R. B. Lyubovskiĭ

Institute of Chemical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Russia; International Laboratory of Strong Magnetic Fields and Low Temperatures, 53-529 Wrocław, Poland

N. D. Kushch

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

M. V. Kartsovnik, W. Biberacher, and K. Andres

Walther Meissner Institute, D-85748, Garching, Germany

H. Kobayashi

Institute for Molecular Science, Okazaki 444, Japan

A. Kobayashi

Department of Chemistry, School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

(Submitted 2 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 205–210 (January 1999)

Quantum de Haas–van Alphen oscillations and semiclassical angular magnetoresistance oscillations are observed in the quasi-two-dimensional organic metal κ-$(BETS)_2GaCl_4$. The behavior of these oscillations attests to the existence of two cylindrical Fermi surface sheets with axes perpendicular to the conducting plane. The cross-sectional areas of these cylinders in the conducting plane are equal to about 20 and 100% of the corresponding cross section of the Brillouin zone. It is shown that the many-particle interaction can be suppressed significantly in this compound. © *1999 American Institute of Physics.* [S1063-7761(99)01801-6]

All the known quasi-two-dimensional organic conductors that have been synthesized from the organic molecule of bis(ethylenedithio)tetrathiafulvalene (ET) are organized in the same manner: the ET molecules form conducting cation-radical layers, which are separated from one another by nonconducting anion layers, so that the conductivity along the ET layers is several orders of magnitude higher than the conductivity between them.[1] The ET molecules can be packed in the conducting layers in different ways, one of which is κ-type packing.[1] Organic metals of the κ type have been known for more than a decade and have aroused special attention, particularly because superconductors with the highest known critical temperatures have been discovered among them.[2] Although κ-type metals differ from one another with respect to the chemical composition of the anions and are not always isostructural, the Fermi surfaces in them are always similar (Fig. 1).[1,3]

According to theoretical calculations, the original Fermi surface in any κ-type metal is a cylinder, whose axis is perpendicular to the conducting plane and whose cross section in that plane has the form of an ellipse. The area of the ellipse is equal to about 100% of the corresponding cross section of the Brillouin zone, regardless of the composition of the κ-type metal. Consequently, the original ellipse intersects the Brillouin-zone boundary between the M and Z points (Fig. 1), and new Fermi surfaces form: one in the form of a cylinder with a cross-sectional area in the conducting plane equal to about 20% of the area of the Brillouin zone and a second in the form of parallel rippled planes. Only two possible modifications of the Fermi surface are presently known in κ-type metals.

The first modification is exhibited in, for example, the salt κ-$(ET)_2Cu(NCS)_2$, which does not have an inversion center.[4] In such salts an energy gap forms between the small closed Fermi surface and the quasiplanes at the point of intersection on the Brillouin-zone boundary (Fig. 1a). In this case electrons can move in weak magnetic fields in the corresponding closed orbit, which is termed the α orbit, as well as in the open quasiplanes.

In crystals like κ-$(ET)_2I_3$ with an inversion center[5] the gap is essentially absent, and electrons can move in a magnetic field in open orbits, in the α orbit, and in the large closed orbit corresponding to the original ellipse, which has been termed the β orbit. This statement has been confirmed

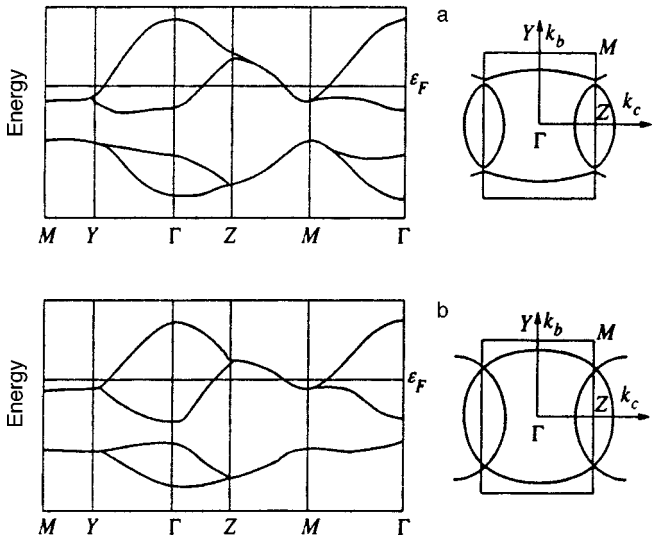


FIG. 1. Calculated energy band structure and Fermi surface in κ-type metals: a — κ-$(ET)_2Cu(NCS)_2$ (Ref. 4); b — κ-$(ET)_2I_3$ (Ref. 5).

experimentally. Quantum oscillations with frequencies corresponding only to the α orbit are observed in κ-$(ET)_2Cu(NCS)_2$ in fields up to 15 T. In stronger fields frequencies associated with the β orbit appear due to magnetic breakdown.[6] Since there is no gap in κ-$(ET)_2I_3$, magnetic breakdown is not required for motion in the β orbit, and frequencies corresponding to the α and β orbits coexist even in weak fields.[7]

Many κ-type organic metals have been synthesized from bis(ethylenedithio)tetraselenafulvalene (BETS), which is a close analog of ET. One particular example is κ-$(BETS)_2GaCl_4$.[8] This compound has an orthorhombic lattice with an inversion center, and BETS layers lie in the **bc** plane and alternate along the **a** axis.[9] The Fermi surface qualitatively resembles the surface shown in Fig. 1b, and it can, therefore, be expected that the galvanomagnetic properties are similar to those observed in κ-$(ET)_2I_3$. For this reason, the report of the discovery of only one frequency of de Haas–van Alphen oscillations corresponding to the β orbit in κ-$(BETS)_2GaCl_4$ in Ref. 9 is unexpected.

This paper presents the result of an investigation of quantum de Haas–van Alphen oscillations and semiclassical angular magnetoresistance oscillations in κ-$(BETS)_2GaCl_4$. It is shown that these effects contain contributions from the motion of electrons in both the β and α orbits.

The crystals measured had the form of parallelepipeds with mean dimensions $2\times2\times0.05$ mm^3, in which the smallest dimension of the samples corresponded to the **a** axis. The resistance measurements were performed by the ordinary four-point method with the measuring current flowing perpendicularly to the conducting layers, i.e., along the **a** axis. The de Haas–van Alphen oscillations were detected in the variation of the rotational moment in a magnetic field.[10] All the measurements were performed in fields up to 15 T and in the temperature range 0.45–4.2 K.

Figure 2 depicts the de Haas–van Alphen oscillations in κ-$(BETS)_2GaCl_4$. The results of Fourier transformation of such oscillations, which are presented in the inset in the figure, exhibit two frequencies, the smaller of which corresponds to the α orbit in the field direction under consider-

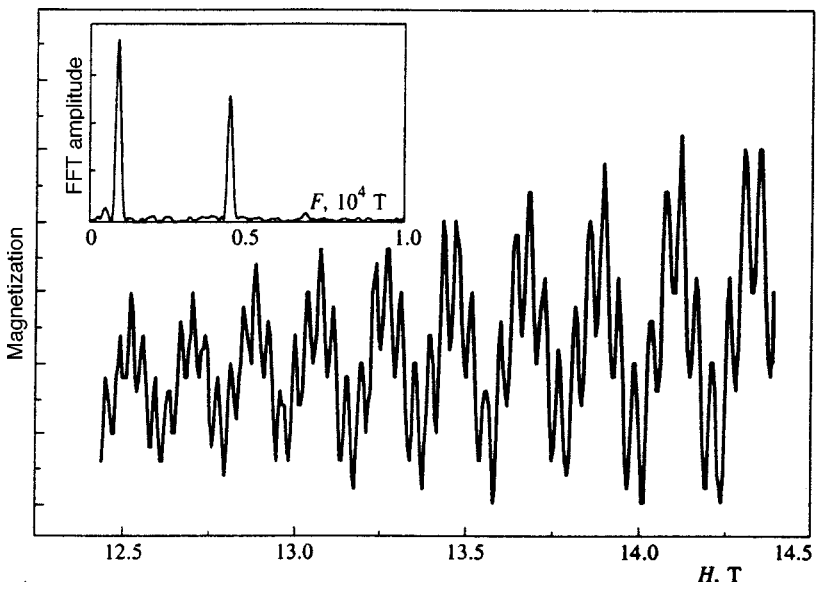


FIG. 2. de Haas–van Alphen oscillations; $\theta=12°$, $T=0.45$ K. Inset: fast Fourier transformation (FFT) of these oscillations.

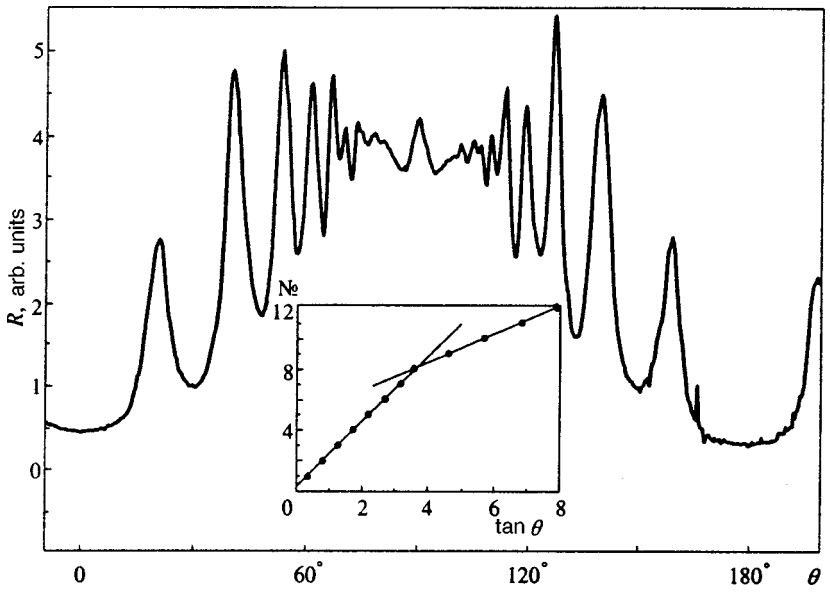


FIG. 3. Angular magnetoresistance oscillations; $H=15$ T, $T=1.45$ K. Inset: Dependence of the number of the magnetoresistance maximum on the tangent of corresponding angle θ.

ation, while the larger frequency corresponds to the β orbit. It should be noted that the two frequencies coexist in fields stronger than about 10 T. The dependences of these frequencies on the angle θ (θ is the angle between the field direction and the **a** axis), as would be expected for quasi-two-dimensional metals, are described by the relations

$$F_\alpha \approx \frac{850\ \mathrm{T}}{\cos\theta}, \quad F_\beta \approx \frac{4400\ \mathrm{T}}{\cos\theta}.$$

Such relations describe two cylindrical Fermi surfaces with axes along the **a** axis. The cross-sectional areas of these cylinders in the conducting **bc** plane are equal to roughly 20% and 100% of the area of the corresponding cross section of the Brillouin zone, in good agreement with the theoretical calculations in Ref. 9.

Figure 3 presents the angular dependence of the magnetoresistance of κ-$(BETS)_2GaCl_4$ with clearly expressed angular oscillations, whose characteristic points are magnetoresistance maxima. These maxima are periodic with respect to the tangent of θ, as can be seen in the inset in Fig. 3. This inset clearly shows that the angular magnetoresistance oscillations contain two sets of oscillations with different periods having a ratio $\Delta_1/\Delta_2 \approx 2.25$. Such behavior corresponds to the motion of electrons on two cylindrical Fermi surface

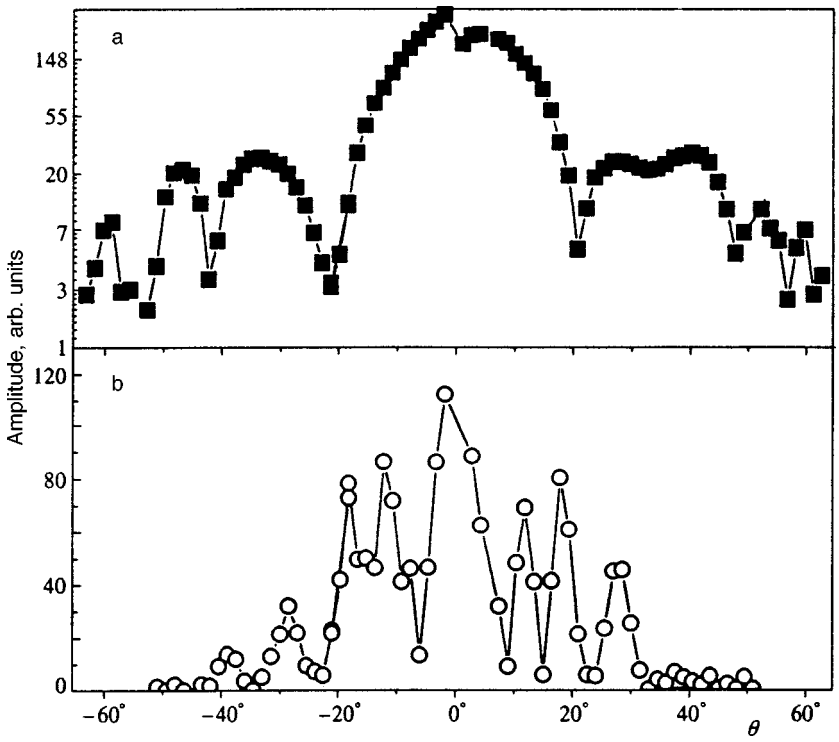


FIG. 4. Angular dependence of the amplitude of de Haas–van Alphen oscillations: a — α oscillations; b — β oscillations.

sheets, which are weakly rippled along their axes.[11] If it is assumed that the cross sections of these cylinders in the **bc** plane are circular, their radii $r_{1,2}$ are specified by the simple relation

$$\Delta_{1,2}=\pi/dr_{1,2},$$

where d is the distance between the conducting planes. In this case the ratio between the areas of the cylinders $S_2:S_1 \sim 5:1$, i.e., it can be assumed in a rough approximation that the periods of the angular magnetoresistance oscillations are determined by the sizes of the cylinders in whose bases the α and β orbits lie. Thus, both the quantum and semiclassical oscillations provide evidence that electrons moving in both of these orbits contribute to these effects.

It possible that Tajima *et al.*[9] did not observe α oscillations because the field direction in their experiment accidentally coincided with a direction in which the amplitude of the α oscillations tends to zero. Such directions do, in fact, exist, and they can be seen in Fig. 4a. The amplitude minima at $\pm 21°$, $-43°$, and $-53°$ correspond to the so-called spin zeros associated with splitting of the Landau levels in a magnetic field.[12] The condition for the existence of spin zeros is

$$\cos(\pi p \mu g/2)=0,$$

where p is the number of the harmonic, μ is the relative effective mass, and g is the g factor. Setting $p=1$ and taking the ordinary angular dependence of the effective mass for a cylindrical Fermi surface, i.e., $\mu(\theta)=\mu(0)/\cos\theta$, we can determine the splitting factor S_α for electrons in the α cylinder:

$$S_\alpha=\mu_\alpha(0)g=(2n+1)\cos\theta\approx 6.6, \quad (1)$$

for the angles indicated above and $n=3$, 4, and 5, respectively. The expression for the splitting factor contains the effective mass and the g factor with corrections for the electron–electron interaction:[12]

$$S_\alpha=\mu_\alpha(0)g=g_s m_{c\alpha}(1+\gamma)/m_0(1+\gamma'),$$

where g_s is the g factor from ESR measurements (as a rule, $g_s=2$ in organic metals[3]), $m_{c\alpha}$ is the band mass for the α orbit, and γ and γ' are the corrections to the mass and the g factor, respectively, for the electron–electron interaction. If the electron–electron interaction is neglected, we obtain $m_{c\alpha}\approx 3.3m_0$ for the band mass, which is very close to the value of the effective mass $m_\alpha^*\approx 3.27m_0$ obtained from temperature dependences of the amplitude of the α oscillations. However, the effective mass determined in this way usually includes corrections for the electron–electron and electron–phonon interactions:[12]

$$m_\alpha^*=m_{c\alpha}(1+\gamma)(1+\lambda),$$

where λ is the correction to the mass for the electron–phonon interaction. One of the reasons for such agreement between m_α^* and $m_{c\alpha}$ may be the significant weakening of the many-particle interactions for electrons in the α cylinder. Taking into account the way in which the Fermi surface is formed in κ-$(BETS)_2GaCl_4$, in which the α and β electrons essentially belong to a single system, we can assume that this weakening is also characteristic of electrons in the β cylinder and should be manifested in the behavior of their oscillations.

Taking into account the value of the effective mass of the β electrons determined from the temperature dependences of the oscillation amplitude of the β electrons, $m_\beta^*\approx 5.3m_0$, and setting it equal to the band mass, we can estimate the splitting factor for the β electrons as $S_\beta\approx 10.6$. It follows from this value and the relation (1) that spin zeros should be observed for the oscillations of the β electrons at angles equal to $\pm 15°$, $\pm 35°$, etc. In fact, minima of the amplitude of the quantum oscillations associated with the β cylinder are observed at these angles (Fig. 4b). Thus, the assumption that the many-particle interactions are suppressed in κ-$(BETS)_2GaCl_4$ is justified. It should be noted that the nature of the remaining extrema (Fig. 4b) is unclear and calls for further research.

This work was carried out with financial support in the form of grants from the Russian Fund for Fundamental Research (Project 96-02-18957), the ''Physics of Quantum and Wave Processes'' Russian State Scientific-Technical Program, and Volkswagen Stiftung (Grant 1/70 206).

*)E-mail: pesot@icp.ac.ru

[1] J. M. Williams, J. R. Ferraro, R. J. Thorn, K. D. Carlson, U. Geiser, H. H. Wang, A. A. Kini, and Myung-Hwan Whangbo, *Organic Superconductors (Including Fullerenes): Synthesis, Structure and Theory*, Prentice Hall, Englewood Cliffs, NJ (1992).
[2] J. M. Williams, A. M. Kini, H. H. Wang, K. D. Carlson, U. Geiser, L. K. Montgomery, G. J. Pyrka, D. M. Watkins, J. M. Kommers, S. J. Boryschuk, A. V. Strieby Crouch, W. K. Kwok, J. E. Schirber, D. L. Overmyer, D. Jung, and Myung-Hwan Whangbo, Inorg. Chem. **29**, 3272 (1990).
[3] J. Wosnitza, *Fermi Surface of Low-Dimensional Organic Metals and Superconductors*, Springer-Verlag, Berlin–Heidelberg (1996).
[4] K. Oshima, T. Mori, H. Inokuchi, H. Urayama, H. Yamochi, and G. Saito, Phys. Rev. B **38**, 938 (1988).
[5] K. Kajita, Y. Nishio, S. Moriyama, W. Sasaki, R. Kato, H. Kobayashi, and A. Kobayashi, Solid State Commun. **64**, 1279 (1987).
[6] T. Sasaki, H. Sato, and N. Toyota, Solid State Commun. **76**, 507 (1990).
[7] K. Oshima, H. Yamazaki, K. Kato, Y. Maruyama, R. Kato, A. Kobayashi, and H. Kobayashi, Synth. Met. **55–57**, 2334 (1993).
[8] A. Kobayashi, T. Udagawa, H. Tomita, T. Naito, and H. Kobayashi, Chem. Lett. 2179 (1993).
[9] H. Tajima, A. Kobayashi, T. Naito, and H. Kobayashi, Solid State Commun. **98**, 755 (1996).
[10] P. Christ, W. Biberacher, H. Muller, and K. Andres, Solid State Commun. **91**, 451 (1994).
[11] Y. Yamaji, J. Phys. Soc. Jpn. **58**, 1520 (1989).
[12] D. Shoenberg, *Magnetic Oscillations in Metals*, Cambridge University Press (1984).

Translated by P. Shelnitz

Tunneling through discrete levels in the continuum

C. S. Kim

Department of Physics, Chonnam National University, Kwangju, Korea

A. M. Satanin*)

N. I. Lobachevskiĭ Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia
(Submitted 6 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 211–230 (January 1999)

We study the ballistic transport in quantum channels containing attractive impurities. We show that coherent interaction between asymptotic resonances may cause resonances to disappear and discrete levels to appear in the continuum at certain (critical) values of the parameters of the system. For the first time the tunneling of an electron through discrete levels is investigated. We find that the transmissivity changes dramatically when the scattered electrons at infinity have an energy coinciding with that of the discrete levels. It is found that a new type of degeneracy may arise in the system at critical values of the parameters, a degeneracy in which one state is described by a localized wave function and the other, by a propagating wave function. We calculate the critical values of the parameters of the structure and discuss ways of experimentally implementing this effect in two-dimensional channels. © *1999 American Institute of Physics.* [S1063-7761(99)01901-0]

1. INTRODUCTION

After the discovery of quantization of conductance, two-dimensional nanostructures can be regarded as a testing ground for demonstrating new quantum coherent effects.[1] The conductance of a nanostructure can be expressed in terms of the transmissivity (transparency) of the channel[2,3] and is determined by electron diffraction in the quantum structure and the interference of the wave in the scattering by impurities. Resonances and dips in transmissivity (or the problem of quantum erosion of conductance) have been studied very thoroughly be experimenters[4–6] and theoreticians.[7–19] In particular, in Refs. 8–17 it was shown that a single attractive impurity in the channel gives rise to an asymmetric resonance in transmissivity, a Fano resonance.[20] When the channel contains two attractive impurities, two types of resonance can be specified: ordinary Breit–Wigner resonances and resonances due to ''quasidonor'' levels formed by virtual potential wells below each of the higher subbands, Fano resonances. As is well known, the width of a Breit–Wigner resonance can vanish only if the quasibound state is separated from the propagating states by infinitely tall walls. The works we have just cited discuss only the situation where the resonance levels have a finite width. However, for Fano resonances there is a new possibility for their widths to vanish.

The present paper examines the coherent effects that accompany the interaction of Fano resonances.

First we show that there are situations in which the widths of asymmetric resonances vanish. As a result, discrete levels appear in the continuum, The possibility that such states exist in quantum theory was discussed in 1929 by von Neumann and Wigner,[21] who used a model potential in their investigation. Similar states have been detected in atomic systems[22–24] (see also the Appendix in Ref. 25). We show that discrete levels may appear in quantum channels at realistic values of the parameters of the system. We also derive the wave function of the discrete levels in explicit form.

Next, we study the tunneling through discrete levels and obtain the transmission amplitude for the case where the energy of the tunneling electron coincides with that of the discrete levels. We also demonstrate that nontrivial degeneracy of the states from the discrete and continuous spectra is possible and that states of different types can be prepared by different selection of the boundary conditions.

Finally, we estimate the parameters of the nanostructure and impurity and discuss the conditions needed for discrete levels to appear in quantum channels and consequences for the problem of impurity erosion.

The plan of the paper is as follows. In Sec. 2 we give an overview of the general approach to describing quantum states in channels. In Sec. 3 we study both analytically and numerically the resonance structure of the scattering matrix in the case of a single impurity. The matrix elements of the impurity potential are calculated in the Appendix. Section 4 is devoted to the study of the coherent interaction of Fano resonances in the case of two attractive impurities. In Sec. 5 we show how discrete levels can appear in the continuum. Section 6 contains a discussion of the structure of the scattering amplitude at critical values of the parameters of the system. Section 7 generalizes the results to the case of levels that are below the higher bands. Finally, in Sec. 8 we sum up the results and discuss the possible applications of the new coherent effects.

2. MODEL AND EQUATIONS

We examine a two-dimensional quantum channel, a quantum waveguide aligned with the x axis. Suppose that the confinement potential acting in the transverse direction is

described by a function $V_c(y)$; this can be e.g., a parabolic or rectangular well. We assume that the waveguide is long in the x direction and is attached to ohmic contacts far from the origin of coordinates. We describe the impurity potential by a function $V(x,y)$. The wave function of the electrons in the waveguide is found by solving the Schrödinger equation

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\Psi(x,y)+V_c(y)\Psi(x,y)+V(x,y)\Psi(x,y)=E\Psi(x,y). \tag{1}$$

For a channel without impurities we have $V(x,y)=0$, and the wave function and energy in this case are

$$\psi^{(0)}_{n,k}(x,y)=\exp(ikx)\,\varphi_n(y), \tag{2}$$

$$E_{nk}=\frac{\hbar^2k^2}{2m}+E_n, \tag{3}$$

where $\varphi_n(y)$ and E_n are the solutions of the equation

$$\left\{-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2}+V_c(y)\right\}\varphi_n(y)=E_n\varphi_n(y). \tag{4}$$

It is convenient to decompose the wave function $\Psi(x,y)$ in the complete set of base functions generated by the solutions of Eq. (4):

$$\Psi(x,y)=\sum_{n=1}^{\infty}\psi_n(x)\varphi_n(y). \tag{5}$$

Substituting (5) in (1), we arrive at an equation for $\psi_n(x)$:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi_n(x)}{\partial x^2}+\sum_{n'=1}^{\infty}V_{n,n'}(x)\psi_{n'}(x)=(E-E_n)\psi_n(x), \tag{6}$$

where the

$$V_{n,n'}(x)=\int\varphi_n(y)V(x,y)\varphi_{n'}(y)\,dy \tag{7}$$

are the matrix elements of the impurity potential.

Here we are interested in the transmission and reflection amplitudes, $t_{n,n'}$ and $r_{n,n'}$, which describe the scattering of electrons from a channel with index n' into a channel with index n. The transmission amplitudes $t_{n,n'}(E)$, examined as functions of the energy E, contain a lot of information about the system. First, the poles of the amplitudes in the complex E plane correspond to levels or resonances. Second, the amplitude determines the conductance G of the structure. In particular, the conductance measured by the two-probe method is determined by the Büttiker–Landauer formula[2,3]

$$G=\frac{2e^2}{h}\sum_{n,n'}T_{n,n'} \tag{8}$$

in terms of the transmission coefficients $T_{n,n'}$ of the system, where n and n' stand for the labels of the incident and scattered waves. The coefficients $T_{n,n'}$ are given by the formula

$$T_{n,n'}=\frac{k_n}{k_{n'}}|t_{n,n'}|^2. \tag{9}$$

The sum in (8) is over all the waves that propagate in the quantum waveguide for a given energy E.

3. SCATTERING BY A SINGLE IMPURITY AND FANO RESONANCES

We model the impurity by a short-range well (in the direction of electron motion) whose center is at the point (X_s,Y_s). Such a potential is specified by matrix elements of the form

$$V_{n,n'}(x)=-\frac{\hbar^2}{m}v_{n,n'}\delta(x-X_s), \tag{10}$$

where $v_{n,n'}\equiv v_{n,n'}(Y_s)$ and $v_{n.n'}>0$. To do estimates and numerical calculations, we use the impurity model proposed in Ref. 18. The parameters of the well and the matrix elements are given in the Appendix.

Equations (6) and (10) imply that a short-range potential is equivalent to the following boundary conditions imposed on the multicomponent functions ($X_s=0$):

$$\psi_n(0^+)-\psi_n(0^-)=0,$$

$$\psi_n'(0^+)-\psi_n'(0^-)=-2\sum_{n'=1}^{\infty}v_{n,n,'}\psi_{n'}. \tag{11}$$

Now we discuss the approximation concerning the matrix elements of the potential, which is used to obtain analytical results. We assume that

$$\frac{\hbar^2}{2m}v^2_{n,n'}\ll|E_n-E_{n'}|,\quad n\neq n', \tag{12}$$

where $|E_n-E_{n'}|$ is the distance between the size quantization levels [in the potential $V_c(y)$]. In this case we can examine the off-diagonal matrix elements $V_{n,n'}$ in (6) by perturbation-theory techniques. If we keep only the diagonal elements $v_{n,n}$, the solution of Eq. (6) can be written

$$\psi_n(x)=\sqrt{v_{n,n}}\exp(-v_{n,n}|x|), \tag{13}$$

$$\epsilon_n=E_n-\frac{\hbar^2v^2_{n,n}}{2m}. \tag{14}$$

Equation (14) shows that in this case levels split away from each size-quantization subband. The corrections to the level ϵ_1 can be found by ordinary perturbation techniques. According to (12), such corrections are small. The states of the higher subbands occupy positions in the continuum of the lowest subbands. To calculate the corrections to ϵ_n for $n\geqslant 2$, we must use the variant of perturbation theory for degenerate levels.[26] Since we are dealing with propagating states, it is convenient to study the poles of the scattering matrix. To find the scattering matrix for a single impurity, we solve Eq. (6) in the regions where the impurity potential is zero:

$$\psi_n=\begin{cases}A_n\exp(ik_nx)+B_n\exp(-ik_nx), & x<0,\\ C_n\exp(ik_nx), & x>0,\end{cases} \tag{15}$$

where $k_n=\sqrt{2m(E-E_n)}/\hbar$. Note that solutions with k_n real belong to propagating states, while states with $k_n=i|k_n|$ (i.e., purely imaginary k_n) represent nonuniform waves. Substituting (15) in (11) yields

$$\mathbf{A}+\mathbf{B}=\mathbf{C}, \quad \mathscr{C}\mathbf{C}=i\mathbf{k}\mathbf{A}, \tag{16}$$

where we have used the matrix notation

$$(\mathscr{C})_{n,n'}=ik_n\delta_{n,n'}+v_{n,n'}, \quad (\mathbf{k})_{n,n'}=k_n\delta_{n,n'}, \tag{17}$$

and the wave amplitudes are considered infinite vectors: $(\mathbf{A})_n=A_n$, etc. We introduce the transmission and reflection matrices; accordingly, $\mathbf{C}=\mathbf{t}\mathbf{A}$ and $\mathbf{B}=\mathbf{r}\mathbf{A}$. Combining these definitions with (16), we get

$$\mathbf{t}=i\mathscr{C}^{-1}\mathbf{k}, \quad \mathbf{r}=i\mathscr{C}^{-1}\mathbf{k}-\mathbf{1}. \tag{18}$$

When the particle energy is close to the bottom of the first band, the one-channel approximation is applicable. The matrix $\mathbf{t}$ has one element for the open channel: $t_{11}(E)=ik_1(ik_1+v_{11})^{-1}$. A pole of the function $t_{11}(E)$ in the complex energy plane occurs at $k_1=-iv_{11}$, or at $E=\epsilon_1$, where ϵ_1 has been defined in (14). If the energy E obeys the inequality $E_1<R<E_2$ and is near ϵ_2, we must examine the two-channel approximation exactly, since in this case the two states have close energies. For instance, the transmission amplitude in the open channel is

$$t_{11}(E)=\frac{ik_1(ik_2+v_{22})}{(ik_1+v_{11})(ik_2+v_{22})-v_{12}^2}. \tag{19}$$

This amplitude has a pole when

$$\tilde{E}=E_2-\frac{\hbar^2}{2m}\left(v_{22}-\frac{v_{12}^2}{ik_1+v_{11}}\right)^2. \tag{20}$$

We can write (20) approximately as

$$\tilde{E}=E_p-i\Gamma, \quad \Gamma\simeq\frac{\hbar^2k_1v_{12}^2v_{22}}{2m(k_1^2+v_{11}^2)}.$$

$$E_p\simeq E_2-\frac{\hbar^2v_{22}}{2m}\left(v_{22}-\frac{2v_{11}v_{12}^2}{k_1^2+v_{11}^2}\right). \tag{21}$$

Comparing this result with (14), we see that the renormalization of the position of the level and the fact that the level acquires a width are due to the resonance interaction of the level and the continuous spectrum of the states belonging to the continuum of band with $n=1$. What is important is that the amplitude $t_{11}(E)$ has a zero for $|k_2|=v_{22}$, or for the energy

$$E_0=E_2-\frac{\hbar^2v_{22}^2}{2m}. \tag{22}$$

From (21) and (22) we see that the energies at the pole ($\tilde{E}$) and the zero (E_0) are close to each other in the complex energy plane, since

$$|\tilde{E}-E_0|\sim\frac{\hbar^2v_{12}^2}{2m}\frac{v_{22}}{\sqrt{k_1^2+v_{11}^2}}\ll E_p\sim E_0. \tag{23}$$

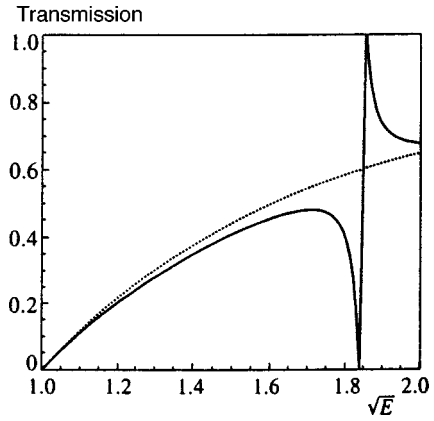


FIG. 1. Transmission coefficient through a waveguide with a single impurity for the energy interval $E_1<E<E_2$, where E is measured in units of $E_1=\pi^2\hbar^2/2mW^2$. The solid curve describes the behavior of the transmission coefficient $T_{11}(E)$ that follows from Eq. (19), and the dotted curve determines the transmissivity in the one-channel approximation.

We conclude that near the zero or pole the amplitude can be written

$$t_{11}(E)\sim\frac{E-E_0}{E-\tilde{E}+i\Gamma}, \tag{24}$$

where E_0, E_p, and Γ are the parameters of a Fano resonance.[20]

The probability of transmission through a channel with an attractive impurity is depicted in Fig. 1 as a function of $\sqrt{E/E_1}$ ($E_1=\pi^2\hbar^2/2mW^2$, with W the channel width) in the case where the impurity parameters $v_{11}=1.261$, $v_{22}=0.785$, and $v_{12}=-0.218$ (the matrix elements are given in units of π/W) were calculated according to Eq. (87) in the Appendix. The expressions for the other impurity parameters are also given in the Appendix. We see that the transmissivity has the structure of a resonance–antiresonance pair. The dotted curve in Fig. 1 depicts a function monotonically increasing with energy, the transmissivity in the one-channel approximation.

Thus, when an electron is scattered by an attractive impurity, the scattering amplitude has the shape of a Fano resonance. If the energy E_0 is real, the transmission coefficient vanishes at $E=E_0$, but near E_0 there is a peak of width Γ. At $E=E_0$ we see that $t_{11}=0$ and $r_{11}=-1$, and the electron is totally reflected from the impurity. Note that for energies that are close to the upper bands, the amplitude can also be represented by a Fano resonance, but generally E_0 is complex-valued, and reflection from the impurity is not total. To illustrate, we examine the case where the energy obeys the inequalities $E_2<E<E_3$. We study in greater detail the amplitude t_{11} found by inversion of $\mathscr{C}$ in the three-channel approximation. The energy at which transmission is zero is given by the expression

$$E_0=E_3-\frac{\hbar^2}{2m}\left(v_{33}-\frac{v_{23}^2}{ik_2+v_{22}}\right)^2. \tag{25}$$

If we have the small parameter (12), we can expand (25) in a perturbation series. Equation (25) implies that the right-hand side for the zero is real for $E_1<E<E_2$, since in this case

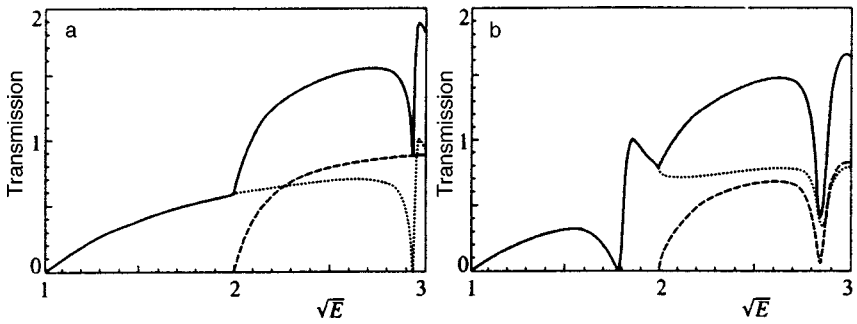


FIG. 2. Transmission coefficient through a waveguide with a single impurity for the energy interval $E_1<E<E_3$, where E is measured in units of $E_1=\pi^2\hbar^2/2mW^2$: (a) the impurity is at the center of the channel, $Y_s=0$; and (b) for $Y_s=0.15W$. The solid curves describe the behavior of the total transmission coefficient (or conductance in units of $2e^2/h$), the dotted curves describe the behavior of $T_{11}(E)$ calculated by (19), and the dashed curves describe the behavior of $T_{22}(E)$.

$ik_2=-|k_2|$ holds. Here the transmission amplitude has the shape of the Fano resonance (24) and reflection from the impurity may be total. When the particle energy satisfies $E_2<E<E_3$, the amplitude t_{11} also has a zero, but the energy at which transmission is zero shifts into the complex E plane:

$$\tilde{E}_0=E_0-i\gamma, \quad \gamma\simeq\frac{\hbar^2k_2v_{23}^2v_{33}}{m(k_2^2+v_{22}^2)},$$

$$E_0\simeq E_3-\frac{\hbar^2v_{33}}{2m}\left(v_{33}-\frac{2v_{23}^2v_{22}}{k_2^2+v_{22}^2}\right), \tag{26}$$

where we have also required $v_{23}\ll v_{33}$, so that the results are manageable. From this reasoning we conclude that total reflection is present when the energy E obeys the inequalities $E_1<E<E_2$. From (21) it follows that the zero-transmission energy is a real quantity if the impurity is in the middle of the waveguide. In this case $v_{23}=0$ holds and reflection is total.

To demonstrate the nature of the Fano resonances in relation to the position occupied by the impurity in the channel, we solved Eq. (16) numerically. Figure 2 depicts the T_{11} vs. E curves for two different positions Y_s of the impurity center, $Y_s=0$ and $Y_s=0.15W$. The impurity parameters were calculated according to formula (87) in the Appendix.

4. COHERENT INTERACTION OF FANO RESONANCES

As noted earlier, an electron is strongly reflected from an impurity if the electron energy is close to the energy E_0 of a Fano-resonance zero. Now we take two impurities separated by a distance L, and study the interaction of Fano resonances. It is convenient to turn to the generalized Fabry–Perot scheme and use the well-known decomposition method. According to this method, the scattering matrix for two impurities can be written

$$\mathbf{t}=\mathbf{t}_2\left(\frac{1}{\mathbf{1}-\mathbf{r}_1\mathbf{r}_2}\right)\mathbf{t}_1, \tag{27}$$

where $\mathbf{t}_1,\mathbf{r}_1$ and $\mathbf{t}_2,\mathbf{r}_2$ are the amplitude matrices of the first and second impurities, respectively. The right-hand side of Eq. (27) takes into account all processes of transmission, reflection, and transformation of waves in the scattering by two impurities. We base our analysis of the passage of an electron through a structure with impurities on Eq. (27).

The scattering matrix for each impurity can be obtained in the same way as in Sec. 3. However, one must allow for the fact that the impurities are shifted in relation to the origin by $x=\pm L/2$, so that the phases of $\mathbf{t}$ and $\mathbf{r}$ differ from those in (18):

$$\mathbf{t}_1=i\mathbf{d}\mathscr{\ell}_1^{-1}\mathbf{k}\mathbf{d}^{-1}, \quad \mathbf{r}_1=\mathbf{d}(i\mathscr{\ell}_1^{-1}\mathbf{k}-\mathbf{1})\mathbf{d}, \tag{28}$$

$$\mathbf{t}_2=i\mathbf{d}^{-1}\mathscr{\ell}_2^{-1}\mathbf{k}\mathbf{d}, \quad \mathbf{r}_2=\mathbf{d}(i\mathscr{\ell}_2^{-1}\mathbf{k}-\mathbf{1})\mathbf{d}. \tag{29}$$

Here the matrices $\mathscr{\ell}_1$ and $\mathscr{\ell}_2$ depend on the impurity parameters and can be found from (17), and $d_{n,n'}=\exp(i\theta_n)\,\delta_{n,n'}$, with $\theta_n=k_nL/2$ (the reader will recall that L is the distance between the impurities).

To simplify the calculations, we transform Eq. (27) to a more convenient form. Substituting (28) and (29) in (27), we find that

$$\mathbf{t}=\mathbf{M}^{-1}\mathbf{k}, \tag{30}$$

where

$$\mathbf{M}=\mathbf{M}_1\mathbf{k}^{-1}\mathbf{M}_2+i(\mathscr{\ell}_1-\mathscr{\ell}_2), \tag{31}$$

$$\mathbf{M}_1=\mathscr{\ell}_1(\mathbf{d}+\mathbf{d}^{-1})-i\mathbf{k}\mathbf{d},$$

$$\mathbf{M}_2=(\mathbf{d}-\mathbf{d}^{-1})\mathscr{\ell}_2-i\mathbf{k}\mathbf{d}. \tag{32}$$

Equation (27) implies that the nontrivial properties of the energy dependence of the transmission amplitude are determined by the properties of the matrix $\mathbf{M}$. The case of identical impurities is the easiest: $\mathscr{\ell}_1=\mathscr{\ell}_2$ and $\mathbf{M}$ can be factorized, or $\mathbf{M}=\mathbf{M}_1\mathbf{k}^{-1}\mathbf{M}_2$. The physical reason for such factorization lies in the symmetry of the system. Since the Hamiltonian is invariant under the transformation $x\to-x$, the solutions of the Schrödinger equation can be chosen as having a definite parity. Then we can easily show that the matrix $\mathbf{M}_s\equiv\mathbf{M}_1$ is responsible for symmetric states in the virtual channels; in the same way the matrix $\mathbf{M}_a\equiv\mathbf{M}_2$ is responsible for antisymmetric states. The matrix elements of $\mathbf{M}_s$ and $\mathbf{M}_a$ can be found from (32):

$$(\mathbf{M}_s)_{n,n'}=2\mathscr{\ell}_{n,n'}\cos\theta_{n'}-ik_n\exp(i\theta_n)\,\delta_{n,n'}, \tag{33}$$

$$(\mathbf{M}_a)_{n,n'}=2i\mathscr{\ell}_{n,n'}\sin\theta_n-ik_n\exp(i\theta_n)\,\delta_{n,n'}. \tag{34}$$

To establish the poles and zeros of the matrix $\mathbf{t}$, we write (30) as

$$\mathbf{t}=\frac{\mathbf{M}_c}{\det\mathbf{M}}\,\mathbf{k}, \tag{35}$$

where $\mathbf{M}_c$ is the adjoint of $\mathbf{M}$ (see Ref. 27). From (35) it follows that the poles of $\mathbf{M}$ are given by the equation

$$\det\mathbf{M}=0, \tag{36}$$

and the zeros by equation of the form

$$\mathbf{M}_c = 0. \tag{37}$$

In view of the fact that $\mathbf{M}$ can be factorized, or $\mathbf{M} = \mathbf{M}_s \mathbf{k}^{-1} \mathbf{M}_a$, we can write Eq. (36) for symmetric and antisymmetric states independently:

$$\det \mathbf{M}_s = \det[\mathscr{C}(\mathbf{d} + \mathbf{d}^{-1}) - i\mathbf{kd}] = 0, \tag{38}$$

and

$$\det \mathbf{M}_a = \det[(\mathbf{d} - \mathbf{d}^{-1})\mathscr{C} - i\mathbf{kd})] = 0, \tag{39}$$

respectively. Note that in contrast to the general equation (36), Eqs. (38) and (39) contain the matrix $\mathscr{C}$, so there is no need to invert matrices when we analyze poles.

Now we determine the matrices for the energy interval $E_1 < E < E_2$ and examine the case where the electron energy is near the unrenormalized level that has split away from E_2. Here the interaction of the wave resonances in the two channels may be strong, so that we must account for it exactly. Below we discuss in detail only the case of symmetric states. Using (38) and (33), we establish the equation for the pole:

$$2(v_{22} - |k_2|)\cosh|\theta_2| + |k_2|\exp(-|\theta_2|) = 4v_{12}^2 \frac{\cosh|\theta_2| \cos\theta_1}{2(ik_1 + v_{11})\cos\theta_1 - ik_1 \exp(i\theta_1)}. \tag{40}$$

We are interested in the solutions of Eq. (40) that are coupled with Fano resonances and are close to the real energy axis. We solve Eq. (40) by expansion in powers of a small parameter, when inequality (12) holds. Suppose that E^{0s} is the solution of the equation

$$2(v_{22} - |k_2|)\cosh|\theta_2| + |k_2|\exp(-|\theta_2|) = 0. \tag{41}$$

In other words, E^{0s} gives the unperturbed position of the pole. The correction can be derived from Eq. (40):

$$\tilde{E}_p^s = E_p^s - i\Gamma^s, \tag{42}$$

where

$$E_p^s = E^{0s} - \frac{2\hbar^2 v_{12}^2 v_{11}}{m} \times \frac{(2k_1 \sin\theta_1 + 2v_{11}\cos\theta_1)\cos\theta_1}{k_1^2 + 4_{11}^2 \cos^2\theta_1 + 4v_{11}k_1 \sin\theta_1 \cos\theta_1}, \tag{43}$$

$$\Gamma^s = \frac{2\hbar^2 v_{12}^2}{m} \frac{k_1 v_{11} \cos^2\theta_1}{k_1^2 + 4v_{11}^2 \cos^2\theta_1 + 4v_{11}k_1 \sin\theta_1 \cos\theta_1}. \tag{44}$$

Now we wish to find the energy at which transmission is zero. The zeros of the amplitude $\mathbf{t}$ can be found from (37). We study the zeros $(\mathbf{M}_c)_{11} = 0$, which are determined by the expression

$$[2(v_{22} - |k_2|)\cosh|\theta_2| + |k_2|\exp(-|\theta_2|)] \times [2(v_{22} - |k_2|)\sinh|\theta_2| - |k_2|\exp(-|\theta_2|)] = 4v_{12}^2 \sin\theta_1 \cos\theta_1. \tag{45}$$

The initial position of a zero follows from (45) with $v_{12} = 0$. For the symmetric case this equation coincides with (41). What is important is that the unrenormalized values of the energies of a zero and pole coincide. The corrections to the position of the zero can be found from (45):

$$E^0 = E_0^0 - 2\frac{\hbar^2 v_{12}^2 v_{11}}{mk_1} \sin\theta_1 \cos\theta_1. \tag{46}$$

Note that the wave functions and phases on the right-hand side of Eq. (43), (44), and (46) depend on the unrenormalized energies of the pole (or zero).

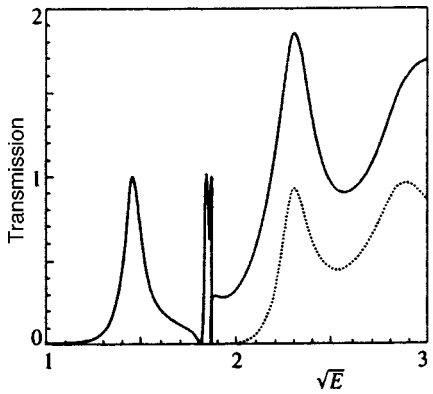


FIG. 3. Transmission coefficient through a symmetric two-impurity system for the energy interval $E_1 < E < E_3$. The distance between the impurities is $L = 1.8v_{22}^{-1}$. The dotted cure represents the contribution to transmissivity from the second subband.

The transmission coefficient T calculated numerically is depicted in Fig. 3 as a function of $\sqrt{E/E_1}$ when the distance L between the impurities equals $1.8v_{22}^{-1}$. The dotted curve in Fig. 3 indicates the contribution to transmission from the second band. The energy interval $E_1 < E < E_2$ clearly contains a Breit–Wigner resonance at $\tilde{E}_1 = 2.120$ with a width $\Gamma_1 = 0.322$ and a pair of Fano resonances with parameters $E_p^s = 3.382$, $\Gamma_s = 0.092$, $E_0^s = 3.290$, and $E_p^a = 3.486$, $\Gamma_a = 0.0045$, $E_0^a = 3.495$ (for the energy unit we have taken E_1). For the energy interval $E_2 < E < E_3$ with the given parameters, there are only Breit–Wigner resonances.

We sum up the results and conclusions of our investigation. Analysis shows that the interaction of resonances gives rise to resonance–antiresonance pairs: the zeros are on the real axis and the poles are in the complex plane. Near the energies related to the symmetric and antisymmetric pairs the amplitude resembles a Fano resonance and can be approximately described by Eq. (24); see Fig. 1. The structure of the ''resonance'' is related to the virtual states in a ''molecule'' in which the coupling and anticoupling levels lie in the continuum and have a finite width. This conclusion agrees with the numerical results of Kumar and Bagwell[10] but not perfectly. Qualitatively, the difference arises because in the impurity model used in Ref. 10 the resonance is ''covered'' by the zero and cannot be seen in the transmission range.

5. DISCRETE LEVELS IN THE CONTINUUM

We wish to show that at certain values of the parameters of the system the width of the resonances vanishes. Equation (42) implies that Γ^s vanishes for

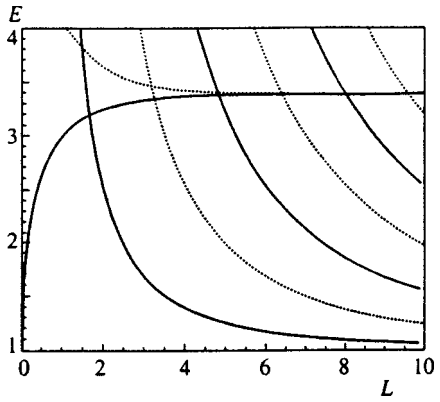


FIG. 4. Two-parameter spectral problem. The solutions $(E(j),L(j))$ are found graphically, where E and L are measured in units of $E_1=\pi^2\hbar^2/2mW^2$ and v_{22}^{-1}, respectively. The points of intersection of the solid curves determine symmetric bound states, and the points at which a dotted curve intersects a solid curve determine antisymmetric states.

$$\cos\theta_1=0. \tag{47}$$

For this to be true, Eqs. (41) and (47) must have a common solution. Thus, we can formulate a two-parameter spectral problem. For instance, if we select the particle energy E and the distance L between the impurities as the parameters, they can be found from Eqs. (42) and (47). First we write (42) and (47) in the form

$$\tanh|\theta_2|=\frac{v_{22}L}{|\theta_2|}-1, \quad \cos\theta_1=0. \tag{48}$$

The first expression is exactly the equation that determines the levels in two wells corresponding to short-range potentials. But the binding energy in a quantum waveguide is higher by E_2 than in free space. If we examine the asymptotes $|\theta_2|\sim v_{22}L/2$ for $v_{22}L\gg 1$ and $|\theta_2|\sim v_{22}L$ for $v_{22}L\ll 1$, we establish that the solution E lies in the interval $E_2-4\hbar^2v_{22}^2/2m<E<E_2-\hbar^2v_{22}^2/2m$.

Correspondingly, the quasibound states are above the energy of the Fano-resonance zero of a single well [Eq. (22)]. The second condition in (48) can be written

$$E=E_1+\frac{\hbar^2\pi^2(2j+1)^2}{2nL^2}, \quad j=0,1,2,\ldots. \tag{49}$$

Thus, Eqs. (48) and (49) determine the spectral characteristics $E(j)$ and $L(j)$. The solution of Eq. (48) can easily be found numerically. A similar analysis was carried for antisymmetric states, but in this case the solution exists only for $v_{22}L>1$. Figure 4 depicts the graphical solution of the equations for the two-parameter problem, where $v_{22}=0.785$ and $v_{12}=0.218$ (the quantity π/W is again chosen as a unit of measurement). According to Eq. (48), the intersection of solid curves makes it possible to determine the discrete levels and the critical distances $[E(j),L(j)]$ for symmetric states. Similarly, the intersection of a dotted curve and a solid curve yields the critical parameters of the antisymmetric states of the system. The values of a few critical parameters are listed in Table I. Note that it may prove more convenient to choose another parameter instead of the distance L, say the width W of the quantum channel.

TABLE I. Spectral values of discrete levels and critical distances $(E(j),L(j))$.

j	$E(j)/E_1$	$L(j)v_{11}$
	Symmetric states	
0	3.1878	0.1613[+1] [a]
1	3.3734	0.4795[+1]
2	3.3828	0.7991[+1]
3	3.3832	0.1129[+2]
	Antisymmetric states	
1	3.3853	0.3195[+1]
2	3.3831	0.6393[+1]
3	3.3832	0.9590[+1]

[a] $[+n]\equiv 10^n$.

Using Eq. (6), we can find the wave function for the discrete levels in explicit form. The wave function of the discrete levels is normalized by the condition

$$\sum_{n=1}^{\infty}\int|\psi_n(x)|^2\,dx=1. \tag{50}$$

Solving Eq. (6) in the two-channel approximation (as in Sec. 3), we find that

$$\psi_1(x)=\begin{cases} a_1\cos k_1x, & |x|<L/2,\\ 0, & |x|>L/2,\end{cases} \tag{51}$$

$$\psi_2(x)=\begin{cases} a_1\cosh|k_2|x, & |x|<L/2,\\ c_2\exp(-|k_2||x|), & |x|>L/2,\end{cases} \tag{52}$$

where a_1, a_2, and c_2 are constants determined by the boundary conditions and normalization. The solution ψ_1 is a standing wave in the open channel, and ψ_2 is a localized state in the closed channel. We see that the standing wave ψ_1 is trapped because of reflection from Fano ''mirrors.'' The wave function for the critical parameters $[L(0),E(0)]$ is depicted in Fig. 5.

Thus, at certain values $E(j),L(j)$ of the parameters, which we call critical, the pole of the transmission matrix ''reaches'' the real energy axis. This means that discrete levels appear in the continuum. What will happen in this event with the energy at which the scattering amplitude t_{11} is zero? According to (46), for critical parameters the corrections to the position of the zero also disappear, i.e., for the critical parameters $[E(j),L(j)]$ the energies of the zero and pole coincide. In other words, the transmission amplitude discards its zero and pole at the same values of the parameters. Generally this means that in the corresponding channels the elements of the adjoint matrix $\mathbf{M}_c$ and the determinant of $\mathbf{M}$ must vanish,

$$\mathbf{M}_c=0, \quad \det\mathbf{M}=0, \tag{53}$$

when the parameter are critical.

We decompose $\mathbf{M}_c$ and det$\mathbf{M}$ at the critical distance $L=L(j)$ and at an energy close to the critical value, $E=E(j)+\varepsilon$, $|\varepsilon|\leqslant E(j)$. Since for critical parameter we have (53), we can write

$$\mathbf{M}_c=\varepsilon\mathbf{M}_c', \quad \det\mathbf{M}=\varepsilon\det\mathbf{M}', \tag{54}$$

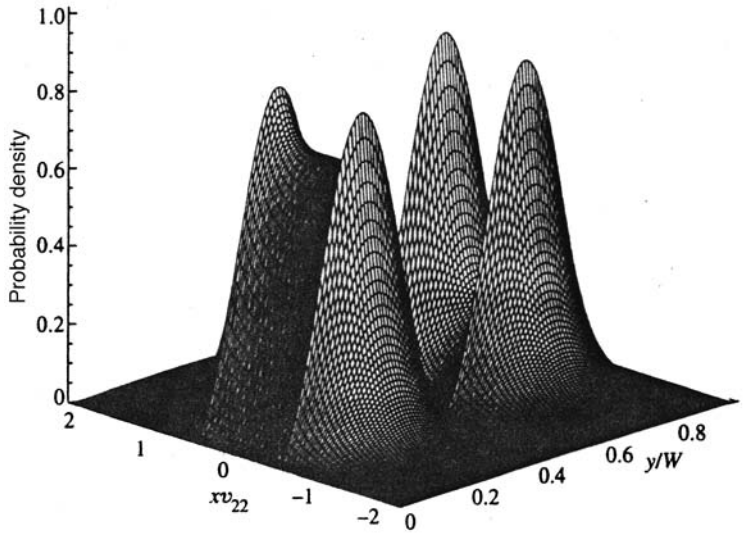


FIG. 5. Square of the absolute value of the wave function of the critical states with $E_c = 3.1878E_1$ and $L_c = 1.613v_{22}^{-1}$.

where $f' \equiv \partial f(E)/\partial E$, with $E = E(j)$. We see that in this case the amplitude must be finite:

$$\mathbf{t} \sim \frac{\mathbf{M}_c'}{\det \mathbf{M}'}\,\mathbf{k}. \tag{55}$$

Thus, as Eqs. (35) and (55) imply, the transmission amplitude changes dramatically at $E = E(j)$. We study the modification of the transmission amplitude more thoroughly in Sec. 6.

6. TUNNELING

To examine the features of tunneling in the critical regime, we study the structure of the scattering matrix for the case where the energy of a tunneling electron coincides with that of a localized state in the channel. But first let us examine the general situation, where for the electron travelling in the channel with $n = 1$ the energy is in the interval $E_1 < E < E_2$. The solution of the Schrödinger equation (6) for $x < -L/2$ is

$$\Psi_1 = A_1 \exp\left[ik_1\left(x + \frac{L}{2}\right)\right] + B_1 \exp\left[-ik_1\left(x + \frac{L}{2}\right)\right],$$
$$\psi_2 = B_2 \exp\left[|k_2|\left(x + \frac{L}{2}\right)\right]. \tag{56}$$

In the region between the impurities, $-L/2 < x < L/2$, the solution is

$$\psi_1 = a_1 \exp(ik_1 x) + b_1 \exp(-ik_1 x),$$
$$\psi_2 = a_2 \exp(-|k_2|x) + b_2 \exp(-|k_2|x), \tag{57}$$

and for $L/2 < x$ we write the solution as

$$\psi_1 = C_1 \exp\left[ik_1\left(x - \frac{L}{2}\right)\right],$$
$$\psi_2 = C_2 \exp\left[-|k_2|\left(x - \frac{L}{2}\right)\right]. \tag{58}$$

(To simplify the notation we redefine the phases of the incident and scattered waves in the formulas that follow.) After substituting the solutions (56)–(58) into the boundary conditions, we arrive at the following equations for the amplitudes:

$$(ik_1 + v_{11})\exp(-i\theta_1)\, a_1 + v_{11}\exp(i\theta_1)\, b_1 + v_{12}[\exp(|\theta_2|)\, a_2 + \exp(-|\theta_2|)\, b_2] = ik_1 A_1, \tag{59}$$

$$(-|k_2| + v_{22})\exp(|\theta_2|)\, a_2 + v_{22}\exp(-|\theta_2|)\, b_2 + v_{12}[\exp(-i\theta_1)\, a_1 + \exp(i\theta_1)\, b_1] = 0, \tag{60}$$

$$ik_1 \exp(i\theta_1)\, a_1 - (ik_1 + v_{11})C_1 - v_{12}C_2 = 0, \tag{61}$$

$$|k_2|\exp(-|\theta_2|)\, a_2 + (-|k_2| + v_{22})C_2 + v_{12}C_1 = 0, \tag{62}$$

$$ik_1 \exp(-i\theta_1)\, b_1 + v_{11}C_1 + v_{12}C_2 = 0, \tag{63}$$

$$-|k_2|\exp(|\theta_2|)\, b_2 + v_{12}C_1 + v_{22}C_2 = 0, \tag{64}$$

where $\theta_1 = k_1 L/2$, and $\theta_2 = k_2 L/2$. Now we turn to the situation in which the localization conditions are met. Below we discuss in detail only the symmetric case. As shown earlier, the critical parameters $[E(j), L(j)]$ can be found from the formulas

$$\exp(i\theta_1) + \exp(-i\theta_1) = 0,$$
$$(-|k_2| + v_{22})[\exp|\theta_2| + \exp(-|\theta_2|)] + |k_2|\exp|\theta_2| = 0. \tag{65}$$

By substituting (65) into (59) and (60) and taking (61)–(64) into account we can easily verify that Eq. (60) is satisfied identically. Under these conditions Eq. (59) yields

$$-(ik_1 + v_{11})C_1 + v_{12}C_2 - v_{11}C_1 - v_{12}C_2 + \frac{2v_{12}^2}{|k_2|}\sinh(|\theta_2|)\, C_1 = -\left[ik_1 + 2v_{11} - v_{22} + \frac{2v_{12}^2}{|k_2|}\sinh|\theta_2|\right]C_1 = ik_1 A_1. \tag{66}$$

What is important is that Eq. (66) does not contain C_2. Thus, the amplitude t_{11} can be written

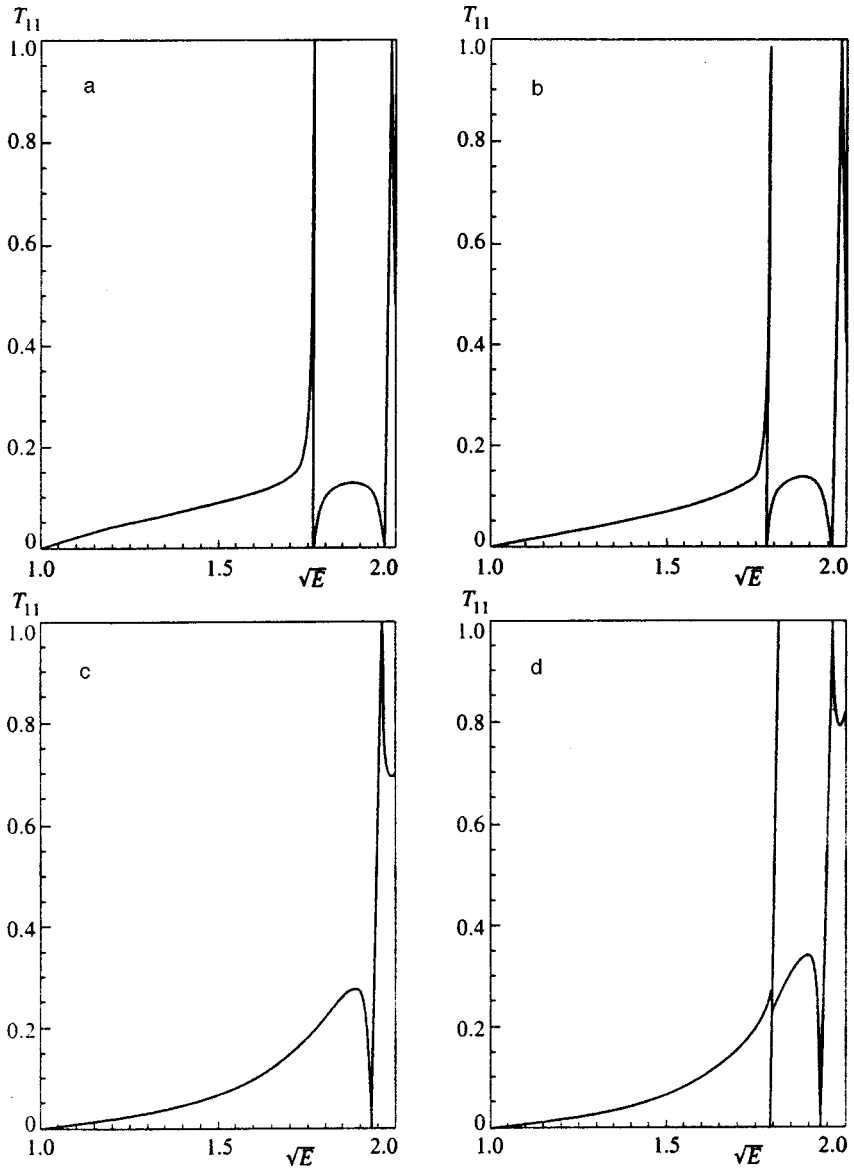


FIG. 6. Transmission coefficient T_{11} as a function of the energy E (in units of E_1) for different values of L (in units of v_{22}^{-1}): (a) $L=1.26$, (b) $L=1.41$, (c) $L=1.613$, and (d) $L=1.73$.

$$t_{11}(E(j))=\frac{-ik_1}{ik_1+2v_{11}-(2v_{12}^2/|k_2|)\sinh|\theta_2|}. \tag{67}$$

Here the incident wave $A_1\exp[ik_1(E(j))x]$ with an energy coinciding with the energy $E(j)$ of a localized state in the waveguide has a finite amplitude and a finite probability of passing through the structure, and the transmissivity of the quantum waveguide undergoes a drastic change, since the zero and the resonance disappear.

Now let us show that for the same energy $E(j)$ another solution of Eqs. (59)–(64) can be found. If we put $A_1=0$ in (59)–(64), Eq. (66) implies $C_1=0$. Then from (61)–(64) we find $a_1=b_1$ and $a_2=b_2$. This solution coincides perfectly with the symmetric localized state (51). Thus, we have shown that Eqs. (59)–(64) yield two types of solution for the critical parameters: (a) waves propagating through the system, and (b) localized states inside the system. Formally this phenomenon is related to the fact that the system of equations for the amplitudes becomes degenerate at critical parameters. It is significant that the wave functions of the degenerate states belong to different types of state, localized and propagating. As is well known, electron states in a realistic potential field belong either to discrete levels with a square-integrable wave function or to levels in the continuum, for which the wave functions cannot be normalized. Usually these states are separated by a definite energy, the mobility edge. In the system considered here, the discrete and propagating states have the same energy, i.e., states belonging to different classes of functions become degenerate.

To illustrate the effect of disappearance of resonances, we show in Fig. 6 the transmission coefficient T_{11} as a function of energy for various distances $L=1.26$, 1.41, 1.613, and 1.73 (for the unit of length we used v_{22}^{-1}), where the distance $L(0)=1.613$ is related to the minimum critical distances. The evolution of a pair of resonances can be seen in Figs. 4a and 4b. When L reaches $L(0)$ the Fano resonance disappears. The transmission coefficient has a finite value of 0.2098 at the energy $E(0)=3.1878$. Figure 4d shows that resonance appear again when the distance exceeds the critical value.

7. MULTICHANNEL APPROXIMATION

Let us first study the passage of the electron through two impurities in the channel. We assume that the electron energy E satisfies $E_1<E<E_2$. For the symmetric case the matrix $\mathbf{M}_s$ is determined by (33). We also assume that the parameters of the system are such that there is a standing wave in the channel $n=1$. This means that the impurities are at a critical distance from each other and the condition for the level width to vanish is $\cos\theta_1=0$. The structure of $\mathbf{M}_s$ suggests that $\det\mathbf{M}_s$ vanishes if

$$\cos\theta_1=0, \quad D_{2,\infty}=0, \tag{68}$$

where $D_{2,\infty}$ is the determinant of a matrix obtained from $\mathbf{M}_s$ by striking out the first column and the first row. The form of $\mathbf{M}_s$ suggests that all the elements determining $D_{2,\infty}$ are real, i.e., Eqs. (68) may have real solutions.

Let us prove this in the three-channel approximation. From (68) we find that

$$\cos\theta_1=0,$$

$$2(v_{22}-|k_2|)\cosh|\theta_2|+|k_2|\exp(-|\theta_2|) = 4v_{23}^2\frac{\cosh|\theta_2|\cosh|\theta_3|}{2(v_{33}-|k_3|)\cosh|\theta_3|+|k_3|\exp(-|\theta_3|)}. \tag{69}$$

If we now use perturbation-theory techniques, in the zeroth approximation we have

$$2(v_{22}-|k_2|)\cosh|\theta_2|+|k_2|\exp(-|\theta_2|)=0. \tag{70}$$

Let E_0^0 be a solution of Eq. (70). The correction to this solution can be found from (69) and has the form $E_0=E_0^0+\delta E$, where

$$\delta E=\frac{2\hbar^2v_{23}^2}{m}\frac{v_{22}\cosh|\theta_3|}{2(v_{33}-|k_3|)\cosh|\theta_3|+|k_3|\exp(-|\theta_3|)}. \tag{71}$$

Calculations that use the higher bands yield real terms on the right-hand side of $D_{2,\infty}$ in a series expansion. Since all the terms in this expansion are real, the levels may shift only along the real axis. We see that in the case $E_1<E<E_2$ we can always find a set of parameters at which discrete levels exist.

Qualitative discrepancies should be observed when the energy is close to the boundaries of the higher bands. In Sec. 3 we found that for the higher bands there is usually no total reflection. For the two-impurity problem it also occurs that in the perturbation-theory setting the expansion of $\det\mathbf{M}_s$ contains complex-valued terms, which shift the levels into the complex energy plane. By way of an example we take the interval $E_2<E<E_3$. Keeping the contributions $\sim v_{n,n'}^2$, we expand $\det\mathbf{M}_s$ in perturbation series and write $\det\mathbf{M}_s=0$ as

$$2(v_{33}-|k_3|)\cosh|\theta_3|+|k_3|\exp(-|\theta_3|) = 4v_{13}^2\frac{\cos\theta_1\cosh|\theta_3|}{2(v_{11}+ik_1)\cos\theta_1-ik_1\exp(i\theta_1)} + 4v_{23}^2\frac{\cos\theta_2\cosh|\theta_3|}{2(v_{22}+ik_2)\cos\theta_2-ik_2\exp(i\theta_2)}. \tag{72}$$

Analysis of the right-hand side of Eq. (72) shows that a solution with a real energy is possible if

$$k_1v_{13}^2\cos^2\theta_1+k_2v_{23}^2\cos^2\theta_2=0. \tag{73}$$

This may happen only if either $\cos^2\theta_1=0$ or $\cos^2\theta_2=0$. Let $\cos^2\theta_1=0$. Then Eq. (72) yields a complex-valued solution in the form $E=E^0+\delta E-i\Gamma$, which E^0 can be found from

$$2(v_{33}-|k_3|)\cosh|\theta_3|+|k_3|\exp(-|\theta_3|)=0, \tag{74}$$

and the level width Γ can be written

$$\Gamma=\frac{2\hbar^2v_{23}^2}{m}\frac{k_2v_{22}\cos^2\theta_2}{k_2^2+4v_{22}^2\cos^2\theta_2+4v_{22}k_2\sin\theta_2\cos\theta_2}. \tag{75}$$

Note that the correction δE has the same structure. Thus, we have a resonance state in the subband $n=3$, which decays into the subband $n=2$. Under these conditions, there is a standing wave in the subband $n=1$. A similar analysis can be done near higher bands, and the characteristics of the corresponding resonances can be obtained.

8. DISCUSSION

We have studied new coherent effects in a quantum waveguide with two attractive impurities. For a pair of impurities we found that the interaction of Fano resonances may change the transmission amplitude dramatically. One consequence of this interaction may be the appearance of discrete levels in the continuum. We have formulated and solved a two-parameter spectral problem to determine the values of the parameters of the system at which resonances disappear and discrete levels appear. We have studied the tunneling through discrete levels and found that the probability of an electron traveling through the waveguide is finite when the electron energy is equal to the energy of the discrete level. We have found that this phenomenon is a consequence of the system of equations that determine the amplitudes of scattering in a multichannel system at critical values of the parameters becoming degenerate. Two types of wave function for the critical parameters are possible here, localized and propagating. The explanation is that two different types of state with the same energy can be prepared by different selection of the boundary conditions.

Modern nanotechnology has shown that artificial impurities[28,29] with fixed parameters can be created in quantum channels. Using the data from Table I, we can easily estimate the smallest values of the critical parameters of a channel of width W: $E(0)=3.18E_1$ and $L(0)=0.41W$, where $E_1=\pi^2\hbar^2/2mW^2$. For instance, if we take a $GaAs/Al_xGa_{1-x}As$-based two-dimensional channel of width $W=300$ nm, for such a structure the minimum critical distance $L(0)$ between the impurities is estimated at roughly 120 nm and the minimum critical distance $E(0)$ at roughly 16 meV.

The authors are grateful to Yong S. Joe for fruitful discussions. The work was supported by grants from the Russian Fund for Fundamental Research (Grant No. 97-02-16923a) and from KOSED, the CNU Reseach Fund, and the Ministry of Education of Korea (Grant No. BSRI-97-2431).

APPENDIX

Here we calculate the matrix elements $V_{n,n'}$ and the parameters $v_{n,n'}$ of an impurity in a quantum channel. To describe the impurity we used the model proposed by Joe and Cosby.[18,19] We assume that the impurity potential is

$$V(x-X_s,y-Y_s)=-V_{\rm att}f(x-X_s)g(y-Y_s), \tag{76}$$

where the functions $f(x)$ and $g(y)$ are defined as

$$f(x)=1, \quad |x|\leqslant L_a, \quad f(x)=0, \quad |x|>L_a,$$
$$g(y)=1, \quad |y|\leqslant W_a, \quad g(y)=0, \quad |y|>W_a, \tag{77}$$

X_s and Y_s are the coordinates of the center of the impurity potential, L_a and W_a are the dimensions of the well, and $V_{\rm att}$ is the well depth. For numerical calculations and estimates we took the model of an infinitely deep well as the confinement potential. In this case the solution of Eq. (4) is

$$\varphi_n(y)=\sqrt{\frac{2}{W}}\sin\left[\pi n\left(\frac{y}{W}+\frac{1}{2}\right)\right], \quad E_n=\frac{\hbar^2\pi^2n^2}{2mW^2}, \tag{78}$$

where W is the width of the waveguide. Using the wave functions of (78), we can easily calculate the matrix elements of the impurity potential (76):

$$V_{n,n'}(x-X_s)=-2V_{\rm att}f(x-X_s)g_{n,n'}(w,y_s), \tag{79}$$

where the diagonal elements are

$$g_{n,n}=\frac{1}{2\pi}\left[w-\frac{1}{n}\sin nw\cos 2ny_s\right], \tag{80}$$

and the off-diagonal elements ($n\neq n'$) have the following form:

$$g_{n,n'}=\frac{1}{\pi(n-n')}\sin\frac{(n-n')w}{2}\cos[(n-n')y_s]$$
$$-\frac{1}{\pi(n+n')}\sin\frac{(n+n')w}{2}\cos[(n+n')y_s], \tag{81}$$

with

$$w=\pi\frac{W_a}{W}, \quad y_s=\pi\left(\frac{Y_s}{W}+\frac{1}{2}\right). \tag{82}$$

Note that the matrix elements (79) rapidly decrease as functions of $|n-n'|$, while in the model discussed in Ref. 8 they are constant.

If the electron wavelength λ_n in an open channel n is much longer than L_a, i.e.,

$$\lambda_n=\frac{2\pi}{k_n}\gg L_a, \tag{83}$$

the short-range potential model can be used to describe the propagation of a wave along the waveguide. Then for $f(x)$ we have

$$f(x)\approx L_a\delta(x). \tag{84}$$

From (83) we see that the approximation (84) is meaningful if

$$|E-E_n|\ll\frac{\hbar^2\pi^2}{mL_a^2}. \tag{85}$$

Now we can write the matrix elements as

$$V_{n,n'}(x-X_s)=2V_{\rm att}L_a\delta(x-X_s)g_{n,n'}(w,y_s). \tag{86}$$

Using the notation

$$v_{n,n'}=\frac{2m}{\hbar^2}V_{\rm att}L_ag_{n,n'}(w,y_s), \tag{87}$$

we write (86) in the form (10). For numerical simulations we used the dimensionless parameters

$$\overline{V}_{\rm att}=\frac{V_{\rm att}}{E_1}, \quad \overline{v}_{n.n'}=\frac{W}{\pi}v_{n,n'},$$
$$\gamma=\pi\frac{L_a}{W}, \quad \overline{v}_{n,n'}=\gamma\overline{V}_{\rm att}g_{n,n'}, \tag{88}$$

the impurity parameters

$$L_a=0.5W, \quad W_a=0.5W, \quad V_{\rm att}=5E_1,$$

and $E_1=\hbar^2k_1^2/2m$ as the energy unit.

*)E-mail: satanin@phys.unn.runnet.ru

[1] *Quantum Transport in Ultrasmall Devices*, Vol. 342 of NATO ASI Series B: Physics, D. K. Ferry, H. L. Grubin, C. Jacoboni, and A.-P. Jauho (Eds.), Plenum, New York (1995).
[2] R. Landauer, Philos. Mag. Philos. Mag. **21**, 863 (1970).
[3] M. Büttiker, Phys. Rev. B **35**, 4123 (1987).
[4] P. L. McEuen, B. W. Alphenaur, R. G. Wheller, and R. N. Sacks, Surf. Sci. **229**, 312 (1990).
[5] M. W. Dellow, P. H. Beton, C. J. G. M. Landerak, T. J. Foster, P. C. Main, L. Eaves, M. Henini, S. P. Beaumont, and C. D. W. Wilkinson, Phys. Rev. Lett. **68**, 1754 (1992).
[6] C.-T. Liang, I. M. Castleton, J. E. F. Frost, C. H. W. Barnes, C. G. Smith, C. J. B. Ford, D. A. Ritchie, and M. Pepper, Phys. Rev. B **55**, 6723 (1997).
[7] C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
[8] P. F. Bagwell, Phys. Rev. B **41**, 10 354 (1990).
[9] E. Tekman and S. Ciraci, Phys. Rev. B **42**, 9098 (1990).
[10] A. Kumar and P. F. Bagwell, Phys. Rev. B **43**, 9012 (1991).
[11] A. Kumar and P. F. Bagwell, Phys. Rev. B **44**, 1747 (1991).
[12] W. Porod, Zhi-an Shao, and C. S. Leng, Appl. Phys. Lett. **61**, 1350 (1992).
[13] P. F. Bagwell and R. K. Lake, Phys. Rev. B **46**, 15 329 (1992).
[14] S. A. Gurvitz and Y. B. Levinson, Phys. Rev. B **47**, 10 578 (1993).
[15] E. Tekman and P. F. Bagwell, Phys. Rev. B **48**, 2553 (1993).
[16] P. J. Price, Phys. Rev. B **48**, 17 301 (1993).
[17] J. U. Nöckel and A. D. Stone, Phys. Rev. B **50**, 17 415 (1994).
[18] Yong S. Joe and R. M. Cosby, Appl. Phys. Lett. **81**, 6217 (1997).
[19] Yong S. Joe and R. M. Cosby, Solid State Commun. **101**, 731 (1997).
[20] U. Fano, Phys. Rev. **104**, 1866 (1961).
[21] J. von Neumann and E. Wigner, Z. Phys. **30**, 465 (1929).
[22] L. Fonda and R. G. Newton, Ann. Phys. (N.Y.) **10**, 490 (1960).
[23] F. H. Stillinger and D. R. Herrick, Phys. Rev. A **11**, 446 (1975).
[24] H. Friedrich and D. Wintgen, Phys. Rev. A **31**, 3964 (1985).
[25] 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 an earlier edition: NASA Techn. Transl. F-510 (1969)].
[26] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[27] F. R. Gantmakher, *The Theory of Matrices*, Chelsea, New York (1959).
[28] Syoji Yamada and Masafumi Yamamoto, Appl. Phys. Lett. **79**, 8391 (1996).
[29] T. Lindberg, J. E. F. Frost, K.-F. Berggern *et al.*, Semicond. Sci. Technol. **12**, 875 (1997).

Translated by Eugene Yankovsky

Dependence of the phonon spectrum of a metal on electron temperature in a nonequilibrium electron-phonon system

D. M. Medvedev*)

Moscow Institute for Physics and Technology, 141700 Dolgoprudnyi, Moscow Region, Russia

Yu. V. Petrov†)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 7 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 231–242 (January 1999)

The dependence of the phonon spectrum of a crystal and the associated thermodynamic functions on electron temperature in the absence of equilibrium between the electrons and the lattice is investigated. The treatment is performed within the Thomas–Fermi approximation for a body-centered cubic crystal at high pressures. © *1999 American Institute of Physics.* [S1063-7761(99)02001-6]

1. INTRODUCTION

One characteristic feature of modern experiments on the interaction of laser radiation with solids using ultrashort laser pulses is violation of the equilibrium between the electronic and ionic subsystems of the solid.[1–4] The temperature of the radiation-absorbing electrons can be several orders of magnitude higher than the lattice temperature, since the pulse duration ($\sim 10^{-14}-10^{-13}$ s) and the thermalization time of the electronic subsystem ($\sim 10^{-15}-10^{-14}$ s) are much shorter than the characteristic time for the exchange of energy between the electrons and the lattice ($\sim 10^{-11}$ s).[5,6] This raises the question of the influence of such a high electron temperature on the phonon spectrum of a solid.[7,8]

The behavior of the elastic constants of crystals with a gap in the electron excitation spectrum, i.e., crystals having diamond-like C, Si, and Ge lattices, as a function of the density of electrons excited into the conduction band was investigated in Refs. 9–11. Dramatic softening of the spectrum of transverse acoustic phonons with increasing density of the electron-hole plasma was obtained. In addition, the presence of a gap in the electron excitation spectrum, especially a large gap, as in the case of carbon, causes any increase in the energy imparted to the crystal to be accompanied by an increase in the number of excited electrons without significant alteration of their temperature. Conversely, in metals, where there is no gap in the spectrum, an increase in the intensity of the laser pulse leads to a dramatic rise in the electron temperature, while the temperature of the phonon subsystem is lower. Here we shall examine the dependence of the phonon characteristics of metals on the electron temperature T_e, assuming that the short thermalization time of the electrons ($\sim 10^{-15}-10^{-14}$ s under normal conditions) allows us to treat their subsystem as a quasiequilibrium system with that temperature.

Within the adiabatic approximation the potential energy for ions is the electron energy E_e, which depends parametrically on the ion coordinates $\{\mathbf{R}(\mathbf{l})\}$ (E_e includes the direct Coulomb interaction of the ions). The equation of motion of an ion of mass M in the equilibrium position at lattice point $\mathbf{l}$ with the radius vector $\mathbf{R}^0(\mathbf{l})$ can be written in the form

$$M\ddot{\mathbf{R}}(\mathbf{l}) = -\frac{\overline{\partial E_e}}{\partial \mathbf{R}(\mathbf{l})}. \tag{1}$$

Here $\mathbf{R}(\mathbf{l})$ is the radius vector of an ion displaced as a result of vibrations. The bar denotes thermodynamic averaging over the electronic state. According to Ref. 12,

$$\frac{\overline{\partial E_e}}{\partial \mathbf{R}(\mathbf{l})} = \left(\frac{\partial F_e}{\partial \mathbf{R}(\mathbf{l})}\right)_{T_e,V}, \tag{2}$$

therefore, the equation of motion of an ion can be written in the form

$$M\ddot{\mathbf{R}}(\mathbf{l}) = -\frac{\partial F_e}{\partial \mathbf{R}(\mathbf{l})}(T_e,V), \tag{3}$$

i.e., the potential energy for ions at an assigned crystal volume V and an assigned temperature of the electronic subsystem T_e is the Helmholtz free energy $F_e(T_e,V)$ of the electrons.

2. HELMHOLTZ FREE ENERGY OF A CRYSTAL AT HIGH PRESSURES

We consider the dynamics of the lattice of a compressed crystal and accordingly take the Helmholtz free energy of the electrons in the Thomas–Fermi approximation. Following Ref. 13, we write the energy of the electrons (in atomic units) in the cell of phase space at the point with the coordinate $\mathbf{r}_j$ in the form

$$\epsilon_e(\mathbf{r}_j) = \sum_{\mathbf{p},\sigma} n(\mathbf{r}_j,\mathbf{p},\sigma)[\epsilon(\mathbf{p}) - \varphi(\mathbf{r}_j)], \tag{4}$$

where the occupation numbers of the states with the momentum $\mathbf{p}$, the kinetic energy $\epsilon(\mathbf{p})$, and the spin projection σ, which are denoted by $n(\mathbf{r}_j,\mathbf{p},\sigma)$, for the electrons are equal

to zero or unity, and $\varphi(\mathbf{r}_j)$ is the electrostatic potential at the point $\mathbf{r}_j$. The total energy of $N_e=NZ$ electrons (N is the number of atoms, Z is their atomic number) equals

$$E_e=\sum_j \epsilon_e(\mathbf{r}_j)-U_{ee}+U_{ii}. \tag{5}$$

Here we took into account that in the summation of $\epsilon(\mathbf{r}_j)$ the energy of the electron–electron interaction U_{ee} is taken into account twice, and we added the direct Coulomb interaction of the nuclei U_{ii}. The expression for the partition function of the electronic system has the form

$$Q_e=\sum_{\{n\}} \exp\left(-\frac{E_e}{T_e}\right)=\sum_{\{n\}} \exp\left\{\frac{1}{T_e}\left[U_{ee}-U_{ii}-\mu N_e - \sum_{j,\mathbf{p},\sigma} n(\mathbf{r}_j,\mathbf{p},\sigma)(\epsilon(\mathbf{p})-\varphi(\mathbf{r}_j)-\mu)\right]\right\}, \tag{6}$$

where μ is the chemical potential, which takes into account the electrostatic interaction. In the Thomas–Fermi approximation it is related to the chemical potential of a free, inhomogeneously distributed electron gas, $\mu_0(\mathbf{r})$, by the expression

$$\mu=\mu_0(\mathbf{r})-\varphi(\mathbf{r}). \tag{7}$$

The summation in (6) is carried out over all the electronic occupation numbers. Considering them, as in Ref. 13, independent and assuming that U_{ee} is determined only by the thermodynamic equilibrium (for the electronic subsystem) electron density, we obtain the partition function in the form

$$Q_e=\exp\left(\frac{U_{ee}-U_{ii}-\mu N_e}{T_e}\right)\prod_{j,\mathbf{p},\sigma}\left[1+\exp\left(\frac{\mu_0(\mathbf{r}_j)-\epsilon(\mathbf{p})}{T_e}\right)\right]. \tag{8}$$

Then, from (8) we obtain the electron Helmholtz free energy:

$$F_e=-T_e\ln Q_e=\int \mu_0(\mathbf{r})n(\mathbf{r})\,d\mathbf{r} - \int \varphi(\mathbf{r})n(\mathbf{r})\,d\mathbf{r}-U_{ee}+U_{ii}+\Omega_0. \tag{9}$$

Here the thermodynamic potential Ω_0 of an inhomogeneous distribution of noninteracting electrons with a density $n(\mathbf{r})$ equals

$$\Omega_0=-T_e\ln\prod_{j,\mathbf{p},\sigma}\left[1+\exp\left(\frac{\mu_0(\mathbf{r}_j)-\epsilon(\mathbf{p})}{T_e}\right)\right]. \tag{10}$$

Using $\varphi_i(\mathbf{r})$ and $\varphi_e(\mathbf{r})$ to denote the potentials created at the point $\mathbf{r}$ by the ions and electrons, we write the electron Helmholtz free energy in the form

$$F_e=\Phi_0-\int \varphi_e(\mathbf{r})n(\mathbf{r})d\mathbf{r}-\int \varphi_i(\mathbf{r})n(\mathbf{r})d\mathbf{r}-U_{ee}+U_{ii}+\Omega_0 = F_0+U_{ei}+U_{ee}+U_{ii}. \tag{11}$$

Here U_{ei} and U_{ee} are, respectively, the energies of the electron–ion and electron–electron interactions, Φ_0 and $F_0=\Phi_0+\Omega_0$ are the Gibbs free energy and the Helmholtz free energy of an inhomogeneous distribution of noninteracting electrons:

$$\Phi_0=\int \mu_0(\mathbf{r})n(\mathbf{r})\,d\mathbf{r}. \tag{12}$$

Going from summation over j and $\mathbf{p}$ to integration over $\mathbf{r}$ and $\mathbf{p}$, for F_0 we obtain

$$F_0=\int f_0(\mathbf{r})\,d\mathbf{r}, \tag{13}$$

where

$$f_0(\mathbf{r})=\frac{\sqrt{2}}{\pi^2}T_e^{5/2}\left[\xi I_{1/2}(\xi)-\frac{2}{3}I_{3/2}(\xi)\right]. \tag{14}$$

Here $\xi(\mathbf{r})=\mu_0(\mathbf{r})/T_e$, and $I_s(\xi)$ is the Fermi integral:

$$I_s(\xi)=\int_0^\infty \frac{x^s\,dx}{e^{x-\xi}+1}. \tag{15}$$

Thus, we obtain the Helmholtz free energy of the electron-nuclear system for the fixed positions of the nuclei $\{\mathbf{R}(\mathbf{l})\}$ as a functional of the electron density for an assigned specific volume per atom v and an assigned electron temperature T_e, which depends parametrically on $\{\mathbf{R}(\mathbf{l})\}$:

$$F_e=\int f_0(\mathbf{r})d\mathbf{r}-Z\sum_{\mathbf{l}}\int\frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{R}(\mathbf{l})|}d\mathbf{r} + \frac{1}{2}\int\int\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}\,d\mathbf{r}'+\frac{1}{2}\sum_{\mathbf{ll}'}\frac{Z^2}{|\mathbf{R}(\mathbf{l})-\mathbf{R}(\mathbf{l}')|}. \tag{16}$$

Here $\xi(\mathbf{r})$ is determined from the condition

$$n(\mathbf{r})=\frac{\sqrt{2}}{\pi^2}T_e^{3/2}I_{1/2}(\xi), \tag{17}$$

whose solution relative to ξ gives

$$\xi(\mathbf{r})=I_{1/2}^{-1}\left(\frac{\pi^2}{\sqrt{2}}\frac{n(\mathbf{r})}{T_e^{3/2}}\right). \tag{18}$$

Here $I_s^{-1}(x)$ is the inverse of $I_s(x)$, so that $I_s(I_s^{-1}(x))=x$. We seek the electron density $n(\mathbf{r})$ by a variational method in the form

$$n(\mathbf{r},\{\mathbf{R}(\mathbf{l})\})=\sum_{\mathbf{l}} Z\frac{\gamma^3}{\pi^{3/2}}\exp[-\gamma^2(\mathbf{r}-\mathbf{R}(\mathbf{l}))^2] \tag{19}$$

with the variational parameter γ, which we determine from the condition that the functional (16) has a minimum at the equilibrium positions of the nuclei $\{\mathbf{R}^0(\mathbf{l})\}$, which we denote by F_e^0. The expression (19) is normalized according to the condition

$$\int n(\mathbf{r})\,d\mathbf{r}=NZ, \tag{20}$$

and explicitly takes into account the screening of the nuclear charges, ensuring the correct acoustic behavior of the body-centered cubic (bcc) crystal which we selected as a high-pressure phase in the long-wavelength limit of phonon frequencies.

In the approximation under consideration it is convenient to introduce the following reduced quantities: $\overline{r}$ for the coordinates, which is defined by the relation $r=\overline{r}Z^{-1/3}$, $\overline{n}$ for the electron density ($n=\overline{n}Z^2$), and $\overline{T}_e$ for the electron temperature ($T_e=\overline{T}_eZ^{4/3}$). Then (16) takes the form

$$F_e=Z^{7/3}\overline{F}_e, \tag{21}$$

where the reduced Helmholtz free energy, which depends parametrically on the reduced coordinates of the nuclei, equals

$$\overline{F}_e=\int \overline{f}_0(\overline{\mathbf{r}})\,d\overline{\mathbf{r}}-\sum_{\mathbf{l}}\int\frac{\overline{n}(\overline{\mathbf{r}})\,d\overline{\mathbf{r}}}{|\overline{\mathbf{r}}-\overline{\mathbf{R}}(\mathbf{l})|}+\frac{1}{2}\int\int\frac{\overline{n}(\overline{\mathbf{r}})\overline{n}(\overline{\mathbf{r}}')}{|\overline{\mathbf{r}}-\overline{\mathbf{r}}'|}d\overline{\mathbf{r}}\,d\overline{\mathbf{r}}'+\frac{1}{2}\sum_{\mathbf{ll}'}\frac{1}{|\overline{\mathbf{R}}(\mathbf{l})-\overline{\mathbf{R}}(\mathbf{l}')|}. \tag{22}$$

Here

$$\overline{f}_0(\overline{\mathbf{r}})=\frac{\sqrt{2}}{\pi^2}\overline{T}_e^{5/2}\left[\overline{\xi}I_{1/2}(\overline{\xi})-\frac{2}{3}I_{3/2}(\overline{\xi})\right], \tag{23}$$

and $\overline{\xi}$ is determined from the relation

$$\overline{\xi}(\overline{\mathbf{r}})=I_{1/2}^{-1}\left(\frac{\pi^2}{\sqrt{2}}\frac{\overline{n}(\overline{\mathbf{r}})}{\overline{T}_e^{3/2}}\right). \tag{24}$$

In this relation

$$\overline{n}(\overline{\mathbf{r}})=\frac{\overline{\gamma}^3}{\pi^{3/2}}\sum_{\mathbf{l}}\exp\{-\overline{\gamma}^2[\overline{\mathbf{r}}-\overline{\mathbf{R}}(\mathbf{l})]^2\} \tag{25}$$

with the reduced parameter $\overline{\gamma}$, which can be defined by the relation $\gamma=\overline{\gamma}Z^{1/3}$.

For each value of the reduced electron temperature $\overline{T}_e$ and the reduced specific volume $\overline{v}$ ($v=\overline{v}Z^{-1}$) we can find the value of $\overline{\gamma}$ which minimizes the functional $\overline{F}_e^0$ obtained from (22) for the equilibrium positions of the nuclei in the bcc lattice that we selected. The expression for $\overline{F}_e^0$ per atom can be written in the form

$$\overline{F}_e^0=\frac{\overline{v}\,\overline{T}_e^{5/2}}{\pi^2}\int_0^1\int_0^1\int_0^1\left[\overline{\xi}I_{1/2}(\overline{\xi})-\frac{2}{3}I_{3/2}(\overline{\xi})\right]d\zeta_1\,d\zeta_2\,d\zeta_3+\frac{1}{2^{1/3}\overline{v}}\left[\sum_{\mathbf{l}}\frac{2\,\mathrm{erfc}(q(l))-\mathrm{erfc}(q(l)/\sqrt{2})}{l}-\frac{2\lambda}{\sqrt{\pi}}\left(1-\frac{1}{2\sqrt{2}}\right)\right]. \tag{26}$$

In Eq. (26)

$$\mathrm{erfc}(z)=\frac{2}{\sqrt{\pi}}\int_z^\infty\exp(-x^2)\,dx, \tag{27}$$

and when $\overline{\xi}$ is calculated from (24), the reduced electron density can be taken in the form

$$\overline{n}(\zeta_1,\zeta_2,\zeta_3)=\frac{\lambda^3}{2\pi^{3/2}\overline{v}}\sum_{\mathbf{l}}\exp\left\{-\frac{\lambda^2}{4}[3(t_1^2+t_2^2+t_3^2)-2(t_1t_2+t_1t_3+t_2t_3)]\right\}, \tag{28}$$

where $t_i=\zeta_i-l_i$ ($i=1,2,3$), and l_i are the integral components of the vector $\mathbf{R}^0(\mathbf{l})$ in the basis set $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ of a bcc lattice with the edge a of the cubic unit cell:

$$\mathbf{R}^0(\mathbf{l})=\sum_{i=1}^3 l_i\mathbf{a}_i, \tag{29}$$

$$\mathbf{a}_1=\frac{a}{2}(-1,1,1), \tag{30}$$

$$\mathbf{a}_2=\frac{a}{2}(1,-1,1), \tag{31}$$

$$\mathbf{a}_3=\frac{a}{2}(1,1,-1). \tag{32}$$

In (26) $\overline{v}=\overline{a}^3/2$, where $\overline{a}=aZ^{1/3}$; $\overline{\gamma}$ has been replaced by another (dimensionless) variational parameter $\lambda=\overline{\gamma a}$;

$$l=\sqrt{3(l_1^2+l_2^2+l_3^2)-2(l_1l_2+l_1l_3+l_2l_3)}, \tag{33}$$

$$q(l)=\lambda l/2. \tag{34}$$

The expression (26) defines the reduced static Helmholtz free energy of a bcc crystal for assigned values of the reduced specific volume $\overline{v}$ and the reduced electron temperature $\overline{T}_e$. The static Helmholtz free energy of a crystal consisting of atoms with an atomic number Z for a specific volume v and an electron temperature T_e can be found from the scaling relation

$$F^0(T,v)=Z^{7/3}\overline{F}_e^0(\overline{T},\overline{v})=Z^{7/3}\overline{F}_e^0(Z^{-4/3}T_e,Zv). \tag{35}$$

3. LATTICE DYNAMICS

To consider the lattice dynamics, we expand the electron Helmholtz free energy (16) in powers of the deviation of the ions from their positions in an ideal bcc lattice, $\mathbf{u}(\mathbf{l})=\mathbf{R}(\mathbf{l})-\mathbf{R}^0(\mathbf{l})$, confining ourselves to the quadratic terms. The resultant force constant matrix $B_{\alpha\beta}(\mathbf{l})$ ($\alpha,\beta=1,2,3$) is expressed in terms of the corresponding reduced matrix $\overline{B}_{\alpha\beta}(\mathbf{l})$ by the scaling relation

$$B_{\alpha\beta}(\mathbf{l})=Z^3\overline{B}_{\alpha\beta}(\mathbf{l}). \tag{36}$$

Here the reduced force constant matrix is

$$\overline{B}_{\alpha\beta}(\mathbf{l})=\left.\frac{\partial^2\overline{F}_e}{\partial\overline{R}_\alpha(0)\partial\overline{R}_\beta(\mathbf{l})}\right|_{\{\mathbf{R}^0(\mathbf{l})\}} \tag{37}$$

and can be represented in the form

$$\overline{B}_{\alpha\beta}(\mathbf{l})=\overline{B}^k_{\alpha\beta}(\mathbf{l})+\overline{B}^p_{\alpha\beta}(\mathbf{l}), \tag{38}$$

where

$$\overline{B}^{k}_{\alpha\beta}(\mathbf{l})=\frac{\lambda^5}{2^{5/3}\pi\overline{v}^{5/3}\overline{T}_e^{1/2}}\exp\left(-\frac{q^2(l)}{2}\right)\int_{-\infty}^{\infty}\exp(-\rho_1^2)\,d\rho_1$$
$$\times\int_{-\infty}^{\infty}\exp(-\rho_2^2)\,d\rho_2\int_{-\infty}^{\infty}\exp(-\rho_3^2)$$
$$\times\left(\rho_\alpha+\frac{\lambda}{2\sqrt{2}}w_\alpha(\mathbf{l})\right)\left(\rho_\beta-\frac{\lambda}{2\sqrt{2}}w_\beta(\mathbf{l})\right)$$
$$\times\frac{d\rho_3}{I_{-1/2}(\overline{\xi}(\overline{\mathbf{r}}))}. \tag{39}$$

In Eq. (39)

$$\boldsymbol{\rho}=\gamma\sqrt{2}\left[\mathbf{r}-\frac{\mathbf{R}^0(\mathbf{l})}{2}\right], \tag{40}$$

$$w_1(\mathbf{l})=-l_1+l_2+l_3, \tag{41}$$

$$w_2(\mathbf{l})=l_1-l_2+l_3, \tag{42}$$

$$w_3(\mathbf{l})=l_1+l_2-l_3. \tag{43}$$

The second term in (38) can be represented in the form

$$\overline{B}^{p}_{\alpha\beta}(\mathbf{l})=\frac{\lambda^3}{2\overline{v}}\left\{n_\alpha n_\beta\left[\frac{G_2(q(l)/\sqrt{2})-4G_2(q(l))}{2q(l)}\right.\right.$$
$$-\frac{3[G_1(q(l)/\sqrt{2})-2\sqrt{2}G_1(q(l))]}{\sqrt{2}q^2(l)}$$
$$\left.+\frac{3[\mathrm{erfc}(q(l)/\sqrt{2})-2\,\mathrm{erfc}(q(l))]}{q^3(l)}\right]$$
$$+\delta_{\alpha\beta}\left[\frac{G_1(q(l)/\sqrt{2})-2\sqrt{2}G_1(q(l))}{\sqrt{2}q^2(l)}\right.$$
$$\left.\left.-\frac{\mathrm{erfc}(q(l)/\sqrt{2})-2\,\mathrm{erfc}(q(l))}{q^3(l)}\right]\right\}. \tag{44}$$

Here

$$G_1(z)=\frac{d}{dz}\mathrm{erfc}(z)=-\frac{2}{\sqrt{\pi}}\exp(-z^2), \tag{45}$$

$$G_2(z)=\frac{d^2}{dz^2}\mathrm{erfc}(z)=\frac{4}{\sqrt{\pi}}z\exp(-z^2), \tag{46}$$

and $\mathbf{n}$ is a unit vector:

$$\mathbf{n}=(n_1,n_2,n_3)=\mathbf{R}^0(\mathbf{l})/R^0(\mathbf{l}). \tag{47}$$

The reduced force constant matrix (38) can be used to define the reduced dynamic matrix:

$$\overline{D}_{\alpha\beta}(\overline{\mathbf{k}})=\frac{1}{M_0}\sum_{\mathbf{l}}\overline{B}_{\alpha\beta}(\mathbf{l})\exp[i\overline{\mathbf{k}}\overline{\mathbf{R}}^0(\mathbf{l})], \tag{48}$$

which permits finding the reduced phonon frequencies as a function of the reduced wave vector $\overline{\mathbf{k}}$ from the characteristic equation

$$\det[\overline{D}_{\alpha\beta}(\overline{\mathbf{k}})-\overline{\omega}_s^2(\overline{\mathbf{k}})\,\delta_{\alpha\beta}]=0. \tag{49}$$

In (48) M_0 is the atomic mass unit: $M_0=1823$. The phonon spectrum $\omega_s(\mathbf{k})$ $(s=1,2,3)$ of a polyatomic bcc crystal consisting of atoms with an atomic number Z and a mass number A can then be found as a function of electron temperature and specific volume from the universal function $\overline{\omega}_s(\overline{T}_e,\overline{v},\overline{\mathbf{k}})$ using scaling:

$$\omega_s(T_e,v,\mathbf{k})=Z^{3/2}A^{-1/2}\overline{\omega}_s(\overline{T}_e,\overline{v},\overline{\mathbf{k}})$$
$$=Z^{3/2}A^{-1/2}\overline{\omega}_s(Z^{-4/3}T_e,Zv,Z^{-1/3}\mathbf{k}). \tag{50}$$

For not excessively heavy elements the isotopes which are stable against β decay have $A\approx 2Z$. In this case we obtain a scaling relation for the frequencies which depends only on the single parameter Z:

$$\omega_s(T_e,v,\mathbf{k})=\frac{Z}{\sqrt{2}}\,\overline{\omega}_s(\overline{T}_e,\overline{v},\overline{\mathbf{k}})$$
$$=\frac{Z}{\sqrt{2}}\,\overline{\omega}_s(Z^{-4/3}T_e,Zv,Z^{-1/3}\mathbf{k}). \tag{51}$$

Universal plots of the reduced phonon spectrum of a bcc lattice for the high-symmetry directions in the Brillouin zone, viz., ΓP, ΓN, and ΓH, and various values of the reduced specific volume and the reduced electron temperature are presented in Figs. 1 and 2. The reduced lattice constant is $\overline{a}=1.6$ for Fig. 1 and $\overline{a}=2.0$ for Fig. 2. The reduced electron temperature is indicated near the curves in Figs. 1 and 2 and has values equal to 0.1, 4, 20, and 50 au. The phonon frequencies of crystals with assigned Z and A are obtained from the curves in Figs. 1 and 2 using Eq. (50) or, for not excessively heavy elements, (51). An increase in the electron temperature leads to an increase in the phonon frequencies for all directions in the Brillouin zone.

Analysis of the reduced phonon spectrum reveals that the hardest of the phonon modes, which are longitudinal, undergo the greatest changes as the reduced electron temperature is varied in the broad range investigated from 0.1 to 50 au. The height of the maximum in the spectrum for the ΓH direction increases by more than three fold. At the same time, the softer transverse modes of the phonon spectrum vary to a lesser degree in response to significant variation of the electron temperature, especially in the long-wavelength region. When the specific volume is increased, the relative changes in the phonon frequencies increase at a fixed electron temperature.

4. THERMODYNAMIC FUNCTIONS OF THE PHONON SPECTRUM

The reduced phonon spectrum obtained from the solution of Eq. (49) permits calculation of the thermodynamic functions associated with lattice vibrations as a function of the specific volume of the crystal and the lattice temperature $T_i\neq T_e$. These functions include the isochoric specific heat $c_V=(T_i\partial s/\partial T_i)_{T_e,v}$. Calculated per atom, it equals

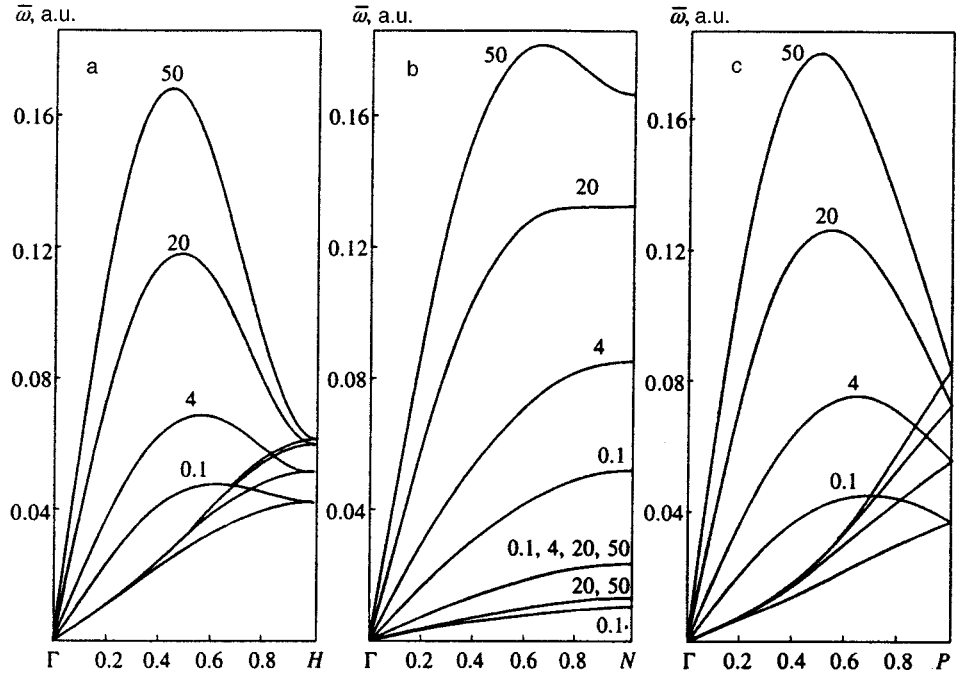


FIG. 1. Reduced phonon frequencies $\bar{\omega}$ of a bcc crystal with the lattice parameter $\bar{a}=1.6$ in the ΓH ([100]) (a), ΓN ([110]) (b), and ΓP ([111]) (c) directions for various values of the reduced electron temperature $\bar{T}_e$, which are indicated (in atomic units) near the curves. The phonon frequencies of a crystal with specified Z and A are found from the values of $\bar{\omega}$ using the scaling relations (50) and (51).

$$c_V=\sum_s \int_{BZ}\left[\frac{\omega_s(\mathbf{k})}{T_i}\right]^2 \frac{\exp[\omega_s(\mathbf{k})/T_i]}{\{\exp[\omega_s(\mathbf{k})/T_i]-1\}^2}\,\frac{v\,d^3k}{(2\pi)^3}. \tag{52}$$

The integration in (52) is carried out over the Brillouin zone. In the low-temperature limit with respect to T_i

$$c_V=\frac{\pi v}{30}T_i^3\sum_s \int_0^{2\pi} d\varphi \int_0^{\pi}\frac{\sin\theta\,d\theta}{c_s^3(\theta,\varphi)}, \tag{53}$$

where θ and φ are the angles in the spherical coordinate system for $\mathbf{k}$, and the velocity of sound $c_s(\theta,\varphi)$ for long waves equals

$$c_s(\theta,\varphi)=\frac{\omega_s(\mathbf{k})}{k}=\frac{\omega_s(k,\theta,\varphi)}{k}. \tag{54}$$

The expression (53) permits finding the Debye temperature $\Theta_D(T_e,v)$, which can be determined from the low-temperature specific heat. It obeys the scaling relation

$$\begin{aligned}\Theta_D(T_e,v)&=Z^{3/2}A^{-1/2}\bar{\Theta}_D(\bar{T}_e,\bar{v})\\&=Z^{3/2}A^{-1/2}\bar{\Theta}_D(Z^{-4/3}T_e,Zv)\end{aligned} \tag{55}$$

with the reduced Debye temperature

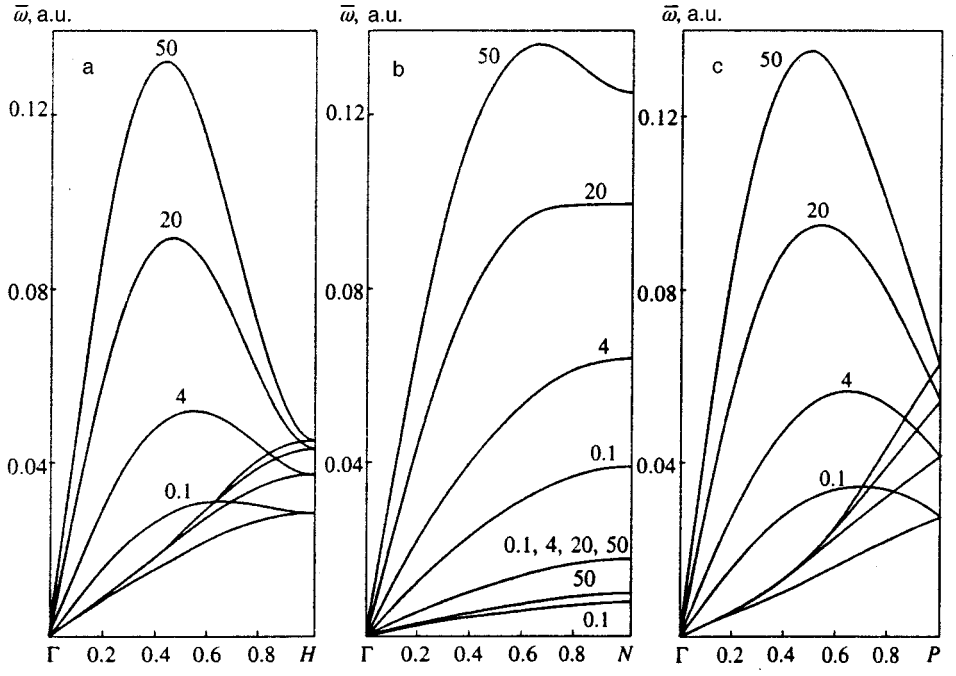


FIG. 2. Same as in Fig. 1 for $\bar{a}=2.0$.

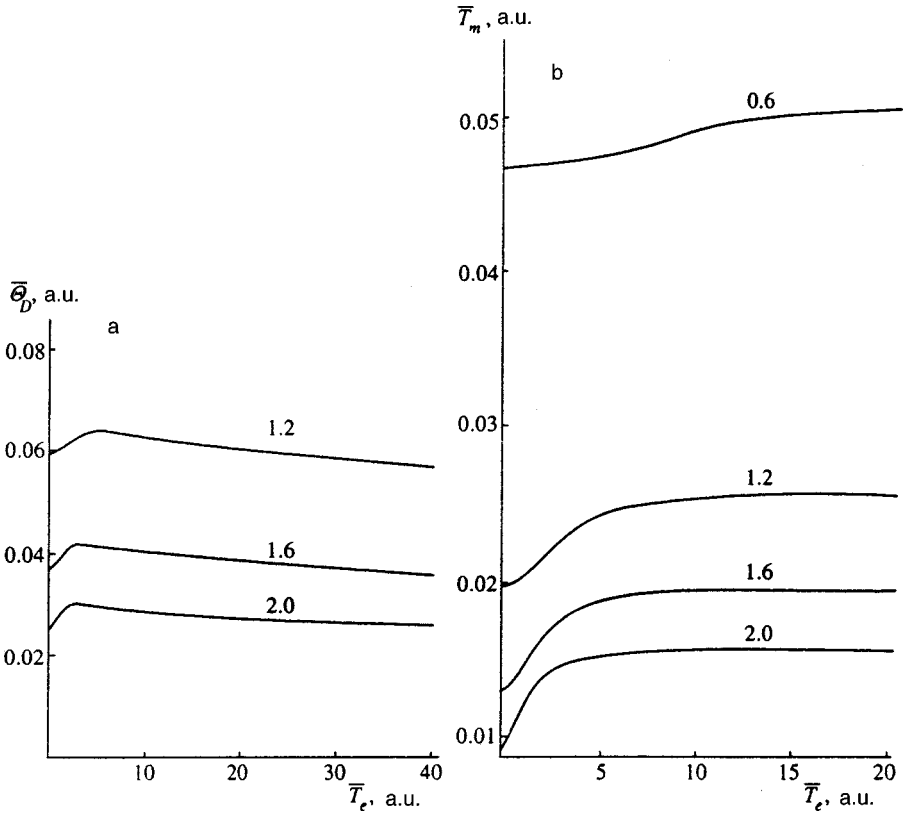


FIG. 3. Reduced Debye temperature $\overline{\Theta}_D$ (a) and melting point $\overline{T}_m$ (b) of a bcc crystal as a function of reduced electron temperature for various values of the reduced lattice constant $\overline{a}$ (which are indicated on the curves). The Debye temperature and the melting point of a crystal with specified Z and A are found from the respective quantities using the relations (55) and (61).

$$\overline{\Theta}_D(\overline{T}_e,\overline{v})=2\cdot 3^{2/3}\pi\left(\overline{v}\sum_s \int_0^{2\pi} d\varphi \int_0^{\pi} \frac{\sin\theta\, d\theta}{\overline{c}_s^3(\theta,\varphi)}\right)^{-1/3}. \tag{56}$$

The reduced Debye temperature is shown in Fig. 3a as a function of the reduced electron temperature for the values of the reduced lattice constant $\overline{a}=1.2$, 1.6, and 2.0. The value of $\overline{\Theta}_D(\overline{T}_e,\overline{v})$ is determined by integrating the phonon specific heat, which is low-temperature with respect to T_i and proportional to T_i^3, over the Brillouin zone. The variation of $\overline{\Theta}_D$ is less than 30% over the entire range of $\overline{T}_e$ considered, i.e., the dependence of the Debye temperature on electron temperature is weak over a broad range of values of $\overline{T}_e$. This is because the phonon modes which increase most strongly with electron temperature make the smallest contribution to the Debye temperature when it is determined in this manner.

In accordance with the stronger variation of the spectrum at larger values of the lattice constant, the variation of the Debye temperature is more significant at larger $\overline{a}$. Being a function of electron temperature, the reduced Debye temperature $\overline{\Theta}_D$ has a maximum, which shifts toward higher temperatures as the lattice constant is diminished, i.e., as the Fermi energy is increased. Since the hardest branches of phonon frequencies make a small contribution to $\overline{\Theta}_D$ as the electron temperature $\overline{T}_e$ increases when the Debye temperature is determined in this manner, the energy of the zero-point vibrations, which is determined mainly by the hardest modes and is equal to $9\overline{\Theta}_D/8$ per atom in the Debye approximation, is poorly described by the Debye approximation as the electron temperature increases. This also applies to the pressure of the zero-point vibrations.

We can use the phonon spectrum obtained to determine the dependence of the melting point of a crystal on electron temperature (we recall that we are dealing with a situation in which there is a lack of equilibrium between the electronic subsystem and the lattice and their temperatures T_e and T_i can differ significantly). Introducing the radius R_s of a sphere with a volume equal to the specific volume per atom v, so that

$$\frac{4}{3}\pi R_s^3=v=\frac{a^3}{2},$$

we determine the relative mean-square displacement of the nuclei from their positions in an ideal lattice:

$$\zeta=\langle u^2\rangle/R_s^2. \tag{57}$$

Here the mean-square displacement $\langle u^2\rangle$ is found by integration over the Brillouin zone:

$$\langle u^2\rangle=\sum_s \int_{BZ} \frac{1}{M\omega_s(\mathbf{k})}\left\{\frac{1}{\exp[\omega_s(\mathbf{k})/T_i]-1}+\frac{1}{2}\right\}\frac{v\,d^3k}{(2\pi)^3}. \tag{58}$$

At large values of T_i near the melting point T_m the expression (58) can be written in the form

$$\langle u^2\rangle=\sum_s \int_{BZ} \frac{T_i}{M\omega_s^2(\mathbf{k})}\frac{v\,d^3k}{(2\pi)^3}. \tag{59}$$

Taking into account the scaling relation (50), we obtain

$$M\omega_s^2(\mathbf{k}) = Z^3 M_0 \overline{\omega}_s^2(\overline{\mathbf{k}}). \quad (60)$$

We determine the melting point T_m, according to the Lindemann criterion, from the critical value of the mean-square displacement ζ_m, which, following Ref. 14, we set equal to $\zeta_m = 0.076$. This gives the following expression for the melting point:

$$T_m(T_e, v) = Z^{7/3}\overline{T}_m(\overline{T}_e, \overline{v}) = Z^{7/3}\overline{T}_m(Z^{-4/3}T_e, Zv), \quad (61)$$

where the reduced melting point $\overline{T}_m$ equals

$$\overline{T}_m(\overline{T}_e, \overline{v}) = \zeta_m \left(\frac{3\overline{v}}{4\pi}\right)^{2/3} \left[\sum_s \int_{BZ} \frac{\overline{v}\, d^3\overline{k}}{(2\pi)^3 M_0 \overline{\omega}_s^2(\overline{\mathbf{k}})}\right]^{-1}. \quad (62)$$

The variation of the reduced melting point of a crystal as a function of $\overline{T}_e$ over the broad range of variation of $\overline{T}_e$ considered is shown in Fig. 3b. As in the case of the Debye temperature, this variation is fairly small, not exceeding 40% in the specific volume range considered. In the expression (62), as in (54), the phonon branches, whose frequencies increase most strongly with increasing electron temperature make the smallest contribution to the integral over the Brillouin zone, while, as follows from the phonon spectra in Figs. 1 and 2, the softer modes, which make the main contribution to this integral, depend weakly on the temperature of the electronic subsystem. This determines the general variation of the melting point of the lattice as the electron temperature increases, which is not as significant as that for the hard branches of the phonon spectrum. Thus, a significant increase in the electron temperature, which can differ from the lattice temperature by several orders of magnitude, causes an increase in the phonon frequencies of a metal, primarily of the hardest modes in the spectrum. At the same time, thermodynamic characteristics associated with the phonon spectrum, such as the Debye temperature and the melting point, vary to a significantly smaller extent with electron temperature.

This work was performed with assistance from the Russian Fund for Fundamental Research and the Program for Supporting Leading Scientific Schools.

*)Present address: Department of Chemistry, University of California, Davis, California 95616, USA.
†)E-mail: petrov@landau.ac.ru

[1] S. I. Anisimov, B. L. Kapeliovich, and T. I. Perel'man, Zh. Éksp. Teor. Fiz. **66**, 776 (1974) [Sov. Phys. JETP **39**, 375 (1975)].
[2] S. I. Anisimov, N. A. Inogamov, and Yu. V. Petrov, Phys. Lett. A **55**, 449 (1976).
[3] H. E. Elsayed-Ali, T. B. Norris, M. A. Pessot, and G. A. Mourou, Phys. Rev. Lett. **58**, 1212 (1987).
[4] D. F. Price, R. M. More, R. S. Walling, G. Guethlein, R. I. Shepherd, R. E. Stewart, and W. E. White, Phys. Rev. Lett. **75**, 252 (1995).
[5] W. M. Knox, C. Hirlimann, D. A. B. Miller, J. Shah, D. S. Chemla, and C. V. Shank, Phys. Rev. Lett. **56**, 1191 (1986).
[6] J. R. Goldman and J. A. Prybyla, Phys. Rev. Lett. **72**, 1364 (1994).
[7] M. Ross, Rep. Prog. Phys. **48**, 1 (1985).
[8] R. B. Laughlin, Phys. Rev. A **33**, 510 (1986).
[9] J. A. Van Vechten, R. Tsu, and F. W. Saris, Phys. Lett. A **74**, 41 (1979).
[10] P. Stamphli and K. H. Bennemann, Phys. Rev. B **42**, 7163 (1990).
[11] P. Stamphli and K. H. Bennemann, Phys. Rev. B **49**, 7299 (1994).
[12] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Vol. 1*, 3rd. ed., Pergamon Press, Oxford–New York (1980).
[13] R. Latter, Phys. Rev. **99**, 1854 (1955).
[14] H. Nagara and T. Nakamura, Phys. Rev. B **31**, 1844 (1985).

Translated by P. Shelnitz

Isotope effect for the thermal expansion coefficient of germanium

V. I. Ozhogin,*) N. A. Babushkina, L. M. Belova, and A. P. Zhernov

Institute of Molecular Physics, ''Kurchatov Institute'' Russian Science Center, 123182 Moscow, Russia

E. E. Haller

University of California at Berkeley and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

K. M. Itoh

Department of Applied Physics, Keio University, Yokohama, 223-8522 Japan

(Submitted 11 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 243–248 (January 1999)

The first experimental and theoretical investigation of the difference in the temperature behavior of the linear expansion coefficients of single crystals grown from isotopically highly enriched and natural germanium is reported. A comparison of the data for ^{70}Ge and ^{nat}Ge crystals reveals the significant influence of isotopic composition over a wide range of temperatures 30–230 K. © *1999 American Institute of Physics.* [S1063-7761(99)02101-0]

Many problems in the theory of the thermal expansion of crystal lattice have been thoroughly studied (see, e.g., Ref. 1). To the best of our knowledge, however, the thermal expansion coefficient α for crystals differing solely in their isotopic composition has not investigated. In this paper we report the first experimental and theoretical study of the behavior of the coefficient α as a function of the isotopic composition over a wide range of temperatures. We are specifically concerned with natural germanium (^{nat}Ge) and highly enriched (99.99%) germanium (^{70}Ge) samples.

The dependence of the thermal expansion coefficient α on the temperature T is dictated by the temperature behavior of the heat capacity in many cases. This is because the total Grüneisen factor γ (which is sensitive to the anharmonic interatomic force parameters) usually exhibits a weak dependence on T. For germanium, on the other hand, the dependence of γ on T is very strong. Here, since $\gamma(T)$ is a sign-indefinite function, the behavior of $\alpha(T)$ for Ge is qualitatively determined primarily by the Grüneisen factor and not by the heat capacity.[1]

In regard to kinetic phenomena two types of isotope effects are possible, which differ in that one depends linearly, and the other quadratically on the difference in the masses of the isotopes. Linear effects are governed by the variation of the phonon spectrum as the isotopic composition changes. Quadratic effects are associated with the irregular distribution of the isotopes and induce an additional relaxation mechanism of phonon (and electron) scattering. Their role has been investigated in application to the thermal conductivity of germanium in, for example, Refs. 2 and 3. As for thermal expansion, on the other hand, both the linear and the quadratic dependences on the isotopic mass are entirely the result of the variation of the phonon spectrum.

EXPERIMENTAL OBSERVATION OF THE ISOTOPE EFFECT IN THE THERMAL EXPANSION OF GERMANIUM

We have performed for the first time measurements of the temperature dependence of the difference in the linear expansion coefficients of chemically pure and structurally perfect germanium ^{70}Ge and ^{nat}Ge single crystals over a wide temperature range 30–230 K. According to mass-spectrometer measurements, the content of the primary isotope in ^{70}Ge single crystals is at least 99.99%. In this connection a measurement of the Hall conductivity in it has shown that the total concentration of electrically active impurities does not exceed $2\times10^{12}\,\mathrm{cm}^{-3}$ (Ref. 2). The ^{nat}Ge single crystal is a mixture of five isotopes having an average mass of 72.59.

The thermal expansion of the samples was investigated by means of a strain-gauge dilatometer. The recording part of the instrument was a bridge circuit. The customary approach in this method is to place the investigated sample and a standard sample on the instrument mounting, with identical standard resistance gauges (Ni–Cu–Cr wire of diameter $30\,\mu$m, $R=100\,\Omega$) attached to them. As the temperature is varied, the deformation of the gauges differs because the samples expand differently, producing different variations of their resistances and thus creating a bridge unbalance. This method has a sensitivity of 5×10^{-7} cm. The temperature dependence of the thermal expansion was measured with the samples heated at a rate not greater than 0.3 K/min. The temperature sensor was a copper–iron (Cu+0.05%Fe)–copper thermocouple. The mounting of the samples is shown in Fig. 1. In principle, the experimental error can depend on the gauge bonding technology and on the temperature interval in which the measurements are performed. Additional measurements have shown that the error is less than 5%.

The quality of the measurement system was tested using a standard sample of pure copper, for which the values of the

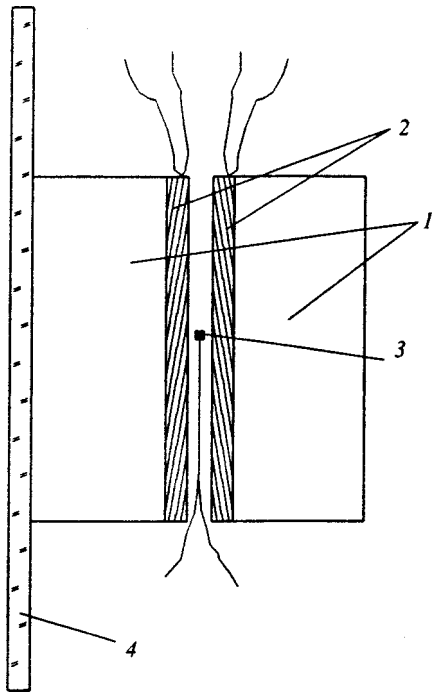


FIG. 1. Schematic view of the sample mounting. (*1*) Sample and standard; (*2*) resistance gauges; (*3*) thermocouple; (*4*) quartz rod, to which the block with the samples is attached by thin nylon thread.

thermal expansion coefficient are well known. Our data for copper (in measurement relative to quartz) differ by at most 5% from established data at temperatures of 20–80 K (Ref. 1).

To improve the reliability of the results in the present study, we performed direct measurements of the difference in the thermal expansion coefficients of chemically pure ^{nat}Ge and ^{70}Ge single crystals. The samples of natural and isotopically pure germanium were mounted in two arms of the measurement bridge. The measured bridge unbalance in this case is proportional to the difference in the thermal expansions of the two samples, ^{nat}Ge and ^{70}Ge.

The samples were cut from germanium single crystals in the shape of $5\times4\times2$-mm parallelepipeds with their longest edges parallel to the [100] axis of the crystal.

This setup was used to measure the relative thermal elongation $\Delta l/l$ of the sample. Our object of interest, the thermal expansion coefficient $\alpha=(1/l)(dl/dT)$, was calculated by numerical differentiation of the graph of $\Delta l/l$ as a function of T.

The results of measurements of the difference $\Delta\alpha$ in the thermal expansion coefficients of the two single crystals are shown in Fig. 2. The scatter of the values of $\Delta\alpha$ as a function of T was found to be $\sim$20% after differentiation of the experimental data with respect to T. The scatter of $\Delta\alpha(T)$ is large because the measured quantity is the temperature dependence of the small difference between the thermal expansions $\Delta l/l$ of the two single crystals (^{nat}Ge and ^{70}Ge). The $\Delta\alpha(T)$ curve in Fig. 2 is smoothed by a polynomial. The absolute rms error of the values so obtained for $\Delta\alpha(T)$ is $\sim$20% in the temperature range 60–200 K and $\sim30-40\%$ in the low-temperature range. Errors can also be attributed to

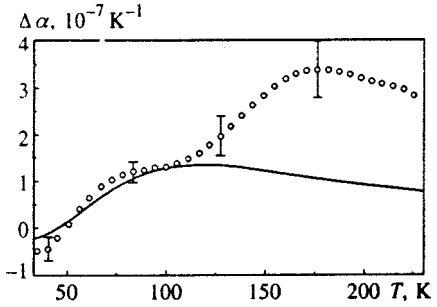


FIG. 2. Temperature dependence of the factor $\Delta\alpha=\alpha_c(M_{c1})-\alpha_c(M_{c2})$, where $M_{c1}=72.59$ and $M_{c2}=70$, theoretical (solid curve) and experimental (dots).

limited capabilities for regulating the temperature regime of the experimental setup, for example, an insufficiently slow temperature scanning rate in measuring the temperature dependence of the relative thermal expansion of germanium.

THEORETICAL ANALYSIS OF THERMAL EXPANSION TO FIRST ORDER IN THE DIFFERENCE IN ISOTOPIC MASSES

In the quasiharmonic approach to the linear isotopic mass-difference approximation for cubic crystals the linear expansion coefficient α is given by the relations[4]

$$\alpha(T)=\frac{1}{3\Omega_0 B_0}\sigma(T), \tag{1}$$

$$\sigma(T)=\sum_l \gamma(l)C_l(T), \quad C_l(T)=\frac{1}{T^2}\omega^2(l)n(\omega(l))\times[n(\omega(l))+1]. \tag{2}$$

Here $\omega(l)$ is the phonon frequency of the lth mode with quasimomentum $\mathbf{f}$ and polarization j, i.e., $l=\{\mathbf{f},j\}$, $n(\omega)$ is the Planck distribution, $\gamma(l)$ is the partial (mode) Grüneisen factor, i.e., by definition

$$\gamma(l)=-\left.\frac{\partial\ln\omega(l)}{\partial\ln\Omega}\right|_{\Omega=\Omega_0}, \tag{3}$$

C_l denotes the heat capacity of the lth mode, Ω_0 is the equilibrium unit cell volume of the lattice, and B_0 is the hydrostatic compression modulus at $T=0$. The Boltzmann and Planck constants are set equal to unity.

Equation (1) can be written in the form

$$\alpha(T)=\frac{1}{3\Omega_0 B_0}\gamma(T)C_L(T), \tag{1'}$$

where $\gamma(T)$ is the total Grüneisen factor, and $C_L(T)=\Sigma_l C_l(T)$ is the lattice heat capacity.

Note the following relation, which holds for any isotopic composition in any mode:

$$M_c\omega^2(l)=\varphi(l), \tag{4}$$

where the effective force parameter $\varphi(l)$ does not depend on the average mass M_c. By definition,

$$M_c = \sum_i c_i M_i ,$$

where c_i is the concentration and M_i is the mass of isotopes of the ith species.

We now fix a specific isotopic composition by the index c_0. Making use of Eqs. (1) and (4), we then have a universal relation for an arbitrary isotopic composition, which we label with the index c:

$$\alpha_c(T) = \alpha_{c_0}(T'), \quad T' = T\sqrt{M_c/M_{c_0}}. \tag{5}$$

We close this section with a few words about the general case of an anharmonic crystal lattice. In such a lattice the mode frequency $\tilde{\omega}(l)$ consists of the sum of the harmonic and anharmonic contributions:

$$\tilde{\omega}(l, T) = \omega(l, \Omega) + \Delta^{(a)}(l, \Omega). \tag{6}$$

The increment $\Delta^{(a)}$ is determined by standard third- and fourth-order anharmonic processes.[5,6] The expression for the anharmonic correction to the Grüneisen factor $\gamma(l)$ now has the form

$$\Delta\gamma^{(a)}(l, T) \approx \Delta\Omega(T)\frac{\partial\gamma(l)}{\partial\Omega_0} - \frac{\partial\{\Delta^{(a)}(l)/\omega(l)\}}{\partial\Omega_0}. \tag{7}$$

Here $\Omega(T) = \Delta\Omega(T) + \Omega_0$ is the unit cell volume.

Second-order effects with respect to the difference in the isotopic masses are not discussed in this paper, because an estimate shows that their role is of little consequence in application to germanium.

COMPARISON OF THEORY WITH EXPERIMENT

Using relations (1)–(3) and (5), we calculate the coefficients $\alpha(T)$ for germanium crystals with average masses of 72.59 and 70. Here we determine the frequencies of the phonon modes $\omega(l)$ from the Born–von Kármán theory. We use the force parameters obtained previously[7] by the fitting of experimental inelastic neutron scattering data. We also use the partial Grüneisen factors γ_l determined in Ref. 8 in the microscopic bond-charge model. In addition, we assign the following values to the lattice constant a_0 and the compression modulus B_0: $a_0 = 5.658$ Å; $B_0 = 0.772\times10^{12}$ dyn/cm^2 (Ref. 9).

The results of the calculations are shown in the same figure as the experimental data (Fig. 2). The difference curves $\Delta\alpha = \alpha(M_{c1}) - \alpha(M_{c2})$ are shown for germanium crystals with masses $M_{c1} = 72.59$ and $M_{c2} = 70$, respectively. It is evident at once that theory and experiment are in reasonably good agreement in the temperature range $\sim 30 - 100$ K. The agreement is unsatisfactory at higher temperatures.

In our work we have observed that the isotope effect is large in the thermal expansion of germanium. In the temperature interval from 75 K to 125 K the relative difference $\Delta\alpha/\alpha$ reaches 10%. The data for α in ^{nat}Ge are borrowed from Ref. 1. According to the above calculations, $\Delta\alpha$ should not be more than a few percent at higher temperatures.

We also estimate the anharmonic corrections to the partial Grüneisen factors in the model of Ref. 8 on the basis of Eqs. (6) and (7). We note that for the majority of the modes the parameters $\gamma^{(a)}(l) = \gamma(l) + \Delta\gamma^{(a)}(l)$ increase with the temperature. The anharmonicity-induced renormalizations, for example, at $T \approx 150$ K, are approximately 4%. On the other hand, a nonstandard situation is encountered for soft transverse modes. The corresponding values of $\gamma^{(a)}$ are approximately 1.3 times smaller in absolute value as a result of anharmonicity at the same temperature of 150 K. Here the calculations yield $\Delta\alpha = \alpha_c(M_{c1}) - \alpha_c(M_{c2}) = 0.142 \times 10^{-6}$ K^{-1} instead of the value of 0.121×10^{-6} K^{-1} obtained in the quasiharmonic approach. Consequently, when anharmonicity is taken into account, the quantitative description of the experimental data improves, but the discrepancy is still substantial.

We note that the theoretical calculations give results too low in comparison with the experimental data. The disparity could stem from the small difference in the chemical purity of the germanium ^{nat}Ge and ^{70}Ge single crystals. According to Ref. 10, the thermal expansion coefficient for Ge can differ for samples having different impurity concentrations. Moreover, the model in Ref. 8 is also in need of further development and refinement.

To summarize, for the first time we have measured the temperature dependence of the difference in the linear expansion coefficients of chemically pure germanium ^{nat}Ge and ^{70}Ge single crystals. We have found that theory and experiment concur reasonably well in the temperature range $\sim 30 - 100$ K. The agreement falls short at higher temperatures. Allowance for the anharmonicity of interaction between phonon modes brings theory and experiment somewhat closer together.

The authors are grateful to E. A. Chistotina for assistance in processing the experimental results.

This work has received financial support from the INTAS Foundation (Project 96-0546) and from the Russian Fund for Fundamental Research (Project 96-15-96738).

*)E-mail: ozhogin@imp.kiae.ru

[1] S. I. Novikova, *Thermal Expansion of Solids* [in Russian], Nauka, Moscow (1974).
[2] V. I. Ozhogin, A. V. Inyushkin, A. N. Taldenkov *et al.*, JETP Lett. **63**, 490 (1996).
[3] M. Asen-Palmer, K. Bartkowski, E. Gmelin *et al.*, Phys. Rev. B **56**, 9431 (1997).
[4] G. Leibfried, ''Gittertheorie der mechanischen und thermischen Eigenschaften der Kristalle (Lattice theory of the mechanical and thermal properties of crystals),'' in *Handbuch der Physik*, S. Flügge (Ed.), Vol. 7, Part 1, Springer-Verlag, Berlin (1955), pp. 104–324.
[5] M. A. Krivoglaz, *Theory of X-Ray and Thermal Neutron Scattering by Real Crystals*, Plenum Press, New York (1969).
[6] J. A. Reissland, *The Physics of Phonons*, Wiley, New York (1973).
[7] A. D. Zdetsis and C. S. Wang, Phys. Rev. B **19**, 2999 (1979).
[8] R. Eryiğit and I. P. Herman, Phys. Rev. B **53**, 7775 (1996).
[9] G. Leibfried and N. Breuer, *Point Defects in Metals, I: Introduction to the Theory*, Springer-Verlag, Berlin (1978).
[10] V. V. Zhdanova, Fiz. Tverd. Tela (St. Petersburg) **5**, 3341 (1963) [Sov. Phys. Solid State **5**, 2451 (1963)].

Translated by James S. Wood

Exactly solvable two-dimensional quantum spin models

D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov*)

N. M. Émanuél Institute of Biochemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia
(Submitted 19 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 249–267 (January 1999)

A method is proposed for constructing an exact ground-state wave function of a two-dimensional model with spin 1/2. The basis of the method is to represent the wave function by a product of fourth-rank spinors associated with the nodes of a lattice and the metric spinors corresponding to bonds between nearest neighbor nodes. The function so constructed is an exact wave function of a 14-parameter model. The special case of this model depending on one parameter is analyzed in detail. The ground state is always a nondegenerate singlet, and the spin correlation functions decay exponentially with distance. The method can be generalized for models with spin 1/2 to other types of lattices. © *1999 American Institute of Physics.* [S1063-7761(99)02201-5]

1. INTRODUCTION

There has been growing interest lately in quantum spin systems with frustrated interactions.[1–10] Of special importance are models in this category for which it is possible to construct an exact ground state. The first example of such a representation was the well-known Majumdar–Ghosh model.[11] It comprises a chain of spins 1/2 with antiferromagnetic interactions J_1 and J_2 of nearest neighbor and next-nearest neighbor spins, where $J_2=J_1/2$. The ground state of this model is doubly degenerate and consists of dimerized singlets; moreover, there is a gap in the spectrum of excitations. Another example of an exactly solvable model is the one-dimensional model with bilinear and biquadratic interactions and spin 1, investigated by Affleck, Kennedy, Lieb, and Tasaki[12] (the AKLT model). Its ground state has a structure of the type where each neighboring pair of spins has valence bonds. It is not degenerate, the spin correlation functions in the ground state decrease exponentially with distance, and there is a gap in the spectrum of excitations. This model therefore has properties predicted by Haldane[13] for the one-dimensional Heisenberg antiferromagnetic model with spin 1. The valence-bond ground state is also exact for systems with many dimensions, but with spin $d/2$ (d is the coordination number of the lattice).[14] The one-dimensional AKLT model has subsequently been generalized and investigated in a number of papers,[15] where it has also been shown that the wave function of the ground state can be represented by the trace of the product of matrices describing the spin states of nodes of a chain (the ''matrix'' form). These two examples are characterized by the fact that the total Hamiltonian of the model is written as a sum of cell Hamiltonians (which are not mutually commuting), and the exact ground-state wave function of the total system is the eigenfunction having the lowest energy of each cell Hamiltonian.

We have previously[16] investigated an exactly solvable, one-dimensional, frustrated model with spin 1/2, whose properties by and large are similar to those of the AKLT model. The ground-state wave function has a special recursion formula, and we have shown that it can be reduced to matrix form. Note, however, that both the recursive form and the matrix form are essentially one-dimensional constructions and cannot be extended directly to higher-dimensional systems. We cite Ref. 17 in this regard, where a method has been proposed for constructing an exact wave function of the ground state for models with spin 3/2 on a hexagonal lattice. The same method is applicable to other systems with spin $d/2$.

In this paper we consider a class of models with spin 1/2 for which the exact wave function of the ground state can be represented in an alternative form. In the one-dimensional case this wave function reduces to a wave function that we have found previously,[16] but it admits generalization to higher-dimensional systems. The present study is devoted primarily to an analysis of the two-dimensional model.

The article is organized as follows. In Sec. 2 we discuss the method of constructing the exact wave function for a one-dimensional model with $s=1/2$. In Sec. 3 we formulate an exactly solvable two-dimensional model. In Sec. 4 we investigate the properties of this model with the aid of numerical calculations. In Sec. 5 we discuss the possibility of generalizing our treatment to other types of lattices. The Appendix gives a proof of the nondegeneracy of the ground state of the two-dimensional model in the presence of cyclic boundary conditions.

2. ONE-DIMENSIONAL MODEL

We have previously[16] investigated a one-dimensional, one-parameter model containing two spins 1/2 in the unit cell and admitting exchange interactions between nearest neighbor spins and spins separated by two nodes of the lattice. The exact ground-state wave function of the cyclic chain can be written in the form

$$\Psi_0=\mathrm{Tr}[A(1,2)\,A(3,4)\,\ldots\,A(N-1,N)], \tag{1}$$

where $A(2i-1,2i)$ is a 2×2 matrix associated with the ith unit cell.

Below we write the wave function Ψ_0 in a form more suitable for subsequent generalization to other types of lattices and give the general form of the Hamiltonian for which Ψ_0 is an exact wave function of the ground state.

We consider a chain of $N=2M$ spins 1/2. The wave function of this system is described by the Nth-rank spinor

$$\Psi=\Psi^{\lambda\mu\nu\ldots\tau}, \tag{2}$$

where the indices $\lambda, \mu, \nu, \ldots, \tau=1,2$ correspond to different projections of the spin 1/2.

We partition the system into pairs of nearest neighbor spins. The wave function can then be written as the product of M second-rank spinors

$$\Psi=\Psi^{\lambda\mu}(1)\Psi^{\nu\rho}(2)\ldots\Psi^{\sigma\tau}(M). \tag{3}$$

We now form a scalar from Eq. (3), simplifying the latter with respect to index pairs:

$$\Psi_s=\Psi^{\lambda}_{\nu}(1)\Psi^{\nu}_{\kappa}(2)\ldots\Psi^{\sigma}_{\lambda}(M). \tag{4}$$

Here subscripts correspond to the covariant components of the spinor, which are related to the contravariant components (superscripts) through the metric spinor

$$g_{\lambda\mu}=g^{\lambda\mu}=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{5}$$

$$\Psi_{\lambda}=g_{\lambda\mu}\Psi^{\mu}, \quad \Psi^{\lambda}=g^{\mu\lambda}\Psi_{\mu}. \tag{6}$$

The scalar function (4) can thus be written in the form

$$\Psi_s=\Psi^{\lambda\mu}(1)g_{\mu\nu}\Psi^{\nu\rho}(2)g_{\rho\kappa}\ldots\Psi^{\sigma\tau}(M)g_{\tau\lambda}. \tag{7}$$

The scalar function Ψ_s does not depend on the angle of rotation of the coordinate system and therefore corresponds to the singlet state.

The second-rank spinor describing the pair of spins 1/2 can be written in the form

$$\Psi^{\lambda\mu}=c_t\Psi_t^{\lambda\mu}+c_s\Psi_s^{\lambda\mu}, \tag{8}$$

where $\Psi_t^{\lambda\mu}$ and $\Psi_s^{\lambda\mu}$ are symmetric and antisymmetric second-rank spinors, respectively, and c_t and c_s are arbitrary constants. We know that the symmetric second-rank spinor describes a system with spin 1, so that the pair of spins 1/2 in this case forms a triplet. If $\Psi^{\lambda\mu}$ is an antisymmetric second-rank spinor reducible to a scalar multiplied by $g_{\lambda\mu}$, the spin pair exists in the singlet state. Consequently, the ratio of the constants c_t and c_s determines the relative weights of the triplet and singlet components on the pair of spins $s=1/2$ and is a parameter of the model. In particular, for $c_s=0$ the wave function (8) contains only a triplet component, and for $c_t=0$ it contains only a singlet component.

In general, we can make the ratio of the constants c_s/c_t different in different pairs, but to preserve translational symmetry, we confine the discussion to the case in which this ratio is the same in every pair.

We note that the wave function (4) has the matrix form (1), the matrices $A(2i-1,2i)$ representing a mixed second-rank tensor:

$$A_{\lambda\nu}(1,2)=\Psi^{\lambda}_{\nu}(1)$$
$$=c_t\begin{pmatrix} \frac{1}{2}(\alpha_1\beta_2+\beta_1\alpha_2) & \beta_1\beta_2 \\ -\alpha_1\alpha_2 & -\frac{1}{2}(\alpha_1\beta_2+\beta_1\alpha_2) \end{pmatrix}$$
$$-c_s\frac{1}{2}(\alpha_1\beta_2-\beta_1\alpha_2)\,I, \tag{9}$$

where α_i and β_i denote the up and down projections of the spin $\mathbf{s}_i$, respectively, and I is the unit matrix.

We now choose a Hamiltonian H for which the wave function (7) is an exact ground-state wave function. To do so, we consider the part of the system (cell) consisting of two nearest neighbor spin pairs. In the wave function (7) the factor corresponding to the two spin pairs is a second-rank spinor:

$$\Psi^{\lambda\mu}(i)g_{\mu\nu}\Psi^{\nu\rho}(i+1). \tag{10}$$

In the general case, therefore, only two of the six multiplets forming two pairs of spin 1/2—one singlet and one triplet—are present in the wave function (10). Inasmuch as four spins 1/2 form two singlets and three triplets, the specific form of the singlet and triplet components present in the wave function (10) depends on the ratio c_s/c_t. The cell Hamiltonian acting in the spin space of nearest neighbor spin pairs can be written as the sum of the projectors onto the four missing multiplets with arbitrary positive coefficients $\lambda_1, \lambda_2, \lambda_3, \lambda_4$:

$$H_{i,i+1}=\sum_{k=1}^{4}\lambda_k P_k^{i,i+1}, \tag{11}$$

where $P_k^{i,i+1}$ is the projector onto the missing multiplets in the corresponding cell Hamiltonian.

The wave function (7) is now an exact wave function of the ground state of the cell Hamiltonian $H_{i,i+1}$ with zero energy, because

$$H_{i,i+1}|\Psi_s\rangle=0, \tag{12}$$

and $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the excitation energies of the corresponding multiplets.

The total Hamiltonian of the entire system can be written as the sum of mutually noncommuting cell Hamiltonians:

$$H=\sum_{i=1}^{N}H_{i,i+1}, \tag{13}$$

and since each term $H_{i,i+1}$ in (13) yields zero in its action on Ψ_s, we have

$$H|\Psi_s\rangle=0. \tag{14}$$

The nondegeneracy of the ground state of this Hamiltonian has been rigorously proved.[16]

Since the specific form of the existing and missing multiplets in the wave function (7) on every two nearest neighbor spin pairs depends on the model parameter c_s/c_t, the projectors in (11) also depend on c_s/c_t. Each projector can be written in the form

$$P_k^{1,2}=J_{12}^{(k)}(\mathbf{s}_1\cdot\mathbf{s}_2+\mathbf{s}_3\cdot\mathbf{s}_4)+J_{13}^{(k)}(\mathbf{s}_1\cdot\mathbf{s}_3+\mathbf{s}_2\cdot\mathbf{s}_4)+J_{14}^{(k)}\mathbf{s}_1\cdot\mathbf{s}_4+J_{23}^{(k)}\mathbf{s}_2\cdot\mathbf{s}_3+J_1^{(k)}(\mathbf{s}_1\cdot\mathbf{s}_2)(\mathbf{s}_3\cdot\mathbf{s}_4)+J_2^{(k)}(\mathbf{s}_1\cdot\mathbf{s}_3)(\mathbf{s}_2\cdot\mathbf{s}_4)+J_3^{(k)}(\mathbf{s}_1\cdot\mathbf{s}_4)(\mathbf{s}_2\cdot\mathbf{s}_3)+C^{(k)}, \tag{15}$$

and this representation is unique for a fixed value of the parameter c_s/c_t.

Substituting the above expressions for the projectors into Eq. (11), we obtain the general form of the cell Hamiltonians $H_{i,i+1}$. Inasmuch as the Hamiltonians $H_{i,i+1}$ have exactly the same form for any i, it suffices here to give the expression for $H_{1,2}$:

$$H_{1,2}=J_{12}(\mathbf{s}_1\cdot\mathbf{s}_2+\mathbf{s}_3\cdot\mathbf{s}_4)+J_{13}(\mathbf{s}_1\cdot\mathbf{s}_3+\mathbf{s}_2\cdot\mathbf{s}_4)+J_{14}\mathbf{s}_1\cdot\mathbf{s}_4+J_{23}\mathbf{s}_2\cdot\mathbf{s}_3+J_1(\mathbf{s}_1\cdot\mathbf{s}_2)(\mathbf{s}_3\cdot\mathbf{s}_4)+J_2(\mathbf{s}_1\cdot\mathbf{s}_3)(\mathbf{s}_2\cdot\mathbf{s}_4)+J_3(\mathbf{s}_1\cdot\mathbf{s}_4)(\mathbf{s}_2\cdot\mathbf{s}_3)+C, \tag{16}$$

where all volume integrals depend on the model parameter and the spectrum of excited states $J_i=J_i(c_s/c_t,\lambda_1,\lambda_2,\lambda_3,\lambda_4)$. In particular, for $c_s=0$, choosing $\lambda_2=\lambda_3=\lambda_4$ and $\lambda_1/\lambda_2=3$, we obtain an expression for $H_{1,2}$ in the form

$$H_{1,2}=\mathbf{L}_1\cdot\mathbf{L}_2+\frac{1}{3}(\mathbf{L}_1\cdot\mathbf{L}_2)^2+\frac{2}{3}, \tag{17}$$

where $\mathbf{L}_1=\mathbf{s}_1+\mathbf{s}_2$ and $\mathbf{L}_2=\mathbf{s}_3+\mathbf{s}_4$.

The Hamiltonian (17) has the form of the AKLT Hamiltonian, a result that is not too surprising, because for $c_s=0$ two spins 1/2 in a pair effectively form spin 1. Note, however, that for $c_s=0$ a set of different forms of the Hamiltonian $H_{1,2}$ exists, corresponding to a different choice of coefficients λ_k.

In general, the Hamiltonian (16) contains both bilinear and four-spin interactions. The latter can be excluded by setting $J_1=J_2=J_3=0$ and solving these equations for $\lambda_1,\lambda_2,\lambda_3,\lambda_4$. However, since the condition $\lambda_1,\lambda_2,\lambda_3,\lambda_4>0$, generally speaking, is not satisfied over the entire range of the parameter c_s/c_t, the simplified Hamiltonian will also have a ground state described by the wave function (7) only in the region where $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ are positive. The nonzero exchange integrals $J_{12},J_{13},J_{14},J_{23}$ and the constant C depend only on the parameter c_s/c_t. The explicit form of this dependence is given in Ref. 16, in which we have also calculated the ground-state spin correlation function $\langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle$, which decays exponentially with correlation length ~ 1.

We emphasize that the spin correlation functions $\langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle$ do not depend on the choice of $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ for a fixed parameter c_s/c_t, because the ground-state wave function of the four-parameter set of Hamiltonians is the same.

3. TWO-DIMENSIONAL MODEL

We consider an $M\times M$-node square lattice with cyclic boundary conditions. We replace each node of the lattice by a square (Fig. 1) with spins $s=1/2$ at its corners, making the total number of spins equal to $4M^2$. To avoid misunderstanding, however, from now on we continue to refer to these squares as nodes. The wave function of the system is described by the product of fourth-rank spinors

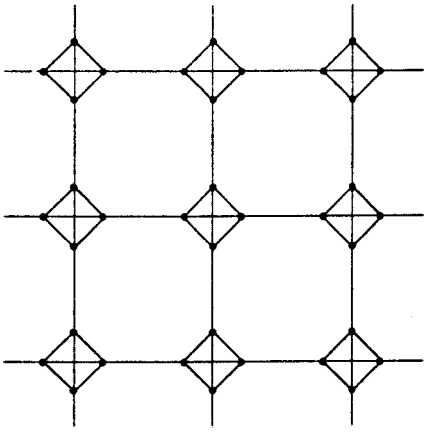


FIG. 1. Two-dimensional lattice on which the spin model is defined.

$$\Psi=\prod_{\mathbf{n}}\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n}). \tag{18}$$

By analogy with (7), from Eq. (18) we form the scalar

$$\Psi_s=\prod_{\mathbf{n}}\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}}g_{\rho_{\mathbf{n}}\mu_{\mathbf{n}+\mathbf{b}}}. \tag{19}$$

where **a** and **b** are unit vectors in the x and y directions.

The singlet wave function (19) is conveniently identified graphically with a square lattice, each node corresponding to a fourth-rank spinor $\Psi^{\lambda\mu\nu\rho}$ (whose form is identical for all nodes), and each segment linking nodes corresponds to a metric spinor $g_{\lambda\mu}$ (Fig. 2).

To completely define the wave function (19), it is necessary to know the form of the node spinor $\Psi^{\lambda\mu\nu\rho}$. For this purpose we classify an arbitrary fourth-rank spinor, simplifying and symmetrizing it with respect to different pairs of indices. We have the following types of spinors as a result:

1) a fourth-rank spinor $Q^{\lambda\mu\nu\rho}$ symmetric with respect to all indices;

2) three linearly independent products of a symmetric and an antisymmetric second-rank spinor: $\varphi^{\lambda\mu}g_{\nu\rho}$, $\varphi^{\lambda\nu}g_{\mu\rho}$, and $\varphi^{\lambda\rho}g_{\mu\nu}$;

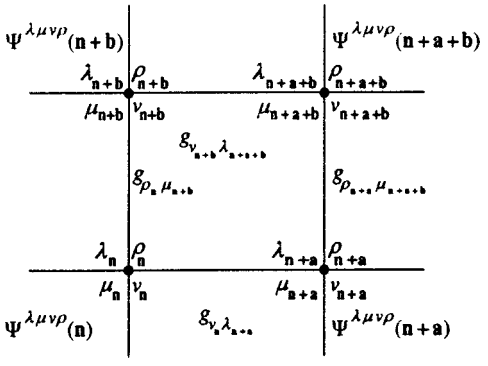


FIG. 2. Graphical correspondence of the model wave function. The indices of the node spinors depend on the node index (not shown in the figure).

3) two linearly independent products of two metric spinors and a scalar function: $g_{\lambda\mu}g_{\nu\rho}\chi$ and $g_{\lambda\nu}g_{\mu\rho}\chi$.

According to this classification, any fourth-rank spinor can be written in the form

$$\Psi^{\lambda\mu\nu\rho}=c_1Q^{\lambda\mu\nu\rho}+c_2\varphi_1^{\lambda\mu}g_{\nu\rho}+c_3\varphi_2^{\lambda\nu}g_{\mu\rho}+c_4\varphi_3^{\lambda\rho}g_{\mu\nu}+c_5g_{\lambda\mu}g_{\nu\rho}\chi_1+c_6g_{\lambda\nu}g_{\mu\rho}\chi_2. \tag{20}$$

We note, however, that because the system of four spins 1/2 contains one quintet, three triplets, and two singlets, Eq. (20) still does not completely determine the form of $\Psi^{\lambda\mu\nu\rho}$, and it is necessary to determine the specific form of the spinors $\varphi_1^{\lambda\mu}$, $\varphi_2^{\lambda\nu}$, and $\varphi_3^{\lambda\rho}$ and the scalar functions χ_1 and χ_2.

Each symmetric second-rank spinor $\varphi^{\lambda\mu}$ describes a triplet state of the system, representing a linear combination of the three basis triplet functions $\varphi_{t1}^{\lambda\mu}$, $\varphi_{t2}^{\lambda\mu}$, and $\varphi_{t3}^{\lambda\mu}$. We can now specify nine linearly independent spinors describing triplet states of four spins $s=1/2$:

$$\begin{aligned}&\varphi_{t1}^{\lambda\mu}g_{\nu\rho}, \quad \varphi_{t2}^{\lambda\mu}g_{\nu\rho}, \quad \varphi_{t3}^{\lambda\mu}g_{\nu\rho},\\ &\varphi_{t1}^{\lambda\nu}g_{\mu\rho}, \quad \varphi_{t2}^{\lambda\nu}g_{\mu\rho}, \quad \varphi_{t3}^{\lambda\nu}g_{\mu\rho},\\ &\varphi_{t1}^{\lambda\rho}g_{\mu\nu}, \quad \varphi_{t2}^{\lambda\rho}g_{\mu\nu}, \quad \varphi_{t3}^{\lambda\rho}g_{\mu\nu}.\end{aligned} \tag{21}$$

The products of two metric spinors and a scalar function $g_{\lambda\mu}g_{\nu\rho}\chi_1$ and $g_{\lambda\nu}g_{\mu\rho}\chi_2$ describe singlets states of four spins $s=1/2$. Since there are two independent singlet functions χ_{s1} and χ_{s2}, we have four linearly independent scalars describing singlet states of four spins $s=1/2$:

$$\begin{aligned}&g_{\lambda\mu}g_{\nu\rho}\chi_{s1}, \quad g_{\lambda\mu}g_{\nu\rho}\chi_{s2},\\ &g_{\lambda\nu}g_{\mu\rho}\chi_{s1}, \quad g_{\lambda\nu}g_{\mu\rho}\chi_{s2}.\end{aligned} \tag{22}$$

As a result, the specific form of the fourth-rank spinor $\Psi^{\lambda\mu\nu\rho}$ [and, hence, the wave function (19)] describing the system of four spins $s=1/2$ is governed by $1+9+4=14$ quantities, which are parameters of the model.

We now choose a Hamiltonian H for which the wave function (19) is an exact ground-state wave function. As in the one-dimensional case, we seek the required Hamiltonian in the form of a sum of cell Hamiltonians acting in the space of two nearest neighbor spin quartets:

$$H=\sum_{\mathbf{n}} H_{\mathbf{n},\mathbf{n}+\mathbf{a}}+\sum_{\mathbf{n}} H_{\mathbf{n},\mathbf{n}+\mathbf{b}}. \tag{23}$$

The first term in Eq. (23) is the sum of the cell Hamiltonians in the horizontal direction, and the second term is the same for the vertical. The cell Hamiltonians along each direction have the same form, but the ''horizontal'' and ''vertical'' Hamiltonians differ in general. In the ensuing discussion, therefore, we consider only the Hamiltonians $H_{1,2}$ and $H_{1,3}$ (Fig. 3), which describe interactions of ''nodes'' in the x and y directions, respectively.

For the wave function (19) to be an exact eigenfunction of the Hamiltonian H, it is sufficient that the sixth-rank spinors

$$\Psi^{\lambda_1\mu_1\nu_1\rho_1}(1)\Psi^{\lambda_2\mu_2\nu_2\rho_2}(2)g_{\nu_1\lambda_2}, \tag{24}$$

$$\Psi^{\lambda_1\mu_1\nu_1\rho_1}(1)\Psi^{\lambda_3\mu_3\nu_3\rho_3}(3)g_{\rho_1\mu_3}, \tag{25}$$

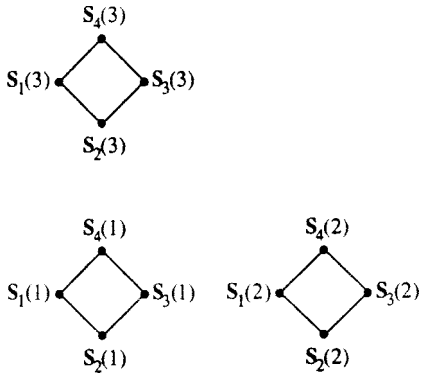


FIG. 3. Lattice nodes associated with interactions $H_{1,2}$ and $H_{1,3}$.

be eigenfunctions of the corresponding cell Hamiltonians $H_{1,2}$ and $H_{1,3}$.

In general, when the node spinor $\Psi^{\lambda\mu\nu\rho}$ is not symmetric with respect to any indices, the possible states of two quartets of spins $s=1/2$ consist of 70 multiplets. A wave function represented by a sixth-rank spinor contains only 20 of them. Accordingly, the cell Hamiltonians $H_{1,2}$ and $H_{1,3}$ can be represented by the sum of projectors onto the 50 missing multiplets:

$$H_{1,2}=\sum_{k=1}^{50}\lambda_kP_k^{1,2}, \quad H_{1,3}=\sum_{k=1}^{50}\mu_kP_k^{1,3}, \tag{26}$$

where the positive constants λ_k and μ_k are the excitation energies of $H_{1,2}$ and $H_{1,3}$, and the specific form of the projectors depends on 14 model parameters.

Inasmuch as

$$H_{\mathbf{n},\mathbf{n}+\mathbf{a}}|\Psi_s\rangle=0, \quad H_{\mathbf{n},\mathbf{n}+\mathbf{b}}|\Psi_s\rangle=0, \tag{27}$$

for the total Hamiltonian (23) we have the expression

$$H|\Psi_s\rangle=0. \tag{28}$$

Consequently, Ψ_s is the ground-state wave function of the total Hamiltonian H, because it is a sum of nonnegative definite cell Hamiltonians. Also, it can be rigorously proved (see the Appendix) that the ground state of H is nondegenerate.

As mentioned above, the specific form of the projectors depends on 14 model parameters, and in general the cell Hamiltonians (26), expressed in terms of scalar products of the type $\mathbf{s}_i\cdot\mathbf{s}_j$, $(\mathbf{s}_i\cdot\mathbf{s}_j)(\mathbf{s}_k\cdot\mathbf{s}_l)$, etc., have an extremely cumbersome form. We therefore consider a few special cases.

When the node spinor $\Psi^{\lambda\mu\nu\rho}$ is a symmetric fourth-rank spinor $Q^{\lambda\mu\nu\rho}$ (corresponding to the two-dimensional AKLT model[12]), only the quintet component out of the six multiplets on each spin quartet is present in the wave function (19). The sixth-rank spinors (24) and (25) are symmetric with respect to two triplets of indices and, hence, contain four multiplets with $S=0,1,2,3$ formed from two quintets. Consequently, the cell Hamiltonian ($H_{1,2}$ and $H_{1,3}$ coincide in this case) has the form

$$H_{1,2}=\sum_{k=1}^{66}\lambda_kP_k^{1,2}. \tag{29}$$

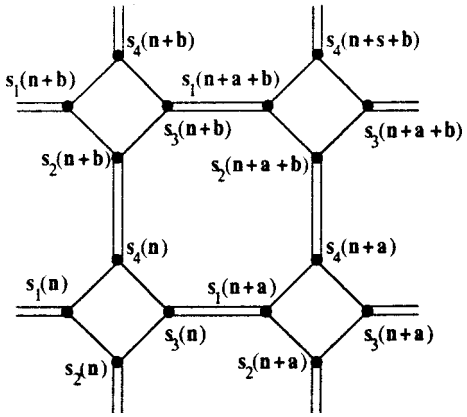


FIG. 4. Pattern of independent singlet pairs (double lines).

If we set $\lambda_k=1$ $(k=1,66)$, we can write Eq. (29) in the form

$$H_{1,2}=P_4(\mathbf{S}_1+\mathbf{S}_2)+[1-P_2(\mathbf{S}_1)P_2(\mathbf{S}_2)], \tag{30}$$

where $\mathbf{S}_i$ is the total spin of the quartet of spins $s=1/2$ on the ith node, $\mathbf{S}_i=\mathbf{s}_1(i)+\mathbf{s}_2(i)+\mathbf{s}_3(i)+\mathbf{s}_4(i)$, and $P_l(\mathbf{S})$ is the projector onto the state with spin $S=l$.

If the four spins $s=1/2$ at each node are replaced by a single spin $S=2$ and if the wave function (19) is treated as a wave function describing a system of M^2 spins $S=2$, the second term in the Hamiltonian (30) vanishes, and we arrive at the Hamiltonian of the two-dimensional AKLT model:

$$H_{1,2}=P_4(\mathbf{S}_1+\mathbf{S}_2)=\frac{1}{28}\mathbf{S}_1\cdot\mathbf{S}_2 + \frac{1}{40}(\mathbf{S}_1\cdot\mathbf{S}_2)^2+\frac{1}{180}(\mathbf{S}_1\cdot\mathbf{S}_2)^3+\frac{1}{2520}(\mathbf{S}_1\cdot\mathbf{S}_2)^4. \tag{31}$$

Another interesting special case is encountered when the system decomposes into independent one-dimensional chains. This happens if the node spinor $\Psi^{\lambda\mu\nu\rho}$ reduces to a product of two second-rank spinors, each describing two spins 1/2. For example,

$$\Psi^{\lambda\mu\nu\rho}(s_1,s_2,s_3,s_4)=\varphi^{\lambda\nu}(s_1,s_3)\varphi^{\mu\rho}(s_2,s_4). \tag{32}$$

In this case the Hamiltonians $H_{1,2}$ and $H_{1,3}$ contain interactions of four rather than eight spins 1/2 and have the form (16).

The simplest case is when the node spinor $\Psi^{\lambda\mu\nu\rho}$ is a product of four first-rank spinors:

$$\Psi^{\lambda\mu\nu\rho}(s_1,s_2,s_3,s_4)=\varphi^{\lambda}(s_1)\varphi^{\mu}(s_2)\varphi^{\nu}(s_3)\varphi^{\rho}(s_4). \tag{33}$$

Now the system decomposes into independent singlet pairs (Fig. 4), and the total Hamiltonian of the system has the form

$$H=\sum_{i,j}\left(\mathbf{s}_i\cdot\mathbf{s}_j+\frac{3}{4}\right), \tag{34}$$

where $\mathbf{s}_i$ and $\mathbf{s}_j$ are the spins forming the singlet pairs.

4. SPIN CORRELATION FUNCTIONS IN THE GROUND STATE

We now look at the problem of calculating the norm and the correlation function of the model described by the wave function (19). The expression for the norm of the wave function $G=\langle\Psi_s|\Psi_s\rangle$ has the form

$$G=\prod_{\mathbf{n}}\langle\Psi^{\lambda'_{\mathbf{n}}\mu'_{\mathbf{n}}\nu'_{\mathbf{n}}\rho'_{\mathbf{n}}}(\mathbf{n})|\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})\rangle \times g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}}g_{\rho_{\mathbf{n}}\mu_{\mathbf{n}+\mathbf{b}}}g_{\nu'_{\mathbf{n}}\lambda'_{\mathbf{n}+\mathbf{a}}}g_{\rho'_{\mathbf{n}}\mu'_{\mathbf{n}+\mathbf{b}}} =\prod_{\mathbf{n}}R^{\lambda'_{\mathbf{n}}\mu'_{\mathbf{n}}\lambda'_{\mathbf{n}+\mathbf{a}}\mu'_{\mathbf{n}+\mathbf{b}}}_{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{b}}}=\prod_{\mathbf{n}}R_{\alpha_{\mathbf{n}}\beta_{\mathbf{n}}\alpha_{\mathbf{n}+\mathbf{a}}\beta_{\mathbf{n}+\mathbf{b}}},$$

$$\alpha_i,\beta_i=\{1,2,3,4\}, \tag{35}$$

where $R_{\alpha_{\mathbf{n}}\beta_{\mathbf{n}}\alpha_{\mathbf{n}+\mathbf{a}}\beta_{\mathbf{n}+\mathbf{b}}}$ is a $4\times4\times4\times4$ matrix.

According to the selection rules for the projection of the total spin S^z, only 70 of the 256 elements in the expression $\langle\Psi^{\lambda'_{\mathbf{n}}\mu'_{\mathbf{n}}\nu'_{\mathbf{n}}\rho'_{\mathbf{n}}}(\mathbf{n})|\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})\rangle$ are nonvanishing. Consequently, the matrix R also contains at most 70 elements. If we regard the elements of R as Boltzmann vertex weights, the problem of calculating the norm reduces to the classical 70-vertex model.

Since the exact solution for the 70-vertex model is unknown, numerical methods must be used to calculate the norm and the expected values.

To calculate the above-indicated expected values, we carry out Monte Carlo calculations on 20×20-node lattices. As mentioned, the ground-state wave function of the model depends on 14 parameters and, of course, cannot possibly be analyzed completely. We confine the numerical calculations to the case in which the spinor $\Psi^{\lambda\mu\nu\rho}$ depends on one parameter α:

$$\Psi^{\lambda\mu\nu\rho}=\cos\alpha\cdot Q^{\lambda\mu\nu\rho}+\sin\alpha\cdot(A^{\lambda\mu\nu\rho}-Q^{\lambda\mu\nu\rho}), \tag{36}$$

where $\alpha\in[-\pi/2;\ \pi/2]$, the spinor $Q^{\lambda\mu\nu\rho}$ is symmetric with respect to all indices, and

$$A^{\lambda\mu\nu\rho}=\varphi^{\lambda}(s_1)\varphi^{\mu}(s_2)\varphi^{\nu}(s_3)\varphi^{\rho}(s_4). \tag{37}$$

In this case we have a one-parameter model with two well-known limiting cases. One corresponds to $\alpha=\pi/4$, for which $\Psi^{\lambda\mu\nu\rho}=A^{\lambda\mu\nu\rho}$, and the system decomposes into independent singlet pairs (Fig. 4); the other limiting case corresponds to $\alpha=0$ (our model reduces to the two-dimensional AKLT model in this case, the spins at each node forming a quintet).

In the given model there are four spins $s=1/2$ at each node, and the enumeration of each spin is determined by the order number of the lattice node to which it belongs and by its own number at this node. The spin correlation function therefore has the form

$$f_{ij}(\mathbf{r})=\langle\mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n}+\mathbf{r})\rangle. \tag{38}$$

In determining the spin structure of the ground state, however, it is more practical to consider the more straightforward quantity $F(\mathbf{r})$:

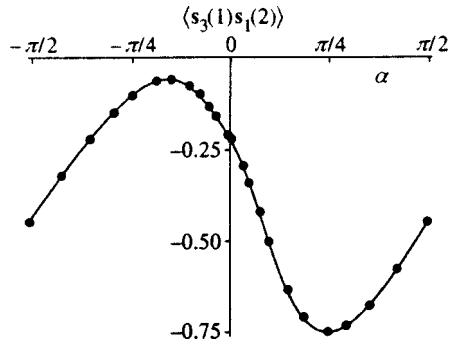


FIG. 5. Dependence of the spin correlation function $\langle \mathbf{s}_3(1)\mathbf{s}_1(2)\rangle$ on the parameter α.

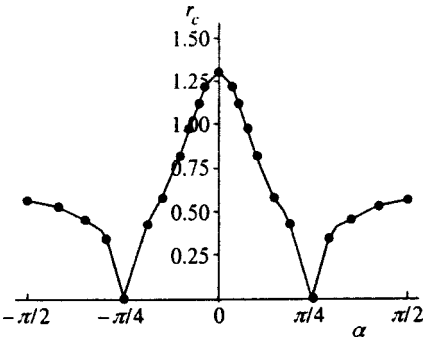


FIG. 7. Dependence of the correlation length on the parameter α.

$$F(\mathbf{r})=\sum_{i,j=1}^{4}\langle \mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n}+\mathbf{r})\rangle=\langle \mathbf{S}(\mathbf{n})\cdot\mathbf{S}(\mathbf{n}+\mathbf{r})\rangle. \quad (39)$$

The function $F(\mathbf{r})$ is left unchanged by a change of sign of α. This invariance is attributable to the fact that the spinor $(A^{\lambda\mu\nu\rho}-Q^{\lambda\mu\nu\rho}$ does not contain a quintet component, so that all the functions of this spinor are orthogonal to all functions of the symmetric spinor

$$\langle Q^{\lambda'\mu'\nu'\rho'}|(A^{\lambda\mu\nu\rho}-Q^{\lambda\mu\nu\rho})\rangle=0 \quad (40)$$

for all λ, μ, ν, ρ and λ', μ', ν', ρ'.

In addition, since the total spin operator $\mathbf{S}$ at a node commutes with $\mathbf{S}^2=\Sigma_{i,j=1}^{4}\mathbf{s}_i\cdot\mathbf{s}_j$, we then have

$$\langle Q^{\lambda'\mu'\nu'\rho'}|\sum_{i=1}^{4}\mathbf{s}_i|(A^{\lambda\mu\nu\rho}-Q^{\lambda\mu\nu\rho})\rangle=0. \quad (41)$$

It follows from Eqs. (35), (40), and (41) that $\sin\alpha$ and $\cos\alpha$ enter into the norm and into the expected value $\langle\Psi|\mathbf{S}(\mathbf{n})\cdot\mathbf{S}(\mathbf{n}+\mathbf{r})|\Psi\rangle$ only in even powers, so that $F(\mathbf{r})$ is invariant under a change of sign of α. We note, however, that only the total correlation function, and not $f_{ij}(\mathbf{r})$, possesses symmetry under a change of sign of α. This assertion is evident, for example, in Fig. 5, which shows the dependence of $f_{31}(\mathbf{a})$ on α as an illustration.

Figure 6 shows plots of $F(\mathbf{r})$ for certain values of the parameter α. In every case it is found that the correlation function decays exponentially as $\mathbf{r}$ increases, differing from the one-dimensional model in that the preexponential factor also depends on $\mathbf{r}$. Figure 7 shows the dependence of the correlation length r_c on the parameter α. The correlation length is a maximum at the point $\alpha=0$ (two-dimensional AKLT model), decreases as α increases, and at $\alpha=\pi/4$, when the system decomposes into independent singlet pairs (Fig. 4), it is equal to zero. With a further increase in α the correlation length increases and attains a second maximum at $\alpha=\pi/2$. Like the correlation function $F(\mathbf{r})$, the function $r_c(\alpha)$ is symmetric with respect to α. It is evident from Fig. 7 that the parameter α has two rangess corresponding to states with different symmetries. In the range $|\alpha|<\pi/4$ the correlation function $F(\mathbf{r})$ exhibits antiferromagnetic behavior:

$$F(\mathbf{r})\propto(-1)^{r_x+r_y}e^{-|\mathbf{r}|/r_c}, \quad (42)$$

whereas the spins at one node are coupled ferromagnetically, $\langle \mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n})\rangle>0$. On the other hand, in the range $\pi/4<|\alpha|<\pi/2$ the correlation function $F(\mathbf{r})$ is always negative:

$$F(\mathbf{r})\propto -e^{-|\mathbf{r}|/r_c} \quad (43)$$

and all the correlation functions at one node are also negative (Fig. 8).

These ranges have two end points in common, $\alpha=\pm\pi/4$, where $r_c=0$. Whereas $\alpha=\pi/4$ corresponds to the trivial partition of the system into independent singlet pairs, the case $\alpha=-\pi/4$ is more interesting.

In this case we have

$$\Psi^{\lambda\mu\nu\rho}=2Q^{\lambda\mu\nu\rho}-A^{\lambda\mu\nu\rho}, \quad (44)$$

and the matrix $\langle\Psi^{\lambda'\mu'\nu'\rho'}|\Psi^{\lambda\mu\nu\rho}\rangle$, which enters into the equation for the norm (35) and the expected values, is transformed into

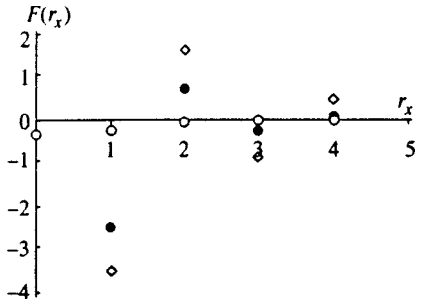


FIG. 6. Dependence of the spin correlation function $F(r_z)$ on the distance along the x axis for various values of the parameter α: ($\diamond$) $\alpha=0$; ($\bullet$) $\alpha=\pi/10$; ($\circ$) $\alpha=\pm\pi/2$.

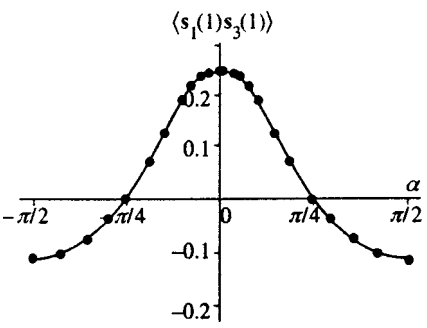


FIG. 8. Dependence of the spin correlation function at one node on the parameter α.

$$\langle \Psi^{\lambda'\mu'\nu'\rho'}|\Psi^{\lambda\mu\nu\rho}\rangle = 4\langle Q^{\lambda'\mu'\nu'\rho'}|Q^{\lambda\mu\nu\rho}\rangle - 2\langle A^{\lambda'\mu'\nu'\rho'}|Q^{\lambda\mu\nu\rho}\rangle - 2\langle Q^{\lambda'\mu'\nu'\rho'}|A^{\lambda\mu\nu\rho}\rangle + \langle A^{\lambda'\mu'\nu'\rho'}|A^{\lambda\mu\nu\rho}\rangle. \quad (45)$$

The symmetry of the spinor $Q^{\lambda\mu\nu\rho}$ with respect to all the indices leads to the relation

$$\langle Q^{\lambda'\mu'\nu'\rho'}|Q^{\lambda\mu\nu\rho}\rangle = \langle A^{\lambda'\mu'\nu'\rho'}|Q^{\lambda\mu\nu\rho}\rangle = \langle Q^{\lambda'\mu'\nu'\rho'}|A^{\lambda\mu\nu\rho}\rangle. \quad (46)$$

Equation (45) therefore acquires the form

$$\langle \Psi^{\lambda'\mu'\nu'\rho'}|\Psi^{\lambda\mu\nu\rho}\rangle = \langle A^{\lambda'\mu'\nu'\rho'}|A^{\lambda\mu\nu\rho}\rangle = \delta_{\lambda\lambda'}\delta_{\mu\mu'}\delta_{\nu\nu'}\delta_{\rho\rho'}. \quad (47)$$

From the equation for the norm (35) we then have

$$G = \prod_{\mathbf{n}} \langle \Psi^{\lambda'_{\mathbf{n}}\mu'_{\mathbf{n}}\nu'_{\mathbf{n}}\rho'_{\mathbf{n}}}(\mathbf{n})|\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})\rangle \times g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}} g_{\rho_{\mathbf{n}}\mu_{\mathbf{n}+\mathbf{b}}} g_{\nu'_{\mathbf{n}}\lambda'_{\mathbf{n}+\mathbf{a}}} g_{\rho'_{\mathbf{n}}\mu'_{\mathbf{n}+\mathbf{b}}} = \prod_{\mathbf{n}} \delta_{\lambda_{\mathbf{n}}\lambda'_{\mathbf{n}}}\delta_{\mu_{\mathbf{n}}\mu'_{\mathbf{n}}}\delta_{\lambda_{\mathbf{n}+\mathbf{a}}\lambda'_{\mathbf{n}+\mathbf{a}}}\delta_{\mu_{\mathbf{n}+\mathbf{b}}\mu'_{\mathbf{n}+\mathbf{b}}} = 2^{2M^2}. \quad (48)$$

The latter equation has been derived with allowance for the relation $\delta_{\nu\nu'}g_{\nu\lambda}g_{\nu'\lambda'} = \delta_{\lambda\lambda'}$.

We now calculate the expected value $\langle\Psi|\mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n}+\mathbf{r})|\Psi\rangle$. If nodes $\mathbf{n}$ and $\mathbf{n}+\mathbf{r}$ are not nearest neighbors, $\langle\Psi|\mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n}+\mathbf{r})|\Psi\rangle$ decomposes into the product of the expected values

$$\langle\Psi|\mathbf{s}_i(\mathbf{n})\cdot\mathbf{s}_j(\mathbf{n}+\mathbf{r})|\Psi\rangle = 2^{2M^2-8}\langle\Psi^{\lambda'\mu'\nu'\rho'}(\mathbf{n})|\mathbf{s}_i(\mathbf{n})|\Psi^{\lambda\mu\nu\rho}(\mathbf{n})\rangle \times\langle\Psi^{\lambda'''\mu'''\nu'''\rho'''}(\mathbf{n}+\mathbf{r})|\mathbf{s}_j(\mathbf{n}+\mathbf{r})|\Psi^{\lambda''\mu''\nu''\rho''}(\mathbf{n}+\mathbf{r})\rangle \times \delta_{\lambda\lambda'}\delta_{\mu\mu'}\delta_{\nu\nu'}\delta_{\rho\rho'}\delta_{\lambda''\lambda'''}\delta_{\mu''\mu'''}\delta_{\nu''\nu'''}\delta_{\rho''\rho'''} = 0. \quad (49)$$

Consequently, for $\alpha = -\pi/4$ all the correlation functions at non-nearest neighbor nodes are equal to zero. But if nodes $\mathbf{n}$ and $\mathbf{n}+\mathbf{r}$ are nearest neighbors, the corresponding correlation function assumes the form

$$\langle\Psi|\mathbf{s}_i(1)\cdot\mathbf{s}_j(2)|\Psi\rangle = 2^{2M^2-7}\langle\Psi^{\lambda'\mu'\nu'\rho'}(1)|\mathbf{s}_i(1)|\Psi^{\lambda\mu\nu\rho}(1)\rangle \times\langle\Psi^{\lambda'''\mu'''\nu'''\rho'''}(2)|\mathbf{s}_j(2)|\Psi^{\lambda''\mu''\nu''\rho''}(2)\rangle \times g_{\nu\lambda''}g_{\nu'\lambda'''}\delta_{\lambda\lambda'}\delta_{\mu\mu'}\delta_{\rho\rho'}\delta_{\mu''\mu'''}\delta_{\nu''\nu'''}\delta_{\rho''\rho'''}. \quad (50)$$

The exact calculation of the latter expression yields the following results (Fig. 3):

$$\langle\mathbf{s}_i(1)\cdot\mathbf{s}_j(2)\rangle = -\frac{25}{768}, \quad i=1,2,4, \quad j=2,3,4,$$

$$\langle\mathbf{s}_i(1)\cdot\mathbf{s}_1(2)\rangle = \langle\mathbf{s}_3(1)\cdot\mathbf{s}_j(2)\rangle = -\frac{15}{256}, \quad (51)$$

$$\langle\mathbf{s}_3(1)\cdot\mathbf{s}_1(2)\rangle = -\frac{27}{256}.$$

It follows from Eqs. (51) that

$$\sum_{i,j=1}^{4}\langle\mathbf{s}_i(1)\cdot\mathbf{s}_j(2)\rangle = -\frac{3}{4},$$

as in the case of independent singlets ($\alpha=\pi/4$). It can also be shown that all the correlations functions at one node are equal to zero.

In order to write the cell Hamiltonian $H_{1,2}$ in explicit form for $\alpha=-\pi/4$, we introduce the notation

$$\begin{cases}\mathbf{l}_1=\mathbf{s}_1(1)+\mathbf{s}_2(1)+\mathbf{s}_4(1),\\ \mathbf{l}_2=\mathbf{s}_2(2)+\mathbf{s}_3(2)+\mathbf{s}_4(2),\end{cases} \quad \begin{cases}\mathbf{s}_1=\mathbf{s}_3(1),\\ \mathbf{s}_2=\mathbf{s}_1(2),\end{cases} \quad \begin{cases}h_1=\mathbf{l}_1\cdot\mathbf{s}_1+\mathbf{l}_2\cdot\mathbf{s}_2,\\ h_2=\mathbf{l}_1\cdot\mathbf{s}_2+\mathbf{l}_2\cdot\mathbf{s}_1.\end{cases} \quad (52)$$

Accordingly, choosing $\lambda_k=1$ $(k=1,50)$, we can write the cell Hamiltonian $H_{1,2}$ in Eq. (26) in the form

$$H_{1,2} = P_{1/2}(\mathbf{l}_1)P_{1/2}(\mathbf{l}_2)P_1(\mathbf{s}_1+\mathbf{s}_2) + P_{3/2}(\mathbf{l}_1)P_{3/2}(\mathbf{l}_2)h_3 + P_{3/2}(\mathbf{l}_1)P_{1/2}(\mathbf{l}_2)h_4 + P_{1/2}(\mathbf{l}_1)P_{3/2}(\mathbf{l}_2)h_5, \quad (53)$$

where

$$h_3 = \frac{207}{256} + \frac{49}{64}\mathbf{s}_1\cdot\mathbf{s}_2 + \frac{3}{64}\mathbf{l}_1\cdot\mathbf{l}_2 + \frac{1}{16}(\mathbf{s}_1\cdot\mathbf{s}_2)(\mathbf{l}_1\cdot\mathbf{l}_2) - \frac{15}{64}h_2 - \frac{1}{32}h_2^2 + \frac{1}{64}[6h_1(\mathbf{l}_1\cdot\mathbf{l}_2) + 4h_1^2(\mathbf{l}_1\cdot\mathbf{l}_2) - 14h_1^2(\mathbf{l}_1\cdot\mathbf{l}_2)^2 + \text{h.c.}],$$

$$h_4 = \frac{3}{4} - \frac{7}{8}\mathbf{s}_1\cdot\mathbf{s}_2 + \frac{1}{4}\mathbf{l}_1\cdot\mathbf{s}_2 + \frac{1}{4}[(\mathbf{l}_1\cdot\mathbf{s}_1)(\mathbf{l}_1\cdot\mathbf{s}_2) + \text{h.c.}],$$

$$h_5 = \frac{3}{4} - \frac{7}{8}\mathbf{s}_1\cdot\mathbf{s}_2 + \frac{1}{4}\left[\mathbf{l}_2\cdot\mathbf{s}_1 + \frac{1}{4}(\mathbf{l}_2\cdot\mathbf{s}_2)(\mathbf{l}_2\cdot\mathbf{s}_1) + \text{h.c.}\right]. \quad (54)$$

The cell Hamiltonian $H_{1,3}$ has the same form (53) but with a change of notation according to Fig. 3:

$$\begin{cases}\mathbf{l}_1=\mathbf{s}_1(1)+\mathbf{s}_2(1)+\mathbf{s}_3(1),\\ \mathbf{l}_2=\mathbf{s}_1(3)+\mathbf{s}_3(3)+\mathbf{s}_4(3),\end{cases} \quad \begin{cases}\mathbf{s}_1=\mathbf{s}_4(1),\\ \mathbf{s}_2=\mathbf{s}_2(3).\end{cases} \quad (55)$$

Of special interest is the case corresponding to $\alpha=\pm\pi/2$. Unfortunately, exact expressions for the correlation function cannot be obtained in this case, but the Hamiltonian can be written in explicit form. Since the node spinor $\Psi^{\lambda\mu\nu\rho}$ does not contain a quintet component for $\alpha=\pm\pi/2$, the wave function of two nearest neighbor nodes (24) and (25) will lack a component with $S=3$. A more detailed analysis shows that 19 multiplets are present in the wave function of two nearest neighbor nodes. In this case, therefore, the cell Hamiltonian has the general form

$$H_{1,2} = \sum_{k=1}^{51}\lambda_k P_k^{1,2}. \quad (56)$$

For a definite choice of λ_k in Eq. (56) the cell Hamiltonian assumes the form

$$H_{12}=P_2(\mathbf{l}_1+\mathbf{s}_1)+P_2(\mathbf{l}_2+\mathbf{s}_2)+P_{1/2}(\mathbf{l}_1)P_{1/2}(\mathbf{l}_2)P_1 \times(\mathbf{s}_1+\mathbf{s}_2)+P_{3/2}(\mathbf{l}_1)P_{1/2}(\mathbf{l}_2)P_{1/2}(\mathbf{l}_1+\mathbf{s}_1+\mathbf{s}_2) +P_{3/2}(\mathbf{l}_2)P_{1/2}(\mathbf{l}_1)P_{1/2}(\mathbf{l}_2+\mathbf{s}_2+\mathbf{s}_1), \quad (57)$$

where the notations (52) and (55) are used for nearest neighbor nodes along the horizontal and along the vertical, respectively.

Our results suggest that the spin correlation functions decay exponentially with a correlation length ~ 1 for an arbitrary parameter α. We also assume that the decay of the correlation function is of the exponential type for the 14-parameter model as well, i.e., for any choice of node spinor $\Psi^{\lambda\mu\nu\rho}$. This assumption is supported in special cases: 1) the partition of the system into one-dimensional chains with exactly known exponentially decaying correlation functions; 2) the two-dimensional AKLT model, for which the exponential character of the decay of the correlation function has been rigorously proved.[14] Further evidence of the stated assumption lies in the numerical results obtained for various values of the parameter in the one-parameter model.

5. GENERALIZATION OF THE MODEL TO OTHER TYPES OF LATTICES

The wave function (7), (19) can be generalized to any type of lattice. The general principle of wave function construction for a system of spins 1/2 entails the following:

1) Each bond on a given lattice has associated with it two indices running through the values 1 and 2, one at each end of the bond.

2) Each bond has associated with it a metric spinor $g_{\lambda\mu}$ with the indices of the ends of this bond.

3) Each node of the lattice (a node being interpreted here, of course, in the same sense as in Sec. 3) with m outgoing bonds has associated with it an mth-rank spinor with the indices of the bonds adjacent to the node.

4) The wave function is the product of all spinors at nodes of the lattice and all metric spinors.

It is obvious that each index in the formulated wave function is encountered twice, so that the wave function is scalar and, hence, singlet.

The wave function so constructed describes a system in which each lattice node contains as many spins $s=1/2$ as the number of bonds emanating from it.

To completely define the wave function, it is necessary to determine the specific form of all node spinors. The coefficients that determine their form are then parameters of the model.

The Hamiltonian of such a model is the sum of the cell Hamiltonians acting in the spin space of the subsystem formed by the spins at two mutually coupled nodes:

$$H=\sum_{\langle ij\rangle} H_{ij}. \quad (58)$$

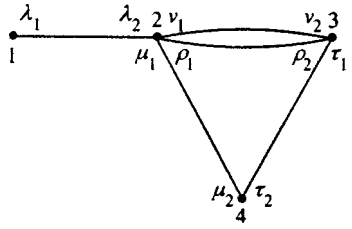


FIG. 9. Example of a graph corresponding to the wave function (60).

Each cell Hamiltonian is the sum of the projectors with arbitrary positive coefficients onto all multiplets possible in the corresponding two-node subsystem except those present in the constructed wave function:

$$H_{i,j}=\sum_k \lambda_k P_k^{i,j}. \quad (59)$$

Then we have $H_{i,j}|\Psi_s\rangle=0$ and, accordingly, $H|\Psi_s\rangle=0$.

Consequently, Ψ_s is an exact ground-state wave function.

We note that any two lattice nodes can be joined by two, three, or more bonds, because this does not contradict the principle according to which the wave function is constructed. Moreover, the general construction principle of the wave function is valid not only for translationally symmetric lattices, but for any graph in general. As an example, let us consider the system shown in Fig. 9. The wave function of this system has the form

$$\Psi_s=\Psi^{\lambda_1}(1)\Psi^{\lambda_2\mu_1\nu_1\rho_1}(2)\Psi^{\rho_2\nu_2\tau_1}(3)\Psi^{\mu_2\tau_2}(4) \times g_{\lambda_1\lambda_2}g_{\mu_1\mu_2}g_{\nu_1\nu_2}g_{\rho_1\rho_2}g_{\tau_1\tau_2} \quad (60)$$

and describes a system containing ten spins 1/2.

If the given lattice has dangling bonds (as occurs for systems with open boundary conditions), the resulting wave function represents a spinor of rank equal to the number of loose ends. The ground state of this kind of system is therefore 2^l-fold degenerate (where l is the number of loose ends). For an open one-dimensional chain, for example, the ground state corresponds to four functions—one singlet and three triplet components. For higher-dimensional lattices this degeneracy depends on the size of the lattice and increases exponentially as its boundaries grow.

6. CONCLUSION

We have proposed a method for the construction of an exact wave function for a class of two-dimensional spin models. In general this model depends on 14 parameters, and its Hamiltonian is written as the sum of the Hamiltonians of nearest neighbor spin quartets. The exact ground-state wave function of the total system is also the exact wave function of each cell Hamiltonian. Since 20 of the 70 multiplets of two nearest neighbor quartets are present in the exact wave function, the cell Hamiltonians are the sums of the projectors with positive coefficients onto the other 50 multiplets. These coefficients are the excitation energies of the corresponding multiplets. Different values of the coefficients correspond to different Hamiltonians. In this case, however, the ground-

state wave function itself and the spin correlation functions in the ground state are identical for all Hamiltonians. This means that the ground-state wave function, as defined by Shastry and Sutherland,[18] is superstable.

We have carried out Monte Carlo calculations of the spin correlation functions in the ground state for the special case of a model that depends on one parameter. For all values of the parameter the spin correlation functions decay exponentially with distance despite the complicated dependence of the correlation functions of nearest neighbor spins on the model parameter. It is justifiable to expect the spin correlations to decay exponentially in the general 14-parameter model as well.

In closing, we are pleased to acknowledge Prof. M. Ya. Ovchinnikov for helpful discussions of the problems treated in the article. This work has received financial support from the Russian Fund for Fundamental Research (Grants No. 96-03-32186 and No. 97-03-33727) and from the Program for Support of Leading Scientific Schools (Grant No. 96-15-97492).

APPENDIX

In Sec. 3 we have constructed the singlet wave function

$$\Psi_s=\prod_{\mathbf{n}} \Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n}) g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}} g_{\rho_{\mathbf{n}}\mu_{\mathbf{n}+\mathbf{b}}} \tag{A1}$$

for a system of $4M^2$ spins $s=1/2$ on a square lattice. The following Hamiltonian was specially chosen for the resulting wave function:

$$H=\sum_{\mathbf{n}} H_{\mathbf{n},\mathbf{n}+\mathbf{a}}+\sum_{\mathbf{n}} H_{\mathbf{n},\mathbf{n}+\mathbf{b}}, \tag{A2}$$

for which the wave function (A1) is the zero-energy ground state:

$$H|\Psi_s\rangle=0. \tag{A3}$$

We now show that the ground state of the system is nondegenerate, i.e., the wave function satisfying Eq. (A3) is unique.

Inasmuch as the Hamiltonian (A2) is a sum of nonnegative definite cell Hamiltonians, any function satisfying Eq. (A3) must satisfy all the cell equations

$$H_{i,j}|\Psi\rangle=0. \tag{A4}$$

This means that Eqs. (A3) and (A4) are equivalent.

We prove the nondegeneracy of the ground state of the Hamiltonian (A2) as follows. We first write the general form of the wave function for the system in question. We then determine the general form of the wave function satisfying one of the cell equations (A4). Making note of the conditions imposed on the general form of the wave function by each cell equation and, at the same time, simultaneously satisfying these conditions for all the cell equations, we obtain the general form of the wave function satisfying all the equations (A4) and, hence, satisfying the total Hamiltonian (A2).

Any wave function of the given system can be written in the form

$$\Psi=\sum_{\lambda\mu\nu\rho} c(\lambda\mu\nu\rho)\cdot\prod_{\mathbf{j}} \Phi^{\lambda_{\mathbf{j}}\mu_{\mathbf{j}}\nu_{\mathbf{j}}\rho_{\mathbf{j}}}(\mathbf{j}), \tag{A5}$$

where the summation is over the $4M^2$ indices λ_i, μ_i, ν_i, ρ_i, $c(\lambda\mu\nu\rho)$ denotes coefficients that depend on these indices, and $\Phi^{\lambda_j\mu_j\nu_j\rho_j}(\mathbf{j})$ are arbitrary fourth-rank node spinors (in general, spinors at different nodes can differ).

We require that the wave function (A5) obey the cell equation

$$H_{\mathbf{n},\mathbf{n}+\mathbf{a}}|\Psi\rangle=0. \tag{A6}$$

By the construction of the singlet wave function (A1), which is matched by the cell Hamiltonian $H_{\mathbf{n},\mathbf{n}+\mathbf{a}}$, any wave function at nodes $\mathbf{n}$ and $\mathbf{n}+\mathbf{a}$ that satisfies condition (A6) is a linear combination of the 64 functions contained in the expression

$$\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})\Psi^{\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{a}}\nu_{\mathbf{n}+\mathbf{a}}\rho_{\mathbf{n}+\mathbf{a}}}(\mathbf{n}+\mathbf{a}) g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}}, \tag{A7}$$

because the cell Hamiltonian $H_{\mathbf{n},\mathbf{n}+\mathbf{a}}$ by definition is the sum of the projectors onto all multiplets [$\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})$ and $\Psi^{\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{a}}\nu_{\mathbf{n}+\mathbf{a}}\rho_{\mathbf{n}+\mathbf{a}}}(\mathbf{n}+\mathbf{a})$ are definite node spinors occurring in the wave function (A1)]. We note that these 64 functions can be linearly dependent (as is the case, for example, for the two-dimensional AKLT model).

Thus, the general form of the wave function satisfying Eq. (A6) can be written

$$\Psi=\sum_{\lambda\mu\nu\rho} c(\lambda\mu\nu\rho|\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}) g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}} \Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n}) \times\Psi^{\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{a}}\nu_{\mathbf{n}+\mathbf{a}}\rho_{\mathbf{n}+\mathbf{a}}}(\mathbf{n}+\mathbf{a}) \prod_{\mathbf{j}\neq\mathbf{n},\mathbf{n}+\mathbf{a}} \Phi^{\lambda_{\mathbf{j}}\mu_{\mathbf{j}}\nu_{\mathbf{j}}\rho_{\mathbf{j}}}(\mathbf{j}), \tag{A8}$$

where $c(\lambda\mu\nu\rho|\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}})$ are coefficients that depend on the indices λ_i, μ_i, ν_i, ρ_i exclusive of the indices $\nu_{\mathbf{n}}$ and $\lambda_{\mathbf{n}+\mathbf{a}}$, and $\Phi^{\lambda_j\mu_j\nu_j\rho_j}(\mathbf{j})$ are arbitrary node spinors.

Comparing the functions (A5) and (A8), we deduce the following conditions that must be met by the function (A5) to obtain the general form of the wave function satisfying Eq. (A6):

1. The spinors at nodes $\mathbf{n}$ and $\mathbf{n}+\mathbf{a}$ must coincide with the node spinors of the wave function (A1):

$$\Phi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})=\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n}),$$
$$\Phi^{\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{a}}\nu_{\mathbf{n}+\mathbf{a}}\rho_{\mathbf{n}+\mathbf{a}}}(\mathbf{n}+\mathbf{a})=\Psi^{\lambda_{\mathbf{n}+\mathbf{a}}\mu_{\mathbf{n}+\mathbf{a}}\nu_{\mathbf{n}+\mathbf{a}}\rho_{\mathbf{n}+\mathbf{a}}}(\mathbf{n}+\mathbf{a}). \tag{A9}$$

2. The coefficients $c(\lambda\mu\nu\rho)$ have the form

$$c(\lambda\mu\nu\rho)=c(\lambda\mu\nu\rho|\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}) g_{\nu_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{a}}}. \tag{A10}$$

From the equation

$$H_{\mathbf{n},\mathbf{n}+\mathbf{b}}|\Psi\rangle=0 \tag{A11}$$

we deduce analogous conditions on the general form of the wave function (A5):

$$\Phi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n})=\Psi^{\lambda_{\mathbf{n}}\mu_{\mathbf{n}}\nu_{\mathbf{n}}\rho_{\mathbf{n}}}(\mathbf{n}),$$
$$\Phi^{\lambda_{\mathbf{n}+\mathbf{b}}\mu_{\mathbf{n}+\mathbf{b}}\nu_{\mathbf{n}+\mathbf{b}}\rho_{\mathbf{n}+\mathbf{b}}}(\mathbf{n}+\mathbf{b})=\Psi^{\lambda_{\mathbf{n}+\mathbf{b}}\mu_{\mathbf{n}+\mathbf{b}}\nu_{\mathbf{n}+\mathbf{b}}\rho_{\mathbf{n}+\mathbf{b}}}(\mathbf{n}+\mathbf{b}), \tag{A12}$$

$$c(\lambda\mu\nu\rho)=c(\lambda\mu\nu\rho|\rho_{\mathbf{n}}\mu_{\mathbf{n}+\mathbf{b}})g_{\rho_{\mathbf{n}}\lambda_{\mathbf{n}+\mathbf{b}}}.$$

The simultaneous satisfaction of all the cell equations (A4) requires consolidation of the conditions imposed by these equations on the general form of the wave function (A5). Combining these conditions in succession, in each step we obtain the general form of a wave function satisfying the equations corresponding to these conditions. Upon satisfying all the conditions, we obtain the general form of the wave function satisfying all the cell equations (A4) and, hence, satisfying Eq. (A3):

$$\Psi_s=\sum_{\lambda\mu\nu\rho} c(\lambda\mu\nu\rho|\lambda\mu\nu\rho) \times\prod_{\mathbf{j}} \Psi^{\lambda_{\mathbf{j}}\mu_{\mathbf{j}}\nu_{\mathbf{j}}\rho_{\mathbf{j}}}(\mathbf{j})g_{\nu_{\mathbf{j}}\lambda_{\mathbf{j}+\mathbf{a}}}g_{\rho_{\mathbf{j}}\mu_{\mathbf{j}+\mathbf{b}}}, \quad \text{(A13)}$$

where $c(\lambda\mu\nu\rho|\lambda\mu\nu\rho)=c$ is a constant.

Comparing the wave functions (A1) and (A13), we readily perceive that, to within an arbitrary factor, the general form of the wave function satisfying Eq. (A3) coincides with the wave function Ψ_s. Consequently, Ψ_s is the nondegenerate ground-state wave function of the Hamiltonian (A2).

*)E-mail: ovchin@glas.apc.org

[1] E. Dagotto, Int. J. Mod. Phys. B **5**, 907 (1991).
[2] P. Sen and B. K. Chakrabarti, Int. J. Mod. Phys. B **6**, 2439 (1992).
[3] T. Tonegawa and I. Harada, J. Phys. Soc. Jpn. **56**, 2153 (1987).
[4] A. V. Chubukov and T. Jolicoeur, Phys. Rev. B **44**, 12050 (1991).
[5] A. V. Chubukov, Phys. Rev. B **44**, 4693 (1991).
[6] D. J. J. Farnell and J. B. Parkinson, J. Phys.: Condens. Matter **6**, 5521 (1994); R. Bursill, G. A. Gehring, F. J. J. Farnell, J. B. Parkinson, Tao Xiang, and Chen Zeng, J. Phys.: Condens. Matter **7**, 8605 (1995).
[7] V. Ya. Krivnov and A. A. Ovchinnikov, Phys. Rev. B **53**, 6435 (1996).
[8] K. Nomura and K. Okamoto, Phys. Lett. A **169**, 433 (1992).
[9] C. E. I. Carneiro, M. J. de Oliveira, and W. F. Wreszinski, J. Stat. Phys. **79**, 347 (1995).
[10] S. R. White and I. Affleck, Phys. Rev. B **54**, 9862 (1996).
[11] C. K. Majumdar and D. K. Ghosh, J. Math. Phys. **10**, 1388, 1399 (1969).
[12] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987); Commun. Math. Phys. **115**, 477 (1988).
[13] F. D. M. Haldane, Phys. Lett. A **93**, 464 (1983); Phys. Rev. Lett. **50**, 1153 (1983).
[14] T. Kennedy, E. H. Lieb, and H. Tasaki, J. Stat. Phys. **53**, 383 (1988).
[15] A. Klumper, A. Schadschneider, and J. Zittartz, Z. Phys. B **87**, 281 (1992); Europhys. Lett. **24**(4), 293 (1993); C. Lange, A. Klumper, and J. Zittartz, Z. Phys. B **96**, 267 (1994).
[16] D. V. Dmitriev, V. Ya. Krivnov, and A. A. Ovchinnikov, Phys. Rev. B **56**, 5985 (1997).
[17] H. Niggemann, A. Klumper, and J. Zittartz, Z. Phys. B **104**, 103 (1997).
[18] B. Sutherland and S. Shastry, J. Stat. Phys. **33**, 477 (1983).

Translated by James S. Wood

Resistive transition and upper critical field in underdoped $YBa_2Cu_3O_{6+x}$ single crystals

V. F. Gantmakher*) and G. E. Tsydynzhapov

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

L. P. Kozeeva and A. N. Lavrov

Institute of Inorganic Chemistry, Siberian Department of Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 16 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 268–284 (January 1999)

A superconducting transition in the temperature dependence of the ab-plane resistivity of underdoped $YBa_2Cu_3O_{6+x}$ crystals in the range $T_c \lesssim 30$ K has been investigated. Unlike the case of samples with the optimal level of doping, the transition width increased insignificantly with magnetic field, and in the range $T \lesssim 13$ K it decreased with increasing magnetic field. The transition point $T_c(B)$ was determined by analyzing the fluctuation conductivity. The curves of $B_{c2}(T)$ measured in the region $T/T_c \gtrsim 0.1$ did not show a tendency to saturation and had a positive second derivative everywhere, including the immediate neighborhood of T_c. The only difference among the curves of $B_{c2}(T)$ for different crystal states is the scales of T and B, so they can be described in terms of a universal function, which fairly closely follows Alexandrov's model of boson superconductivity. © *1999 American Institute of Physics.* [S1063-7761(99)02301-X]

1. INTRODUCTION

The nature of high-temperature superconductivity is presently one of the most interesting subjects of the solid state physics. An important topic of research in this field is the temperature dependence of the upper critical field B_{c2}. In conventional (low-temperature) superconductors, in accordance with the BCS theory, the curve $B_{c2}(T)$ is described by a universal function $b_{\mathrm{BCS}}(t)$ in terms of reduced variables: the temperature is scaled by the zero-field transition temperature, $t = T/T_c$, and the magnetic field is scaled by the product of T_c and the derivative of $B_{c2}(T)$ at T_c: $b = B/[T_c(-dB_{c2}/dT)_{T=T_c}]$.[1] The function $b_{\mathrm{BCS}}(t)$ is linear in the neighborhood of T_c and saturates to $b \simeq 0.7$ at $t = 0$. In high-temperature superconductors (HTSC) the behavior of $B_{c2}(T)$ is radically different. In $Tl_2Ba_2CuO_6$[2] and $Bi_2Sr_2CuO_y$[3] films, and in $K_{0.4}Ba_{0.6}BiO_3$ single crystals,[4,5] a positive second derivative and a sharp increase in $B_{c2}(T)$ at low temperature have been detected. Similar properties of function $B_{c2}(T)$ have been observed in other HTSC systems, namely, in $YBa_2(Cu_{1-y}Zn_y)_3O_{6+x}$ with a critical temperature lowered by the strong scattering[6] and $Sm_{1.85}Ce_{0.15}CuO_{4-y}$ with n-type conductivity.[7]

HTSC is not the only class of materials where the upper critical field does not follow the BCS universal function $b_{\mathrm{BCS}}(t)$. But, as concerns HTSC, such deviations are probably present in all materials of the family, and magnitudes of these deviations are enormous.[2,3] Therefore, there is every reason to seek fundamental causes of these deviations, which are general for all HTSC.

Several models have been suggested. Ovchinnikov and Kresin[8] focused attention on magnetic impurities, which, as they assumed, cause pair breaking and effectively suppress superconductivity near T_c. The tendency to magnetic ordering at lower temperatures results in a lower spin-flip scattering amplitude, thus enhancing superconductivity. The presence of magnetic impurities is a common feature of HTSC, since current carriers in most of them are due to doping, which generates magnetic defects at the same time.

Spivak and Zhou[9] studied the role of Landau quantization combined with a random potential. The quantization leads to a higher density of states on Landau levels, whereas the random potential brings to the Fermi level Landau sublevels with opposite spins at points close to one another in space. In this case, the random potential must satisfy two opposite conditions: its variation over the coherence length ξ should be larger than the Zeeman splitting, on the other hand, scattering by this potential should not wipe away peaks in the density of states. The HTSC structure favors both these conditions: fluctuations in the concentration of dopants, which are at the same time scattering centers, should occur even in high-quality crystals, but these scatterers and current carriers are separated in space.

It is possible that there are more fundamental causes of the peculiar shape of $B_{c2}(T)$ curves that can be put down to an exotic nature of superconductivity in HTSC. One example is the ''bipolaron'' or, in a more general approach, the ''boson'' model of superconductivity suggested by Alexandrov and Mott.[10] The model assumes that pairs (charged bosons, e.g., bipolarons) are preformed, and the superconducting transition consists in Bose-condensation of these pairs. In the presence of a random potential, the curve of $B_{c2}(T)$ has a positive curvature. The conventional superconductivity in a Fermi liquid can transform to the boson superconductivity if

the electron–phonon coupling is strong and the carrier density is low. Again, HTSC materials are good candidates for realization of such a scenario. Their carrier concentration is lower than in conventional metals and drops further with decreasing doping level, whereas the coupling constant $\lambda \gtrsim 1$.

Abrikosov suggested for HTSC a model whose central component is a saddle-like singularity in the electron spectrum. This model predicts, in particular, a positive curvature of the $B_{c2}(T)$ curve[11] because the problem becomes effectively one-dimensional due to the saddle point; as a result, the magnetic field's capability of destroying superconductivity is limited considerably. In the absence of the paramagnetic limit, the model yields the divergent function $B_{c2}(T)$, but if the paramagnetic limit is taken into account, the critical field is limited to a finite value.

The experimental data accumulated over recent years are insufficient for making an ultimate choice of one of these model. Further research is needed, and the present paper is a step in this direction. We present an investigation of the effect of a magnetic field on the resistivity of $YBa_2Cu_3O_{6+x}$ single crystals at doping levels below the optimal one. The aim of this work was to measure the temperature dependence of B_{c2} in this material at x such that $T_c < 30\,\mathrm{K}$ and derive from these data changes in parameters that control B_{c2} when $T_c \to 0$.

The paper is organized as follows. Section 2 presents basic theoretical concepts concerning the superconductor phase diagram in a magnetic field and the behavior of conductivity around the superconducting transition point; they are essential in the analysis of experimental data. Section 3 describes sample fabrication techniques and experimental procedures, Sec. 4 reports on experimental results. The curves $\rho(T)$ and their evolution induced by the magnetic field are discussed in Sec. 4.1. The derivation of $B_{c2}(T)$ from resistance-versus-temperature data for HTSC has remained a controversial issue,[12,13] therefore this topic is given special treatment in Sec. 4.2. Since the transition broadening induced by magnetic field is insignificant, qualitative conclusions concerning the behavior of $B_{c2}(T)$ are not affected by the specific routine employed in determination of the superconducting transition point. Nonetheless, in determining $B_{c2}(T)$ quantitatively, we analyzed the fluctuation conductivity in the normal state as a function of temperature. Section 4.3. discusses $B_{c2}(T)$ derived from experimental data: the curvature of $B_{c2}(T)$ curves proved to be positive throughout the available temperature range, including the close neighborhood of T_c; no signs of saturation in the low-temperature range have been detected; the experimental data are compared with existing models.

2. BASIC THEORETICAL CONCEPTS

2.1. Phase diagram

The phase diagram of a type-II superconductor in the $B-T$ plane in the mean-field approximation contains a Meissner region, where magnetic field is fully ejected from a sample, a mixed state region, where a lattice of Abrikosov's flux lines exists, and a normal metal region. These regions are separated by lines of second-order phase transitions: $B_{c1}(T)$ between the Meissner and mixed phases and $B_{c2}(T)$ between the mixed state and normal metal.

Beyond the mean-field approximation, thermal fluctuations of the order parameter slightly change the phase diagram configuration. Now, it contains a region of ''vortex liquid,'' where fluctuations change largely the order parameter phase (which can be interpreted in terms of free motion of Abrikosov's flux lines), and a region of critical fluctuations close to $B_{c2}(T)$, where the order parameter amplitude fluctuates and its mean value changes rapidly with the temperature or magnetic field intensity. There are superconducting fluctuations above $B_{c2}(T)$ also, but their amplitude is small and decreases away from the line of $B_{c2}(T)$. The phase transition to the superconducting state with a long-range order established occurs on the boundary between the vortex liquid and vortex lattice [melting line $B_m(T)$], whereas the curve of $B_{c2}(T)$ determined in the mean-field approximation defines the line of a crossover from the normal metal, where the order parameter fluctuation amplitude is low, to the vortex liquid, where the magnitude of the order parameter is almost unity.[14–16]

In conventional superconductors, the regions of critical fluctuations and vortex liquid are quite narrow and essentially unobservable. The melting line $B_m(T)$ coincides with $B_{c2}(T)$, therefore, the mean-field approximation adequately describes the phase diagram. In HTSC the situation is different. Owning to the high critical temperature, small coherence length, and high anisotropy, fluctuations play a more important part, and the vortex liquid phase occupies a considerable region of the phase diagram, so B_m and B_{c2} are separated. Since fluctuations broaden features of field dependencies of transport and thermodynamic properties at point B_{c2}, it is most difficult to determine this point in experiment. Nonetheless, the value of $B_{c2}(T)$ is still very important since this is the parameter that controls the behavior of thermodynamic quantities in the region far from the line of transition, where the mean-field approximation is valid.

In materials with strong pinning, the phase diagram is further modified: the pinning destroys the order in the vortex lattice and transforms it to a vortex glass. The melting line is replaced by the ''irreversibility line'' $B^*(T)$, above which vortices are depinned by thermal fluctuations and move freely even at very low current densities, which results in a finite resistivity and reversible dc magnetization. Below $B^*(T)$ vortices are pinned in the low-current limit, and the magnetization curve shows a hysteresis.

2.2. Resistive transition

In high-temperature superconductors with optimal doping, curves of $\rho(T)|_B$ form a fan with a common transition onset point, so the positions of the transition onset are almost independent of the magnetic field.[14,17] The drop in the resistivity around the transition onset is controlled by the contribution of superconductive fluctuations to the conductivity. The characteristic field of fluctuation suppression is B_{c2}, hence the shift of the transition onset should follow the function $B_{c2}(T)$. On the low-temperature side, the resistivity

should vanish when the vortex motion is frozen. Qualitatively, the line on the $B-T$ phase diagram where the vortex mobility becomes significant is the ''irreversibility line'' $B^*(T)$. Thus, the resistive transition is confined by the lines $B_{c2}(T)$ and $B^*(T)$ and is associated with the vortex liquid region on the phase diagram so that the fan-like appearance of resistivity curves is due to broadening of this region with the magnetic field while the line of $B_{c2}(T)$ is almost vertical.

The breadth of the vortex liquid region, hence the transition width, is determined by the relation between pinning and fluctuations. The vortex depinning is favored by the small coherence length ξ, high temperatures, and weak coupling between neighboring superconducting layers of CuO_2, i.e., by the high anisotropy. Variations in the doping level (carrier density n) to both sides from the optimal doping $n_{\rm opt}$ lead to lower T_c and larger ξ. On the other hand, the anisotropy is stronger at lower doping and weaker at higher doping levels. The resistivity curves of overdoped HTSC samples with high carrier densities and low anisotropy are similar to those of conventional superconductors with strong pinning.[2,18]

The difference between over- and underdoped states was demonstrated by comparing $La_{2-x}Sr_xCuO_4$ samples with different x.[18] Whereas a magnetic field of $B=8$ T broadened by 15–20 K the resistive transition in an underdoped sample with $x=0.08$ and $T_c\approx 30$ K, the transition curve in an overdoped sample with $x=0.20$ and approximately the same T_c was shifted by magnetic field without changing its shape.[18] This observation was confirmed by other researchers,[19,20] who also reported that decreasing the oxygen content in $YBa_2Cu_3O_{6+x}$ thin films and single crystals considerably enhances effects originated from vortex motion, in particular, increases transition broadening in the magnetic field. All these experiments, however, used samples with $T_c \gtrsim 40$ K, and it remained unclear whether this tendency should persist in the range of low T_c.

There is an alternative interpretation of the resistive transition in cuprates, which attributes most of the change in the resistivity to a phase transition between the vortex liquid and vortex lattice (vortex glass) at $B_m(T)$.[14,21,22] In this case, the resistive transition is decomposed into a resistivity jump on the $B_m(T)$ line [well below $B_{c2}(T)$] and a crossover on line $B_{c2}(T)$,[21] which can produce only slight changes in resistivity.

The high conductivity in the normal state of overdoped cuprates might in fact mask the transition from the normal to vortex liquid state.[2] But changes in transport characteristics around B_{c2} are evident even in high-quality $YBa_2Cu_3O_{6+x}$ crystals with optimal doping and very weak pinning.[23] They should be the much more notable in underdoped samples, whose conductivity in the normal state is essentially lower.

3. EXPERIMENTAL

$YBa_2Cu_3O_{6+x}$ single crystals were grown by slow cooling the melt containing 10.0 to 11.4 wt.% of $YBa_2Cu_3O_{6+x}$ and eutectic mixture of 0.28 BaO and 0.72 CuO as a solvent with subsequent decanting of the residual flux. For our experiments, we selected single crystals without visible signs of block structure and shaped as plates 20 to 40 μm thick with areas of several square millimeters. After oxygenating at 500 °C, they had $T_c \approx 90$–92 K and fairly narrow resistive superconducting transitions with $\Delta T < 1$ K.

In $YBa_2C_3O_{6+x}$, current carriers (holes) are generated in CuO_2 planes as a result of capturing electrons in layers of CuO_x chains. The hole density depends on the oxygen content x and configuration of oxygen atoms in chains in CuO_x layers. Consequently, the carrier density in $YBa_2Cu_3O_{6+x}$ (along with the superconducting transition temperature) can be varied by two methods: changing the oxygen content and varying its ordering in CuO_x layers.

The technique for changing the oxygen content is the high-temperature annealing, and it allows one to produce the whole range of states from antiferromagnetic insulator to optimally doped superconductor. The annealing temperature at a given partial pressure of oxygen controls the oxygen content in a crystal and is a convenient technological parameter in processing superconducting samples.[24] In order to reduce the oxygen content to $x=0.37$–0.47, we annealed crystals in air at 700–800 °C and then quenched them in liquid nitrogen to prevent exchange of oxygen with the atmosphere during cooling.

The second technique allows us to vary the carrier density over a relatively narrow interval by changing the average length of oxygen chains at constant x.[25,26] In chains of finite lengths, there are q oxygen atoms per $q+1$ copper atoms, hence, one has $(q+1)/q$ electrons per oxygen atom. For this reason, oxygen atoms in shorter chains are less efficient in capturing electrons from CuO_2 planes. The average chain length can change owing to the high diffusion mobility of oxygen in CuO_x layers at the room temperature and above. Longer chains have lower energy, but they contribute less to the entropy, which makes them less preferable at high temperatures. The balance between these two factors determines the average chain length in equilibrium (hence, the number of holes) as a function of temperature. The relaxation time strongly depends on temperature, so rapid cooling freezes the oxygen configuration, thus fixing the carrier density. In real experiments, we heated crystals to 120–140 °C and then quenched them in liquid nitrogen. This procedure notably reduced the number of holes in the sample, hence lowered T_c. After that samples could be stored in liquid nitrogen for indefinitely long times without any changes whatsoever. If a sample was exposed to the room temperature, the carrier concentration increased gradually owing to oxygen coagulation in longer chains. This aging process could be monitored continuously by measuring the sample resistance at a constant temperature and interrupted at any moment by cooling the sample, thus we could obtain any intermediate value of T_c. The aging of a sample at the room temperature for several days returns it to its initial equilibrium state. Since all restructuring processes in the oxygen subsystem proceed at relatively low temperatures, this method allows one to obtain a sequence of sample states with minimal differences in configurations of defects and pinning centers.

All in all, we have studied three crystals at several carrier densities in each. The sample parameters are listed in Table I. The different crystals are numbered 1 to 3, their

TABLE I.

Samples					
Sample	ρ_{room}/ρ_{50} K	x	Degree of quenching	T_c, K	B_{sc}, T
			quenched	16.5	3.0
1a	3	0.43	intermediate	20.5	3.8
			aged	25.5	8.9
2a	8	0.41	aged	19	2.8
2b	10	0.47	quenched	38.5	120
			aged	44.5	240
3a	3	$\approx$0.37	quenched	0	–
			aged	6.3	0.61
3b	3	$\approx$0.37	aged	$\approx$3	–

states with different oxygen contents are labeled a and b, and the quenching states are referred to as itquenched, itintermediate, and itaged. The ratio between resistances at the room temperature and 50 K, when the free path is largely controlled by defect scattering, is a characteristic of crystal purity. This parameter of sample 2 is a factor of about three higher than in samples 1 and 3. Parameter B_{sc} will be discussed in Sec. 4.3.

We measured the resistance in the ab plane using a four-terminal circuit. Since $YBa_2Cu_3O_{6+x}$ crystals with low oxygen contents are highly anisotropic, it is very important that the current be uniformly distributed over the sample thickness, so that only one component of the resistivity tensor is measured. To this end, the current contacts were fabricated over the entire surfaces of two opposite crystal faces. The contacts were made by a silver paste and fixed by annealing before all thermal manipulations designed to vary the hole concentration. The resistance was measured by the standard technique using a nanovolt-range lock-in amplifier at 23 Hz. The probe current was weak enough to ensure the linear regime and avoid overheating even at the lowest temperatures. The uncertainty in the geometrical factor restricted the accuracy of absolute measurements of conductivity to 10–20%, nonetheless, note that the geometrical factor of each sample was the same in all conducting states.

Most of experiments were performed in a cryostat with a ^{3}He pumping system, which allowed us to vary the temperature between 0.3–300 K.[27] At temperatures of 0.3–1.2 K the sample was immersed in liquid ^{3}He, at higher temperatures it was in the ^{3}He atmosphere at a pressure of several torr serving as a heat-exchange gas. The temperature was measured by a carbon resistance thermometer calibrated by a reference platinum thermometer, ^{4}He vapor pressure, and cerium–magnesium nitrate in appropriate temperature ranges. The magnetic field of up to 8.25 T was applied along the c-axis.

Sample 3b in the aged state with low T_c was tested in a dilution refrigerator at temperatures down to 30 mK and magnetic fields of up to 14 T.

4. RESULTS

4.1. Temperature dependence of resistivity

In our experiments on samples with $T_c \geqslant 30$–35 K (samples 2b quenched and aged), we record fans of $\rho(T)|_B$ curves similar to those reported by other authors.[19,20] In samples with lower T_c, the effect of magnetic field on the resistive transition is radically different, and in this publication we concentrate on these effects, namely, the behavior of $YBa_2Cu_3O_{6+x}$ samples in states with $T_c \leqslant 30$ K (samples 1a, 2a, 3a, and 3b in all quenching states). In these samples, magnetic field shifts the transition without a notable broadening (Fig. 1), which indicates that the effect of vortex motion on the shape of transition curve no longer dominates. Nonetheless, the shape of the transition curve is affected by the magnetic field, and one can see on curves of temperature derivatives $\partial\rho/\partial T$ plotted on the right of Fig. 1 that these changes are nonmonotonic. Since the normal state resistivity is almost constant with temperature, the peak amplitude on the derivative curve is inversely proportional to the resistive transition width. These graphs clearly show that, irrespective of T_c ($\leqslant 30$ K), the transition width is maximum at about 13–14 K. If the zero-field T_c is higher, the transition first shifts to lower temperatures with magnetic field and broadens (Fig. 1a). Then, below 13–14 K, the transition narrows concurrently with its shift to lower temperatures. If T_c is initially lower than 13–14 K (Fig. 1b and 1c), the transition is narrowed by magnetic field concurrently with its shift to lower temperatures from the start, and the slope of the transition curve in magnetic field becomes steeper than at zero field.

The comparison between samples 1a and 2a demonstrates that the nonmonotonic change in the transition width with magnetic field is a reproducible property and is little affected by the sample quality. The superconducting transition temperatures of these two samples were driven to one value by annealing (Fig. 1b and 1c), but their parameters in the normal state were notably different. Sample 2a contained less impurities and structural defects, as a result, its resistivity around T_c was twice as small (Fig. 1), it dropped more rapidly in the process of cooling from the room temperature to 50 K (Table I) and showed a smaller increase in the range of lower temperatures. Nonetheless, irrespective of all these differences, both the transition shift rate in magnetic field and the evolution of transition curves of these samples are similar. Narrowing of the resistive transition in an underdoped $YBa_2Cu_3O_{6+x}$ with increasing magnetic field in this temperature range was detected by Seidler *et al.*,[25] but, since their measurements were presented in a different form, it is difficult to compare them directly with our results.

Such a behavior of transition curves is observed for all samples with $T_c \geqslant 6$ K. In states with lower transition temperatures, we were not able to achieve sufficiently narrow transitions at zero magnetic field to measure T_c and transition width. Therefore, the measurement data for sample 3b will be given and discussed separately in Sec. 4.3.

4.2. Derivation of $B_{c2}(T)$ from resistance measurements

The absence of the notable transition broadening in magnetic field in $YBa_2Cu_3O_{6+x}$ samples with low T_c indicates that, unlike samples with $T_c \geqslant 30$–35 K, they have a narrower region of the ''vortex liquid'' on the phase diagram. The transition width, however, is not so small that it could be neglected in determining $B_{c2}(T)$. Since the point B_{c2} is not

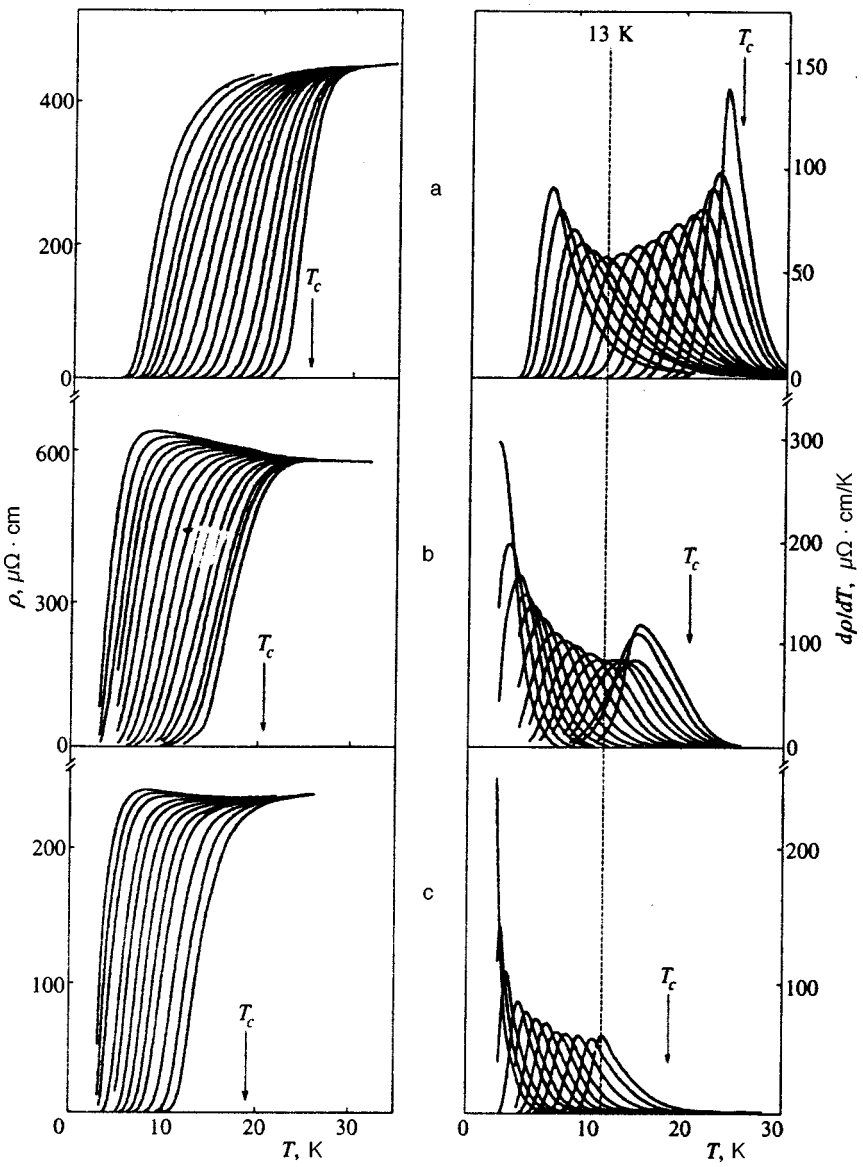


FIG. 1. Conductivity ρ_{ab} (on the left) and its derivative (on the right) at various magnetic fields aligned with the *c*-axis. (a) Sample 1a in the aged state; applied fields (from right to left): 0, 0.06, 0.12, 0.23, 0.35, 0.6, 0.8, 1.2, 1.6, 2.2, 3.0, 3.8, 4.6, 5.5, 6.7, and 8.2 T. (b) Sample 1a in the intermediate state; applied fields: 0, 0.06, 0.12, 0.23, 0.35, 0.6, 0.8, 1.2, 1.6, 2.2, 3.0, 3.8, 4.6, 5.5, 6.7, and 8.2 T. (c) Sample 2a in the aged state; applied fields: 0.12, 0.23, 0.5, 0.8, 1.2, 1.6, 2.2, 3.0, 3.8, 5.5, 6.7, and 8.2 T.

marked by a sharp feature on curves of $\rho(T)$, there is a good reason to determine this point by fitting a theoretical curve describing the crossover between the normal metal and vortex liquid to the experimental data. In developing this approach, let us consider the sample conductivity as a sum of the normal and fluctuation components: $\sigma(T)=\sigma_n(T)+\sigma_{fl}(T)$.

The fluctuation conductivity σ_{fl} in quasi-two-dimensional systems in zero field is usually described by the Lawrence–Doniach formula:

$$\sigma_{fl}=\frac{1}{16}\frac{e^2}{\hbar d\epsilon}\left[1+\left(\frac{2\xi_c(0)}{d}\right)^2\frac{1}{\epsilon}\right]^{-1/2}, \quad \epsilon\equiv\ln\frac{T}{T_c}, \qquad (1)$$

where d is the interplane separation. Friedman *et al.*[28] show that, even in analyzing optimally doped $YBa_2Cu_3O_{6+x}$ crystals with the resistivity anisotropy no higher than 30–100, one can neglect the factor in brackets which takes into account effects of the third dimension and use Aslamazov–Larkin's expression for two dimensions:

$$\sigma_{fl}=\frac{1}{16}\frac{e^2}{\hbar d}\epsilon^{-1}. \qquad (2)$$

In oxygen deficient crystals, the anisotropy is up to $(5-10)\times10^3$,[24] therefore Eq. (2) is *a fortiori* valid throughout the temperature range in question, except the neighborhood of T_c.

There is no consistent theoretical description of $\sigma_{fl}(T,B)$ in nonzero magnetic field for arbitrary $B_{c2}(T)$. Ullah and Dorsey[16] analyzed σ_{fl} in a system with strong fluctuations in magnetic field and suggested a scaling expression for the fluctuation conductivity, which is often used in describing the resistive transition and determining $B_{c2}(T)$ of cuprate superconductors.[29–31] Since their approach is based on the mean-field approximation and assumes a linear dependence $B_{c2}(T)$ near T_c, it does not apply when $B_{c2}(T)$ is strongly nonlinear. (It will be shown below that this is the case in our samples.) Nonetheless, in the region well above $T_c(B)$ ($\epsilon_B\gtrsim0.1$), where Gaussian fluctuations dominate, a

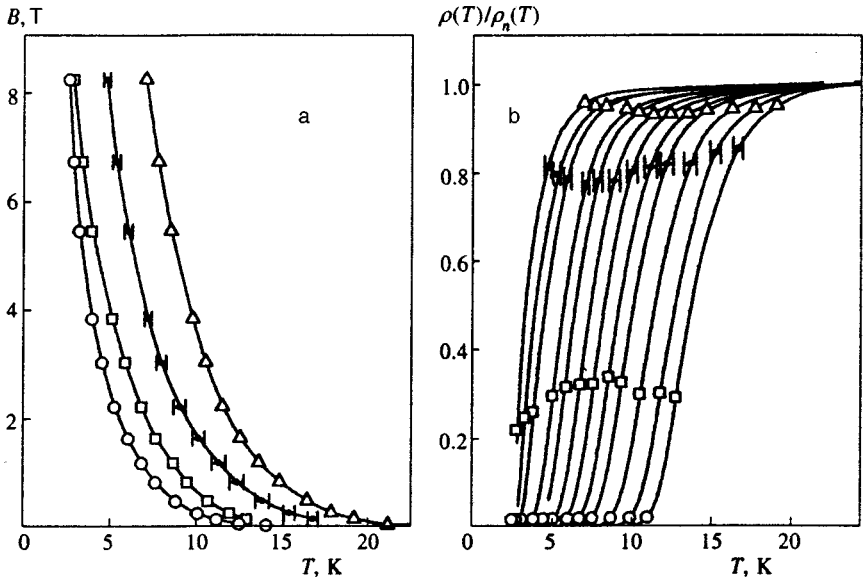


FIG. 2. Characteristic points of the superconducting transition in sample 2a plotted in (a) $B-T$ and (b) $\rho-T$ planes (the resistivity is normalized to function $\rho_n(T)$ in Eq. (4), which was used in determination of the fluctuation conductivity): (open circles) ''irreversibility line,'' $\rho=5\,\mu\Omega\cdot$cm; (squares) peak of derivative $\partial\rho/\partial T$; (full circles) $B_{c2}(T)$; (triangles) ''transition onset,'' $\partial\sigma_{fl}/\partial T=10^2\ (\Omega\cdot\text{cm})^{-1}/\text{K}$.

formula similar to that suggested by Aslamazov and Larkin can be used:

$$\sigma_{fl}\propto\frac{1}{\epsilon_B},\quad \text{where}\ \ \epsilon_B=\ln\frac{T}{T_c(B)}, \tag{3}$$

in both zero and finite magnetic fields (see Ref. 16 and references therein). Here $T_c(B)$ is the functional inverse of $B_{c2}(T)$. This formula also assumes, generally speaking, a linear dependence $B_{c2}(T)$, but a possible change in the exponent of this function should lead only to a small systematic shift of the resulting curve $T_c(B)$.

In contrast to the case of optimal doping, the normal conductivity in our samples is low, of order of $e^2/\hbar d$ (Fig. 1), if d is assumed to be of order of the lattice constant, 11.7 Å. Simple estimates based on the Aslamazov–Larkin formula (2) with a reasonable value of d indicate that the contribution of fluctuations, σ_{fl}, should be several percent of σ_n even at $\epsilon_B\gtrsim 0.5$. This makes determination of $\sigma_n(T)$ more difficult. The difficulties are exacerbated by the fact that the normal state resistivity has a minimum in the region of 30–40 K and increases at lower temperatures. Therefore, we decide to select *a priori* the function $\sigma_n(T)$ with several fitting parameters. The fitting to experimental data is performed by varying all parameters in both $\sigma_{fl}(T)$ and $\sigma_n(T)$.[29] This procedure could hardly produce sensible results if each curve $\rho(T)$ were described by a different set of parameters. Fortunately, the magnetoresistance of $YBa_2Cu_3O_{6+x}$ crystals in the discussed region of fields and temperatures is negligible in the normal state, i.e., the shape of $\sigma_n(T)$ is constant with the magnetic field.

Our previous investigations of $YBa_2Cu_3O_{6+x}$ single crystals near the boundary of the superconducting region of the phase diagram[32] revealed that the normal resistivity of such samples at $T<20$ K is well described by a logarithmic function. In a broader temperature range ($0.5\ \text{K}<T<150$ K) the conductivity is very closely described by the empirical function

$$\sigma_n(T)=\rho_n^{-1}=[\alpha-\beta\log T+\gamma T]^{-1}. \tag{4}$$

This function with three fitting parameters is used in processing our experimental data.

By approximating the conductivity in zero magnetic field by a sum of σ_n from Eq. (4) and σ_{fl} from Eq. (2), T_c and d being fitting parameters, along with α, β, and γ, we obtain reasonable values $d=8-15$ Å, which are in fair agreement with the $YBa_2Cu_3O_{6+x}$ lattice constant along the c-axis. This indicates that the Aslamazov–Larkin formula yields a correct estimate of the fluctuation conductivity in CuO_2 layers and its application is justified. The normal conductivity is fitted so as to obtain the best approximation of the fluctuation conductivity throughout the range of magnetic field. Nonetheless, the uncertainty in the normal resistivity was quite considerable. It turned out, however, that calculations of the transition temperature are little affected by admissible variations in $\sigma_n(T)$. The resulting uncertainties in the transition temperature are shown in Fig. 2.

This procedure enable us to derive $B_{c2}(T)$ in the mean-field approximation from our measurements. Since the resulting curve of $B_{c2}(T)$ is nonlinear and it casts doubt on the applicability of Eq. (3), we deem it necessary to demonstrate that, on the qualitative level, the shape of the $B_{c2}(T)$ curve is not affected by subtleties of the data processing, owing to the absence of considerable transition broadening. Figure 2a shows the curve of $B_{c2}(T)$ for sample 2a, along with its other characteristic fields, namely, the ''irreversibility line'' determined at $\rho=5\ \mu\Omega\cdot$cm, positions of the peak of derivative $\partial\rho/\partial T$, and the line of ''transition onset,'' which was defined as a point were $\partial\sigma_{fl}/\partial T=10^2\ \ (\Omega\cdot\text{cm})^{-1}/\text{K}$. These lines are plotted in the $B-T$ diagram in Fig. 2a, and Fig. 2b shows positions of these points on the transition curves. (It is noteworthy that the values of B_{c2} are fairly close to those which would be obtained by defining the transition point at a constant resistivity level $\rho/\rho_n=0.8$.) Figure 2a clearly shows that all curves in the $B-T$ plane have positive curvature throughout the range of studied magnetic fields, including the region of low fields. This leads us to a conclusion that, even if the data processing procedure yields erroneous values

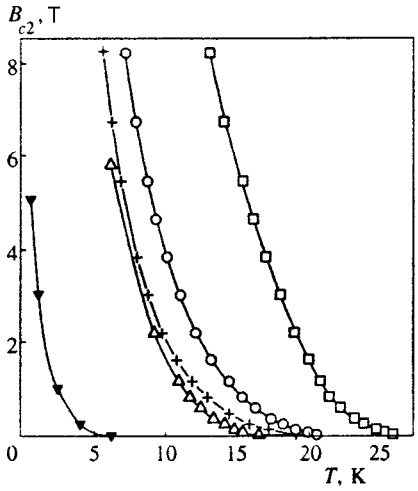


FIG. 3. Temperature dependence $B_{c2}(T)$ in different states. Data for sample 1a in (empty triangles) quenched, (open circles) intermediate, and (squares) aged states; aged states of (crosses) sample 2a and (full inverted triangles) sample 3a.

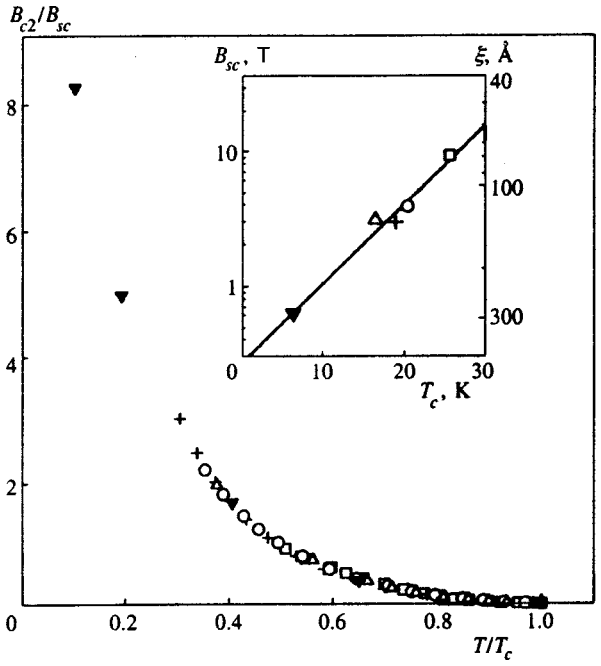


FIG. 4. Dependencies $B_{c2}(T)$ for different samples reduced to the universal function $b_{sc}(t)$ using variables (5). The notation is the same as in Fig. 3. The inset plots the parameters B_{sc} (left-hand axis) and correlation length ξ_0 calculated by Eq. (6) (right-hand axis).

of B_{c2}, the temperature dependence of this parameter is qualitatively correct. Our further analysis, however, will be based on the values derived from measurement data for the fluctuation conductivity.

4.3. Universal temperature dependence of the upper critical field

Measurements of $B_{c2}(T)$ in three samples and five different states (all the states of samples 1a and 2a and the aged state of sample 3a) are given in Fig. 3. It turned out that the curves for all the states can be brought to coincidence by varying the scales of the magnetic field and temperature, i.e.,

$$B_{c2}=B_{sc}b_{sc}(t), \quad t=T/T_c, \tag{5}$$

where B_{sc} is the parameter characterizing the state and $b_{sc}(t)$ is a universal function (Fig. 4). Function $b_{sc}(t)$ contains an arbitrary numerical factor. In Fig. 4 parameter B_{sc} is defined as B_{c2} at a specific reduced temperature equal for all samples, namely, $T_c/2$, i.e., the curves of $B_{c2}(T)$ were brought to coincidence at two points, namely, at $t=1$ and $t=0.5$. The values of B_{sc} for different states are listed in Table 1 and plotted in the inset to Fig. 4 as a function of the zero-field transition temperature. These points lie on one smooth curve, even though they are derived from measurements of the three different samples. The characteristic scale of magnetic field decreases (accordingly, the coherence length increases) with decreasing doping level more rapidly than T_c, i.e., B_{sc} is a superlinear function of T_c. This may be the main cause of the narrowing of the vortex-liquid region in the $B-T$ phase diagram. As a result, $YBa_2Cu_3O_{6+x}$ crystals with a high degree of underdoping with $T_c \lesssim 30$ K do not display notable broadening of the resistive transition due to magnetic field.

The function $B_{c2}(T)$ was measured on sample 2b in a very narrow temperature range, $T/T_c \gtrsim 0.9$, owing to the limit on available magnetic fields. Its second derivative in this interval is also positive and all measurements of $B_{c2}(T)$ can be fitted to function (5). But, since no data for lower temperatures are available and the expected critical fields are very high, the measurements of sample 2b have not been analyzed in this context.

Function $b_{sc}(t)$ is much different from $b_{\mathrm{BCS}}(t)$. First, it has no linear section near $t=1$. This statement relies on Eq. (5), since for each $B_{c2}(T)$ curve the limited precision allows one to draw a straight line of a small slope in the region within 1–2 K near T_c, but if we consider the samples with higher T_c, this linear region would be more narrow, and the slope of function $b_{sc}(t)$ at $t=1$ is smaller, which leads us to a conclusion that the universal curve has no linear section near $t=1$.

Second, b_{sc} continues to rise as $t\to 0$. Figure 4 shows this tendency in the region down to $t=0.1$. In order to test the range of lower t, we investigated sample 3b with $T_c \approx 3$ K at millikelvin temperatures. Its transition curve is too wide to determine quantitatively T_c and $B_{c2}(T)$. Nonetheless, the measurements yield important qualitative information. Figure 5 shows the sample resistance versus magnetic field obtained at temperatures of 50 and 36 mK normalized to the resistance at a magnetic field of 14 T. It is clear that a drop in temperature shifts the magnetoresistance curve to lower fields, i.e., $B_{c2}(T)$ still grows with decreasing temperature even at $T/T_c \sim 0.01$. We can obtain the following estimate: on the level $\rho/\rho_n=0.8$, which approximately corresponds to $B_{c2}(T)$ according to Fig. 2b, the magnetic field increases by 0.6 T; this yields a derivative of 40 T/K. Unfortunately, we cannot plot these points in Fig. 4 for the lack of T_c and B_{sc}.

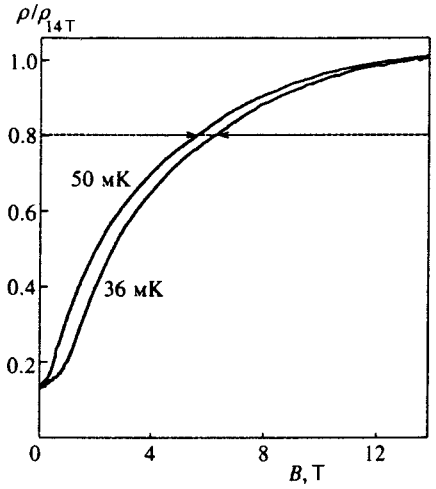


FIG. 5. Reduced resistivity of sample 3b in the aged state at temperatures of 50 and 36 mK as a function of magnetic field. The arrows indicate the difference between magnetic fields at which the curves cross the level $\rho/\rho_{14T}=0.8$.

This observation of $B_{c2}(T)$ increasing even at very low temperatures is in accord with measurements of other materials, e.g., $Tl_2Ba_2CuO_6$,[2] where the upper critical field continues to grow at temperatures down to $T/T_c=0.001$.

Our data indicate that function $B_{c2}(T)$ in underdoped $YBa_2Cu_3O_{6+x}$ is nonlinear in the neighborhood of T_c, and $(\partial B_{c2}/\partial T)_{T_c}=0$. This conclusion contradicts most theoretical models based on the BCS model or the Ginzburg–Landau functional, which either predict a linear behavior of this curve near T_c or assume its existence *a priori*. This issue was not discussed in previous publications of experimental investigations,[2–7] but they all reported very low, if not zero, values of $\partial B_{c2}/\partial T$ at T_c.

The increase in the critical field owing to weakening of the spin-flip scattering predicted by Ovchinnikov and Kresin[8] should occur in the range of low temperatures, so it leaves the linearity of $B_{c2}(T)$ near T_c essentially unaffected. The mechanism suggested by Spivak and Zhou[9] is effective only in high magnetic fields, where Landau quantization is significant, i.e., it also should not affect $B_{c2}(T)$ near T_c. Abrikosov[11] derived $B_{c2}(T)$ from the Ginzburg–Landau functional based on his model, which leads, naturally, to a linear dependence of B_{c2} in the first order in $1-t$.

The nonlinearity of $B_{c2}(T)$ near T_c follows at present only from the model of bipolaron superconductivity[10,33] which yields positive curvature of $B_{c2}(T)$ for a charged Bose-liquid in a localizing potential, this throughout the entire temperature range. At temperatures that are not overly low, the model predicts[33]

$$B_{c2}(T)=B_d^*\left(\frac{T_c}{T}\right)^{3/2}\left[1-\left(\frac{T}{T_c}\right)^{3/2}\right]^{3/2},$$

$$B_d^*=\frac{\Phi_0}{2\pi\xi_0^2}\left(1-\frac{n_L}{2n}\right)^{1/2}. \tag{6}$$

Here ξ_0 is the correlation length and $n_L/2n$ characterizes the random potential. Equation (6) defines a universal function in reduced variables without free parameters. The only normalization parameter B_d^* corresponds to parameter B_{sc} introduced in Eq. (5). It follows from Eq. (6) that $B_d^*=0.68B_{sc}$. Comparison between our data and calculations by Eq. (6) (Fig. 6a) shows excellent agreement in the region $T/T_c>0.3$. At lower reduced temperatures experimental points deviate from the theoretical curve, but note that in this range we have only measurements of one state (sample 3a aged).

The factor $(1-n_L/2n)^{1/2}$ in Eq. (6) is unknown, but, since neither in state 3a nor in state 3b have we detected a re-entrant behavior of $B_{c2}(T)$ predicted by Alexandrov,[33] it should be rather close to unity. Assuming this, we can derive from Eq. (6) the correlation length ξ_0 (Fig. 4, right-hand axis in the inset). The length ξ_0 varies between 70 and 300 Å. The notable increase in the correlation length may be the main cause of the narrowing of the vortex-liquid region on the $B-T$ diagram.

In the low-temperature region $0.1<t<0.6$ the function $b_{sc}(t)$ can be empirically described by the exponential

$$b_{sc}=b_0\exp(-t/t_0), \tag{7}$$

with parameters $b_0=15$ and $t_0=5.4$ (Fig. 6b). Such an

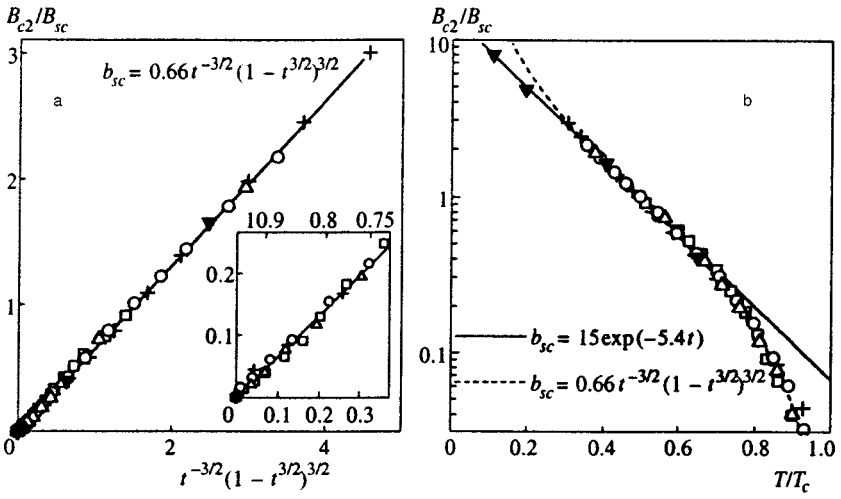


FIG. 6. Function $b_{sc}(t)$ plotted in different coordinates: (a) the coordinates are selected in accordance with the boson model, Eq. (6); the inset shows the section close to $t=T/T_c=1$ on the extended scale; (b) semilogarithmic coordinates; the dashed line follows Eq. (6).

unusual temperature dependence in the region of low temperatures was detected previously[25,34] in measurements of the irreversibility line $B^*(T)$. We suppose that the region of vortex liquid in phase diagrams of samples with low T_c is a very narrow strip between B^* and B_{c2}, hence B^* closely follows $B_{c2}(T)$, especially at low temperatures.

5. CONCLUSIONS

Our investigation has supplemented the list of materials displaying anomalous temperature dependence of the upper critical field $B_{c2}(T)$ with underdoped cuprate $YBa_2Cu_3O_{6+x}$. We have studied samples with different carrier concentrations and T_c ranging between 6 and 30 K. Throughout the studied temperature range, the curve of $B_{c2}(T)$ for these samples have positive curvature and does not saturate at low temperatures. The curves for states with different T_c can be brought to coincidence in reduced coordinates T/T_c and $B/B_{sc}(T_c)$. A fundamental feature of the universal function $b_{sc}(t)$ obtained in this manner is the tendency of its first derivative $\partial B_{c2}/\partial T$ to zero as $T \to T_c$. Such a behavior can be interpreted at present only in terms of the model[10,33] treating the superconducting transition as Bose-condensation of preformed pairs. Other models designed to interpret the anomalous shape of the $B_{c2}(T)$ curve predict a linear temperature dependence of B_{c2} near T_c.

In the low-temperature range $T/T_c < 0.3$, experimental points deviate from function (6). On the other hand, measurements in the temperature interval between the lowest accessible values and $t \approx 0.6$ follow function (7). The combination of Eqs. (6) and (7) analytically describes function $b_{sc}(t)$.

However, the ''universality'' of function $b_{sc}(t)$ is limited. We tested this function on our measurements of $K_{0.4}Ba_{0.6}BiO_3$,[5] and the experimental curve after renormalization of variables according to Eq. (5) was different from function $b_{sc}(t)$ plotted in Fig. 4.

We are indebted to V. T. Dolgopolov and A. A. Shashkin for the opportunity to conduct low-temperature measurements in the dilution refrigerator.

This work was supported by RFBR–PICS (Grant 98-02-22037), by RFBR–INTAS (Grant 95-02-302), and by the Statistical Physics Program sponsored by the Russian Ministry of Science and Technology.

*)E-mail: gantm@issp.ac.ru

[1] N. R. Werthamer, E. Helfand, and C. Hohenberg, Phys. Rev. **147**, 295 (1966).
[2] A. P. Mackenzie, S. R. Julian, G. G. Lonzarich *et al.*, Phys. Rev. Lett. **71**, 1238 (1993).
[3] M. S. Osofsky, R. J. Soulen, Jr., S. A. Wolf *et al.*, Phys. Rev. Lett. **71**, 2315 (1993).
[4] M. Affronte, J. Marcus, C. Escribe-Filippine *et al.*, Phys. Rev. B **49**, 3502 (1994).
[5] V. F. Gantmakher, L. A. Klinkova, N. V. Barkovskii *et al.*, Phys. Rev. B **54**, 6133 (1996).
[6] D. D. Lawrie, J. P. Franck, J. R. Beamish *et al.*, J. Low Temp. Phys. **107**, 491 (1997).
[7] Y. Dalichaouch, B. W. Lee, C. L. Seaman, J. T. Markert, and M. B. Maple, Phys. Rev. Lett. **64**, 599 (1990).
[8] Yu. N. Ovchinnikov and V. Z. Kresin, Phys. Rev. B **54**, 1251 (1996).
[9] B. Spivak and Fei Zhou, Phys. Rev. Lett. **74**, 2800 (1995).
[10] A. S. Alexandrov and N. F. Mott, Rep. Prog. Phys. **57**, 1197 (1994).
[11] A. A. Abrikosov, Phys. Rev. B **56**, 446, 5112 (1997).
[12] A. S. Alexandrov, V. N. Zavaritsky, W. Y. Liang, and P. V. Nevsky, Phys. Rev. Lett. **76**, 983 (1996).
[13] A. V. Nikulov, Phys. Rev. Lett. **78**, 981 (1997).
[14] D. S. Fisher, M. P. A. Fisher, and D. A. Huse, Phys. Rev. **43**, 130 (1991).
[15] G. Blatter, M. V. Feigelman, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Rev. Mod. Phys. **66**, 1125 (1994).
[16] S. Ullah and A. T. Dorsey, Phys. Rev. B **44**, 262 (1991).
[17] M. Tinkham, Phys. Rev. Lett. **61**, 1658 (1988).
[18] M. Suzuki and M. Hikita, Phys. Rev. B **44**, 249 (1991).
[19] A. Carrington, D. J. C. Walker, A. P. Mackenzie, and J. R. Cooper, Phys. Rev. B **48**, 13051 (1993).
[20] S. Fleshler, W. K. Kwok, U. Welp *et al.*, IEEE Trans. Appl. Supercond. **3**, 1483 (1993).
[21] G. W. Crabtree and D. R. Nelson, Phys. Today **4**, 38 (1997).
[22] A. Carrington, A. P. Mackenzie, and A. Tyler, Phys. Rev. B **54**, R3788 (1996).
[23] W. K. Kwok, S. Fleshler, U. Welp *et al.*, Phys. Rev. Lett. **69**, 3370 (1992).
[24] A. N. Lavrov and L. P. Kozeeva, Physica C **248**, 365 (1995).
[25] G. T. Seidler, T. F. Rosenbaum, D. L. Beinz *et al.*, Physica C **183**, 333 (1991).
[26] A. N. Lavrov and L. P. Kozeeva, Physica C **253**, 313 (1995).
[27] S. I. Dorozhkin, G. V. Merzlyakov, and V. N. Zverev, Prib. Tekh. Éksp. No. 2, 165 (1996).
[28] T. A. Friedmann, J. P. Rice, J. Giapintzakis, and D. M. Ginsberg, Physica C **39**, 4258 (1989).
[29] S. B. Han, C. C. Almasan, M. C. de Andrade *et al.*, Phys. Rev. B **46**, 14290 (1992).
[30] M. A. Crusellas, J. Fontcuberta, and S. Piñol, Physica C **213**, 403 (1993).
[31] B. Iwasaki, S. Inaba, K. Sugioka *et al.*, Physica C **290**, 113 (1997).
[32] V. F. Gantmakher, L. P. Kozeeva, A. N. Lavrov *et al.*, JETP Lett. **65**, 870 (1997).
[33] A. S. Alexandrov, Phys. Rev. B **48**, 10571 (1993).
[34] K. E. Gray, D. B. Kim, B. W. Veal *et al.*, Phys. Rev. B **45**, 10071 (1992).

Translation provided by the Russian Editorial office.

Orientational dependence of the tails of dipole-broadened NMR spectra in crystals

V. E. Zobov*) and Yu. N. Ivanov

L. V. Kirenskiĭ Institute of Physics Siberian Division, Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

M. A. Popov

Krasnoyarsk State University, 660041 Krasnoyarsk, Russia

A. I. Livshits

Institute of Chemistry and Chemical Technology, Siberian Branch, Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 3 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 285–305 (January 1999)

This paper describes experimental and theoretical studies of the tails of the dipole-broadened nuclear magnetic resonance (NMR) absorption spectra of ^{19}F in isomorphic single crystals of BaF_2 and CaF_2 with the magnetic field directed along three crystallographic axes. The results obtained by directly measuring the derivative of the tail of the NMR absorption spectrum and the falloffs of the Engelsberg–Lowe free precession after Fourier transformation qualitatively agree. It is shown that the shape of the tail is well described by an exponential function in which the orientational dependence of the exponent does not reduce to variation of the second moment. The observed shape of the tail and the orientational dependence of its parameters are explained on the basis of a self-consistent fluctuating-local-field theory. Nonlinear integral equations are derived for the correlation functions, taking into account the changes of the actual number of nearest neighbors caused by the anisotropy of the dipole–dipole interaction and the contribution of lattice sums with loops. The equations are solved numerically. Good agreement is obtained for the computed dropoffs of the free precession, the NMR spectra, and the cross-polarization rates with the experimental results. © *1999 American Institute of Physics.* [S1063-7761(99)02401-4]

1. INTRODUCTION

The continued interest in the problem of the absorption line shape and the spectra of other correlation functions measured by nuclear magnetic resonance (NMR) in the solid state has two causes: first, the important applied significance of NMR for studying the properties of solids at the microlevel, and second, as a typical many-body problem. An indisputable advantage of model crystals such as CaF_2 or BaF_2 is the simplicity of the known laws governing the interactions in their nuclear magnetic subsystems (the main one of which is the dipole–dipole subsystem) and the possibility of experimentally verifying the theoretical derivations. The central part of the spectrum is ordinarily used in applied problems in this case, whereas information concerning the fundamental multiparticle dynamic properties of the system is included in the tails of the spectrum. This is because, in a homogeneous regular system, a response to an effect with a frequency many times as great as the rms precessional frequency in a local field is impossible unless a large number of spins participate. The distant region of the spectrum (the tail) is of the greatest practical interest when one is studying processes involving the establishment of equilibrium in a spin system consisting of strongly differing resonance frequencies of the subsystems (the reservoirs)—cross-relaxation processes. This is shown by the large number of experimental papers on measuring the rates of these processes (see the citations and their analyses in Ref. 1). The study of such processes in turn is closely associated with the general problem of mixing in nonlinear mechanics.

Because of this multifrequency behavior, calculation of the tails of the spectra of the correlation functions imposes requirements on the theory unlike those of the calculation of the central part. It is very difficult to experimentally measure the tails, because they are small and are therefore strongly affected by noise, nonideal properties of the apparatus, etc. For these reasons, the tails of the spectra have been insufficiently studied both theoretically and experimentally. This is also very true for the tails of the NMR absorption line. The experimental papers we are aware of measured either the central part or the Fourier transform—the falloff of the free precession. The former relates to the work of Bruce,[2] and the latter to that of Engelsberg and Lowe,[3] which is of tremendous interest among theoreticians because of the oscillations of the falloff of the free precession in CaF_2, measured with high accuracy. In fact, it became the cornerstone of theories concerning the NMR line shape (see, for example, Refs. 4–10).

The exponential form of the tail of the NMR spectrum follows from the results of Ref. 3 (see Appendix A), and this agrees with the results of a number of experiments[1,11–14] and

TABLE I. Parameters of the NMR spectrum of ^{19}F in BaF_2 for three directions of the magnetic field.

Field direction	M_2^{theor}, Oe^2	M_2^{exp}, Oe^2	$2H_m$, Oe	N_s	S_3/S_1^2	K_6	K_8
[111]	1.055	1.219	0.4	25	0.12	0.10	0.4
[110]	2.284	2.324	0.5	20	0.17	0.18	0.5
[100]	5.966	5.798	0.6	12	0.09	0.05	0

of the theory constructed in the approximation of a self-consistent fluctuating field.[15–19] Other papers on the theory of the line shape did not pay proper attention to the tail. Thus, for example, it falls off more quickly in the constant-local-field approximation[4,5] than for a Gaussian function, whereas, when the field fluctuations are specified by a random Markov process,[7,10] the tail becomes a power function. In the theory that we developed,[16–18] in which the approximation of a self-consistent fluctuating field is chosen, corresponding to the limit of systems of large dimensionality, all the parameters are expressed in terms of one scale parameter, the second moment. However, the variations of the parameters of the tails of the spectra of the experimental falloffs of the free precession in Ref. 3 are not described by the variation of only the second moment when the magnetic field is directed along the crystallographic axes [100], [110], and [111].

This paper derives nonlinear integral equations for the correlation functions in the self-consistent fluctuating-local-field approximation,[1,16–19] taking into account the characteristics of actual lattices, which, as a consequence of the anisotropy of the dipole–dipole interaction, depend on the magnetic-field orientation. At the same time, this paper reports the direct measurement of the tail of the NMR absorption line of ^{19}F in a BaF_2 crystal isomorphous with CaF_2, with the same magnetic-field directions. Such an experiment seems important to us, since the fraction of high frequencies in the spectrum is exponentially small, and they can easily be distorted during observation in the dropoff of the free precession in a mixture with the central part of the spectrum. The orientational dependences of the parameters of the tail, measured by two methods, are in qualitative agreement. These results are explained from the position of the theory that we developed.

2. EXPERIMENT

The single crystal of BaF_2 studied here was grown at the Crystallography Institute, Russian Academy of Sciences, by the Bridgman method. The quality of the crystal was monitored by x-ray phase analysis and by NMR. The long spin-lattice relaxation time is evidence that the concentration of paramagnetic impurities in the test sample is low. The single crystal was oriented on an x-ray diffractometer. The lattice parameter of 6.2001 Å in BaF_2 (Ref. 20) is a factor of 1.14 greater than in CaF_2. The experiment was run on a modified RYa-2310 spectrometer with an autodyne sensor in a 12-kOe field at room temperature. The first derivative of the NMR absorption line was digitally measured by a microprocessor device with field scanning of the spectrum. The long-term stability of the spectrometer parameters was monitored by simultaneously measuring the signal from a mark placed in part of the coil of the NMR sensor separately from the sample. Particular attention was paid to choosing the optimum rf field so that the saturation effect was below the noise level. The SNR was substantial increased by accumulating the NMR signal by multiple scanning of the spectrum (the number N_s of scans is shown in Table I). The time for one scan was 20 min.

The NMR lines were measured with a constant magnetic field oriented along the crystallographic axes. Because the spectra are symmetric, Fig. 1a shows only half of them. Figure 1b shows the tails of these derivatives on a semi-log plot. The curves in the figures are normalized to unit area of the absorption line. The deviation of the field from the center of the spectrum in each orientation is expressed in units of $M_2^{1/2}$, where M_2 is the second moment of the spectrum. This eliminates the difference of the scales of the spectra and allows their shapes to be compared.

The experimental values of M_2 were calculated by extrapolating the ratio of the integrals of the product of the measured first derivatives of the spectrum and the cube of the detuning and the triple detuning to larger values of the upper limit of integration.[21] Table I also shows the theoretical values of the second moments for BaF_2. The latter were calculated using lattice sums from Refs. 12 and 22, taking into account the small contribution of the magnetic isotopes ^{135}Ba and ^{137}Ba, whose maximum is reached in the [111] orientation and equals 3% of the contribution of the ^{19}F nuclei. Moreover, because the NMR line is broadened by modulating the constant magnetic field with an amplitude of H_m (see Table I), $H_m^2/4$ should be added to these values of the second moment.[23] The remaining differences of the theoretical and experimental values of the moments are associated with imprecision in the orientation of the crystal in the magnetic field. Since we consider spectra normalized to M_2, a slight discrepancy of the moments does not prevent the shapes of the spectra from being compared. Therefore, we shall pay no attention to these differences in what follows, nor to the contribution of the Ba nuclei and the field modulation to the broadening. According to our estimates, the possible shape distortions of the tail are below the experimental accuracy.

Figures 1a and 1b also show the derivatives of the spectra obtained by Fourier-transforming the function

$$f(t)=\exp\{C[A-(A^2+t^2)^{1/2}]\}\prod_{n=1}^{61}(1-\alpha_n^2t^2), \qquad (1)$$

which Engelberg and Lowe[3] used to accurately describe their experimental dropoffs of the free precession in CaF_2. They determined the parameters which determine this function for the same three magnetic field orientations. When the curves

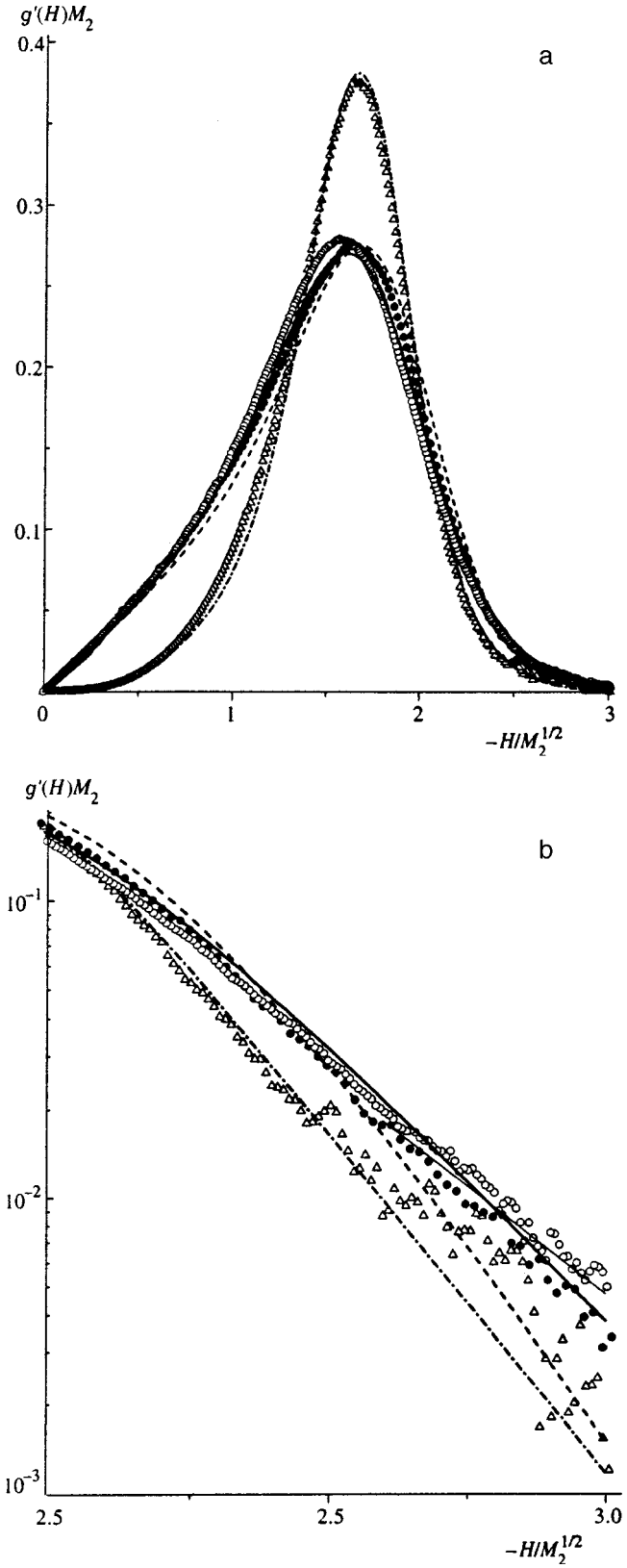


FIG. 1. Derivatives of the NMR absorption spectra of ^{19}F in BaF_2 [(a) central part, (b) tail] as a function of the detuning from the center of the spectrum, with the magnetic field directions along the crystallographic axes [100] (triangles), [110] (closed circles), and [111] (open circles). The dot-dashed, dashed, and solid curves show the derivatives of the Fourier spectra of the Engelsberg–Lowe function, Eq. (1), in the corresponding orientations. A thin line segment is drawn in (b) according to the asymptotic formula, Eq. (2). All the curves are normalized to unit area of the absorption spectrum and unit second moment.

in the figure were calculated, these parameters were expressed in units of the experimental values of the second moments for CaF_2 given in this paper. It can be seen from the figure that the Fourier transform of the function given by Eq. (1) generally describes our experimental NMR absorption spectrum. The small differences can be associated with the noncoincidence of the orientations of the crystals and the instrumental functions of the two methods.[24,25] together with the replacement of the actual falloffs of the free precession in Ref.3 by the simple function given by Eq. (1). We shall return to this question below.

We proceed to an analysis of the shape of the tail of the NMR spectrum. To describe it, we turn to the theory that we developed,[1,16–19] based on the self-consistent fluctuating-local-field approximation, by means of which, in the limit $H \gg M_2^{1/2}$ (the H field is measured from the center of the spectrum), the desired tail is determined from

$$g(H) \simeq c_0 |H|^{\chi} \exp(-|H|\tau_0), \tag{2}$$

where τ_0 is the coordinate of the closest two singular points, symmetrically placed relative to the coordinate origin on the imaginary time axis, and c_0 and χ are characteristics of the singular points. In the limit of a large number of nearest neighbors,[16,17]

$$\tau_0 = 3.72/M_2^{1/2}, \quad c_0 \approx 29.3 M_2, \quad \chi = 1. \tag{3}$$

A section of the curve corresponding to the derivative of Eq. (2) is shown in Fig. 1b. It passes fairly close to the experimental tail in the [111] orientation. In the other two orientations, the tails of the experimental spectra fall more steeply.

We now turn to the Engelsberg–Lowe function given by Eq. (1). As can be seen from Fig. 1b, its spectrum decreases more quickly in all three orientations. The asymptotic expression for the tail of the spectrum of this function, obtained in Appendix A, has the form of Eq. (2) with $\chi = -1/2$ and $\tau_0 = A$. An unexpected orientational dependence is detected in the exponential in this case: A is larger in the [110] orientation than in the [100] orientation.

Our analysis of the curves in Fig. 1b thus shows that, first, the shape of the spectrum at the tail is close to exponential, given by a straight line in the semi-log coordinates chosen in the figure. Second, the slope of the corresponding straight lines depends on the orientation of the crystal in the magnetic field. Since the change in the width of the spectrum with orientation is already taken into account in Fig. 1b after transforming to dimensionless fields measured in units of $M_2^{1/2}$, the remaining change of the slope of the straight lines is evidence of an additional orientational dependence of the argument of the exponential.

3. THEORY

To explain the observed orientational dependence of the tail of the NMR spectrum, let us consider the system of spins ($I = 1/2$) of the ^{19}F nuclei of the fluorite crystal, which form a simple cubic lattice. We write the Hamiltonian of the secular part of the dipole–dipole interaction in a strong constant magnetic field[25] as

$$\mathscr{H}_d = \sum_{i \neq j} b_{ij}[I_i^z I_j^z - \xi(I_i^x I_j^x + I_i^y I_j^y)], \tag{4}$$

where $b_{ij} = \gamma^2\hbar[1-3\cos^2\theta_{ij}]/2r_{ij}^3$, θ_{ij} is the angle made by the internuclear vector $\mathbf{r}_{ij}$ with the constant magnetic field $\mathbf{H}_0$, and $\xi = 1/2$ is a parameter that we introduced for convenience in the theoretical analysis. We shall describe the dynamics of the spin system by the correlation functions

$$\Gamma_p(t) = \mathrm{Tr}\{\exp(i\mathscr{H}_d t) I_p \exp(-i\mathscr{H}_d t) I_p\}/\mathrm{Tr}\{(I_p)^2\}, \tag{5}$$

where the subscript $p = 1,2,3$ indicates the three correlation functions: $\Gamma_1(t) = M_x(t)$ is the correlation function of the x projection of the total spin of the system or the transverse component of the magnetization, coinciding with the falloff of the free precession; $\Gamma_2(t) = \Gamma_x(t)$ and $\Gamma_3(t) = \Gamma_z(t)$ are the autocorrelation functions of the x and z components of an individual spin of the system, respectively.

In the self-consistent fluctuating-local-field approximation, corresponding to the limit $d \to \infty$, the system of equations for the correlation functions (4) is obtained in the form[16–18]

$$\frac{d}{dt}\Gamma_p(t) = -\int_0^t G_p(t-t')\Gamma_p(t')dt'. \tag{6}$$

The kernels $G_p(t)$ of the integral equations (the memory functions) can be represented as a series over irreducible dressed skeletal diagrams, each term of which is expressed via a multiple time integral of the products of the functions $\Gamma_x(t')$ and $\Gamma_z(t'')$. As shown in Refs. 16 and 17, the equations for the autocorrelation functions are the equations for the precession of the magnetic moment in a three-dimensional Gaussian random local field. These equations have a complex form because the rotations around the time-varying instantaneous field directions are noncommutative. In this approximation, all the coefficients in $G_p(t)$ are expressed in terms of M_2, and therefore, in the solutions of the equations, the orientational dependence repeats the dependence of M_2 and reduces to a variation of the time scale in Eqs. (5).

For three-dimensional lattices, Refs. 15, 26, and 27 proposed to introduce correction terms in the kernel $G_p(t)$, the number of which rapidly increases as the number of vertices on the diagrams increases. Such an equation is hard to use in practice. It is necessary to regroup the series for the kernel so that its first several terms are sufficient to describe the experiments.

To do this, we separate out from the dipole–dipole interaction of Eq. (4) the longitudinal part, consisting of the spin components parallel to the external constant magnetic field.[1,4–6,8,10,18,27] Although the coefficients of the two parts in Eq. (4) differ by only a factor of two, the longitudinal part is distinguished by the axial symmetry of the Hamiltonian, which causes the projection of the total spin onto the z axis to be conserved in time. It is also important that for $\xi = 0$ the autocorrelation function given by Eq. (5) for the x projection of spin i is easy to compute:[4,25]

$$\Gamma_0(t) = \prod_j \cos(b_{ij}t), \tag{7}$$

and describes the independent precession of one of the spins of the system in its constant longitudinal local field, $2\Sigma_j b_{ij} I_j^z$.

The transverse part of the interaction given by Eq. (4), consisting of the spin components perpendicular to the external constant magnetic field, as is well known,[25] plays an important role in transporting polarization from node to node (spin diffusion). Taking into account the transport of the transverse polarization, Refs. 4 and 5 derived an equation in first order in the transverse interaction:

$$M_x(t) = \Gamma_\lambda(t) + K\int_0^t \frac{d\Gamma_\lambda(t')}{dt'} M_x(t-t')dt', \tag{8}$$

where

$$K = 9/4\lambda^2 - 1, \tag{9}$$

and $\Gamma_\lambda(t)$ is the correlation function given by Eq. (7) with coefficient b_{ij} increased by factor of λ. This equation, which we shall call the basic approximate equation, gave a good description of the falloff of the free precession in CaF_2 for $\lambda = 1.225$ (Ref. 4) and $\lambda = 1.19$.[5] Note that the factor λ in Refs. 4–6 and 16 has a different physical basis. We shall regard it as a renormalization parameter of the longitudinal local field, defined in terms of the moment of the spectrum.

The success of Eq. (8) in describing the falloff of the free precession suggests that, after the terms in Eq. (6) corresponding to Eq. (8) are singled out, the rest of the series for the kernel will play the role of a small correction. We carry out the indicated transformation by the following formal procedure. We represent $\Gamma_\lambda(t)$ as the solution of an integral equation of the form (6) with kernel $Q(t)$, which can be given by series $G_2(t)$ if the terms with vertices corresponding to interaction between transverse spin projections are discarded from it. By combining the Laplace transforms of this equation and Eqs. (6) and (8), we find

$$M_x(t) = \Gamma_\lambda(t) + K\int_0^t \frac{d\Gamma_\lambda(t')}{dt'} M_x(t-t')dt' - \int_0^t \Phi(t-t')M_x(t')dt', \tag{10}$$

where

$$\Phi(t) = \int_0^t \Gamma_\lambda(t-t')\{G_1(t') - (1+K)Q(t')\}dt'.$$

The resulting equation makes it possible to find the necessary correction terms, since it is formally exact when the complete series for $G_1(t)$ and $Q(t)$ are retained.

Another important consequence of the transverse interaction is the time variation of the spin orientation, which causes I_j^z to be replaced by $I_j^z(t)$ in the expression for the longitudinal local field. The basic approximate Eq. (8) does not reflect such fluctuations, whose presence follows not only from theory but also from experiments, for example, from the cross-polarization of the rare nuclei ^{43}Ca,[11] in which the spectrum of these fluctuations is measured. Therefore, although the falloff of the free precession is successfully described by this equation at short times, discrepancies

with experiment appear at long times. In particular, beats appear in the oscillations of the falloff of the free precession and are especially appreciable in the [100] orientation in the region of the 5–7th zeros.[4] The tails of the Fourier spectra decrease more rapidly in the calculated falloffs of the free precession than with Gaussian functions.

We will include fluctuations of the longitudinal local field in the basic approximate equation (8), having replaced $\Gamma_\lambda(t)$ with a new autocorrelation function $P(t)$. The procedure for deriving Eq. (10) allows us to make such a replacement in this equation. To determine $P(t)$, we consider the correlation function of the longitudinal local field at spin i,[16,17]

$$2\lambda\sum_j b_{ij}I_j^z(t).$$

Interaction with this spin is excluded in the time evolution of its neighboring spins:

$$\langle\omega_i(t)\omega_i\rangle=\lambda^2\sum_j b_{ij}^2\Gamma_{zj/i}(t)+\lambda^2\sum_{j,k} b_{ij}b_{ik}\Gamma_{zjk/i}(t). \tag{11}$$

The first term contains the autocorrelation function of the z projection of spin j. The second term is the overlap correlation function of the two spins j and k. The slash indicates that interaction with the selected spin i is excluded, as mentioned above. The contributions to the local field from the different spins of the neighborhood are not independent. Such independence appears only in the limit $d\to\infty$.[1,16,17] In fact, in this limit, lattice sums with loops composed of bonds become negligible by comparison with lattice sums that contain no loops and that are expressed in terms of the power of the second moment. Other model systems where there are no loops are systems on Bethe lattices.[19] The contributions of adjacent spins to the local field will also be independent in these systems, since interaction with spin i is excluded in them. Bethe lattices have an advantage over hypercubic lattices of infinite dimensionality in that the number Z of neighbors in them can be arbitrary.

The contribution of the second term in Eq. (11) is comparatively small for a cubic lattice, although it does not disappear. To estimate it, we expand Eq. (11) in powers of time:

$$\langle\omega_i(t)\omega_i\rangle=\lambda^2S_1+2\xi^2\lambda^4S_1^2(1-S_2/S_1^2-S_3/S_1^2)t^2+O(t^4), \tag{12}$$

where

$$S_1=\sum_j b_{ij}^2,\quad S_2=\sum_j b_{ij}^4,\quad S_3=\sum_{j,k} b_{ij}b_{ik}b_{jk}^2 \tag{13}$$

are known lattice sums.[12,22] The term with S_2 in Eq. (12) results from excluding interaction with the selected spin, while the term with S_3 characterizes the correlation of the contributions. The ratio S_3/S_1^2 in a cubic lattice varies from 0.17 in a [110] orientation to 0.09 in a [100] orientation.

It follows from Eq. (12) that correlation in the motion of the spins that create the local field weakens its fluctuations. The same conclusion can be drawn from the expression for the correlation function of the local field of a heteronuclear system proposed in Ref. 28. This paper, besides the expansion for short times, treated the diffusion asymptotics of the autocorrelation function of the field as $t\to\infty$. We do not do this, because the spectral tail of interest to us is determined by the singular points on a comparatively small time interval from the beginning, and diffusion can not develop. Neglecting diffusion tails allows us to write for Eq. (11) the following expression, which is simpler than that in Ref. 28:

$$\langle\omega_i(t)\omega_i\rangle=\lambda^2\sum_j b_{ij}^2\Gamma_{zj/i}^\nu(t), \tag{14}$$

where the attenuation of the fluctuations indicated above is introduced via the exponent $\nu<1$. In particular, when

$$\nu=\nu_0\equiv 1-S_3/S_1^2,$$

the first two terms of the time expansion of Eq. (14) coincide with Eq. (12). At long times, additional attenuation of the fluctuations from more complicated loops should be expected, as well as bulk interaction of the branches of the trees formed by the b_{ij} bonds.[19] An estimate of the latter for the Heisenberg model by numerical modelling of the placement of the trees on a cubic lattice gave $\nu=\nu'\approx 2/3$.[19] If this value of ν' is used and both these effects are taken into account, the index $\nu=\nu'\nu_0$ changes from 0.55 to 0.61 in different orientations. Bearing in mind that this is a rough estimate, we shall set $\nu=1/2$ in subsequent calculations.

The main advantage of the approximation given by Eq. (14) is that it keeps the contributions of different spins to the longitudinal local field independent when the fluctuations of the latter are taken into account. Such an approximation makes it possible to obtain equations that are simple enough to be used in practice. As a result,

$$\Gamma_{xi}(t)\approx\left\langle\exp\left[2i\lambda\sum_j b_{ij}\int_0^t I_j^z(t')dt'\right]\right\rangle$$
$$=\prod_j\left\langle\exp\left[2i\lambda b_{ij}\int_0^t I_j^z(t')dt'\right]\right\rangle,$$

and the product of cosines in Eq. (7) is replaced by the product

$$P_i(t)=\prod_j F_{ij}(t) \tag{15}$$

of functions that satisfy the equations

$$\frac{d}{dt}F_{ij}(t)=-\int_0^t G_{Fij}(t')F_{ij}(t-t')dt'. \tag{16}$$

The memory function in Eq. (16) can be determined as a series, as was done in Eq. (6). The first term of this series,

$$G_{Fij}^{(1)}(t)=\lambda^2b_{ij}^2\Gamma_{zj/i}^\nu(t) \tag{17}$$

is the contribution to Eq. (14) from spin j. The appearance of the remaining terms of the series is associated with the noncoincidence of the correlation function of the product of the operators

$$\prod_{p=1}^{2n} I_j^z(t_p)$$

with the product of the two-spin correlation functions. In the basic approximate equation, we restrict ourselves to the first term of this series, Eq. (17). Its remaining part is implied in the correction given by Eq. (10). As is to be expected, when the fluctuations are neglected [for $\Gamma_{zj/i}(t)=1$], Eq. (15) gives a product of cosines, Eq. (7), whereas, in the limit of a large number of neighbors, Eqs. (15)–(17) transform into an expression for $\Gamma_{xi}(t)$ with a Gaussian random field.[15–19,25,29]

An autocorrelation function $P(t)$ that takes into account the fluctuations of the longitudinal local field is thereby obtained. We next need to derive an equation for autocorrelation function $\Gamma_{zj/i}(t)$. We take Eq. (6) for the corresponding function, while keeping only the first term in the series for its kernel[1,16,17,26,27] [we recall that the remaining part of this series is meant to be treated as correction terms of Eq. (10)]:

$$\frac{d}{dt}\Gamma_{zj}(t)=-2\xi^2\sum_k b_{jk}^2\int_0^t\Gamma_{xj}(t')\Gamma_{xk}(t')\Gamma_{zj}(t-t')dt'. \tag{18}$$

To clarify the subsequent transformations, we have written out the nodal indices of the interacting spins in Eq. (18). As pointed out above, the equation has such a form in the limit of a large number of neighbors. When the number of neighbors is limited, it becomes important to exclude from the autocorrelation functions the interaction with spins already explicitly included via b_{jk}^2. Carrying out such a procedure and replacing $\Gamma_{xj}(t)$ with $P_j(t)$, we obtain

$$\frac{d}{dt}\Gamma_{zj/i}(t)=-2\xi^2\sum_{k(\neq i)} b_{jk}^2\int_0^t P_{j/ik}(t')\times P_{k/ij}(t')\Gamma_{zj/i}(t-t')dt', \tag{19}$$

where we recall that the indices of the spins with which interaction is excluded are shown after a slash in the symbols of the functions.

The system of Eqs. (15)–(17) and (19) determines the desired autocorrelation functions self-consistently. If function $P(t)$ is then substituted into Eq. (8) in place of $\Gamma_\lambda(t)$, we get the basic approximate equation for $M_x(t)$, taking into account the fluctuations of the longitudinal fields. For the first two moments of the NMR spectrum [the coefficients of the expansion of $M_x(t)$ in powers of time], we get from these equations

$$M_2=(1+K)\lambda^2 S_1,$$
$$\frac{M_4}{M_2^2}=1+\frac{2}{1+K}\left(1+\frac{\nu\xi^2}{\lambda^2}\right)-2\left(1+\frac{\nu\xi^2}{\lambda^2}\right)\frac{S_2}{(1+K)S_1^2}. \tag{20}$$

This result should be compared with the exact expressions for the moments:[25,30]

$$M_2=(1+\xi)^2 S_1,$$
$$\frac{M_4}{M_2^2}=3-B-\frac{(2-B)S_2}{S_1^2}+\frac{BS_3}{S_1^2}, \tag{21}$$
$$B=\frac{4\xi}{1+\xi}-\frac{6\xi^2}{(1+\xi)^2}.$$

Comparing the expressions for M_2, we find that

$$K=(1+\xi)^2/\lambda^2-1,$$

which transforms into Eq. (9) when $\xi=1/2$. Equating the coefficients in front of the lattice sums for M_4 in Eqs. (20), we find

$$\nu=(1+2\xi^2-\lambda^2)/\xi^2. \tag{22}$$

In particular, the values $\nu=1/2$ and $\xi=1/2$ correspond to $\lambda^2=11/8$.

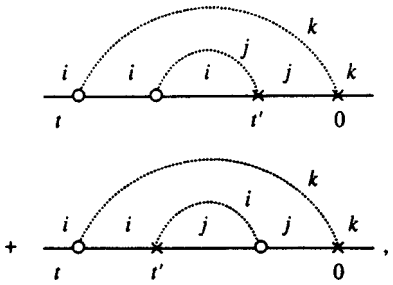
To restore the missing contribution from S_3 in Eqs. (20) for M_4, we go overfrom Eq. (8) to Eq. (10), replace $\Gamma_\lambda(t)$ in it by $P(t)$, and write correction $\Phi(t)$ in the form

$$\Phi_4(t)=\frac{3BS_3M_2}{4S_1^2\lambda^2}\int_0^t\varphi(t-t')\Gamma_z^\nu(t')\{\dot P(t-t')\times P(t')+P(t-t')\dot P(t')\}dt', \tag{23}$$

where

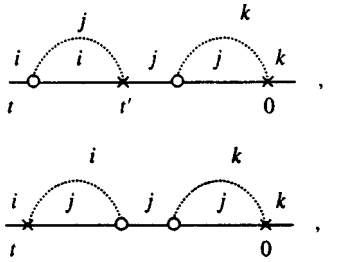
$$\varphi(t)=\int_0^t\Gamma_z^\nu(t')dt',\quad \dot P(t)=\frac{dP(t)}{dt},$$

while the correlation functions under the integral are determined without limitations on the interaction. For clarity, we show this correction in the diagram representation of Refs. 16–18:



(24)

where $\times$ indicates a transverse vertex and $\bigcirc$ indicates a longitudinal one, and the lines show the autocorrelation functions of spins i, j and k (the x projections are shown by solid curves, and the z projections by dotted curves). Let us turn our attention to the approximate replacement $\Gamma_z^\nu(t'')\approx\Gamma_z^\nu(t''-t')\Gamma_z^\nu(t')$, made when we go from Eq. (24) to Eq. (23) to simplify the calculations. The two successive diagrams with two vertices already taken into account in Eq. (8),



(25)

differ from $\Phi_4(t)$ in Eq. (24) in the placement of the vertices. The physical meaning of Eq. (24) is that the polarization can be transferred from spin i to spin k not only via the two-spin correlations given by Eq. (25) but also via three-spin correlations that have the form of a loop composed of bonds and therefore do not reduce to the square of the two-spin correlations. Function $\Phi(t)$ also contains diagrams with another placement of four vertices. To simplify the equations, we do not exhibit all of them, since they have the same qualitative effect as those already shown, and their contribution was taken into account by the choice of the coefficient in Eq. (23).

Besides four-vertex corrections, $\Phi(t)$ contains corrections with a larger number of vertices. Since they have a weaker role, to simplify the calculations we take them in simpler form than in $\Phi_4(t)$:

$$\Phi_{2n}(t)=K_{2n}D_n(t), \tag{26}$$

$$D_n(t)=\int_0^t dt_1\dot{P}(t-t_1)\Gamma_z^{\nu}(t_1)D_{n-1}(t_1), \quad D_1(t)=\dot{P}(t). \tag{27}$$

We choose the coefficients K_{2n} for $n>2$ by fitting $M_x(t)$ to the experimental dropoffs of the free precession.

4. CALCULATION AND DISCUSSION

The system of Eqs. (15)–(17) and (19) consists of an enormous number of nonlinear equations, which makes it hard to solve. Fortunately, the main contribution to the determination of the form of the spectrum comes from the interaction with a comparatively small number Z of nearest neighbors.[6,9] Thus, in the case of CaF_2 and BaF_2, we choose $Z=20$ when a strong constant magnetic field is along the [111] crystallographic axis, $Z=8$ when it is along [110], and $Z=6$ when it is along [100]. This variation of Z results from the strong anisotropy of the dipole interaction constants of the magnetic moments of the fluorine nuclei.[6,9] Because of the symmetry of the field orientations considered here, the interaction constants with the Z chosen neighbors take no more than three values. We denote the three corresponding coefficients b_{ij}^2 in Eqs. (17) and (19) as b_q ($q=1,2,3$) and express them in units of M_2. We denote the number of neighbors with interaction coefficient b_q as n_q. For the [100] orientation we get

$$b_1=d_c/27, \quad b_2=4b_1,$$

$$n_1=4, \quad n_2=2, \quad d_c=0.898,$$

for the [110] orientation we get

$$b_1=d_c/36, \quad b_2=4b_1, \quad b_3=2b_1,$$

$$n_1=4, \quad n_2=n_3=2, \quad d_c=0.791,$$

and for the [111] orientation we get

$$b_1=4m/9, \quad b_2=4m, \quad b_3=27m/8,$$

$$n_1=6, \quad n_2=2, \quad n_3=12, \quad m=8d_c/921, \quad d_c=0.825.$$

The ratio of the contribution from the remaining Z neighbors to the total second moment—the constant d_c—was determined by means of the lattice sums from Ref. 12.

Keeping only these interactions in Eqs. (15)–(17) and (19) and taking $\xi=1/2$, we get the following system of nonlinear equations for the autocorrelation functions:

$$\frac{d}{dt}F_q(t)=-\lambda^2 b_q\int_0^t\Gamma_{z/q}^{\nu}(t-t')F_q(t')dt', \tag{28}$$

$$\frac{d}{dt}\Gamma_{z/q}(t)=-\frac{1}{2}\int_0^t\Gamma_{z/q}(t-t')\left\{\frac{b_1n_1}{F_1^2(t')}+\frac{b_2n_2}{F_2^2(t')}+\frac{b_3n_3}{F_3^2(t')}-\frac{b_q}{F_q^2(t')}\right\}\frac{R_c^2(t')dt'}{F_q(t')}, \tag{29}$$

where

$$R_c(t)=F_1^{n_1}(t)F_2^{n_2}(t)F_3^{n_3}(t). \tag{30}$$

At the same time, Eq. (10) for the correlation function of the x projection of the total spin takes the form

$$M_x(t)=P(t)+K\int_0^t\frac{dP(t')}{dt'}M_x(t-t')dt'-\int_0^t\Phi(t-t')M_x(t')dt', \tag{31}$$

where

$$\Phi(t)=\sum_{n=2}\Phi_{2n}(t), \tag{32}$$

$$P(t)=R_c(t)R_f(t). \tag{33}$$

In Eq. (33), we have combined the contribution of a large number of distant spins in the form of the autocorrelation function of the spin in a random Gaussian field:

$$R_f(t)=\exp\left\{-\frac{4}{9}\lambda^2(1-d_c)\int_0^t(t-t')\Gamma_z^{\nu}(t')dt'\right\}, \tag{34}$$

where $\Gamma_z(t)$ is determined from an equation that differs from Eq. (29) in having A kernel does not contain the divisor $F_q(t)$ and the subtrahend $b_q/F_q^2(t)$. Finally, in Eq. (32) we determine $\Phi_4(t)$ from Eq. (23), and $\Phi_{2n}(t)$ with $n>2$ from Eq. (26).

Applying to the system of nonlinear equations (28) and (29) the same analysis methods as in Refs. 1 and 16–19, it can be shown that its solution has singular points on the imaginary time axis (see Appendix B). Consequently, the Fourier spectrum of this solution has exponential high-frequency asymptotics determined by the nearest singular points. Since the detunings achieved in experiment are not large enough for us to limit ourselves to the first term of the asymptotic series, we shall not dwell on an analysis of the singular points but immediately proceed to a numerical solution of the resulting equations.

The system of Eqs. (28)–(31) was solved by the method of finite differences. The falloffs of the free precession were accurately calculated on the time interval from $t=0$ to

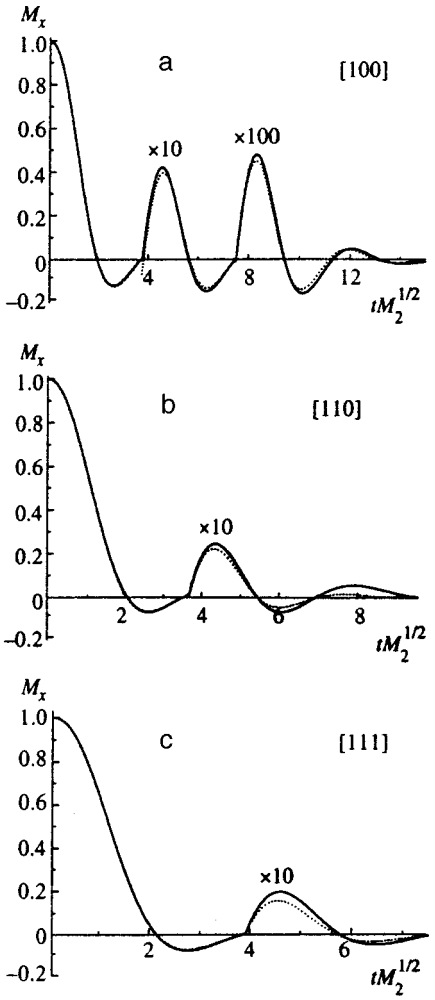


FIG. 2. Falloffs of the free precession $M_x(t)$ with the magnetic field directions along the crystallographic axes [100] (a), [110] (b), and [111] (c), increased at long times by factors of 10 and 100. The solid curves are the theoretical results, and the dashed curves are the Engelsberg–Lowe functions, Eq. (1).

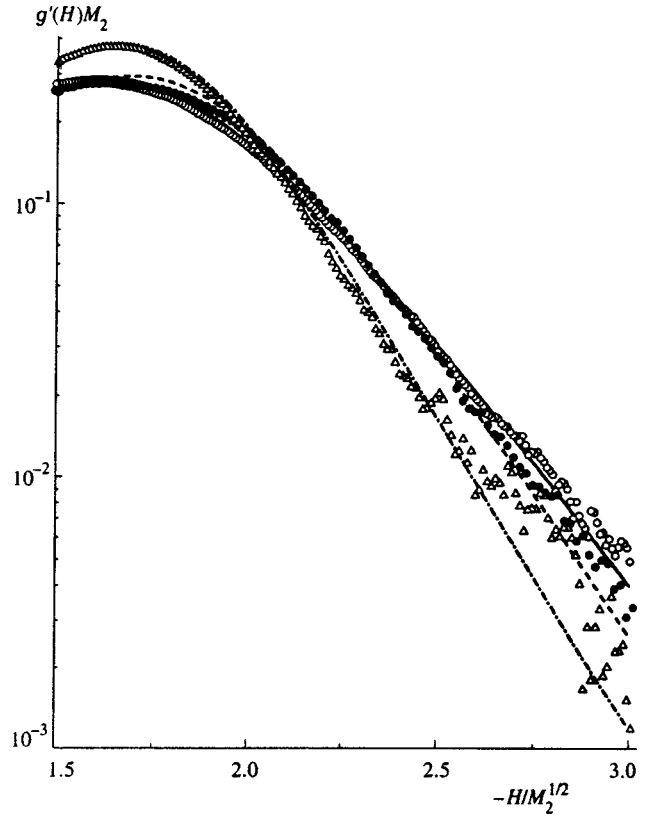


FIG. 3. Tails of the derivatives of the Fourier spectra of the theoretical curves shown in Fig. 2, in comparison with the experimental tails of the NMR absorption spectra of ^{19}F in BaF_2 shown in Fig. 1. The theoretical curves are solid for the [111] orientation, dashed for [110], and dot–dashed for [100].

$t=20M_z^{-1/2}$, broken up into 2000 points. The results are shown in Fig. 2, while the derivatives of their Fourier spectra are shown in Fig. 3. The calculation uses $\lambda^2=11/8$, $\nu=1/2$, and the values of the orientation-dependent parameters shown in Table I. A numerical analysis showed that the basic approximate equation without corrections gives oscillating falloffs of the free precession with an oscillation frequency less than the experimental value. The addition of $\Phi_4(t)$ increases the oscillation frequency, but excessively raises the amplitude of the first maximum (between the second and the third zeros). The correction $\Phi_6(t)$ made it possible to correct this distortion. The correction $\Phi_8(t)$ was also included in the [110] and [111] orientations, since the role of the complex correlations in the transfer of polarization is large in these orientations. This is reflected on the experimental falloffs of the free precession, in particular, in the inequivalence of the zeros (their approximation). One basic approximate equation gives the falloff of the free precession with equidistant zeros and a rapidly damped amplitude. Agreement with experiment can be achieved only by adding correction terms. In particular, the difference remaining at long times in Fig. 2b can be eliminated by adding $\Phi_{10}(t)$. Since the authors of Refs. 4–6 and 10 neglected polarization transfer via complex correlations and restricted themselves to the basic approximate equation, the calculated falloffs of the free precession that they obtained shows significantly worse agreement with experiment in these orientations.

Let us proceed to the results for the tail of the NMR spectrum. As can be seen from Fig. 3, the approximation chosen to describe the local-field fluctuations and expressed in Eqs. (28) and (29) correctly describes the shape of the tail and its orientational dependence. It follows from this that the damping of the tail speeds up as one goes from field orientation [111] to [110] and then to [100] mainly because the number Z of neighbors decreases. This can be explained qualitatively by noting that the field is created by Z neighbors, but it varies because of the interaction with the $Z-1$ spins. In the self-consistent approach this occurs each time more new spins are involved in the interaction with the passing of time. Consequently, the ratio for the higher-order moments can be expected to be

$$M_{2n}(Z)/[M_2(Z)]^n \sim M_{2n}(\infty)/[M_2(\infty)]^n[(Z-1)/Z]^n.$$

From this, the parameter in the exponential for the tail should be estimated as

$$\tau_0(Z) \sim \tau_0(\infty)[Z/(Z-1)]^{1/2}.$$

Figure 3 shows that the calculated tail decreases somewhat more steeply than the experimental one. This can show that the fluctuations of the longitudinal field in fact are attenuated to a smaller degree than in our calculation when the parameter $\nu=1/2$ is chosen, or rather that more neighbors should be included in the system of nonlinear equations. At the same time, we should point out that the correction terms added to Eq. (31), as shown by calculation, change the center of the spectrum, in particular the position of the maxima of the derivative, but have virtually no effect on the tail of the spectrum.

The resulting equations for the spin-system dynamics make it possible to describe other experiments as well as the NMR absorption spectra. As an example consider the experiment noted above, in which the rate of cross-polarization of an impurity of the rare isotope ^{43}Ca from the dipole–dipole reservoir of ^{19}F nuclei in a CaF_2 crystal is measured.[11] The dependence of the rate of this process on the rf field amplitude H_1 is determined by[11,31]

$$1/T_{IS}=M_{2IS}g(H_1)/\pi,$$

where M_{2IS} is the second moment at the impurity nucleus from the dipole interaction with the fluorine nuclei, and $g(H_1)$ is the spectrum of the correlation function of the longitudinal local field of Eq. (11) at the ^{43}Ca nucleus from the fluorine nuclei, normalized to unit area. As can be concluded from the values of the lattice sums,[12,31] the contribution with loops is even smaller in the [111] and [110] experimental field orientations than it was in the field at the ^{19}F nucleus. Therefore, $g(H_1)$ coincides with the spectrum of the correlation function $\Gamma_z(t)$ with high accuracy. The equation for calculating this function with the total second moment can be obtained from Eq. (29) after eliminating the division by $F_q(t')$ and adding in the brackets, in place of the subtractive term $b_q/F_q^2(t')$, the contribution $4(1-d_c)/9$ from distant spins. The functions $F_q(t)$ in this equation are calculated from the previous nonlinear equations. Because of the slow damping of $\Gamma_z(t)$, the time interval was increased to $40/M_2^{1/2}$ and broken up into 64 000 points. The results of the calculation of the spectra are shown in Fig. 4 along with the experimental data. A comparison shows that Eqs. (28) and (29) gave a good description of the cross-polarization and, consequently, of the fluctuations of the longitudinal local field. To be fair, it must be said that an equation with a Gaussian memory function[31] gave even better agreement. The reason is that the central part of the $\Gamma_z(t)$ spectrum, strongly narrowed by fluctuations, was in fact experimentally observed, as is evidenced by the large ratio of its moments, M_{4z}/M_{2z}^2. Therefore a self-consistent description of the fluctuations had no advantage over describing them by a Gaussian function, while a decrease appeared in M_{4z} because the interaction with distant spins was neglected in the nonlinear Eqs. (28)–(30).

We have thus convinced ourselves that the equations obtained here correctly describe the experiment in terms of cross-polarization and the tail of the NMR absorption line, measured by a continuous method. If we turn to the results in Fig. 1, obtained after Fourier-transforming the Engelsberg–Lowe formula, Eq. (1), for the falloffs of the free precession,

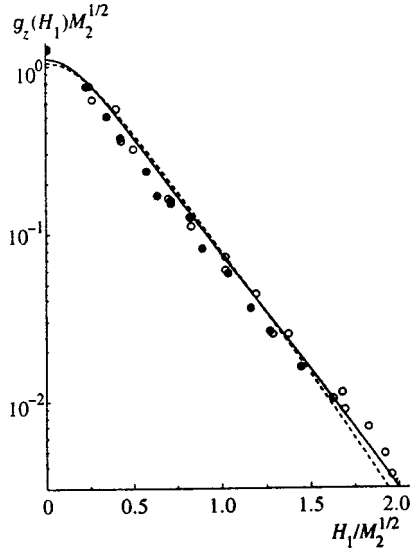


FIG. 4. Cross-polarization spectra for ^{43}Ca–^{19}F in CaF_2 for two magnetic field orientations. The experimental data of McArthur, Hahn, and Walstedt[11] are shown by the circles (open for $\mathbf{H}_0\|[111]$ and closed for $\mathbf{H}_0\|[110]$). The Fourier spectra of the correlation functions $\Gamma_z(t)$ are shown by a solid curve for $\mathbf{H}_0\|[111]$ and by a dashed curve for $\mathbf{H}_0\|[110]$.

we can conclude that this function gives a fairly good description of the tails of the spectra in the [100] and [111] orientations, but makes the decreasing tail in the [110] orientation appreciably steeper.

Let us now analyze the shape of the tail that follows from the theories cited above. The authors of Refs. 4 and 5 in general failed to take into account the fluctuations of the longitudinal local field, and therefore the tail of the NMR spectrum falls off even more steeply in their theory than does the tail of a Gaussian function. Reference 6 introduced a substantial improvement: Instead of considering the entire longitudinal local field to be unchanged, they considered the contribution to it from the close-lying spins (the spins of the cell) to be unchanged, while the field of the distant spins is described by a Gauss–Gauss random process. These changes brought the tail of the theoretical NMR spectrum closer to the experimental spectrum, but the description of the center of the spectrum became even worse. The approach in which a cell was distinguished was developed further in Ref. 9. However, since the contribution of the distant spins is introduced into the falloff of the free procession by multiplying by the exponential multiplier from the Engelsberg–Lowe function, Eq. (1), the same tail is obtained as in the spectrum of that function. Finally, Ref. 10 assumed that the longitudinal local field from all the spins fluctuated. A discontinuous Markov process is used to specify the field variations in time; this should work well for describing the changes of the NMR spectra, because of the mobility of the atoms and molecules.[25,32] This is by no means a successful approximation of the actual local-field fluctuations in a rigid lattice, since it gives a Lorentzian tail for the spectrum and consequently an infinite value for all the spectral moments, whereas they should have finite values in a rigid lattice.[25]

Another approach that does not use the concept of longitudinal local field was given in Ref. 8. In that paper, the

effect of the dipole–dipole interaction between the projections of the spins on the external magnetic field was mainly taken into account in a continued-fraction formalism. It was shown that, when the interaction between the transverse spin components is truncated in the dipole–dipole Hamiltonian, the coefficients in the continued fraction increase linearly as the number increases. When the truncated interaction is included, the increase of the coefficients accelerates. To obtain a closed expression that could be used for a calculation, the authors had to make an assumption concerning the form of this dependence. It was proposed to extrapolate the quadratic dependence on the number, established from the exact values of the first four coefficients. A similar dependence was detected earlier in the anisotropic Heisenberg model.[33] The unusual properties of continued fractions with such coefficients were discussed in Refs. 33 and 34. It is interesting for us that the tail of the spectrum is obtained as an exponential for such an approximation, in agreement with the result of our theory of a self-consistent fluctuating field. The formal transition in the continued fractions from a linear to a quadratic dependence of the coefficients on the number thereby obtained a physical explanation in our theory as a transition from constant local fields to fluctuating fields. To simplify the calculations, instead of a quadratic dependence, the same paper[8] later postulates that the coefficients, beginning with the fifteenth, are constant. In this case, a spectrum is obtained with truncated tails. The truncations, it is true, are rather far from the center.

This review of the work shows that the main advantage of the proposed theory over other theories is that, when the correlation functions are computed from self-consistent equations, it becomes unnecessary to postulate their shape or the shape of the memory function in the equations for them. Other advantages that made it possible to achieve better agreement with experiment are that the theory takes into account the finiteness of the number of nearest neighbors and polarization transfer via complex correlations. At the same time, the estimate of the attenuation of the field fluctuations still needs to be refined in order to more consistently take into account the contribution of the distant spins, as well as the contribution of complex loops.

5. CONCLUSION

Thus, both pulsed and continuous NMR studies have revealed that the tails of the spectrum have an exponential dependence. Varying the parameters of this dependence by changing the orientation of the crystal in a magnetic field does not alter the second moment of the spectrum. We explain this fact by means of nonlinear equations for the correlation functions, derived in the approximation of a self-consistent fluctuating field, taking into account the properties of the actual lattice. It has been shown that a tail of exponential shape results from self-consistent local-field fluctuations. The number of nearest neighbors and the contribution of the complex correlations change with orientation because of the anisotropy of the dipole–dipole interaction, and this changes the intensity of the local-field fluctuations and causes a dependence of the parameters of the tail in addition to that involving the second moment. On the other hand, if one restricts oneself to the approximation of constant local fields, one can arrive at the erroneous conclusion, drawn, for example, by Waugh,[35] that the spectrum will have a limit; if the rf field is detuned beyond this limit, the field ceases to heat the spin system. As shown above, this is not so. The spectrum, although exponentially weak, extends to virtually infinite frequencies. This conclusion is important for the theory of the establishment of equilibrium in spin systems.

The authors are grateful to P. P. Fedorov for providing a single crystal of BaF_2 and to V. A. Atsarkin, F. S. Dzheparov, and A. A. Lundin for discussing the results of the work.

This work was carried out with the financial support of the Krasnoyarsk Regional Science Fund (Grant 5F0068).

APPENDIX A

From the theory for computing the asymptotic forms of integrals,[36] the Fourier transform of the Engelsberg–Lowe function, Eq. (1), is determined for sufficiently large frequencies by its behavior on the imaginary time axis close to the branch point $t=iA$. In this region, we substitute the variable $t=i\tau$ and rewrite the product in the function in Eq. (1) in a new form:

$$\prod_{n=1}(1+\alpha_n^2\tau^2)=\frac{\sinh(b\tau)}{b\tau}\prod_{n=1}\frac{1+\alpha_n^2\tau^2}{1+\tau^2/(n\tau')^2}, \qquad \text{(A1)}$$

where the factors with the first nonequidistant zeros of the falloff of the free precession have been retained in the product, while the infinite product with equidistant zeros ($t_n=n\tau'$, a prime that was absent in Ref. 3 is added to prevent it from being confused with imaginary time) is collected into the function $\sin(bt)/bt$ (Ref. 3) with parameter $b=\pi/\tau'$. The product on the right-hand side of Eq. (A1) varies insignificantly on the interval $(iA, i\infty)$ of the imaginary axis of interest to us, and therefore we substitute into Eq. (A1) its value D at point $t=iA$ ($\tau=A$), which has the following values in the three orientations: 0.883 in [100], 0.514 in [110] and 0.690 in [111]. After this, the desired derivative of the spectrum is expressed in terms of the modified Bessel function of the second kind. Limiting ourselves to the first terms of the asymptotic series of this function, we get

$$\frac{d}{d\omega}g(\omega)\approx\frac{DC}{2b}\left(\frac{A}{2\pi}\right)^{1/2}\Omega^{-3/2}\exp\{A(C-\Omega)\}, \qquad \text{(A2)}$$

where

$$\Omega=[(\omega-b)^2+C^2]^{1/2}.$$

For $\omega=2M_2^{1/2}$, the value of Eq. (A2) is 15% less than the calculated spectrum of the function in Eq. (1), whereas, beginning with $2.5M_2^{1/2}$ it virtually coincides with it.

APPENDIX B

Let us determine the principal part of the solution of the system of Eqs. (28) and (29) in the neighborhood of the singular point with coordinate τ_0, using a method analogous

to the Painlevé analysis of the movable singularities of nonlinear ordinary differential equations. In order to do this, we write it in the form

$$F_q(t)\approx c_q(it+\tau_0)^{-\delta_q},\quad \Gamma_{z/q}(t)\approx a_q(it+\tau_0)^{-\zeta_q},\quad \text{(B1)}$$

substitute in Eqs. (28) and (29), and keep only the principal terms on the right-hand sides. From the condition of the equality of the left- and right-hand sides of the resulting equations, we find for the singularity indices

$$\begin{aligned}&\delta_1=\delta_2=\delta_3=2(1+\nu)/[\nu(2Z-3)-1],\\ &\zeta_1=\zeta_2=\zeta_3=(2+\delta_1)/\nu=\delta_1(2Z-3)-2,\end{aligned}\quad \text{(B2)}$$

while we get for the amplitudes a system of algebraic equations:

$$a_q\zeta_1(1+\zeta_1)=\frac{1}{2}\frac{c^2}{c_q}\left(\frac{n_1b_1}{c_1^2}+\frac{n_2b_2}{c_2^2}+\frac{n_3b_3}{c_3^2}-\frac{b_q}{c_q^2}\right),$$

$$\lambda^2 b_q a_q^{\nu}=c_q\delta_1(1+\delta_1),\quad c=c_1^{n_1}c_2^{n_2}c_3^{n_3}.$$

In the same way, for the characteristics of the function

$$\Gamma_z(t)\approx a(it+\tau_0)^{-\zeta}\quad \text{(B3)}$$

we find

$$\begin{aligned}&\zeta=\zeta_1+\delta_1=2(2Z-1+\nu)/[\nu(2Z-3)-1],\\ &a\zeta(1+\zeta)=\frac{c^2}{2}\left(\frac{n_1b_1}{c_1^2}+\frac{n_2b_2}{c_2^2}+\frac{n_3b_3}{c_3^2}\right).\end{aligned}\quad \text{(B4)}$$

Finally, if we substitute Eq. (B3) into Eq. (34), we find that the singularity in function $R_f(t)$ is stronger than in the other functions. To correct this disagreement, the power ν should be replaced by the smaller β, determined by the condition $\beta\zeta=2$. The reason is that changing the orientations of the spins far from the selected spin produces a smaller change of the local field than does the reorientation of the neighboring spins. In exactly the same way, ν should be replaced by β in the correction terms $\Phi_{2n}(t)$. We did not make these replacements in the text above, since they do not appreciably change the calculated curves in the regions under consideration.

*)E-mail: root@iph.krasnoyarsk.su

[1] V. E. Zobov, A. A. Lundin, Zh. Éksp. Teor. Fiz. **106**, 1097 (1994) [JETP **79**, 595 (1994)].
[2] C. R. Bruce, Phys. Rev. **107**, 43 (1957).
[3] M. Engelsberg and I. J. Lowe, Phys. Rev. B **10**, 822 (1974).
[4] K. W. Becker, T. Plefka, and G. Sauermann, J. Phys. C **9**, 4041 (1976).
[5] G. Sauermann and M. Wiegand, Physica B **103**, 309 (1981).
[6] G. E. Karnaukh, A. A. Lundin, B. N. Provotorov, and K. T. Summanen, Zh. Éksp. Teor. Fiz. **91**, 2229, (1986) [Sov. Phys. JETP **64**, 1324 (1986)].
[7] R. N. Shakhmuratov, J. Phys.: Condens. Matter **3**, 8683 (1991).
[8] J. Jensen, Phys. Rev. B **52**, 9611 (1995).
[9] A. A. Lundin, Zh. Éksp. Teor. Fiz. **110**, 1378 (1996) [JETP **83**, 759 (1996)].
[10] B. N. Provotorov, T. P. Kulagina, and G. E. Karnaukh, Zh. Éksp. Teor. Fiz. **113**, 967 (1998) [JETP **86**, 527 (1998)].
[11] D. A. McArthur, E. L. Hahn, and R. E. Walstedt, Phys. Rev. **188**, 609 (1969).
[12] H. T. Stokes and D. C. Ailion, Phys. Rev. B **15**, 1271 (1977).
[13] C. Cusumano and G. J. Troup, Phys. Status Solidi B **65**, 655 (1974).
[14] V. A. Atsarkin, G. A. Vasneva, and V. V. Demidov, Zh. Éksp. Teor. Fiz. **91**, 1523 (1986) [Sov. Phys. JETP **64**, 898 (1986)].
[15] M. Blume and J. Hubbard, Phys. Rev. B **1**, 3815 (1970).
[16] V. E. Zobov, Preprints 514F and 518F, Inst. Fiz. Sib. Otdel. Ross. Akad. Nauk, Krasnoyarsk (1988).
[17] V. E. Zobov, Teor. Mat. Fiz. **77**, 426 (1988); **84**, 111 (1990).
[18] A. A. Lundin, A. V. Makarenko, and V. E. Zobov, J. Phys.: Condens. Matter **2**, 10 131 (1990).
[19] V. E. Zobov, M. A. Popov, Teor. Mat. Fiz. **112**, 479 (1997).
[20] R. W. G. Wyckoff, *Crystal Structures*, vol. 1 (Wiley, New York, 1963).
[21] É. I. Fedin, I. K. Shmyrev, and A. I. Kitaĭgorodskiĭ, in *Paramagnetic Resonance* (Nauka, Moscow, 1971), p. 283.
[22] G. W. Canters and C. S. Johnson, Jr., J. Magn. Reson. **6**, 1 (1972).
[23] E. R. Andrew, Phys. Rev. **91**, 425 (1953).
[24] K. W. Vollmers, I. J. Lowe, and M. Punkkinen, J. Magn. Reson. **30**, 33 (1978).
[25] A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon Press, Oxford, 1961; Inostr. Lit., Moscow, 1963).
[26] P. Borckmans and D. Walgraef, Physica (Amsterdam) **35**, 80 (1967); Phys. Rev. **167**, 282 (1968).
[27] P. Borckmans and D. Walgraef, Phys. Rev. B **7**, 563 (1973).
[28] M. I. Bulgakov, A. D. Gul'ko, F. S. Dzheparov *et al.*, JETP Lett. **58**, 592 (1993).
[29] P. W. Anderson and P. R. Weiss, Rev. Mod. Phys. **25**, 269 (1953).
[30] A. Lösche, *Kerninduktion* (Deutsche Verlag der Wissenschaften, Berlin, 1957; Inostr. Lit., Moscow, 1963).
[31] D. E. Demco, J. Tegenfeldt, and J. S. Waugh, Phys. Rev. B **15**, 4133 (1975).
[32] V. E. Zobov, in *The rf Spectroscopy of Solids*, Part 1 (Inst. Fiz. Sib. Otdel. Ross. Akad. Nauk, Krasnoyarsk, 1974), p. 22.
[33] M. Böhm, H. Leschke, M. Henneke *et al.*, Phys. Rev. B **49**, 417 (1994).
[34] M. H. Lee, J. Kim, W. P. Cummings, and R. Dekeyser, J. Phys.: Condens. Matter **7**, 3187 (1995).
[35] J. S. Waugh, in *NMR and More in Honour of Anatole Abragam*, Ed. M. Goldman and M. Porneuf (Les Editions de Physique Les Ulis, France,1994), p. 353.
[36] M. V. Fedoryuk, *Asymptotics: Integrals and Series* (Nauka, Moscow, 1987).

Translated by W. J. Manthey

Two-photon interaction of coherent radiation with a thin film of resonant atoms

A. M. Basharov, A. I. Maĭmistov,*) and S. O. Elyutin

Moscow State Engineering-Physics Institute (Technical University), 115409 Moscow, Russia
(Submitted 21 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 30–42 (January 1999)

The problem of the interaction of an ultrashort optical pulse and a thin film of resonant atoms under the conditions of two-photon absorption, third-harmonic generation, and the inverse effect of the latter on the pump pulse via Raman scattering is studied. The fact that the field acting on an atom differs from the macroscopic field in the film is also taken into consideration. It is shown that the polarization of the film undergoes dynamic relaxation even in the absence of irreversible relaxation, suppressing Rabi oscillations and establishing stationary values of the populations of the resonant energy levels and of the polarization of the film at the pump and the third-harmonic frequencies. © *1999 American Institute of Physics.* [S1063-7761(99)00301-7]

1. INTRODUCTION

Low-dimensional systems have been attracting a great deal of attention for many years. A simple example of such a system is a thin film, whose thickness is less than the wavelength of optical radiation, on the interface between two dielectric media. The first investigations of nonlinear phenomena in this system indicated the possibility of coherent propagation of an optical pulse of a surface wave along the interface, containing a thin film of two-level atoms, between dielectrics[1] and established the conditions of reflection and refraction of a plane wave passing through such an interface.[2–4] However, the models considered neglected the corrections due to the Lorentz field,[5] which are large precisely in the case of a thin film,[6] and for this reason they did not reflect the true picture of the interaction between the field of an ultrashort optical pulse and a nonlinear interface.[7–9]

Many other interesting phenomena are possible in a thin film besides nonlinear surface waves. There are numerous treatments of the optical bistability accompanying diffraction and reflection of an electromagnetic wave by a thin film of resonant two-level atoms. It has been shown that a thin film of resonant atoms can be described by the same system of equations that arises in the model of a nonlinear Fabry–Perot resonator.[10] Therefore it is natural to expect optical bistability and self-pulsation phenomena in the reflection of an ultrashort pulse from a thin film of resonant atoms.[8,11–14] Optical bistability, taking account of the finite thickness of the substrate, has been studied in Refs. 10 and 15–17. It was noted that in this case there exists an analogy with the problem of the passage of a wave through a system of optical resonators.

The parametric interaction of several waves is a typical example of nonlinear optical phenomena. Three-wave interaction in a thin film of a nonlinear dielectric on the interface between two linear dielectrics was studied in Ref. 18. It was shown that coherent responses, specifically, photon echoes[19] and superluminescent pulses,[20,21] can be generated in a thin film of resonant two-level atoms in response to an external pulse.

Two-photon resonance in a multilevel medium or double resonance in a three-level medium is a simple generalization of the model of a resonant medium forming a thin film on an interface. The case of double resonance has been studied only in Ref. 4, but the Lorentz field was neglected in the model investigated. Two-photon resonance was studied in Refs. 22 and 23. In Ref. 23, in contrast to Ref. 22, the passage of light through a thin film of resonant atoms was studied taking account of the local field, and it was shown that optical bistability can appear in the quasistationary state.

It is known[24–26] that if a monochromatic wave interacts with a medium under two-photon resonance conditions, then a response at the frequency of this wave and at the tripled frequency of the wave necessarily arises in such a medium. Such third-harmonic generation also occurs when quasimonochromatic waves, which correspond to optical pulses, are used. The basic problem studied in the present paper is to take account of correctly two-photon absorption and third-harmonic generation as well as the inverse of these processes in a thin film on the interface between linear dielectrics.

The problem is formulated in Sec. 1. A distinguishing feature of the formulation is that third-harmonic generation and its inverse effect on the interaction with the atoms in the film are taken into account. The effective Hamiltonian for the problem under study is obtained in Sec. 2. The Bloch and coupling equations are formulated in Secs. 3 and 4. The latter equations express in terms of the incident field and the parameters of the medium the local field acting on an atom. An example of the application of the approach developed for taking account of third-harmonic generation and its inverse effect on the passage of an ultrashort pulse is discussed in Sec. 5.

2. THIN FILM ON THE INTERFACE BETWEEN MEDIA

Let a thin film of atoms interacting resonantly with the electromagnetic field of a light wave be present on the interface between two dielectric media in the $x=0$ plane. The dielectric media surrounding the film are characterized by the permittivities ε_1 for $x<0$ and ε_2 for $x>0$. The z axis is

chosen to lie in the plane of the interface. The resonant atoms are described by an effective-Hamiltonian model.[25,26] The duration of the light pulse is assumed to be short compared with the relaxation times of the polarization and the population difference but much longer than the optical period, so that the approximation of slowly varying complex envelopes of such ultrashort pulses would be applicable. On account of the presence of a planar interface the system of Maxwell's equations decomposes into two independent systems describing *TE* waves,

$$\mathbf{E}=(0,E_y,0), \quad \mathbf{H}=(H_x,0,H_z)$$

and *TM* waves,

$$\mathbf{E}=(E_x,0,E_z), \quad \mathbf{H}=(0,H_y,0).$$

The film thickness l is assumed to be much shorter than the wavelength of the resonant radiation.

Let us consider an ultrashort TE pulse incident on the interface from the $x<0$ region. The reflected wave propagates back into the $x<0$ region, and the refracted wave propagates into the $x>0$ region. The passage of an ultrashort pulse through the interface will be studied as in Refs. 2 and 3.

It is convenient to represent the field intensities **E** and **H** and the polarization **P** of the resonant atoms inside the thin film as

$$\mathbf{E}(x,z,t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{d\beta}{2\pi}\exp(-i\omega t+i\beta z)\tilde{\mathbf{E}}(x,\beta,\omega),$$

$$\mathbf{H}(x,z,t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{d\beta}{2\pi}\exp(-i\omega t+i\beta z)\tilde{\mathbf{H}}(x,\beta,\omega),$$

$$\mathbf{P}(z,t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{d\beta}{2\pi}\exp(-i\omega t+i\beta z)\tilde{\mathbf{P}}(\beta,\omega).$$

Outside the film the Fourier components $\tilde{\mathbf{E}}(x,\beta,\omega)$ and $\tilde{\mathbf{H}}(x,\beta,\omega)$ of the field vectors are determined by Maxwell's equations, and the components at $x=0$ are determined by continuity so that for the *TE* case under consideration we obtain the system of equations

$$\frac{d^2\tilde{E}}{dx^2}+(k^2\varepsilon_j-\beta^2)\tilde{E}=0, \tag{1a}$$

$$\tilde{H}_z=-\frac{\beta}{k}\tilde{E}, \quad \tilde{H}_z=-\frac{i}{k}\frac{d\tilde{E}}{dx}, \quad \tilde{E}=\tilde{E}_y$$

with the boundary conditions

$$\tilde{E}(x=0-)=\tilde{E}(x=0+),$$

$$\tilde{H}_z(x=0+)-\tilde{H}_z(x=0-)=4\pi ik\tilde{P}_y(\beta,\omega). \tag{1b}$$

Here $j=1,2$ and $k=\omega/c$. Outside the thin film the solution of Eq. (1a), taking account of the behavior of the field far from the film, has the form

$$\tilde{E}(x,\beta,\omega)=\begin{cases}A(\beta,\omega)\exp(iq_1x)+B(\beta,\omega)\exp(-iq_1x),\\ x<0,\\ C(\beta,\omega)\exp(iq_2x),\\ x>0,\end{cases}$$

and

$$\tilde{H}(x,\beta,\omega)=\begin{cases}q_1k^{-1}\{A(\beta,\omega)\exp(iq_1x)-B(\beta,\omega)\exp(-iq_1x)\},\\ x<0,\\ q_2k^{-1}C(\beta,\omega)\exp(iq_2x),\\ x>0,\end{cases}$$

where $q_j=\sqrt{k^2\varepsilon_j-\beta^2}$, $j=1,2$. The boundary conditions (1b) at $x=0$ give relations between the amplitudes of the incident A, reflected B, and refracted C waves and the polarization $P_s=\tilde{P}_y$ of the film

$$C(\beta,\omega)=\frac{2q_1}{q_1+q_2}A(\beta,\omega)+i\frac{4\pi k^2}{q_1+q_2}P_s(\beta,\omega),$$

$$B(\beta,\omega)=\frac{q_1-q_2}{q_1+q_2}A(\beta,\omega)+i\frac{4\pi k^2}{q_1+q_2}P_s(\beta,\omega). \tag{2}$$

It is convenient to introduce notation for the (Fresnel) transmission coefficient T and the coupling constant κ:

$$T(\beta,\omega)=\frac{2q_1}{q_1+q_2}, \quad \kappa(\beta,\omega)=\frac{4\pi k^2}{q_1+q_2}.$$

We shall now focus our attention on the refracted wave. We shall consider only the case $\varepsilon_1<\varepsilon_2$, where total internal reflection does not occur for any angle of incidence $\theta_{in}=\cos^{-1}(q_1/k\sqrt{\varepsilon_1})$ and the refraction angle θ_{tr} is determined by Snell's law

$$\sin\theta_{tr}=\beta/k\sqrt{\varepsilon_2}=\sqrt{\varepsilon_1/\varepsilon_2}\sin\theta_{in}.$$

If the polarization of the film is determined, then the expressions (2) determine the field in all space. We emphasize that they are in no way related with the assumption of slowly varying envelopes of the optical pulses and are exact. To find the polarization of the atoms in the film a model of the resonant system must be specified. In addition, let the envelopes of the optical pulses and the polarization of the atoms vary slowly. The optical pulses are assumed to be ultrashort.

3. EFFECTIVE HAMILTONIAN AND POLARIZATION OF A THIN FILM

Let us consider two-photon resonance, where the doubled frequency of the carrying wave of the ultrashort pulse equals approximately the atomic transition frequency, but the transition itself is dipole-forbidden. It was shown in Refs. 24–26 that in this case the wave with the carrying frequency generates polarization of the resonant medium at the frequencies ω_0 and $3\omega_0$. For this reason, if we takeaccount of the inverse effect of the medium on the atoms in the film, two fields with intensity

$$E_{a1}=\mathscr{A}_1\exp(-i\Phi_0)+\text{c.c.},$$
$$E_{a3}=\mathscr{A}_3\exp(-3i\Phi_0)+\text{c.c.}, \tag{3}$$
$$\Phi_0=\omega_0 t-\beta z,$$

with amplitudes $\mathscr{A}_1$ and $\mathscr{A}_2$, and carrying frequencies ω_0 and $3\omega_0$ act on the atoms in the film, and the two-photon resonance conditions

$$2\omega_0\approx\omega_{ca},$$

where $\omega_{ca}\approx(E_c-E_a)/\hbar$ is the transition frequency between the energy levels E_a and E_c, are satisfied.

We now obtain an expression for the atomic polarization induced in the film by a quasimonochromatic wave with carrying frequencies ω_0 and $3\omega_0$. The atomic polarization is given by the standard formula

$$P(t,z)=N_0\mathrm{Tr}(\rho d), \tag{4}$$

where N_0 is the surface density of the atoms, and the density matrix ρ of the atoms satisfies the standard equation

$$i\hbar\frac{d\rho}{dt}=[H_0-E_{\text{atom}}d,\rho]. \tag{5a}$$

Here H_0 is the Hamiltonian of a resonant atom in the film, d is the dipole moment operator of the atom, and $E_{\text{atom}}=E_{a1}+E_{a2}$ is the intensity of the electric field acting on an atom. The square brackets denote a commutator.

The resonant energy levels E_a and E_c of the Hamiltonian and the nonresonant energy levels E_σ are eigenvalues of the Hamiltonian H_0:

$$H_0|\alpha\rangle=E_\sigma|\alpha\rangle,\quad \alpha=a,c,\sigma,\sigma',\quad E_c-E_a\approx 2\omega_0.$$

To write down a system of equations describing only resonant levels and to construct the effective interaction Hamiltonian and the effective dipole moment operator of an atom, we shall follow the approach described in Ref. 26. We use the unitary operator $\exp(iS)$ to transform the atomic density matrix as

$$\tilde{\rho}=e^{-iS}\rho e^{iS}.$$

The equation for the transformed density matrix

$$i\hbar\frac{\partial}{\partial t}\tilde{\rho}=[\tilde{H},\tilde{\rho}] \tag{5b}$$

is determined by the Hamiltonian

$$\tilde{H}=e^{-iS}H_0e^{iS}-e^{-iS}E_{\text{atom}}de^{iS}-i\hbar e^{-iS}\frac{\partial}{\partial t}e^{iS},$$

which we expand in the standard manner

$$\tilde{H}=H_0-i[S,H_0]-\frac{1}{2}[S,[S,H_0]]-\ \ldots\ -E_{\text{atom}}d$$
$$+i[S,E_{\text{atom}}d]+\frac{1}{2}[S,[S,E_{\text{atom}}d]]$$
$$+\ \ldots\ -i\hbar e^{-iS}\frac{\partial}{\partial t}e^{iS}.$$

We represent S and $\tilde{H}$ as series expansions in powers of the electric field intensity

$$S=S^{(1,1)}+S^{(2,0)}+S^{(0,2)}+\ \ldots,$$
$$\tilde{H}=\tilde{H}^{(0,0)}+\tilde{H}^{(1,0)}+\tilde{H}^{(0,1)}+\tilde{H}^{(2,0)}+\ \ldots \tag{6}$$

($S^{(n,m)}$ and $\tilde{H}^{(n,m)}$ are terms of order n in the field E_{a1} and order m in the field E_{a2}). Then

$$\tilde{H}^{(0,0)}=H_0,$$

$$\tilde{H}^{(1,0)}=-E_{a1}d-i[S^{(1,0)},H_0]+\hbar\frac{\partial}{\partial t}S^{(1,0)},$$

$$\tilde{H}^{(0,1)}=-E_{a2}d-i[S^{(0,1)},H_0]+\hbar\frac{\partial}{\partial t}S^{(0,1)},$$

$$\tilde{H}^{(2,0)}=\frac{i}{2}[S^{(1,0)},E_{a1}d]-\frac{i}{2}[S^{(1,0)},\tilde{H}^{(1,0)}]$$
$$-i[S^{(2,0)},H_0]+\hbar\frac{\partial}{\partial t}S^{(2,0)},$$

$$\tilde{H}^{(1,1)}=\frac{i}{2}[S^{(1,0)},E_{a2}d]+\frac{i}{2}[S^{(0,1)},E_{a1}d]$$
$$-\frac{i}{2}[S^{(0,1)},\tilde{H}^{(1,0)}]-\frac{i}{2}[S^{(1,0)},\tilde{H}^{(0,1)}]$$
$$-i[S^{(1,1)},H_0]+\hbar\frac{\partial}{\partial t}S^{(1,1)},$$

$$\tilde{H}^{(0,2)}=\frac{i}{2}[S^{(0,1)},E_{a2}d]-\frac{i}{2}[S^{(0,1)},\tilde{H}^{(0,1)}]$$
$$-i[S^{(0,2)},H_0]+\hbar\frac{\partial}{\partial t}S^{(0,2)},$$

$\ldots$

Let $\tilde{H}^{(1,0)}=\tilde{H}^{(0,1)}=0$. Hence, and from the condition that the field is switched on adiabatically, we find the following expressions for the matrix elements of the hermitian operators:

$$S^{(1,0)}_{aa'}=-\frac{id_{aa'}}{\hbar}\left(\frac{\mathscr{A}_1e^{-i\Phi_0}}{\omega_{aa'}-\omega_0}+\frac{\mathscr{A}_1^*e^{i\Phi_0}}{\omega_{aa'}+\omega_0}\right), \tag{7}$$

$$S_{aa'}^{(0,1)}=-\frac{id_{aa'}}{\hbar}\left(\frac{\mathscr{A}_3 e^{-i3\Phi_0}}{\omega_{aa'}-3\omega_0}+\frac{\mathscr{A}_3^* e^{i3\Phi_0}}{\omega_{aa'}+3\omega_0}\right).$$

We take as the effective Hamiltonian

$$H^{\text{eff}}=\tilde{H}^{(0,0)}+\tilde{H}^{(2,0)}+\tilde{H}^{(1,1)}+\tilde{H}^{(0,2)}.$$

Following Ref. 26, it is not difficult to obtain elements of the effective Hamiltonian matrix in the form

$$H_{ca}^{\text{eff}}=-\left[\frac{1}{2}\mathscr{A}_1^2\Pi_{ca}(\omega_0)+\mathscr{A}_3\mathscr{A}_1^*\Pi_{ca}(-\omega_0)\right]e^{-i2\Phi_0}$$
$$=H_{ca}^{\text{eff}*},\quad H_{\alpha\alpha}^{\text{eff}}=E_\alpha+E_\alpha^{St},\tag{8}$$
$$E_\alpha^{St}=|\mathscr{A}_1|^2\Pi_\alpha(\omega_0)+|\mathscr{A}_3|^2\Pi_\alpha(3\omega_0),\quad \alpha=a,c,\sigma,$$

where

$$\Pi_{ca}(\omega)=\sum_\sigma \frac{d_{c\sigma}d_{\sigma a}}{\hbar}\left(\frac{1}{\omega_{\sigma c}+\omega}+\frac{1}{\omega_{\sigma a}-\omega}\right),$$

$$\Pi_\alpha(\omega)=\sum_{\alpha'} \frac{|d_{\alpha\alpha'}|^2}{\hbar}\left(\frac{1}{\omega_{\alpha\alpha'}+\omega}+\frac{1}{\omega_{\alpha\alpha'}-\omega}\right).$$

The polarization of the film can be expressed in terms of the transformed density matrix $\tilde{\rho}$ and the effective dipole moment operator D

$$P=N_0\text{Tr}(\rho d)=N_0\text{Tr}(e^{iS}\tilde{\rho}e^{-iS}d)=N_0\text{Tr}\tilde{\rho}D,$$

$$D=e^{-iS}de^{iS}=d-i[S,d]-\frac{1}{2}[S,[S,d]].$$

Taking account of Eq. (7), we obtain the following expressions for the matrix elements of the effective dipole moment:

$$D_{ac}=\mathscr{A}_1\Pi_{ca}^*(-\omega_0)e^{-i\Phi_0}+\mathscr{A}_1^*\Pi_{ca}^*(\omega_0)e^{i\Phi_0}$$
$$+\mathscr{A}_3\Pi_{ca}^*(-3\omega_0)e^{-i3\Phi_0}+\mathscr{A}_3^*\Pi_{ca}^*(3\omega_0)e^{i3\Phi_0},$$
$$D_{aa}=-\mathscr{A}_1\Pi_a(\omega_0)e^{-i\Phi_0}-\mathscr{A}_3\Pi_a(3\omega_0)e^{-i3\Phi_0}+\text{c.c.},\tag{9}$$
$$D_{cc}=-\mathscr{A}_1\Pi_c(\omega_0)e^{-i\Phi_0}-\mathscr{A}_3\Pi_c(3\omega_0)e^{-i3\Phi_0}+\text{c.c.}$$

To calculate the polarization of the medium we shall neglect the nonresonant terms as well as the terms proportional to $\mathscr{A}_3$ in the effective dipole moment operator, making the assumption that

$$|\mathscr{A}_3|\ll|\mathscr{A}_1|.\tag{10}$$

Then, expressing the elements of the density matrix in terms of the slowly varying amplitudes we obtain

$$P(t,z)=\mathscr{P}_1(t,z)\exp(-i\Phi_0)+\mathscr{P}_3(t,z)$$
$$\times\exp(-3i\Phi_0)+\text{c.c.},$$
$$\mathscr{P}_1(t,z)=-(\Pi_a(\omega_0)\tilde{\rho}_{aa}+\Pi_c(\omega_0)\tilde{\rho}_{cc})\mathscr{A}_1$$
$$+\Pi_{ca}^*(\omega_0)R\mathscr{A}_1^*,$$
$$\mathscr{P}_3(t,z)=\Pi_{ca}^*(-\omega_0)R\mathscr{A}_1,\tag{11}$$

$$R=\tilde{\rho}_{ca}\exp(2i\Phi_0).$$

We introduce the variable $N=\tilde{\rho}_{aa}-\tilde{\rho}_{cc}$. It will be shown below that $\tilde{\rho}_{aa}+\tilde{\rho}_{cc}=$ const. Choosing the constant to be 1, the envelope of the polarization at the frequency ω_0 can be expressed as

$$\mathscr{P}_1(t,z)=-(\Pi_+(\omega_0)+\Pi_-(\omega_0)N)\mathscr{A}_1$$
$$+\Pi_{ca}^*(\omega_0)R\mathscr{A}_1^*,\tag{12}$$

where

$$\Pi_+(\omega_0)=\frac{1}{2}(\Pi_a(\omega_0)+\Pi_c(\omega_0)),$$

$$\Pi_-(\omega_0)=\frac{1}{2}(\Pi_a(\omega_0)-\Pi_c(\omega_0)).$$

4. THE BLOCH EQUATIONS

The equations for the elements $\tilde{\rho}_{aa},\tilde{\rho}_{cc}$, and $\tilde{\rho}_{ca}=\tilde{\rho}_{ac}^*=R\exp(-2i\Phi_0)$ of the transformed density matrix $\tilde{\rho}$ form a closed system:

$$\frac{\partial\tilde{\rho}_{aa}}{\partial t}=i(\Lambda^*R-\Lambda R^*),$$

$$\frac{\partial\tilde{\rho}_{cc}}{\partial t}=-i(\Lambda^*R-\Lambda R^*),$$

$$\left(\frac{\partial}{\partial t}-i\Delta\right)R=i\Lambda(R_{aa}-R_{cc}),$$

where

$$\Lambda=\left[\frac{1}{2}\mathscr{A}_1^2\Pi_{ca}(\omega_0)+\mathscr{A}_3\mathscr{A}_1^*\Pi_{ca}(-\omega_0)\right]\hbar^{-1},$$
$$\Delta=2\omega_0-\omega_{ca}-\{|\mathscr{A}_1|^2[\Pi_c(\omega_0)-\Pi_a(\omega_0)]\tag{13}$$
$$+|\mathscr{A}_3|^2[\Pi_c(3\omega_0)-\Pi_a(3\omega_0)]\}\hbar^{-1}.$$

One can see that $\tilde{\rho}_{aa}+\tilde{\rho}_{cc}=$const in the effective Hamiltonian approximation and neglecting relaxation, so that in tersm of the variables R and N we have

$$\frac{\partial R}{\partial t}=i\Delta R+i\Lambda N,\quad \frac{\partial N}{\partial t}=2i(\Lambda^*R-\Lambda R^*).\tag{14}$$

5. THE COUPLING EQUATIONS

The intensity E_{atom} of the electric field acting on an atom is determined by the field E_f in the film and by the polarization P of the medium:

$$E_{\text{atom}}=E_f+\xi P,$$

where the parameter ξ accounts for the environment. It is often assumed that $\xi=4\pi/3$ for an isotropic medium. If we take into account third-harmonic generation, this expression assumes the following form in terms of slowly varying variables:

$$\mathscr{A}_1=\mathscr{E}_{f1}+\xi\mathscr{P}_1,\quad \mathscr{A}_3=\mathscr{E}_{f3}+\xi\mathscr{P}_3.$$

We neglect other harmonics (fifth and so on) and the spread in the parameter ξ.

We shall obtain from Eq. (2) relations for the slowly varying amplitudes. Neglecting the dispersion of the linear media we have

$$\mathscr{E}_{f1}(t)=T(\beta_0,\omega_0)\mathscr{E}_{in}(t)+i\kappa(\beta_0,\omega_0)\mathscr{P}_1(t),$$
$$\mathscr{E}_{f3}(t)=i\kappa(3\beta_0,3\omega_0)\mathscr{P}_3(t). \tag{15}$$

Using the expression (12) for the polarization of the medium at the pump frequency, we obtain the following dependence of the amplitude of the field acting on an atom on the amplitude of the incident field and the parameters of the medium:

$$\mathscr{A}_1=T(\beta_0,\omega_0)U^{-1}\{\mathscr{E}_{in}+(\xi-i\kappa)(\Pi_+(\omega_0)+\Pi_-(\omega_0)N)\mathscr{E}_{in}+(\xi+i\kappa)\Pi^*_{ca}(\omega_0)R\mathscr{E}^*_{in}\}, \tag{16}$$

where

$$U=(1-\xi(\Pi_+(\omega_0)+\Pi_-(\omega_0)N))^2+\kappa^2(\Pi_+(\omega_0)+\Pi_-(\omega_0)N)^2-(\xi^2+\kappa^2)|\Pi_{ca}(\omega_0)R|^2.$$

The standard condition of phase matching for bulk media will appear in the present problem as a rule determining the angles $\theta^{3\omega}_{ref}$, $\theta^{3\omega}_{tr}$, and $\theta'^{3\omega}_{tr}$ for which the harmonic wave will propagate into the medium surrounding the thin film. The relations for these angles follow from the requirement $\beta(3\omega_0)=3\beta(\omega_0)$:

$$n_1(3\omega_0)\sin\theta^{3\omega}_{ref}=n_1(\omega_0)\sin\theta^{\omega}_{in},$$
$$n_2(3\omega_0)\sin\theta^{3\omega}_{tr}=n_1(\omega_0)\sin\theta^{\omega}_{in}, \tag{17a,b}$$
$$n_2(3\omega_0)\sin\theta'^{3\omega}_{tr}=n_2(\omega_0)\sin\theta^{\omega}_{tr}. \tag{17c}$$

6. PROPERTIES OF THE REFRACTION OF AN ULTRASHORT PULSE

Let the elements of the matrix $\Pi_{jk}(\omega)$ be real (otherwise, their constant phases can be included in the phase of the slowly varying amplitude of the density matrix). We neglect the dispersion of these quantities and write

$$\frac{\Pi_{cc}(\omega)}{\Pi_{aa}(\omega)}=k,\quad \frac{\Pi_{aa}(\omega)}{\Pi_{cc}(\omega)}=2m.$$

Introducing the normalized amplitude A_0 of the optical pulses and the characteristic time $t_0=2\hbar(\Pi_{ac}A_0^2)^{-1}$ so that

$$\mathscr{E}_1=A_0a_1,\quad \mathscr{E}_2=A_0a_2,\quad \mathscr{E}_{in}=A_0a_{in},\quad t=t_0\tau,$$

the Bloch equations and the coupling equations can be written in the normalized form

$$\frac{\partial\sigma}{\partial\tau}=i\Delta\omega t_0\sigma+4im(k-1)(|a_1|^2+|a_3|^2)\sigma+iF\rho, \tag{18a}$$

$$\frac{\partial\rho}{\partial\tau}=2i(\sigma F^*-\sigma^*F), \tag{18b}$$

and

$$a_1(\tau)=T(\beta_0,\omega_0)a_{in}(\tau)+g(\alpha+i\beta)\times\{-m(k+1)a_1+m(k-1)\rho a_1+\sigma a_1^*+\nu\sigma^*a_3\}, \tag{19a}$$

$$a_3(\tau)=g(\alpha+3i\beta)\{-m(k+1)a_3+m(k-1)\rho a_3+\sigma a_1\}, \tag{19b}$$

where $F=a_1^2+2\gamma a_1^*a_3$ and $\gamma\approx1$, $g=4\pi n_A|\Pi_{ac}|\sqrt{\varepsilon_1}$. Here γ and ν are the ratios of the elements of the matrix $\Pi_{\alpha'\alpha''}(\omega)$, which differ only by the value of the argument. These ratios are approximately 1, since the dispersion of $\Pi_{\alpha'\alpha''}(\omega)$ is neglected. The parameters α and β are given by

$$a=\frac{\xi\sqrt{\varepsilon_1}}{3},\quad \beta=\frac{(\omega/c)l}{\cos\theta+\sqrt{\cos^2\theta+(\varepsilon_2-\varepsilon_1)/\varepsilon_1}},$$

and the transmission coefficient is

$$T(\beta_0,\omega_0)=\frac{2\cos\theta}{\cos\theta+\sqrt{\cos^2\theta+(\varepsilon_2-\varepsilon_1)/\varepsilon_1}},$$

where the angle of incidence θ is employed.

Certain sources in the literature give

$$\Pi_{ac}\approx1.4\times10^{-24}\text{ emu},\quad \Pi_{aa}\approx2\times10^{-24}\text{ emu},$$
$$\Pi_{cc}\approx4\times10^{-24}\text{ emu},$$

so that $k\approx2$ and $m\approx1.5-1.4$. The parameter g depends on the density of resonant atoms and is 1.76×10^{-3} emu, if $n_A=10^{20}$ cm^{-3}. In the subsequent numerical calculations the refractive indices of the media surrounding the film were held constant, $\varepsilon_1=1$ and $\varepsilon_2=2.25$. The parameters in Eqs. (18) and (19) were taken as $\xi=1$, $g=0.5$, $2\pi(l/\lambda)=0.05$, $k=2$, and $m=1$.

Let us consider first an ultrashort pulse, having a duration of the order of one period of the Rabi oscillations, incident normally on a thin film. To solve the system of equations (18) and (19) numerically it was assumed that the normalized envelope of the incident pulse is described by the function

$$a_{in}(\tau)=a_0\operatorname{sech}\left(\frac{\tau-\tau_m}{\tau_p}\right), \tag{20}$$

where a_0 is the amplitude, τ_p is the normalized duration ($\tau_p=t_p/t_0$), and τ_m is the normalized time shift of this ultrashort pulse.

The numerical solution of Eqs. (18) and (19) showed that the shape of the transmitted pump pulse is the same as that of the incident ultrashort pulse. The transmission coefficient differs by less than 1% from the coefficient calculated using the Fresnel formula. This is because the shift in the resonant transition frequency due to the high-frequency Stark effect and the Lorentz field drive the system out of resonance, and we observe a weak perturbation of the atoms in the thin film and, in consequence, a negligibly small response of the atoms. Here it should be noted that for a two-photon resonance (in contrast to a one-photon resonance) the high-frequency Stark effect plays the dominant role in this

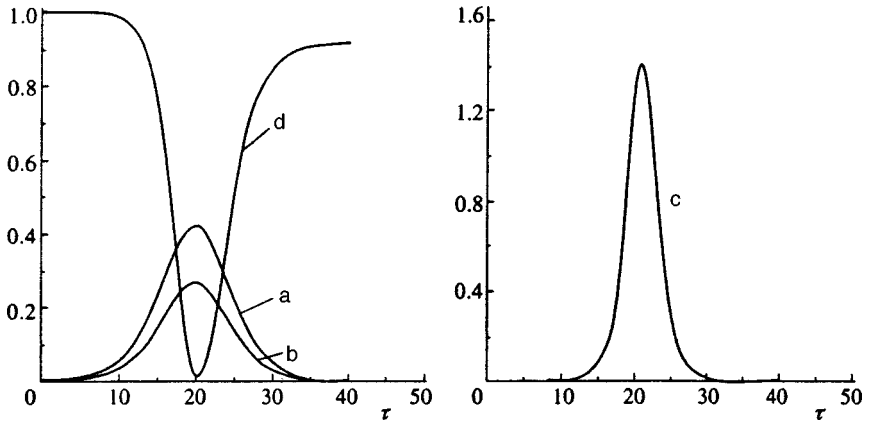


FIG. 1. Envelopes of the incident pulse (a), transmitted pulse (b), harmonic pulse (c), and population difference (d). The pump pulse had the form (20) with $a_0=0.65, \tau_p=6$, and $\tau_m=20$.

dynamic destruction of resonance. In order of magnitude the harmonic signal is 10^{-2} times the pump amplitude (Fig. 1). Increasing the amplitude of the incident signal does not change the Fresnel character of the reflection of the pulse. The harmonic signal acquires an oscillatory character (Fig. 2), attesting to rapid evolution of the Bloch vector (the normalized effective field in the Bloch equations, which determines the angular rotation rate of the Bloch vector, can be estimated as $F\sim a_1^2$).

Let us now consider a normally incident ultrashort pulse persisting much longer than the period of the Rabi oscillations. Here, in solving numerically the system of equations (18) and (19) it was assumed that the normalized envelope of the incident pulse has the form of a ''plateau''

$$a_{in}(\tau)=a_0\left[\tanh\left(2\frac{\tau-\tau_m}{\tau_p}\right)-\tanh\left(2\frac{\tau-\tau_m-\tau_w}{\tau_p}\right)\right], \quad (21)$$

where $\tau_w=t_{\text{width}}/t_0$ is the normalized width of the ''plateau.'' In the calculations the slopes of the edges of this pulse were fixed.

A characteristic example of the numerical calculation is displayed in Fig. 3. Once again, the transmitted signal had the same shape as the incident ultrashort pulse but was weaker than the latter in accordance with the Fresnel formulas. This case differs considerably from the preceding one in that the oscillations of the populations of the atoms in the film decay and the population reaches a stationary value different from the equilibrium value. It is important to emphasize that here the irreversible relaxation times in the system are once again much longer than the duration of the ultrashort pulse. This behavior becomes understandable if the Bloch equation is rewritten in the form

$$\frac{\partial R}{\partial t}\sim i(\Delta+\mathrm{Re}(f)N)R+i(g_1+g_2N)N-\mathrm{Im}(f)NR, \quad (22)$$

where f, g_1, and g_2 are certain functions that depend on the incident field and the parameters of the medium according to the general equations (13), (14), and (16). It is evident that the Lorentz field and the characteristic features of the two-photon interaction together produce an effective relaxation mechanism whereby the polarization and the population difference of the atoms in the film reach stationary values. This is seen especially clearly in the shape of the third-harmonic signal (Fig. 4): The amplitude modulation near the leading edge of the third harmonic is replaced by stationary harmonic generation. We emphasize once again that in the case considered the phase matching determines the direction of emission of the harmonic from the film but in no way influences the harmonic generation efficiency.

This behavior of a normally incident ultrashort pulse is manifested with increasing angle of incidence of the fundamental wave on the interface containing a thin film of resonant atoms, between two dielectric media. Since the effective field penetrating into the film decreases with increasing angle of incidence, the period of the Rabi oscillations changes, so that such a pulse with a thin film can exhibit both types of interaction for the same pulse length.

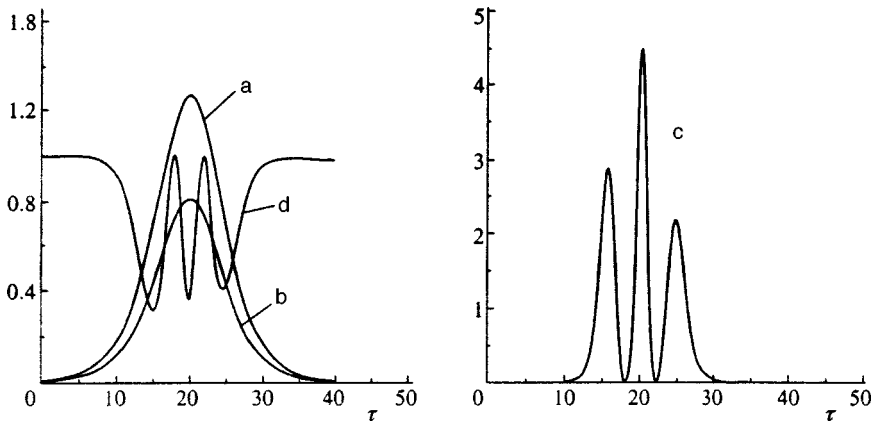


FIG. 2. Envelopes of a — incident pulse of the form (20), b — transmitted pulse, c — harmonic pulse, and d — population difference. $a_0=1.13, \tau_p=6$, and $\tau_m=20$.

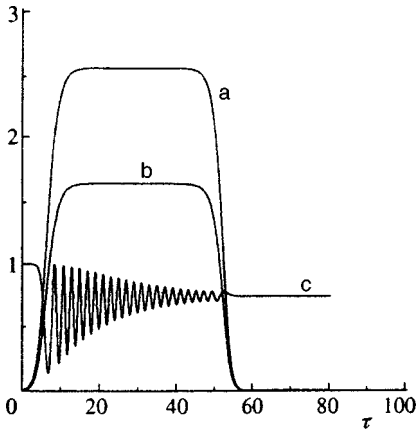


FIG. 3. Envelopes of the incident pulse (a), transmitted pulse (b), and population difference (c) for a plateau-shaped pulse of the fundamental wave (21) with the parameters $a_0=0.75, \tau_p=6, \tau_m=6$, and $\tau_w=8$.

Since the density of resonant atoms in the film was chosen to be low, a large change in the shape of the pump pulse was not observed in the series of investigations performed. For films with resonant atom density $n_A=10^{22}$ cm^{-3} or giant dipole moments ~ 10 D, the shape of an ultrashort pulse interacting with the film will be strongly deformed because of the inverse effect of the film. As the amplitude of the incident pulse increases, the transmitted pulse can be expected to split, just as in the case of an ultrashort pulse propagating in an extended medium in the two-photon self-induced transparency regime.[27] As the angle of incidence of the pump radiation increases, the additional peaks in the envelope of the transmitted pulse will vanish. Since the effective field acting on an atom decreases, the number of Rabi oscillations, each of which corresponds to a peak in the envelope of the transmitted pulse, decreases. The harmonic signal will exhibit all of these features of the pump pulse dynamics. This case must be investigated in greater detail.

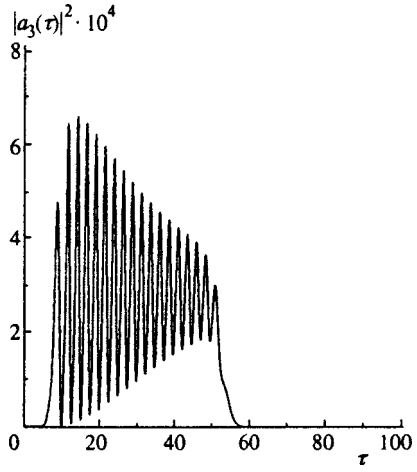


FIG. 4. Envelope of the harmonic pulse produced by a pulse of the fundamental wave (21) with the same parameters.

7. CONCLUSIONS

We have examined effects associated with the passage of an ultrashort electromagnetic pulse, under two-photon resonance conditions through the interface, containing a thin film of resonant atoms, between two dielectric media. The fundamental aspect of our analysis is that the Raman interaction of the fundamental wave and the generated third harmonic is taken into account. The matching conditions determine only the direction and not the generation efficiency of the harmonic radiation.

When the local Lorentz field is taken into account, a dynamic relaxation of the polarization of the film of resonant atoms occurs in times much shorter than the irreversible relaxation times of the medium. As a result, a stationary population of the energy levels of the resonant atoms which is different from the equilibrium value is established in the medium. In the future it would be best to focus attention on optical bistability, which is a characteristic phenomenon in the situation considered.

The effects discussed can be conveniently observed in the third harmonic, since its signal can be easily distinguished from the powerful pump wave by using filters and/or the matching conditions.

It is worth noting that if a harmonic signal comparable to the pump signal can be obtained on account of a high density of resonant atoms in the film, then parametric bleaching,[24] where the population of resonant atoms remains constant during the application of the electromagnetic pulses, can be expected. However, generation of the fifth and higher harmonics is now possible, since for them phase matching in the film is not required.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17429). One of us (A. M. B.) thanks the ZAO ''Russkoe Zoloto'' for assisting in this work.

*)E-mail: maimistov@pico.mephi.ru

[1] V. M. Agranovich, V. I. Rupasov, and V. Ya. Chernyak, JETP Lett. **33**, 185 (1981).
[2] V. I. Rupasov and V. I. Yudson, Kvant. Élektron. **9**, 2179 (1982) [Sov. J. Quantum Electron. **12**, 415 (1982)].
[3] V. I. Rupasov and V. I. Yudson, Zh. Éksp. Teor. Fiz. **93**, 494 (1987) [Sov. Phys. JETP **66**, 282 (1987)].
[4] A. M. Basharov, A. I. Maĭmistov, and É. A. Manykin, Zh. Éksp. Teor. Fiz. **97**, 1530 (1990) [Sov. Phys. JETP **70**, 864 (1990)].
[5] Y. Ben-Aryeh, C. M. Bowden, and J. C. Englund, Phys. Rev. A **34**, 3917 (1986).
[6] M. G. Benedikt, A. I. Zaĭtsev, V. A. Malyshev, and E. D. Trifonov, Opt. spektrosk. **66**, 726 (1989) [Opt. Spectrosc. **66**, 424 (1989)].
[7] M. G. Benedict and E. D. Trifonov, Phys. Rev. A **38**, 2854 (1988).
[8] M. G. Benedict, V. A. Malysheev, E. D. Trifonov, and A. I. Zaitsev, Phys. Rev. A **43**, 3845 (1991).
[9] M. G. Benedict and E. D. Trifonov, Potsdam. Forsch. B, No. 64, 9 (1990).
[10] A. M. Basharov, Zh. Éksp. Teor. Fiz. **94**, 12 (1988) [Sov. Phys. JETP **67**, 1741 (1988)].
[11] M. G. Benedikt, A. I. Zaĭtsev, V. A. Malyshev, and E. D. Trifonov, Opt. Spektrosk. **68**, 812 (1990) [Opt. Spectrosc. **68**, 473 (1990)].
[12] S. M. Zakharov and E. A. Manykin, Poverkhnost', fiz., khim., mekhan., No. 2, 137 (1988).
[13] C. L. P. Lambruschini, J. Mod. Opt. **37**, 1175 (1990).
[14] A. M. Basharov, Zh. Éksp. Teor. Fiz. **108**, 842 (1995) [JETP **81**, 459 (1995)].

[15] Yu. A. Logvin and A. M. Samson, Zh. Éksp. Teor. Fiz. **102**, 472 (1992) [Sov. Phys. JETP **75**, 250 (1992)].
[16] Yu. A. Logvin, A. M. Samson, and S. I. Turovets, Kvant. Élektron. **17**, 1521 (1990) [Sov. J. Quantum Electron. **20**, 1425 (1990)].
[17] V. A. Goryachev and S. M. Zakharov, Kvant. Élektron. **24**, 193 (1997).
[18] D. J. Kaup and B. A. Malomed, Phys. Lett. A **169**, 335 (1992).
[19] S. M. Zakharov É. A. Manykin, Opt. Spektrosk. **63**, 1069 (1987) [Opt. Spectrosc. **63**, 630 (1987)].
[20] A. M. Samson, Yu. A. Logvin, and S. I. Turovets, Opt. Commun. **78**, 208 (1990).
[21] É. Vanagas, Pis'ma Zh. Tekh. Fiz. **19**(6), 6 (1993) [Tech. Phys. Lett. **19**, 158 (1993)]
[22] S. M. Zakharov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **95**, 800 (1989) [Sov. Phys. JETP **68**, 457 (1989)].
[23] S. M. Zakharov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **105**, 1053 (1994) [JETP **78**, 566 (1994)].
[24] E. A. Manykin and A. M. Afanas'ev, Zh. Éksp. Teor. Fiz. **52**, 1246 (1967) [Sov. Phys. JETP **25**, 828 (1967)].
[25] A. M. Basharov, A. I. Maĭmistov, and É. A. Manykin, Zh. Éksp. Teor. Fiz. **84**, 487 (1983) [Sov. Phys. JETP **57**, 282 (1983)].
[26] A. M. Basharov, A. I. Maĭmistov, and É. A. Manykin, *Photonics. Nonlinear Coherent Processes* (Moscow Engineering Physics Institute, Moscow, 1986); A. M. Basharov, *Photonics. The Unitary Transformation Method in Nonlinear Optics* (Moscow Engineering Physics Institute, Moscow, 1990).
[27] S. O. Elyutin, A. I. Maĭmistov, and É. A. Manykin, Opt. spektrosk. **50**, 354 (1981) [Opt. Spectrosc. **50**, 191 (1981)].

Translated by M. E. Alferieff

Two-channel resonance scattering of waves and particles by point and planar defects

A. M. Kosevich*)

B. I. Verkin Institute of Low-Temperature Physics and Technology, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine

(Submitted 21 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 306–317 (January 1999)

The shattering of a wave (quasiparticle) with a dispersion curve consisting of two quadratic branches by a planar defect is discussed. The analog of such a process is the scattering of a similar wave (quasiparticle) in a one-dimensional system by a point defect. It is shown that even when the defect is passive, i.e., has no internal degrees of freedom, scattering may become resonant. The physical explanation of this effect is that a wave with a lower-lying spectrum scattered by the defect is in resonance with a localized (bound) state emerging because of the interaction between the defect and a wave with a higher-lying spectrum. © *1999 American Institute of Physics.* [S1063-7761(99)02501-9]

1. INTRODUCTION

The study of the scattering of various types of wave in a medium with point or planar defects has a long history and is interesting not only from the scientific standpoint but also with regard to applications. The theoretical aspects of scattering by a planar defect are interesting because analyzing this process reduces to studying one-dimensional dynamical problems and often allows for a simple analytical description. It is this fact that makes it possible to easily record some features of scattering in a resonance situation. For this reason we will deal in this paper with the scattering of waves by a planar defect.

The occurrence of resonance scattering by a defect can easily be explained in two cases: (a) when the defect has a structure and the wavelength λ of the scattered wave matches the geometrical size h of this structure, e.g., $\lambda = 2nh$, where n is an integer (geometric resonance); and (b) when the defect has an internal dynamic degree of freedom characterized by some frequency, and the frequency of the scattered wave coincides with the natural frequency of the defect. The common approach to studying resonance scattering in quantum mechanics is to examine the scattering of particles with a quadratic dispersion law by a double-humped potential, which in the one-dimensional case has the shape schematically depicted in Fig. 1. If the defect potential has a quasistationary level, the scattered particles with a quadratic dispersion law and an energy E satisfying $E_0 < E < E_m$ may resonantly interact with this level. All the features of such resonance scattering in the one-dimensional case are determined by the shape of the potential and have been thoroughly described (see, e.g., Ref. 1). When the defect is planar, the coordinate x is measured along the normal to the defect plane.

If the planar defect is very thin and the wavelength is much longer than its thickness ($\lambda \gg h$), geometric resonance is impossible. On the other hand, if the defect is passive, i.e., has no internal vibrational modes (or quasistationary energy levels in the quantum problem), ordinary (frequency) resonance is also impossible. In this case the reflection coefficient (or the transmission coefficient) has no singularities.

Nevertheless, in a recent paper,[2] Darinskii and Maugin described resonance reflection of a transverse elastic wave from a thin passive planar defect in conditions where the phase velocity of the incident wave is close to the velocity of the longitudinal elastic wave. This result has stimulated our studies of the physical reasons for the resonance effect.

To clarify the statement of the problem, we take a simple example of scattering by a point defect in a one-dimensional system. This example has been described in many textbooks on wave and particle scattering, but its analysis will allow us to introduce and recall necessary notation and terminology.

We consider a one-dimensional system whose state is described by a field $\Psi(x,t)$ obeying, in the case of steady-state vibrations ($\Psi(x,t) = \psi(x)e^{-i\varepsilon t}$), the equation

$$\varepsilon\psi = \varepsilon_0\psi - \frac{1}{2m}\frac{d^2\psi}{dx^2}. \tag{1}$$

The eigensolution of this equation, $\psi(x) = \psi_0 e^{ikx}$, corresponds to an energy ε with a dispersion law

$$\varepsilon = \varepsilon_0 + \frac{k^2}{2m}. \tag{2}$$

When there is a point defect localized at point $x = 0$, the right-hand side of Eq. (1) acquires a local-potential term. In the long-wavelength approximation, the local potential can always be written as

$$U(x) = U_0 h\,\delta(x), \tag{3}$$

where h is the effective width of the localized perturbation ($kh \ll 1$).

Solving the scattering problem amounts to finding solution of the form

$$\psi(x) = e^{ikx} + Ae^{-ikx}, \quad x < 0,$$

$$\psi(x) = Be^{ikx}, \quad x > 0. \tag{4}$$

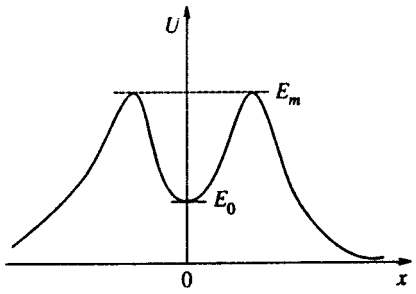


FIG. 1.

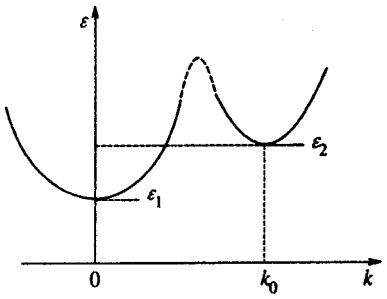


FIG. 2.

The amplitudes A and B determine the reflection coefficients $R=|A|^2$ and transmission coefficient (transparency) $T=|B|^2$, with $R+T=1$. These coefficients can easily be found; in particular, the transmission coefficient is

$$T=\frac{k^2}{k^2+(mhU_0)^2}. \tag{5}$$

For $U_0\neq 0$ the transmission coefficient T has no singularities for k real, i.e., in the region where plane waves exist ($\varepsilon>\varepsilon_0$), and T is always positive but smaller than unity.

The transmission coefficient T may acquire a singularity only for $\varepsilon<\varepsilon_0$ and U_0 negative (a local attractive potential), at $k=i\kappa$, with $\kappa=\sqrt{2m(\varepsilon_0-\varepsilon)}$. However, a value $\varepsilon<\varepsilon_0$ gives rise to a localized state,

$$\psi(x)=\psi_0 e^{-\kappa|x|}, \tag{6}$$

which corresponds to an eigenvalue lying in the forbidden band for the natural frequencies of free waves (particles). Usually such a state is not directly related to the scattering of free waves (with $\varepsilon>\varepsilon_0$).

The situation changes if the dispersion law $\varepsilon=\varepsilon(k)$ in a defect-free medium has several branches (or valleys, to use the terminology of electronic semiconductor theory). This means that additional scattering ''channels'' can open due to the existence of several valleys.

Usually the independent branches of vibrations (or group of particles) differ in the parameters of their dispersion laws. In the simplest case of two types of particle (wave), the dispersion law within a certain range of eigenvalues has the form $\varepsilon=\varepsilon_1(k)$ and $\varepsilon=\varepsilon_2(k)$ for the first and second groups, respectively, where

$$\varepsilon_1(k)=\varepsilon_1+\frac{k^2}{2m_1},\quad \varepsilon_2(k)=\varepsilon_2+\frac{(k-k_0)^2}{2m_2}, \tag{7}$$

with $\varepsilon_1<\varepsilon_2$ (Fig. 2).

The excitations of different branches are independent in an ideal medium. When the medium has defects, local hybridization of the excitations is possible, which may have a strong effect on the scattering processes. Let us assume that a passive local attractive potential allows for local states of the form (6) for the second group when $\varepsilon<\varepsilon_2$ holds. Then a scattered particle of the first group in a state of the form (6) with $\varepsilon_1<\varepsilon<\varepsilon_2$ may find itself in resonance with the local eigenstates of excitations belonging to the second branch. In Sec. 2 we will see that in this situation particles of the first type may satisfy the conditions $T=0$ ($R=1$) and also the conditions $T=1$ ($R=0$).

In Sec. 3 we will study the resonance scattering of a transverse acoustic wave by a planar defect under conditions in which the phase velocity of the wave along the defect lies within the interval limited by the velocities of the transverse and longitudinal acoustic waves. Finally, in Sec. 4, we will discuss the scattering of a particle (wave) with a two-valley dispersion law by a point defect in a three-dimensional medium. We will also find the particle energy at which the wave is not scattered by the point defect.

2. ONE-DIMENSIONAL MODEL: SCATTERING BY A POINT DEFECT OF A WAVE HAVING A DISPERSION LAW WITH TWO BRANCHES

Suppose that the elementary excitations of the system have two branches in the dispersion law, similar to those depicted in Fig. 2, where ε is the energy (or the frequency squared), and k is the wave vector. In the energy range $\varepsilon\sim\varepsilon_2-\varepsilon_1$, the dynamics of the system is described by two groups of excitations with dispersion law (7). In an ideal system, these particle-waves do not interact, and their dynamics obeys a wave equation or the Schrödinger equations

$$\begin{aligned}&\varepsilon\psi_1-\varepsilon_1\psi_1+\frac{1}{2m_1}\frac{d^2\psi}{dx^2}=0,\\ &\varepsilon\psi_2-\varepsilon_2\psi_2+\frac{1}{2m_2}\left(\frac{d}{dx}-ik_0\right)^2\psi_2=0.\end{aligned} \tag{8}$$

The quasiparticles interact when there is a defect. The defect generates an interaction that corresponds to the additional energy of the system localized at the defect. Assume that the density of this energy is

$$U_{\rm int}(x)=U_0\{\alpha_1|\psi_1|^2+\alpha_2|\psi_2|^2+\beta(\psi_1^*\psi_2+\psi_1\psi_2^*)\}\delta(x). \tag{9}$$

As a result the dynamical equations for the system with a defect acquire the form

$$\varepsilon\psi_1-\varepsilon_1\psi_1+\frac{1}{2m_1}\frac{d^2\psi_1}{dx^2}=U_0[\alpha_1\psi_1(0)+\beta\psi_2(0)]\delta(x), \tag{10}$$

$$\varepsilon\psi_2-\varepsilon_2\psi_2+\frac{1}{2m_2}\left(\frac{d}{dx}-ik_0\right)^2\psi_2$$
$$=U_0[\alpha_2\psi_2(0)+\beta\psi_1(0)]\delta(x). \tag{11}$$

Clearly, the presence of a local perturbation potential (9) is equivalent to the presence of additional boundary conditions for Eqs. (8) at point $x=0$ (with the functions $\psi_1(x)$ and $\psi_2(x)$ continuous):

$$\frac{d\psi_1^+}{dx}-\frac{d\psi_1^-}{dx}=2m_1U_0[\alpha_1\psi_1(0)+\beta\psi_2(0)],$$
$$\frac{d\psi_2^+}{dx}-\frac{d\psi_2^-}{dx}=2m_2U_0[\alpha_2\psi_2(0)+\beta\psi_1(0)], \tag{12}$$

where

$$\varphi_\alpha^+=\varphi_\alpha(+0),\quad \varphi_\alpha^-=\varphi_\alpha(-0),\quad \alpha=1,2.$$

Note that the adopted potential of the interaction of the fields ψ_1 and ψ_2 at the defect is similar to the free surface energy of a superconductor introduced in Ref. 3, the energy expressed in terms of the order parameter ψ in the presence of a twinning boundary. Finding the spatial distribution of the order parameter ψ in the presence of a planar defect reduces to solving a one-dimensional equation of the form (8) (if we ignore the existence of the vector potential of the magnetic field) satisfying the boundary conditions, and Eqs. (12) can be considered a particular case of these conditions.

Let us take the standard problem of the wave $\psi_1(x)$ with an energy ε in the interval between ε_1 and ε_2 ($\varepsilon_1<\varepsilon<\varepsilon_2$) scattered by such a defect. The wave $\psi_1(x)$ is sought in the standard form (4), and a natural expression for the wave $\psi_2(x)$ is one generalizing (6),

$$\psi_2(x)=M\exp^{\{ik_0x-\kappa|x|\}}, \tag{13}$$

where $k^2=2m_1(\varepsilon-\varepsilon_1)$, and $\kappa^2=2m_2(\varepsilon_2-\varepsilon)$. In (13) we have allowed for the continuity of the function $\psi_2(x)$ at $x=0$ and for natural boundary conditions at infinity.

The continuity of $\psi_1(x)$ connects the amplitudes of the reflected (A) and transmitted (B) waves, $1+A=B$, and Eqs. (12) yield

$$B=\frac{ik(\kappa+\alpha_2m_2U_0)}{\Delta(\varepsilon)}, \tag{14}$$

$$M=-\frac{ik\beta m_2U_0}{\Delta(\varepsilon)}, \tag{15}$$

$$\Delta(\varepsilon)=ik(\kappa+\alpha_2m_2U_0)$$
$$+m_1U_0[\alpha_1\kappa+(\alpha_1\alpha_2-\beta^2)m_2U_0]. \tag{16}$$

We see that $B=0$ if

$$\kappa(\varepsilon)+\alpha_2m_2U_0=0. \tag{17}$$

This relationship is meaningful for $\alpha_2U_0<0$, i.e., if the potential of the defect is attractive for the particle $\psi_2(x)$. In this case the reflection and transmission coefficients are $R=1$ and $T=0$, respectively.

But if

$$\alpha_1\kappa+(\alpha_1\alpha_2-\beta^2)m_2U_0=0, \tag{18}$$

we have $B=1$, and the scattering process is characterized by $R=0$ and $T=1$.

The physical meaning of the resonance conditions is quite clear and can easily be explained by analyzing Eqs. (11) and (12). The amplitude B vanishes if $\psi_1(0)=0$, a condition that means that Eq. (11) becomes autonomous and has a localized solution if condition (17) is met (this follows from (12) if $\psi_1(0)=0$). A localized vibration of the field $\psi_2(x)$ is the dynamical system which is in resonance with the incident wave of the field $\psi_1(x)$ and creates conditions for total reflection of this wave.

The coefficient T is equal to unity if $\alpha_1\psi_0(0)+\beta\psi_2(0)=0$, a condition that means that $\psi_1(x)$ does not feel the perturbation generated by the potential U_0, so that the incident wave passes through the defect without being distorted. As for the field $\psi_2(x)$, it becomes autonomous, since in its right-hand side

$$\alpha_2\psi_2(0)+\beta\psi_1(0)=\left(\frac{\alpha_2-\beta^2}{\alpha_1}\right)\psi_2(0).$$

Hence Eq. (12) again has a localized solution corresponding, however, to another energy, for which condition (18) is met.

Thus, we have established that the resonance features of the scattering of a wave of field $\psi_1(x)$ by a point defect are related, as expected, to the features of the problem of steady eigenstates of the total field $\psi=\psi_1+\psi_2$. Here it is advisable to analyze in detail the part of the eigenstate problem that is related to this aspect (scattering by a point defect). Solving the problem of steady eigenstates of the field $\psi=\psi_1+\psi_2$, we can write

$$\psi_1(x)=\begin{cases}A\sin(kx-\varphi_1), & x<0,\\ B\sin(kx-\varphi_2), & x>0,\end{cases}$$
$$\psi_2(x)=M\exp\{-\kappa|x|+ik_0x\}, \tag{19}$$

where k and κ have been defined earlier.

The existence of the solution (19) implies that the steady state has two terms, a standing wave of the field $\psi_1(x)$ on the entire x axis and a vibration localized near the defect.

The solution (19) is characterized by five parameters: the amplitudes A, B, and M and the phases φ_1 and φ_2. The continuity of the function $\psi_1(x)$ and the boundary conditions (12) lead to the a system of homogeneous algebraic equations for the amplitudes A, B, and M:

$$A\sin\varphi_1-B\sin\varphi_2=0,$$
$$A(2\alpha_1m_1U_0\sin\varphi_1+k\cos\varphi_1)-Bk\cos\varphi_2$$
$$+2\beta m_1U_0M=0, \tag{20}$$
$$A\beta m_2U_0\sin\varphi_1+(\kappa+\alpha_2m_2U_0)M=0.$$

By equating the determinant of the system (20) to zero we arrive at a relationship between the phases φ_1 and φ_2, retaining one parameter (or the difference of the parameters) as a free parameter.

This leads to two conclusions: (a) if the energy ε satisfies $\varepsilon_1<\varepsilon<\varepsilon_2$, a quasilocalized vibration is the steady state of the system; and (b) for ε fixed there is a continuous spectrum of quasilocalized states.

A remarkable feature of the set of quasilocalized vibrations is that there is a steady state of the field $\psi(x)$ in which a standing wave exists only on one semiaxis and a localized vibration exists on both semiaxes. Indeed, setting $B=0$ in (20), we see that φ_1 is automatically equal to zero, i.e., $\psi_1(0)=0$ and, in addition, $\kappa+\alpha_2 m_2 U_0=0$. As for the amplitude M, it is uniquely determined by the formula

$$M=-\frac{kA}{2\beta m_1 U_0}. \tag{21}$$

Naturally, the condition for the existence of a quasilocalized vibration of the type we have just described (with $B=0$) coincides with the condition for total resonance reflection. Hence the standing wave $\psi_1(x)$ with an amplitude A can exist only on one semiaxis ($x<0$) if

$$\varepsilon=\varepsilon_2-\frac{1}{2}\alpha_2^2 m_2 U_0^2 \tag{22}$$

and at the same time there is a localized vibration of the field $\psi_2(x)$ with an amplitude (21).

Finally, a steady state with a single standing wave along the entire x axis with $\psi_1(0)\neq 0$ corresponds to the conditions

$$A=B, \quad \varphi_1=\varphi_2=\frac{\pi}{2}. \tag{23}$$

In this case, as Eqs. (20) imply, we have $M=-\alpha_1 A/\beta$, and the frequency can be found from the condition (18), which yields

$$\varepsilon=\varepsilon_2-\frac{1}{2}\left(\frac{\alpha_2-\beta^2}{\alpha_1}\right)^2 m_2 U_0^2. \tag{24}$$

This is possible if $(\alpha_1\alpha_2-\beta^2)U_0<0$.

Clearly, the results of this section are weakly dependent on the type of boundary conditions ''mixing'' the fields ψ_1 and ψ_2 and on the specific form of the dynamical equations that lead to a ''two-valley'' dispersion law with different in which the valleys have different depths. Moreover, the main conclusions drawn in this section are valid not only for to systems with continuous field distributions but also for discrete models.

3. SCATTERING OF A TRANSVERSE ELASTIC WAVE BY A PLANAR DEFECT IN AN ISOTROPIC MEDIUM

Examples of planar defects in crystals are well known: twinning boundaries, interphase boundaries, flat packing faults, etc. In the long-wavelength approximation (which we consider in this paper), where the crystal dynamics is described by the theory of elasticity, such a defect may assumed to be lying on a plane. In this case, as shown in Sec. 2, the presence of the defect manifests itself in additional boundary conditions imposed on the equations of the field (in our case the field of elastic displacements **u**). We start with a discussion of the possible boundary conditions. Suppose that the defect is entirely in the plane $z=0$. Two types of violation of the properties of the solid may be localized along this plane.

1. The surface $z=0$ coincides with a monatomic layer of isotopes, atoms whose mass M differs from the mass m of the regular atoms of the medium. If σ_{ik} is the elastic stress tensor and ρ is the density of the medium, the dynamical elasticity equation perturbed by the defect plane has the form

$$\frac{\partial\sigma_{ik}}{\partial x_k}=\rho\frac{\partial^2 u_i}{\partial t^2}+\eta\rho\frac{\partial^2 u_i}{\partial t^2}h\,\delta(z), \tag{25}$$

where $\eta=(M-m)/m$, and h is the thickness of the defect layer.

Equation (25) yields the obvious boundary conditions

$$\sigma_{iz}^{+}-\sigma_{iz}^{-}=\eta h\rho\frac{\partial^2 u_i(0)}{\partial t^2}, \quad i=1,2,3. \tag{26}$$

Let us examine the scattering of a plane elastic wave by this defect, assuming that the wave is polarized in the plane of incidence. We select the zx plane as the plane of incidence. Then the conditions (26) that are important are those with $i=x$ and $i=z$.

Allowing for the continuity of the vector **u** and its derivatives with respect to x in the plane $z=0$ and the monochromaticity of the wave, we can reduce the conditions (26) to the following two:

$$c_l^2\left(\frac{\partial u_z^{+}}{\partial z}-\frac{\partial u_z^{-}}{\partial z}\right)=W_0 u_z(0), \tag{27}$$

$$c_t^2\left(\frac{\partial u_x^{+}}{\partial z}-\frac{\partial u_x^{-}}{\partial z}\right)=W_0 u_x(0), \tag{28}$$

where $\mathbf{u}^{+}=\mathbf{u}(z=+0)$, $\mathbf{u}^{-}=\mathbf{u}(z=-0)$, $W_0=-\eta h\omega^2$, with ω the frequency, and c_l and c_t are the velocities of the longitudinal and transverse acoustic waves, respectively.

The conditions (27) and (28) generalize the boundary conditions (12) in the one-dimensional case. Note that the conditions (26)–(28) corresponds to the continuity of displacements in the plane $z=0$ and a discontinuity in the stresses across the opposite sides of the defect plane, and in this sense are similar to (12).

2. Different boundary conditions arise when the interatomic interaction along the defect plane varies.

The ways of deriving the macroscopic boundary conditions vary, but all are based on studying the long-wavelength limit of the crystal lattice dynamics. However, the model most often used here is that of a three-layer sandwich in which the thickness of the middle layer goes to zero. (This was the approach used in Refs. 2 and 4 to obtain the boundary conditions. Other boundary conditions were formulated phenomenologically in Ref. 5).

In the simplest case where the effective elastic moduli of the defect layer are small compared to the elastic moduli of the medium, the boundary conditions presuppose that the normal components of the stress tensor are continuous in the presence of a discontinuity of the corresponding components of the displacement vector:

$$\sigma_{xz}^{+}=\sigma_{xz}^{-}=\frac{\mu}{l_1}(u_x^{+}-u_x^{-}),$$

$$\sigma_{zz}^{+}=\sigma_{zz}^{-}=\frac{\mu}{l_3}(u_z^{+}-u_z^{-}), \tag{29}$$

where l_1 and l_3 are the effective thicknesses of the defect layer, which are inversely proportional to the elastic moduli of this layer, and μ is the shear modulus of the isotropic medium.

The boundary conditions (29) are broadly applicable range if one assumes that the parameters l_1 and l_3 are fixed. They have natural limits: at $l=\infty$ (the appearance of a free layer) both σ_{xz} and σ_{zz} vanish at the defect layer ($z=0$), while at $l=0$ the displacements (and stresses) become continuous at the plane $z=0$.

Having established the boundary conditions (27), (28), or (29), we begin the solution of the problem by examining the scattering of a transverse wave. We assume that the wave is incident at an angle θ to the z axis. Then the dependence of the displacement field on the coordinate x is the same for all the field components and is described by the wave $\exp^{ikx}$, where $k=(\omega/c_t)\sin\theta$. We also assume (this is important!) that the phase velocity c of the wave along the surface is $\omega/k=c_t/\sin\theta<c_l$, i.e., we assume that $\sin\theta>c_t/c_l$.

We write the displacement vector $\mathbf{u}$ for the geometry of the problem [$\mathbf{u}=(u_x=0,u_z)$] in the form

$$\mathbf{u}(x,z,t)=\mathbf{u}(z)\exp(ikx-i\omega t), \tag{30}$$

where the field $\mathbf{u}(z)$ determines the dependence of the sum of the longitudinal and transverse vibrations of the elastic medium on the coordinate z:

$$\mathbf{u}(z)=\mathbf{u}^{(t)}+\mathbf{u}^{(l)}, \quad ku_x^{(t)}=i\frac{du_z^{(t)}}{dx}, \quad ku_z^{(l)}=-i\frac{du_x^{(l)}}{dz}. \tag{31}$$

In the bulk ($z\neq 0$), the longitudinal and transverse components obey the equations

$$(\omega^2-c_t^2k^2)\mathbf{u}_t=c_t^2\frac{d^2\mathbf{u}_t}{dz^2}, \tag{32}$$

$$(\omega^2-c_l^2k^2)\mathbf{u}_l=c_l^2\frac{d^2\mathbf{u}_l}{dz^2}. \tag{33}$$

It should be recalled thatk, by assumption,

$$c_tk<\omega=ck<c_lk, \tag{34}$$

and solving the standard problem of the scattering of a longitudinal wave reduces to finding the solution

$$\mathbf{u}^{(t)}=\begin{cases}\mathbf{a}e^{iqz}+\mathbf{A}e^{-iqz}, & z<0,\\ \mathbf{B}e^{iqz}, & z>0,\end{cases} \tag{35}$$

$$\mathbf{u}^{(l)}=\begin{cases}\mathbf{M}e^{\kappa z}, & z<0,\\ \mathbf{N}e^{-\kappa z}, & z>0,\end{cases} \tag{36}$$

where $ka_x+qa_z=0$ for the polarization vector of the incident transverse wave, with $c_tq=\sqrt{\omega^2-c_t^2k^2}=k\sqrt{c^2-c_t^2}$, and $c_l\kappa=\sqrt{c_l^2k^2-\omega^2}$.

We start with the problem of the scattering of the wave by a defect that leads to the boundary conditions (27) and (28). The conditions for the continuity of the total displacements $u_x(z)$ and $u_z(z)$ together with boundary conditions (27) and (28) constitute a system of four linear algebraic equations for determining the amplitudes **A**, **B**, **M**, and **N**, whose components are linked through relationships that follow from (31). Technically the solution of the problem is difficult, and here we give only the results that are of interest to us. In order to proceed with the discussion of the problem of resonance scattering, we write the expression for the amplitude **B**, whose calculation yields[1)]

$$\begin{aligned}\mathbf{B}&=-\frac{2iq}{\Delta}[2\kappa c_l^2+\eta h\omega^2]\mathbf{a},\\ \Delta&=\eta h\left[(\eta h\omega^2-2\kappa c_l^2)\frac{\omega^2}{c_t^2}+(\kappa^2-q^2)\eta hk^2c_l^2\right]\\ &\quad -2iq[(2\kappa c_l^2-\eta h\omega^2)-(\eta hk)^2c_l^2\kappa],\end{aligned} \tag{37}$$

with $\sigma=c_t^2/c_l^2$.

We see that total reflection ($B=0$) occurs when

$$2c_l^2\sqrt{k^2-\frac{\omega^2}{c_l^2}}=-\frac{\Delta m}{m}h\omega^2. \tag{38}$$

Total reflection is possible only if there is a heavy defect layer ($M>m$). In the long-wavelength approximation ($hk\ll 1$), the phase velocity $c=\omega/k$ of a resonantly reflected wave is close to the velocity c_l of longitudinal acoustic waves and is given by the formula

$$c=c_l\left[1-\frac{1}{2}\left(\frac{\Delta m}{2m}hk\right)^2\right]. \tag{39}$$

As is usual in such situations, $\Delta c/c\sim(\Delta m/m)^2h^2k^2$.

Note that the condition (38) determines the frequency (or phase velocity c) of a steady-state localized longitudinal wave obeying the boundary condition (27) in the absence of a transverse component. In addition to the wave (36), the total steady-state solution incorporates the standing transverse wave

$$\mathbf{u}^{(t)}=\begin{cases}\mathbf{a}\sin qz, & z<0,\\ 0, & z>0,\end{cases}$$

which exists only on one side of the planar defect.

Now we turn to the second case, in which the boundary conditions (29) are employed. The solution of the problem under such boundary conditions is given in Ref. 6. Here we will again list the results related to the calculation of the transmission coefficient $T=|\mathbf{B}|^2$. It turns out that

$$T=\frac{4s^2(s-1)}{\Delta(s)}\left[\frac{(2-s)^2}{kl_1}-\frac{2s\sqrt{1-\sigma s}}{kl_1kl_2}-\frac{4(1-\sigma s)}{kl_3}\right]^2, \tag{40}$$

where $s=(c/c_t)^2=(\omega/kc_l)$, $\sigma=(c_t/c_l)^2$, and $\Delta(s)$ is sure to be a positive function of s for $s>1$, i.e., for $c>c_t$.

The condition for total wave reflection corresponds to $T=0$. Equation (40) implies that this is possible either for $s=1$ ($c=c_t$, which corresponds to a wave gliding along the planar defect) or for

$$2s\sqrt{1-\sigma s}=(2-s)^2kl_3-4(1-\sigma s)kl_1, \tag{41}$$

which is an equation for the phase velocity.

In the limit $l_1,l_3\to 0$, Eq. (41) simplifies and the phase velocity c tends to c_l, i.e.,

$$c^2 = c_l^2 - \frac{\lambda^4 (kl_3)^2}{[2\mu(\lambda+2\mu)]^2}, \quad kl_3 \ll 1, \tag{42}$$

where λ and μ are the Lamé coefficients.

The same dependence on $(kl_3)^2$ for the resonant phase velocity was obtained in Ref. 2. However, Eq. (42) represents a more complicated dependence over a broad range of values of the parameters l_1 and l_3.

In Ref. 6 it was shown that Eq. (41) coincides with the condition for the existence of a steady-state quasilocalized solution, which for $z<0$ has the form of a standing transverse wave and a longitudinal vibration localized near the defect (in full agreement with the results obtained with the other boundary conditions).

If we introduce the notation $\xi = l_1/l_3$, it is convenient to write Eq. (41), which represents the dependence of the phase velocity on, say, kl_3, in the inverse-function form:

$$kl_3 = \frac{2s\sqrt{1-\sigma s}}{(2-s)^2 - 4(1-\sigma s)\xi}. \tag{43}$$

Comparing (43) with the law of dispersion of pseudosurface waves,[4,7] we conclude that the corresponding quasilocalized wave has a phase φ given by the following relationship:

$$\xi = \sqrt{\frac{1-s}{1-\sigma s}} \cot\varphi.$$

Thus, we have again demonstrated the general law for the occurrence of resonance conditions when a transverse wave is scattered by a planar defect, a law related to the ''multichannel'' nature of the dispersion of the scattered wave.

4. THE RESONANCE FEATURES OF SCATTERING BY A POINT DEFECT IN A THREE-DIMENSIONAL MEDIUM

And now let us confirm the above conclusion by analyzing the three-dimensional case. Suppose that the state of the particles with a two-valley dispersion law is described by the sum of the wave functions $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ corresponding to the dispersion laws $\varepsilon = \varepsilon_\alpha(\mathbf{k})$, $\alpha = 1,2$, which differ from the dispersion laws (7) by the simple fact that the wave vectors $\mathbf{k}$ and $\mathbf{k}_0$ are three-dimensional.

A defect at point $\mathbf{r}=0$ ''mixes'' the states ψ_1 and ψ_2. We write the simplest local condition at the defect, a condition that acts as boundary conditions and generalizes (12):

$$\lim_{r\to 0}\left[\frac{\partial}{\partial r}(r\psi_1) - 2m_1 r U_0(\alpha_1\psi_1 + \beta\psi_2)\right] = 0,$$
$$\lim_{r\to 0}\left[\left(\frac{\partial}{\partial r} - i\mathbf{k}_0\cdot\mathbf{n}\right) r\psi_2 - 2m_2 r U_0(\alpha_2\psi_2 + \beta\psi_1)\right] = 0, \tag{44}$$

with $\mathbf{n} = \mathbf{r}/r$. The solution of the problem of the scattering of a particle with an energy $\varepsilon < \varepsilon_0$ reduces in the three-dimensional case to finding the solution

$$\psi_1 = e^{i\mathbf{k}\cdot\mathbf{r}} + \frac{\chi}{r} e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{45}$$

$$\psi_2 = \frac{Q}{r}\exp\{i\mathbf{k}_0\cdot\mathbf{r} - \kappa r\}, \quad r^2 = x^2+y^2+z^2, \tag{46}$$

where $q^2 = 2m_1\varepsilon$, and $\kappa^2 = 2m_2(\varepsilon_0 - \varepsilon)$. The factors χ and Q can be found from the boundary conditions (44):

$$\chi = \frac{\kappa + 2m_2\alpha_2 U_0}{\Delta}, \quad Q = -\frac{2m_2\beta U_0}{\Delta}, \tag{47}$$

$$\Delta = (2m_1\alpha_1 U_0 - i\mathbf{q}\cdot\mathbf{n})(2m_2\alpha_2 U_0 + \kappa) - 4m_1 m_2 \beta^2 U_0^2. \tag{48}$$

We see that if the point defect ensures particle attraction ($\alpha_2 U_0 < 0$), a situation with $\chi = 0$ is possible if

$$\sqrt{2m_2(\varepsilon_0 - \varepsilon)} = 2m_2|\alpha_2 U_0|, \tag{49}$$

which is a condition similar to (38). The condition (49) determines the energy at which the local state (46) would arise in the absence of the wave ψ_1. In the case of the scattering of the wave (45), the amplitude of the local state is determined by the fact that the amplitude of the incident wave $Q = (2m_1\beta U_0)^{-1}$ is equal to unity.

Thus, we have shown that a point defect can be ''resonantly'' transparent in a characteristic two-channel scattering process. This possibility is related to the fact that at a certain energy there is a steady quasilocalized state in the system.

5. CONCLUSION

There are two features that make the results of the present study remarkable. From the standpoint of scattering theory, the resonance effects described in this paper are a manifestation of the scattering of a particle (wave) by a quasisteady state. What sets these effects apart from those commonly described in the literature is that in the case at hand the resonance effect is due entirely to the special features of the two-valley dispersion law for the particles in an ideal (defect-free) medium and is weakly dependent on the shape of the local potential that causes the scattering. From the viewpoint of the theory of quasilocalized states, the above results can be considered an addition to the results of Refs. 4 and 7, where a complete one-parameter system of eigensolutions of the dynamical elasticity theory corresponding to pseudosurface waves was derived.

The author is grateful for the financial support provided by the Ukrainian Ministry of Science and Technology (Project 2.4/163).

*)E-mail: kosevich@ilt.kharkov.ua

1)The necessary calculations were done by D. Semagin,[8] a student of Kharkov State University.

[1] D. Bohm, *Quantum Theory*, Prentice-Hall, New York (1951).
[2] A. N. Darinskii and G. A. Maugin, Wave Motion **23**, 363 (1996).
[3] A. F. Andreev, JETP Lett. **46**, 584 (1987).
[4] A. M. Kosevich and A. V. Tutov, Phys. Lett. A **213**, 265 (1996).
[5] L. J. Pyrak-Nolte, J. Xu, and G. M. Haley, Phys. Rev. Lett. **68**, 3650 (1992).
[6] A. M. Kosevich and A. V. Tutov, Phys. Lett. A **248**, 271 (1998).
[7] A. M. Kosevich and A. V. Tutov, Fiz. Nizk. Temp. **19**, 905 (1993) [Low Temp. Phys. **19**, 905 (1993)].
[8] A. Kosevich and D. Semagin, Cond. Meta Phys. (Lviv) **1**, 409 (1998).

Translated by Eugene Yankovsky

The effect of parametrically excited spin waves on the dispersion and damping of magnetostatic surface waves in ferrite films

G. T. Kazakov, A. V. Kozhevnikov, and Yu. A. Filimonov*)

Saratov Branch, Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 410019 Saratov, Russia

(Submitted 18 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 318–332 (January 1999)

This paper describes an experimental study of variations of the dispersion and damping of magnetostatic surface waves in ferrite films, caused by three- and four-magnon interactions with parametric spin waves excited by an auxiliary surface magnetostatic pump wave with frequency f_p. The variations in the dispersion and damping were identified, respectively, with variations $\Delta k'$ in the real part and $\Delta k''$ in the imaginary part of the wave number of the surface magnetostatic wave. The $\Delta k'$ and $\Delta k''$ values were determined from the ratio of the changes of the phase increment $\Delta\phi$ and the amplitude increment ΔA of the surface magnetostatic wave to the length L of the nonequilibrium section of the film, where the parametric spin waves exist. It is found that, when three-magnon decay processes are allowed for the pump wave and the surface magnetostatic probe wave, the probe wave can substantially alter the distribution of the parametric spin waves in the film. © *1999 American Institute of Physics.* [S1063-7761(99)02601-3]

1. INTRODUCTION

Processes of parametric excitation of spin waves in ferrite films by travelling magnetostatic waves are of significant interest from both scientific and applied viewpoints.[1] One effective way to experimentally study the parametric instability of magnetostatic waves is the two-pump method.[2–6] This method is based on the effects of the interaction of the probe signal of a magnetostatic wave whose power P is less than the parametric instability threshold P_{th} with parametric spin waves created by a pump wave with frequency f_p and power $P_p \geqslant P_{\text{th}}$. When the probe signal passes through the nonequilibrium section of the film, containing the parametric spin waves, changes can be observed in its amplitude A and phase ϕ from which one can judge whether parametric instability will appear in the pump wave and what its behavior will be beyond the threshold.[2–6]

On the other hand, there is interest in the amplitude ΔA and phase variation $\Delta\phi$ of the probe wave themselves, since they carry information concerning the dispersion and damping of the magnetostatic waves on the nonequilibrium section of the film. Finally, there is interest in seeking the conditions under which the probe wave not only experiences the action of the parametric spin waves but also itself affects their distribution in the film. The goal of this paper is to experimentally study the effect of parametric spin waves on the dispersion and damping of magnetostatic waves in an yttrium–iron garnet film and to detect how the magnetostatic waves affect the distribution of the parametric spin waves in the film.

Note that the effect of parametric spin waves on magnetostatic waves in an yttrium–iron garnet film was apparently considered for the first time in Ref. 7. Backward bulk magnetostatic waves propagated in the film, and locally homogeneous parallel pumping served as the source of the parametric spin waves, while the excitation frequency f of the magnetostatic wave and the interval of magnetic fields H_0 were chosen to satisfy the condition

$$f < f_{\text{th}}^{3m}, \tag{1}$$

where f_{th}^{3m} is the limiting frequency for three-magnon decay processes. The inequality (1) signifies[8] that three-magnon processes for magnetostatic waves are forbidden by the conservation laws, and magnetostatic and parametric spin waves can directly interact only via four-magnon processes of type[1]

$$2f = f_1 + f_2, \quad 2\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2, \tag{2}$$

where f and $\mathbf{k}$ are, respectively, the frequency and wave vector of the magnetostatic wave, while $f_{1,2}$ and $\mathbf{k}_{1,2}$ are analogous quantities that characterize the parametric spin waves.

Unlike Ref. 7, this paper discusses changes in the dispersion and damping of Damon–Eshbach surface magnetostatic waves[9] in an yttrium–iron garnet film under the influence of such parametric spin waves, excited by inhomogeneous pumping in the form of an additional magnetostatic surface wave with frequency f_p. Moreover, we consider the changes in the dispersion and damping of a magnetostatic surface wave, caused by interaction with parametric spin waves both in processes of the form (2) and in three-magnon processes, which satisfy the conservation laws[1]

$$f = f_1 + f_2, \quad \mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2. \tag{3}$$

In determining how parametric spin waves affect the dispersion and damping of magnetostatic surface waves, we use an approach in which the phase and amplitude variations $\Delta\phi$ and ΔA of the probe signal passing through the film are related to the overall changes $\Delta k'$ and $\Delta k''$ in the real and

imaginary parts of the wave number $k=k'+ik''$ of the magnetostatic surface wave when the length of the nonequilibrium section of the film is L:

$$\Delta\phi=\int_0^L \Delta k'(x)dx, \tag{4}$$

$$\Delta A=-8.68\int_0^L \Delta k''(x)dx, \tag{5}$$

where $\Delta\phi$ is the phase change of the signal in radians, ΔA is the amplitude change of the magnetostatic surface wave in decibels, and x is the coordinate along the film. The $\Delta k'$ and $\Delta k''$ values averaged over the length of the nonequilibrium section of the film will then be determined by

$$\Delta k'=\frac{\Delta\phi}{L}, \tag{6}$$

$$\Delta k''=-\frac{\Delta A}{8.68L}. \tag{7}$$

It is understood that the length L of the nonequilibrium section must be known in order to implement such an approach, and this depends on the power P_p and the type (three- or four-magnon) of parametric instability process of the pump wave.[4–6] We shall show that this problem can be solved if, as in Refs. 4–6, the pump and probe waves are excited by separate transducers the distance S between which can be varied by displacing one of them along the surface of the film, just as in the movable probe method of Ref. 10.

2. THE INFLUENCE OF PARAMETRIC SPIN WAVES ON THE DISPERSION AND DAMPING OF MAGNETOSTATIC SURFACE WAVES

The studies were carried out on a prototype of a magnetostatic-surface-wave delay line, consisting of three parallel microstrip transducers 4 mm long and 40 μm wide (Fig. 1). Transducer *1* was placed on separate polycore board *4*, while polycore board *5* was common to transducers *2* and *3*, and, as in the movable-probe method of Ref. 10, was displaced relative to polycore board *4*, remaining in the same plane with it. In this case, the distance S_1 between transducers *1* and *2* could vary within the limits $S_1=0.1$–4 mm, while the distance between transducers *2* and *3* was fixed and equalled $S_2=4$ mm. On the transducers was placed an epitaxial yttrium–iron garnet film on a substrate made from gadolinium–gallium garnet with (111) orientation. These studies used 30×5-mm samples in which the film thickness was $h\approx 5$ μm and ≈ 7 μm, the saturation magnetization was $4\pi M_0=1750$ G, and the ferromagnetic resonance line width was $\Delta H=0.2$–0.3 Oe.

A pump signal with frequency $f_p=3455.0$ MHz and power $P_p^{\text{in}}\leqslant 60$ mW was supplied to transducer *2* or *1*. Two other transducers were used to study the amplitude–frequency and phase–frequency characteristics of a probe signal with frequency $f_z=3$–5.6 GHz and power $P_z^{\text{in}}=0.01$–1 μW, where the lower limit was determined by

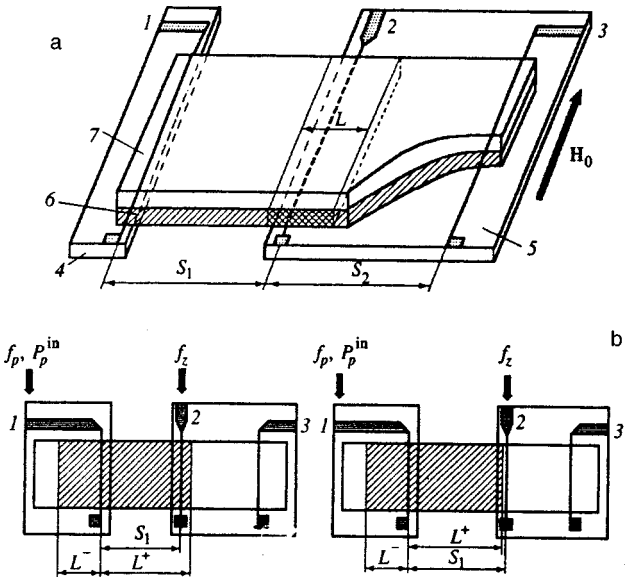


FIG. 1. Prototype of delay line. (a) Design: *1*—input transducer of probe signal f_z; *2*—input transducer of pump signal f_p to create nonequilibrium region L; *3*—output transducer; *4*—board with movable transducer; *5*—board with fixed transducers; *6*—epitaxial film of yttrium–iron garnet; *7*—gadolinium–gallium garnet substrate. (b) Determining the length L^+ of the nonequilibrium region.

the sensitivity of the measurement apparatus. The spectrum of the pump signal transmitted through the film was simultaneously monitored from transducer *3*.

The magnetic field was oriented parallel to the transducers and was varied within the limits $H_0=430$–610 Oe. First, this corresponded to the condition for excitation of magnetostatic surface waves at the frequencies of the pump and the probe signals; second, for $H_0=560$–610 Oe the limiting frequency for three-magnon decays f_{th}^{3m} was within the limits of the frequency band $f_0<f_{\text{th}}^{3m}<f_s$, where f_0 and f_s are, respectively, the long-wavelength and short-wavelength limits of the spectrum of magnetostatic surface waves in the film.[8] The conditions for three-magnon processes with the participation of parametric spin waves were satisfied in the frequency interval f_{th}^{3m}–f_s in this case, whereas four-magnon processes played the main role at frequencies f_0–f_{th}^{3m}. For a magnetostatic surface pump wave with fixed frequency f_p, three-magnon decay processes are possible for $H_0<H_0^{3m}$, while four-magnon processes are possible for $H_0>H_0^{3m}$, where H_0^{3m} corresponds to the condition $f_p=f_{\text{th}}^{3m}(H_0^{3m})$.

The limiting frequency f_{th}^{3m} and field H_0^{3m} were determined experimentally from the bound on the output power of the magnetostatic surface wave when it propagates under the conditions of three-magnon decay:[1,8] The power of the magnetostatic surface wave is bounded for frequencies $f\geqslant f_{\text{th}}^{3m}$. (The typical form of the amplitude–frequency characteristic and the position of the limiting frequency f_{th}^{3m} when frequency f_{th}^{3m} lies within the excitation-frequency band of the magnetostatic surface wave are shown in Fig. 2.) A field of $H_0^{3m}\approx 589$ Oe was obtained for the chosen pump frequency and the test film, in agreement with calculations of H_0^{3m} and f_{th}^{3m} that take into account the effect of the anisotropy and inhomogeneous-exchange fields as indicated in Ref. 11.

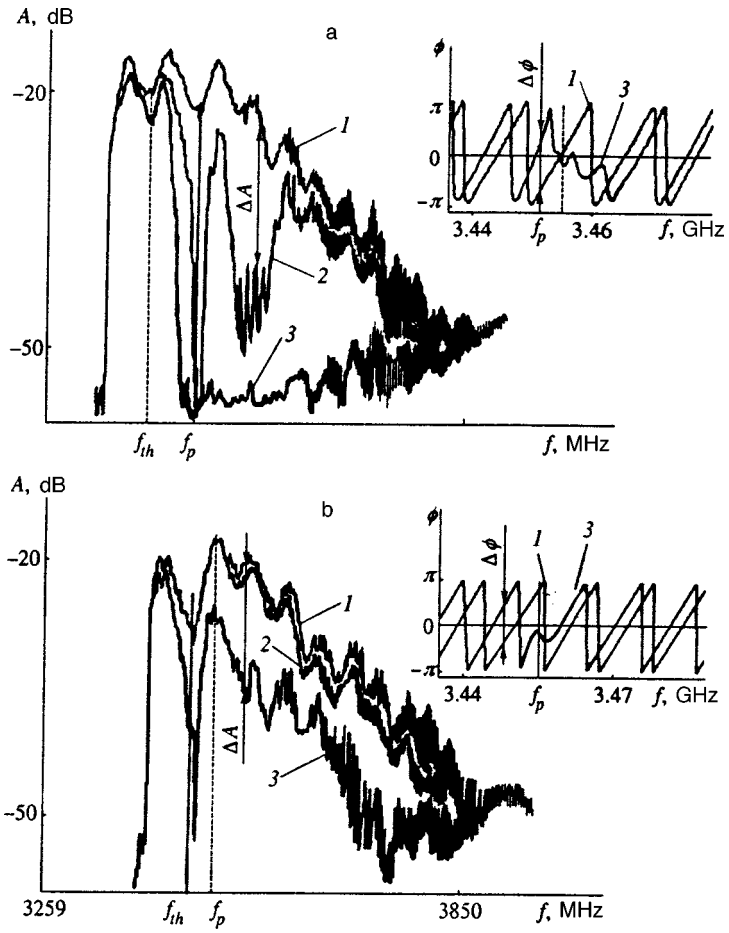


FIG. 2. Amplitude–frequency and phase–frequency (in insets) characteristics of the prototype with a magnetostatic surface wave, recorded for probe signal f_z at various power levels P_p of the pump signal ($f_p = 3455$ MHz): *1*—$P_p < P_{\text{th}}$, *2*—$P_p > P_{\text{th}}$, *3*—$P_p \gg P_{\text{th}}$. (a) $H_0 = 579$ Oe, for three-magnon decay, $f_p > f_{\text{th}}^{3m}$; (b) $H_0 = 598$ Oe, for four-magnon instability, $f_p < f_{\text{th}}^{3m}$; ΔA is the additional damping of the magnetostatic probe wave and $\Delta\phi$ is the phase change of the probe signal at the output of the prototype, both caused by parametric interaction with the parametric spin waves.

To study the effect of parametric spin waves on the dispersion and damping of the probe wave, the pump signal was fed to transducer *2*, and the surface magnetostatic probe wave was excited and detected by transducers *1* and *3*, respectively (Fig. 1a). A pump wave with a power $P_p > P_{\text{th}}$ in the path of the surface magnetostatic probe wave in this case created a nonequilibrium section of the film (see the shaded area in Fig. 1). The overall length L of the section was made up of the lengths of the nonequilibrium sections in the forward (L^+) and reverse (L^-) directions, which correspond to magnetostatic surface pump waves localized on the boundaries between the film and the air and between the film and the substrate and propagating in opposite directions from transducer *2*. When the probe wave passes through the nonequilibrium section, changes analogous to those observed in Refs. 2–6 appeared in its amplitude- -frequency and phase–frequency characteristics.

For cases in which a magnetostatic surface pump wave excites parametric spin waves as a result of three-magnon and four-magnon processes, Fig. 2 shows the behavior of the amplitude–frequency and phase–frequency characteristics as the pump supercriticality $C = 10 \log(P_p/P_{\text{th}})$ increases, where P_p corresponds to the power of the wave under the input transducer. The power P_p was determined from

$$P_p = P_R(H_{01}) - P_R(H_0), \tag{8}$$

where P_R is the power reflected from the input transducer, measured at field H_0 corresponding to the conditions for exciting a magnetostatic surface wave and field $H_{01} > f_p/\gamma$ for which the pump frequency lies below the bottom of the spectrum of spin waves. Typical powers of the magnetostatic surface pump wave were 5–50% of the incident power P_p^{in}. As in Refs. 2–6, the ''threshold power'' P_{th} refers to the pump-wave power P_p at which an absorption band close to frequency f_p is formed in the amplitude–frequency characteristic of the probe signal. Curve *1* in Fig. 3 shows how the threshold power at pump frequency f_p in the film depends on the field H_0.

The amplitude change ΔA and phase change $\Delta\phi$ of the probe wave at the frequency f_z of the probe wave were determined experimentally as shown in Fig. 2. The ΔA and $\Delta\phi$ values thus obtained were then used in Eqs. (6) and (7) to calculate the corrections to the dispersion law, $\Delta k'$, and to the damping law, $\Delta k''$, at the chosen frequency f.

To determine the length L of the nonequilibrium section, the pump signal was fed to transducer *1*, while transducers *2* and *3*, which were used to analyze the amplitude–frequency characteristic of the probe wave, were displaced along the film (Fig. 1b). It is obvious that absorption bands in the amplitude–frequency characteristic of the probe wave of the type shown in Fig. 2 will be observed only until transducer *2* is on the nonequilibrium section of the film ($S_1 < L^+$). The distance S_1 at which the pumping ceased to affect the amplitude–frequency characteristic was taken to be the length L^+ of the nonequilibrium section for a given super-

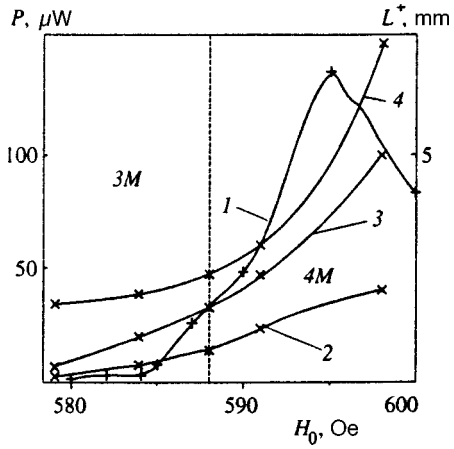


FIG. 3. Dependence on external magnetic field H_0 of the threshold power P_{th} of the decay processes (curve *1*) and of the length L^+ of the nonequilibrium region for supercriticality $C=10$ dB (curve *2*), 20 dB (curve *3*), and 30 dB (curve *4*). The dashed line shows the boundary between the regions of three-magnon and four-magnon processes.

criticality level C of the pumping. To determine the length L^- of the nonequilibrium section corresponding to a surface magnetostatic pump wave propagating along the boundary between the film and the substrate, the direction of field H_0 is changed. Because of the nonreciprocity of magnetosta tic surface waves,[1] the length L^- of the section in the reverse direction is always less than in the forward direction; it satisfied $L^- \leqslant L^+/5$ in our case. Note that, besides the nonreciprocity effect, the ratio of lengths L^+ and L^- can be affected by the difference of the film parameters at the boundaries between the film and the air and between the film and the substrate.

Curves *2–4* in Fig. 3 show the how the length L^+ of the nonequilibrium section in the forward direction depends on the field H_0 at different supercriticality levels of the pumping. It can be seen that the size of the nonequilibrium region increases as C increases and is always greater for four-magnon instability of the magnetostatic surface wave than for three-magnon decays. However, the character of the $L^+ = L^+(C)$ dependence is essentially determined by the type of instability of the magnetostatic surface wave. Under conditions of three-magnon decay of the surface magnetostatic pump wave, the length of the nonequilibrium section begins quickly to increase with increasing C in the region of values $C>20$ dB, for which kinetic instability of the spin waves is generated.[12,13,5,6] For values of $H_0>H_0^{3m}$, the increase of L^+ slows down in the region of values $C>12$ dB, where kinetic instability arises for four-magnon processes with the participation of magnetostatic surface waves.[6]

Note that for both three and four-magnon instability of a magnetostatic surface wave with frequency f_p close to the limiting frequency for three-magnon decay, spin waves are excited, propagating almost parallel to the field H_0.[2–6] Therefore, removal of the interaction products in our case must not cause a substantial difference in the length of the nonequilibrium sections for three-magnon and four-magnon instability of a surface magnetostatic pump wave. The significant increase of the size of the nonequilibrium region observed when $H_0>H_0^{3m}$ should consequently be associated with the smaller damping of the pump wave under the conditions of four-magnon instability than with three-magnon decay processes.

Figures 4 and 5 show the frequency dependence of the corrections to the dispersion law, $\Delta k'$, and the damping law, $\Delta k''$, under conditions of three- and four-magnon instability of a surface magnetostatic pump wave, calculated from the experimental values of ΔA and $\Delta\phi$ using Eqs. (6) and (7), respectively. The same figures show the spectra of the pump signal transmitted by the film.

Most importantly, we compare the form of the frequency dependence of the corrections to $\Delta k'$ and $\Delta k''$ with the form of the spectrum of the output pump signal (see Figs. 4 and 5). Under conditions of four-magnon instability of the pump wave, the character of the $\Delta k'(f)$ and $\Delta k''(f)$ dependence as a whole corresponds to the form of the spectrum: the frequency band in which the corrections to the dispersion and damping law appear corresponds to the frequency band of the existence of the noise spectrum close to the pump frequency (Fig. 5).

It can be seen from Fig. 4 that, under conditions of three-magnon decay, there can be an appreciable difference in the frequencies F_s at which satellites exist in the pump spectrum

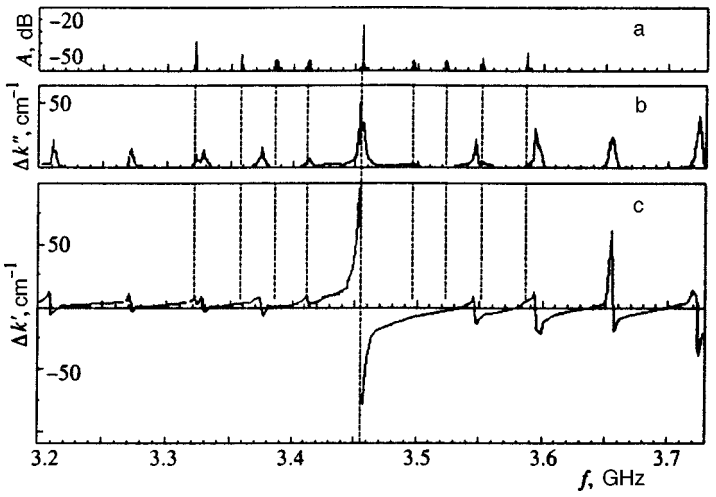


FIG. 4. Spectrum of the output pump signal (a). Frequency dependence of the additional losses $\Delta k''$ (b) and of the dispersion changes $\Delta k'$ of the magnetostatic probe wave (c) caused by interaction of the probe wave with parametric spin waves in the nonequilibrium region for $H_0=490$ Oe and $C=22.5$ dB in the case of three-magnon decay, $f_p>f_{th}^{3m}$.

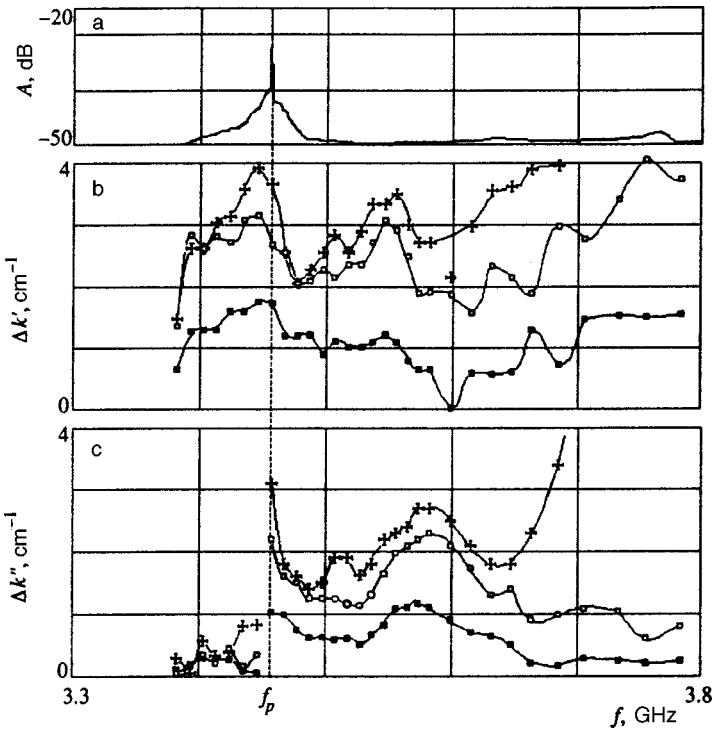


FIG. 5. Spectrum of the output pump signal (a) and frequency dependence of the additional losses $\Delta k''$ (b) and dispersion changes $\Delta k'$ of the magnetostatic probe wave (c) caused by interaction of the latter with parametric spin waves in the nonequilibrium region for $H_0 = 595$ Oe, $C = 11$ dB (filled squares), $C = 16.2$ dB (open squares), and $C = 19$ dB (crosses) in the case of four-magnon instability, $f_p < f_{\text{th}}^{3m}$.

and those at which one finds corrections to the dispersion, $\Delta k'$, and the damping, $\Delta k''$, of the surface magnetostatic probe wave. Since $\Delta k'(f)$ and $\Delta k''(f)$ differ from zero only at frequencies F_a where absorption bands of the amplitude–frequency characteristic exist, this difference is a consequence of the noncoincidence of frequencies F_s and F_a— a fact pointed out earlier in Refs. 2, 4–6. At the same time, it is to be hoped that there is a definite correspondence between the satellites and the absorption bands in the amplitude–frequency characteristic. To establish this correspondence, it was convenient to use the dependence of the frequencies F_a and F_s on the field H_0. Figure 6 shows the experimental field dependence of the frequencies of the satellites and absorption bands lying below pump frequency f_p. It can be seen that those satellites and absorption bands whose frequencies are connected by the relationship $F_a(H_0) \approx 2F_s(H_0) - f_p$ exist in the same range of magnetic fields H_0 (one such pair corresponds to curves b and b' in Fig. 6). It is to be hoped that there is a direct connection between such satellites and absorption bands.

We should point out that the satellites are secondary magnetostatic surface waves, generated on the nonequilibrium section of the film as a result of processes of blending of parametric spin waves of the form[14]

$$f_1 + f_2 = F_s, \quad \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_s. \tag{9}$$

Note that the $F_s(H_0)$ has a characteristic shape.[8,14]

As can be seen in Figs. 2–5, the instability of the pump wave increases the spatial damping rate and changes the wave number of the surface magnetostatic probe wave. The

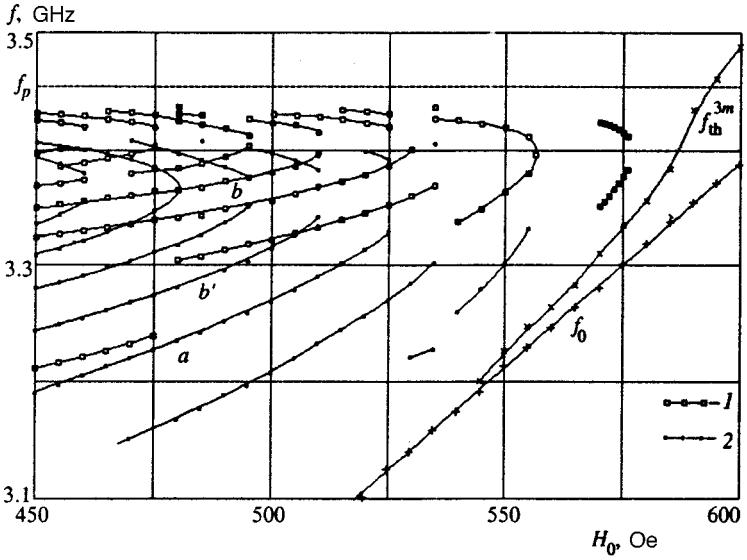


FIG. 6. Dependence on external magnetic field H_0 of the frequencies of the satellites in the spectrum of the output pump signal (*1*) and of the damping bands in the amplitude–frequency characteristic of the magnetostatic probe wave, caused by interaction with packets of parametric spin waves in the nonequilibrium region for a film with $h \approx 7$ μm. Curves b and b' are an example of parametric interdependence of the satellites and damping bands; a and b, b' are the dependences corresponding to Figs. 8a and 8b, respectively.

indicated changes occur only at those frequencies f of the probe wave at which the conditions given by Eq. (2) or (3) are satisfied, corresponding, respectively, to four-magnon or three-magnon interaction with the parametric waves excited by the pump wave. The $\Delta k'$ and $\Delta k''$ values recorded under conditions of three-magnon decay of the pump wave are an order of magnitude greater than for four-magnon instability in this case. Note also that the dispersion of the probe wave assumes an anomalous character in the frequency region where the absorption maximum is observed.

In the linear approximation, the dispersion equation for magnetostatic surface waves in a ferromagnetic film has the form[9]

$$\exp(2kd)=\frac{(2+\chi_1)^2-\chi_2^2}{\chi_1^2-\chi_2^2}, \tag{10}$$

where $\chi_1=\omega_H\omega_m/(\omega_H^2-\omega^2)$ and $\chi_2=\omega\omega_m/(\omega_H^2-\omega^2)$ are the components of the magnetic susceptibility tensor of the ferromagnet. If it is assumed that, for the probe wave, the processes given by Eqs. (2) and (3) do not change the form of the dispersion relation (10), but only affect the susceptibility of the ferromagnet, resulting in a small nonlinear addition $\delta\chi$ to the susceptibility ($\chi \gg \delta\chi$), we get for long-waveleng th ($2kd \ll 1$) magnetostatic surface waves

$$\Delta k' \sim \mathrm{Re}\,\delta\chi, \quad \Delta k'' \sim \mathrm{Im}\,\delta\chi. \tag{11}$$

Comparing the frequency dependence of $\Delta k'$ and $\Delta k''$ with Eqs. (11), we see that in essence they reflect the Kramers–Kronig relations for the real and imaginary parts of the nonlinear additions to the susceptibility.

The increase of the losses and the variation of the dispersion of the magnetostatic surface wave shown in Figs. 4 and 5 can result not only from interaction with the parametric spin waves when they pass through the nonequilibrium section, but also from the effect of the latter on the probe-wave excitation and detection processes[2,4] and from thermal heating of the film by the pump wave.[7] The effect of the parametric spin waves on the excitation and detection processes was eliminated by placing transducers *1* and *3* beyond the limits of the nonequilibrium section: $S_1 > L^-$, $S_2 > L^+$.

To estimate how thermal heating of the film affected the measured results, the dependence of ΔA and $\Delta\phi$ on the inverse duty factor U was studied by modulating the pump signal with square pulses at a frequency of $f_m = 1$ kHz. Note that the time to establish a steady-state temperature distribution over the thickness $s = 0.5$ mm of a structure consisting of a film with a substrate is $\tau \approx 0.1$ sec and can be calculated from $\tau = s^2/D$,[15] where $D = 0.02$ cm/sec^2 is the thermal diffusion coefficient. It was found that varying the inverse duty factor within the limits $1 \leqslant U \leqslant 5000$ does not cause variations of ΔA and $\Delta\phi$ to the accuracy with which these quantities are measured, better than 3% over the entire frequency range in which magnetostatic surface waves were excited.

Of course, the absence of appreciable variations of ΔA and $\Delta\phi$ by no means indicates that there is no heating of the film with microwave power—the temperature increase ΔT of the section of the film close to transducer *2*, measured with the maximum level of decreasing power by means of a thermocouple, was $\Delta T \approx 0.2\,^\circ$C. If we assume that the temperature of the entire film increased by $\Delta T \approx 0.2\,^\circ$C, and note that at room temperatures the rate of variation of the magnetization with varying temperature is ≈ 3.3 G/deg,[15] the magnetization of the entire film changes by ≈ 0.7 G. If it is assumed that the other parameters of the film remained unchanged, the increment of the real and imaginary parts of the wave numbers, obtained by solving the dispersion equation for the magnetostatic surface waves[9] with parameters corresponding to Fig. 2, are $\Delta k' \approx 0.5$–$2\ \mathrm{cm}^{-1}$ and $\Delta k'' \approx 0.001$–$0.002\ \mathrm{cm}^{-1}$. Such changes of the wave numbers in our case will correspond to an amplitude change of $\Delta A \approx 0.1$–0.2 dB and a phase change of $\Delta\phi \approx 30$–$90°$. A comparison of these ΔA and $\Delta\phi$ values with those observed in experiment (see Fig. 2) shows that the contribution of heating to the observed phase change of the magnetostatic surface wave could be $\approx 50\%$ in a number of cases, while the effect of heating on signal attenuation should be negligible. However, no phase change of the magnetostatic surface wave associated with heating appeared in experiments with a modulated pump signal. Such a discrepancy can be explained by assuming that only the part of the film close to transducer *2* is heated. If one starts from an accuracy of 3% in determining $\Delta\phi$, the heated region must not exceed ≈ 1 mm for the case corresponding to Fig. 2, where $\Delta\phi \approx 130°$–$300°$. Since the thermocouples used for the measurements themselves had a size of ≈ 0.5 mm, we could convince ourselves only that the heated region in our case does not exceed 2 mm.

3. EFFECT OF A SURFACE MAGNETOSTATIC PROBE WAVE ON THE DISTRIBUTI ON OF PARAMETRIC SPIN WAVES IN (ω,**k**) SPACE

The interaction of the surface magnetostatic probe wave and the parametric spin waves created by pumping can manifest itself not only in a change in the dispersion and in damping of the magnetostatic surface wave but also in reshaping of the distribution of the parametric spin waves in $(\omega,\mathbf{k})$ space. If it is assumed that, in the absence of a surface magnetostatic probe wave, a pump wave with frequency f_p and power $P_p > P_{\mathrm{th}}$ has established a steady-state distribution $n_0(\omega,\mathbf{k})$ of parametric spin waves in $(\omega,\mathbf{k})$ space, the processes given by Eqs. (2) and (3) with the participation of a probe wave can not only change the number of already existing parametric spin waves, $n(\omega,\mathbf{k}) = n_0(\omega,\mathbf{k}) + \delta n(\omega,\mathbf{k})$, but can also cause them to appear in other regions of $(\omega,\mathbf{k})$ space. When this happens, new frequency components can appear or new amplitude variations of the already existing satellites can be observed in the spectrum of the output pump signal. Of course, the indicated changes in the spectrum of the output signal are possible if the $\delta n(\omega,\mathbf{k})$ values substantially exceed the level of thermal spin waves $n_T(\omega,\mathbf{k})$ [$\delta n(\omega,\mathbf{k}) \gg n_T(\omega,\mathbf{k})$].

When the delay-line prototype shown in Fig. 1a is used, the effect of the probe wave on the distribution of the parametric spin waves created by the pumping will be determined by the power P_z and frequency f_z of the probe signal, as well as by the distance S_1 between the transducers that excite the probe wave and the surface magnetostatic pump

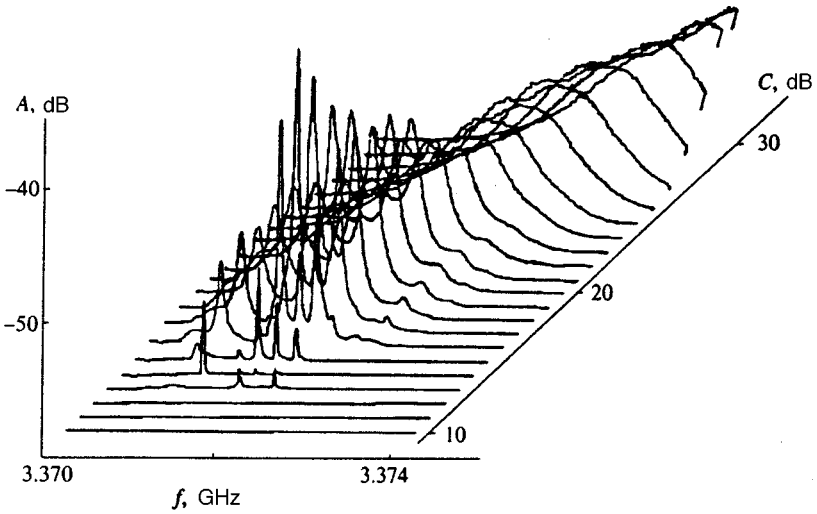


FIG. 7. Variation of the spectral structure of one of the satellites as a function of the pumping supercriticality C at $H_0=480$ Oe.

wave. The value of S_1 determines not only the losses of the probe wave on the way from transducer *1* to the boundaries of the nonequilibrium region of the film but also the mechanism by which the probe wave affects the distribution of the parametric spin waves established by pumping. Actually, when the power of the probe wave is greater than the threshold value close to transducer *1*, a nonequilibrium region of size L_z also arises, containing parametric spin waves created by the probe wave. If $S_1<L_z^+ + L_p^-$ holds the distribution of the parametric spin waves generated by the pump wave can be altered not only by the surface magnetostatic probe wave itself, but also by the parametric spin waves created by it.[16]

Here we consider the results for a film with $h\approx 7\ \mu$m, obtained at such a distance S_1 that the nonequilibrium regions close to transducers *1* and *2* did not overlap no matter what P_p and P_z values were used: $S_1>L_z^+ + L_p^-$. Moreover, the frequencies of the probe wave and the pump wave were chosen to be different, $f_z \neq f_p$, and such that three-magnon decay processes were allowed for a magnetostatic surface wave at these frequencies. The power of the probe wave varied within the limits $P_z=0.1$–$100\ \mu$W.

Note that when $f_z=f_p$ holds adding the powers of the probe wave P_z and the pump wave P_p gives an effect analogous to simply varying the supercriticality level of the pumping. The processes that occur here in the system of parametric spin waves were studied in Refs. 2–6, 8, 14 and 17 for the case of three-magnon decay of a magnetostatic surface wave. It was shown that, for magnetic fields $H_0=430$–600 Oe and a supercriticality level of $C\approx 5$–25 dB, satellites separated from the pump frequency by $\Delta F_s=|f_p-F_s| \approx 5$–200 MHz appear in the spectrum of the signal transmitted by the film.[8] Since these satellites result from threshold-less processes involving the blending of parametric spin waves of the form of Eq. (9), the amplitude A_s of the satellites is associated with the numbers of parametric spin waves in the (ω_1,k_1) and (ω_2,k_2) regions by

$$A_s \propto n_{01}(\omega_1,k_1)n_{02}(\omega_2,k_2). \tag{12}$$

It is significant that, as the field H_0 increases, the satellite widths vary within the range 0.5–8 MHz and that fine structure becomes distinguisha ble in them at some field. In Fig. 7, for C values close to the threshold where a satellite appears, one can clearly see this structure as a set of narrower satellites with width $\Delta\Omega\sim 10$–100 kHz, unstable because of parasitic deviations of the generator frequency and magnetic-field fluctuations.

It can be assumed that, for the chosen experimental parameters ($f_z\neq f_p$, $S_1>L_p^- + L_z^+$), the surface magnetostatic probe wave will have an appreciable effect on processes in the system of parametric spin waves in combination with the pump wave if two conditions are satisfied: First, a region of $(\omega,\mathbf{k})$ space must be found where the spin waves are simultaneously in parametric resonance with both the surface magnetostatic pump wave and the probe wave. Second, the indicated spin waves at a pump level of $C\approx 5$–25 dB must be close to losing stability.

The necessity of the second condition is associated with the chosen experimental technique, in which the three-magnon decay processes are judged from the spectra of the output signal or from the change of the amplitude–frequency and phase–frequency characteristics of the probe wave. Then in the indicated supercriticality interval, on one hand, the pumping excites parametric spin waves in rather narrow regions of $(\omega,\mathbf{k})$ space, and, on the other hand, the interaction of the parametric spin waves with each other is not yet great enough to generate a noise spectrum—the ''kinetic instability'' of Refs. 12, 5 and, 6.

In order to choose experimental parameters at which the first requirement is satisfied, it was sufficient at a fixed value of H_0 to set the frequency of the probe wave within the frequency interval of one of the absorption bands in the amplitude–frequency characteristic, $f_z=F_a$. In this case, for a distance between the transducers of $S_1=0.1$–4 mm, amplitude changes whose character substantially depended on their width and structure were observed at the satellites under the influence of the probe signal.

When the width of the satellites was less than 1–2 MHz, a change was observed in the amplitude of the satellites as a whole. The amplitude of the satellite directly associated with the chosen absorption band in the amplitude–frequency characteristic and separated from the pump frequency by ΔF_s

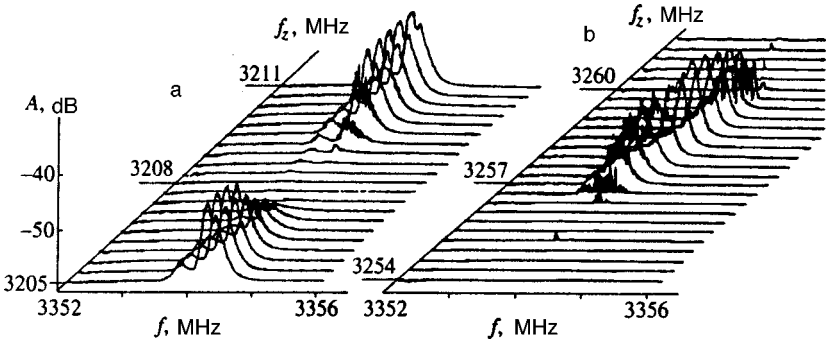


FIG. 8. Behavior of the spectrum of the output pump signal in the neighborhood of the satellite frequency for $H_0=457$ Oe, $f_p=3455$ MHz, $S_1=4$ mm when the probe-signal frequency f_z changes in a neighborhood that does not correspond to the given satellite of the absorption band for $C=20.3$ dB, $C_z=-2$ dB (a); and that does correspond to the satellite of the absorption band for $C=16$ dB, $C_z=-2$ dB (b).

$\approx|F_a-f_p|/2$ could be increased in this case by $\delta A_s\approx 1-18$ dB. The amplitudes of all the other satellites could be decreased at the same time by $\delta A_s\approx 1-18$ dB. To illustrate this, Fig. 8 shows how the signal spectrum of the satellite with central frequency $F_s=3354.6$ MHz depends on frequency f_z in a field of $H_0=457$ Oe (curve *c* in Fig. 6). Figure 8a illustrates the situation in which the frequency f_z passed through a series of values in the neighborhood of the absorption band $F_a\approx 3208$ MHz that had no direct connection with the chosen satellite (curve *a* in Fig. 6). Figure 8b illustrates the character of the changes in the spectrum of the satellite for probe-signal frequencies corresponding to the absorption band with central frequency $F_a\approx 3258$ MHz, which is directly associated with the chosen satellite (curve b' in Fig. 6). Note that an interaction was observed for these parameters even when C_z was reduced all the way to -6 dB.

When the satellite width was within 2.5–8 MHz, the presence of the fine structure rendered the interaction ambiguous, and so the amplitude was observed to increase for $f_z=F_a$ in a frequency interval less than the satellite width. However, for $f_z\neq F_a$ suppression of the satellite also occurred in a narrow frequency interval, whereas the rest of the satellite could increase its amplitude.

4. CONCLUSION

The following conclusions can be drawn from the experiments that have been carried out:

1. The main contribution to the variation of the dispersion and damping of a magnetostatic surface wave on a section of film subjected to the action of microwave pumping in the form of an additional magnetostatic surface wave at a frequency of f_p comes from three-magnon and four-magnon interaction processes involving parametric spin waves excited by the pumping. For identical supercriticality levels of the pumping, the variation of the dispersion and damping of the magnetostatic surface wave is an order of magnitude higher for three-magnon interaction than with four-magnon interaction.

2. Under conditions of three-magnon instability of the pumping and the probe wave, the distribution of the parametric spin waves in $(\omega,\mathbf{k})$ space can change substantially at the same time as the dispersion and damping of the probe wave.

This work was partially supported by the Russian Fund for Fundamental Research (Grant 96-02-16168A) and the Ministry of Science and Technical Policy of the Russian Federation (Grant 3.911/2-96).

*)E-mail: fil@sfire.saratov.su

[1] A. G. Gurevich and G. A. Melkov, *Magnetization Oscillations and Waves*, CRC Press, Boca Raton, FL, 1996 (Russ. orig. Nauka, Moscow, 1984, 464 pp.).
[2] O. A. Chivileva, A. N. Anisimov, A. G. Gurevich *et al.*, Pis'ma Zh. Tekh. Fiz. **13**, 1497 (1987) [Sov. Tech. Phys. Lett. **13**, 627 (1987)].
[3] A. V. Vashkovskiĭ, V. N. Zubkov, E. G. Lokk, and S. A. Nikitov, Fiz. Tverd. Tela **30**, 827 (1988) [Sov. Phys. Solid State **30**, 475 (1988)].
[4] A. N. Anisimov, O. A. Chivileva, and A. G. Gurevich, Fiz. Tverd. Tela **32**, 1622 (1990) [Sov. Phys. Solid State **32**, 946 (1990)].
[5] G. T. Kazakov, A. V. Kozhevnikov, and Yu. A. Filimonov, Pis'ma Zh. Tekh. Fiz. **21**, No. 14, 47 (1995) [Tech. Phys. Lett. **21**, 558 (1995)].
[6] G. T. Kazakov, A. V. Kozhevnikov, and Yu. A. Filimonov, Fiz. Tverd. Tela **39**, 330 (1997) [Phys. Solid State **39**, 288 (1997)].
[7] I. V. Krutsenko, G. A. Melkov, and S. A. Ukhanov, Fiz. Tverd. Tela **26**, 3433 (1984) [Sov. Phys. Solid State **26**, 2061 (1984)].
[8] A. M. Mednikov, Fiz. Tverd. Tela **23**, 242 (1981) [Sov. Phys. Solid State **23**, 136 (1981)].
[9] R. W. Damon and J. R. Eshbach, J. Phys. Chem. Solids **19**, 308 (1961).
[10] P. E. Zil'berman, G. T. Kazakov, and V. V. Tikhonov, Radioélektronika **30**, 1164 (1985).
[11] B. Hillebrands, Phys. Rev. B **37**, 9885 (1988).
[12] A. V. Lavrinenko, V. S. L'vov, G. A. Melkov, and V. B. Cherepanov, Zh. Éksp. Teor. Fiz. **81**, 1022 (1981) [Sov. Phys. JETP **54**, 542 (1981)].
[13] G. M. Dudko, G. T. Kazakov, A. V. Kozhevnikov, and Yu. A. Filimonov, in *Abstracts of Reports of the Regular Conference on Microwave Spin-Wave Electronics*, Krasnodar, 1987, p. 119.
[14] A. G. Temiryazev, Fiz. Tverd. Tela **29**, 313 (1987) [Sov. Phys. Solid State **29**, 179 (1987)].
[15] G. T. Kazakov, M. L. Kats, A. G. Sukharev, and Yu. A. Filimonov, Zh. Tekh. Fiz. **62**, No. 11, 115 (1992) [Sov. Phys. Tech. Phys. **37**, 1103 (1992)].
[16] G. A. Melkov, Fiz. Tverd. Tela **17**, 1728 (1975) [Sov. Phys. Solid State **17**, 1123 (1975)].
[17] G. A. Melkov and S. V. Sholom, Zh. Éksp. Teor. Fiz. **96**, 712 (1989) [Sov. Phys. JETP **69**, 403 (1989)].

Translated by W. J. Manthey

Development of stochastic oscillations in a one-dimensional dynamical system described by the Korteweg–de Vries equation

A. V. Gurevich*) and K. P. Zybkin

I. E. Tamm Division of Theoretical Physics, P. N. Lebedev Physical Institute, Russian Academy of Science, 117924 Moscow, Russia

G. A. Él'

Institute of Earth Magnetism, the Ionosphere, and Radio Wave Propagation, Russian Academy of Sciences, 142092 Troitsk, Moscow Oblast', Russia

(Submitted 15 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 333–360 (January 1999)

The behavior of the solution of the Korteweg–de Vries equation for large-scale oscillating aperiodic initial conditions prescribed on the entire x axis is considered. It is shown that the structure of small-scale oscillations arising in a Korteweg–de Vries system as $t\to\infty$ loses its dynamical properties as a consequence of phase mixing. This process can be called the generation of soliton turbulence. The infinite system of interacting solitons with random phases developing under these conditions leads to oscillations having a stochastic character. Such a system can be described using the terms applied to a continuous random process, the probability density and correlation function. It is shown that for this it suffices to determine from the prescribed initial conditions amplitude distribution function of the solitons and their mean spatial density. The limiting stochastic characteristics of the mixed state for problems with initial data in the form of an infinite sequence of isolated small-scale pulses are found. Also, the problem of stochastic mixing under arbitrary initial conditions in the dispersionless limit (the Hopf equation) is completely solved.
[S1063-7761(99)02701-8]

1. INTRODUCTION

The main features of the structure of the asymptotic solutions of the Korteweg–de Vries (KdV) equation

$$u_t+6uu_x+\varepsilon^2u_{xxx}=0 \tag{1}$$

are determined by the form of the initial potential $u(x,0)=u_0(x)$.

In the classical theory of the integration of this equation by the inverse-scattering method[1] a necessary condition is that the function $u_0(x)$ vanish with sufficient rapidity as $|x|\to\infty$ (localized initial data). The asymptotic solution of this problem in the limit $t\to\infty$ describes an ordered chain of diverging solitons defined by the discrete spectrum for the quantum-mechanical problem of scattering by the potential $-u_0(x)$. The contribution of the continuous spectrum is represented by the oscillating wave ''tail,'' whose amplitude falls as $t^{-1/2}$. In the semiclassical case $\varepsilon^2\ll1$ (Refs. 2 and 3) the number of solitons (of order $1/\varepsilon$), although large, is fixed, and the asymptotic limit of the solution as $t\to\infty$ remains regular regardless of the detailed form of the function $u_0(x)$.

The initial data $u_0(x)$ in the form of a smooth function with different asymptotic limits at infinity $u_0(-\infty)>u_0(+\infty)$ lead to the appearance of a continuously expanding oscillator region.[4]

An important difference from the classical theory is that the number of solitons now grows without bound in time. However, their structure as $t\to\infty$ remains regular.

A fundamental difference from the cases described above is the solution of the periodic problem

$$u_0(x)=u_0(x+X_0),$$

where X_0 is the period. This problem can be integrated exactly only if $u_0(x)$ is chosen from a special class of periodic functions called finite-band potentials.[1,5–8] The asymptotic dynamics of the oscillations in the periodic problem can have a very complicated character (in the case $\varepsilon\ll1$ the number of nonlinear modes is large, of order $1/\varepsilon$), which has the outward appearance of a disordered system of waves of different amplitudes.

Despite this, the exact solution of this problem always manifests its dynamical nature explicitly, strictly conserving its periodicity in space: the condition $u(x,t)=u(x+X_0,t)$ is fulfilled at all times.

The aim of the present work is to examine the asymptotic behavior of the solutions of the KdV equation in the semiclassical limit $\varepsilon\ll1$ for aperiodic initial functions $u_0(x)$ oscillating on the entire x axis. A more precise definition of the class of initial data will be given below.

As will be shown, this case differs fundamentally from all solutions of the KdV equation considered previously. Although at any finite time t the structure of the solution, as before, has a dynamical character, its asymptotic limit as $t\to\infty$ completely loses its dynamical properties as a consequence of ergodic phase mixing. The infinite system of in-

teracting solitons with random phases developing in this case gives rise to oscillations having a stochastic character. Such a system can be described in terms of a continuous random process. Specifically, in the asymptotic limit $t\to\infty$ the exact dynamical value of $u(x,t)$ loses meaning and it is possible only to speak of $f(u;x,t)$, the probability density of finding at the point x,t the given value u in the interval $[u,u+du]$ or of $f(u,u';x,t,x',t')$, the two-point distribution function, i.e., the joint probability density of having values $[u,\ u+du\ ]$ at the point x,t and values $[u',\ u'+du']$ at the point x',t'; or of a three-point distribution function, etc. It is remarkable that the possibility arises here, as we will see below, of a significant simplification in the description of the asymptotic behavior of the solution. In particular, for a spatially homogeneous initial function $u_0(x)$ the probability density $f_\infty(u)$ depends only on the velocity u and the spatial correlation function $\mathbf{K}=\mathbf{K}(s)$, where $s=|x-x'|$.

Note that the evolution of random initial data in integrable systems has been investigated in a number of earlier works; however, they only considered the influence of a small stochastic perturbation on the soliton solutions (see, e.g., Refs. 9 and 10). The question of the emergence of a stochastic regime as a result of the purely dynamical evolution of an integrable system is posed in the present work for the first time.

This paper is organized as follows. Section 2 formulates the initial-va lue problem for a spatially homogeneous oscillating function $u_0(x)$. It then considers its solution in the dispersionless limit, described by the Hopf equation, containing neither dissipation nor dispersion. In this case, nonlinear reversal gives rise to multistream flows. Here the number of streams, and consequently their density in velocity space (u) increases continuously with time. Therefore, as $t\to\infty$ to first order in $O(1/t)$ a steady-state velocity distribution function of the streams, $f(u)$, is set up over all space. A general solution of the problem is found, allowing one to determine the form of the distribution function $f(u)$ for any initial function $u_0(x)$. The asymptotic behavior of the correlation functions is determined, which shows that as $t\to\infty$ in fact a distribution that is completely uncorrelated in space is set up.

Section 3 investigates the semiclassical KdV problem for the initial condition considered, $u_0(x)$. As a consequence of the presence of the dispersion term $\varepsilon^2 u_{xxx}$ multistream flows never arise here. However, in the vicinity of the reversal points of the original profile bands of regular small-scale (with period $\sim\varepsilon$) oscillations appear. With the passage of time the oscillator regions expand and overlap. It is shown that the structure arising as $t\to\infty$ can be represented as a system of random interacting solitons homogeneous in space–time. They are described by a soliton amplitude distribution function $f(a)$, and also by a mean soliton density in space (on the x axis), which defines the intensity of their interaction. Specific examples of the calculation of the function $f(a)$ and the mean soliton density are given.

Section 4 considers the case in which the initial function $u_0(x)$ consists of an infinite sequence of isolated pulses. The steady-state velocity distribution function $f_\infty(u)$ is found, which, on the one hand, has certain features in common with the distribution function $f(u)$ obtained for the same initial-value problem in the Hopf limit but, on the other hand, differs from it significantly. Finally, Sec. 5 determines for the same problem the steady-state correlation function **K(s)** set up as $t\to\infty$ and, correspondingly, the spectral power of the process.

The Conclusion summarizes the main features of the process—the generation of soliton turbulence in the dynamics of a continuous one-dimensional nondissipative medium, described by the integrable KdV equation. In addition, we provide a brief comparison with other known mechanisms for the occurrence of a chaotic state in dynamical systems. Of course, a rigorous basis of the stochastic systems considered here requires a special mathematical study.

Note also that integration of the KdV equation by the inverse-scattering method establishes a direct connection between the dynamical theory considered here and the spectral theory of disordered systems, one subject of which is the quantum-mechanical Schrödinger equation with so-called ''metrically transitive random potentials.''[11,12] It may therefore be supposed that the methods developed here may be of interest not only for nonlinear dynamics, but also for the theory of such quantum systems.

2. DISPERSIONLESS LIMIT

The main features of ergodic mixing can be illustrated by the example of dispersionless evolution, described by the Hopf equation

$$\partial_t u + u\partial_x u = 0, \quad u(x,0)=u_0(x), \tag{2}$$

where $u_0(x)$ is an arbitrary smooth oscillating aperiodic function prescribed on the entire x axis. We also assume that the distances l between the zeros are contained in the interval

$$l_{\min}\leqslant l\leqslant l_{\max}, \tag{3}$$

where $l_{\min}$ and $l_{\max}$ are arbitrary finite values. The function $u_0(x)$ has a finite maximum $u_{\max}$ and a finite minimum $u_{\min}$:

$$u_{0\min}\leqslant u_0(x)\leqslant u_{0\max}. \tag{4}$$

We also assume that $u_0(x)$ is spatially homogeneous, i.e., that there exists a scale, as large as desired but finite, starting from which all the basic properties of the function $u_0(x)$ repeat. This latter condition excludes the possibility of singular behavior of the function $u_0(x)$ at infinity. It is also necessary that the value

$$u_0(x)=u_{0\min} \tag{5}$$

be repeated not less than once on the scale L. Conditions (2)–(5) are satisfied by almost periodic or quasiperiodic functions $u_0(x)$.

The evolution (2), (3), as is well known,[13] leads after a finite time $t=t_k$ to the appearance of a singularity [Fig. 1a]

$$\left|\frac{\partial u}{\partial x}\right|_{x=x_k}\to\infty. \tag{6}$$

In the absence of dissipation and dispersion, multistreaming motion in the system is possible. In this case, for $t>t_k$ in the

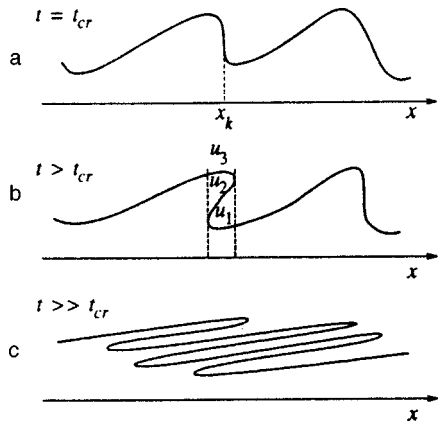


FIG. 1. Appearance of multistream flows in the Hopf equation.

vicinity of the singularity (6) three flows form: $u_i (i=1,2,3)$, each of which as before is described by the Hopf equation

$$\partial_t u_i + V_i \partial_x u_i = 0, \quad V_i = u_i . \tag{7}$$

At the singular points—the caustic points defined by condition (6)—the flows join together (see Ref. 14) and the multistream system closes [Fig. 1(b)]. Such a solution is of course equivalent to the implicit solution of the Hopf equation

$$x = ut + x_0(u), \tag{8}$$

describing the evolution $u(x,t)$ in terms of multivalued functions. Here $x_0(u)$ is the inverse function of $u_0(x)$. Note that another, fundamentally different, approach to the solution of the Hopf equation viewed as a limit of the dissipative Burgers equation was developed by Sinai.[15]

For the initial-value problem (2), as one can easily convince oneself, the developing regions of three-stream flow gradually expand with time t for $t > t_k$; then as a result of intersection of the regions of reversal of the original profile $u_0(x)$ regions of five-stream flow, seven-stream flow, etc., successively appear [Fig. 1c]. To describe the process of multiplication of streams, it is convenient to use the stream number density distribution function $f(u,x,t)$ in the phase space (u,x) of the system (2) satisfying the Liouville equation

$$\frac{\partial f}{\partial t} + u \frac{\partial f}{\partial x} = 0. \tag{9}$$

The dynamical formulation corresponds to initial conditions in the form of a δ-function:

$$f_0(u,x,0) = \delta(u - u_0(x)). \tag{10}$$

In our case $u_0(x)$ is an oscillating aperiodic function (2) obeying conditions (3)–(5).

The system (2), (10) has the obvious solution

$$f(u,x,t) = f_0(u,x,t) = \delta(u - u_0(x - ut)). \tag{11}$$

The zeros of the δ function are the streams u_i in the arising multistream system (7). These streams are defined as the branches of the multivalued curve $u(x,t)$ prescribed by the implicit solution of the Hopf equation (8). The solution (11) in this case can be represented in the form[14]

$$f(u,x,t) = \sum_{i=1}^{N} \rho_i(x,t)\, \delta(u - u_i(x,t)), \tag{12}$$

where N is the total number of streams at the given point x at the time t, and $\rho_i(x,t)$ is the relative weight (or density) of the ith stream:

$$\rho_i(x,t) = \frac{1}{1 + |dx_0/du|_i^{-1} t}. \tag{13}$$

It follows from Eq. (13) that the weight ρ_i at any time t is determined directly by the initial function $u_0(x)$.

We will point out now the main asymptotic properties of the solution (11), (12) for the initial function (2).

1. The number of streams N at any point x as $t \to \infty$ grows in direct proportion to the time:

$$N \propto a\, t, \quad \frac{\langle |u_0| \rangle}{l_{\min}} < a < \frac{\langle |u_0| \rangle}{l_{\max}}, \tag{14}$$

where $\langle |u_0| \rangle$ is the mean value of the magnitude of the initial function and $l_{\min}$ and $l_{\max}$ are the half-periods, defined by condition (3).

2. At any point x for any value u belonging to the interval (4) on which the initial function u_0 is defined, a stream u_i is always found such that

$$|u - u_i| < \delta, \quad \delta = O(1/t). \tag{15}$$

3. The asymptotic limit for each stream u_i for $t \gg 1$ follows from Eqs. (8) and (12):

$$\frac{\partial u_i}{\partial t} = O(1/t), \quad \frac{\partial u_i}{\partial x} = O(1/t).$$

By virtue of the indicated properties, as follows from Eqs. (12)–(15), in the asymptotic limit $t \to \infty$ the distribution function $f(u,x,t)$ can be represented in the form

$$f(u,x,t) = f(u) + O(1/t).$$

Here

$$f(u) = \lim_{t \to \infty} \left\{ \frac{1}{t} \sum_{i=1}^{N} \left| \frac{dx_0}{du} \right|_i \delta(u - u_i) \right\}, \quad N \propto t \ \ (t \to \infty), \tag{16}$$

and the sum is over all roots of Eq. (8) in the limit $t \to \infty$. The asymptotic limit (16) implies that the distribution function (11), prescribed at the initial time $t = 0$ in the form of a δ function, i.e., one stream at each point x, washes out with the passage of time due to the multiplication of streams and, as $t \to \infty$, goes over in the $1/t$ approximation to a distribution that is uniform in space (on the x axis) and smeared over the entire interval of initial velocities (4): $f(u)$. This is the process of ergodic mixing.

To determine the function $f(u)$, it is convenient to average the Liouville equation (9) over the spatial interval $L \gg l_{\max}$. Defining

$$\overline{f}(u,t)_L=\frac{1}{L}\int_0^L f(u,x,t)\,dx, \tag{17}$$

we find from Eq. (9)

$$\frac{\partial\overline{f}(u,t)}{\partial t}=-\frac{u}{L}[f(u,L,t)-f(u,0,t)].$$

Hence it follows that as $L\to\infty$

$$\left.\frac{\partial\overline{f}(u,t)}{\partial t}\right|_{L\to\infty}=0$$

and, consequently, the averaged function does not depend on time:

$$\overline{f}(u,t)_{L\to\infty}=\overline{f}(u). \tag{18}$$

This means that the limiting function $f(u)$ can be calculated directly at the initial time $t=0$ by direct spatial averaging of the initial function (11):

$$\begin{aligned}f(u)=\overline{f_0}(u)&=\lim_{L\to\infty}\frac{1}{L}\int_0^L\delta(u-u_0(x))\,dx\\&=\lim_{L\to\infty}\frac{1}{L}\int_0^L\sum_{i=1}^{N(L)}\frac{\delta(x-x_i)}{|u_0'(x_i)|}dx\\&=\lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{N(L)}\frac{1}{|u_0'(x_i)|}=\left\langle\frac{1}{|u_0'|}\right\rangle_{x_i}.\end{aligned} \tag{19}$$

In the last expression the angle brackets $\langle\ldots\rangle_{x_i}$ denote averaging over the ensemble of roots $x_i=x(u)$ of Eq. (2).

In other words, the values $|u_0'|$ are taken at the points $x_i(u)$ defined by the relation

$$u=u_0(x). \tag{20}$$

We emphasize that the existence of limits (15), (17), and (18) is ensured by ergodicity of the initial function (2)–(5) (for an exact definition of ergodicity, see, for example, Ref. 16).

The two-point distribution function and also higher probability characteristics of the mixed state can be found in an analogous way. Indeed, taking into account that the limiting function $f(u)$ does not depend on x, one can readily convince oneself that the two-point distribution function $F_2(v_1,v_2;x',x'';t)$ also possesses the same property in the mixed state. Indeed, the function F_2 satisfies a Liouville equation analogous to (9):

$$\frac{\partial F_2}{\partial t}+v_1\frac{\partial F_2}{\partial x'}+v_2\frac{\partial F_2}{\partial x''}=0 \tag{21}$$

with initial conditions

$$F_2(v_1,v_2;x',x'';0)=\delta(v_1-u_0(x'))\,\delta(v_2-u_0(x'')). \tag{22}$$

In Eq. (21) it is convenient to transform to the new variables

$$s=x'-x'',\quad x=x''.$$

Thus it takes the form

$$\frac{\partial F_2}{\partial t}+(v_1-v_2)\frac{\partial F_2}{\partial s}+v_2\frac{\partial F_2}{\partial x}=0.$$

By virtue of the spatial homogeneity of the problem, we average the function F_2 over x on the interval $[0,L]$. For the averaged function $\langle F_2\rangle_L$ in this case we obtain the equation

$$\frac{\partial\langle F_2\rangle_L}{\partial t}+(v_1-v_2)\frac{\partial\langle F_2\rangle_L}{\partial s}=O(1/L) \tag{23}$$

with the initial condition

$$\langle F_2\rangle_L(v_1,v_2;s;0)=\frac{1}{L^2}\int_0^L\delta(v_1-u_0(x))\,\delta(v_2-u_0(x+s))\,dx. \tag{24}$$

The solution of Eq. (23) is easily found and has the form

$$\langle F_2\rangle_L(v_1,v_2,s,t)=\frac{1}{L^2}\int_0^L\delta(v_1-u_0(x))\,\delta(v_2-u_0(x+s-(v_2-v_1)t))\,dx+O(1/L). \tag{25}$$

Expression (25) can be represented as a sum of δ functions:

$$\langle F_2\rangle_L(v_1,v_2,s,t)=\frac{1}{L^2}\sum_{a,b}^{N(L)}\frac{\delta(s+x_0^a(v_1)-x_0^b(v_2)-(v_2-v_1)t)}{|u_0'(x_0^a(v_1))|\,|u_0'(x_0^b(v_2))|}, \tag{26}$$

where, as before, $x_0^a(u)$ are the roots of the equation $u_0(x_0^a)=v_1$. Physically, the meaning of the argument of the δ function is obvious. It describes the increase with time of the distance s between correlated neighboring points x_1 and x_2 assuming that at the initial time $t=0$ these points were found a distance $s_{ab}=x_0^a-x_0^b$ apart. For late enough times, if $v_1\neq v_2$, the initial distance s_{ab} can be neglected. Thus from Eq. (26) in the limit $L\to\infty$ we obtain

$$\langle F_2\rangle(v_1,v_2,s,t)=\delta(s-(v_1-v_2)t)f(v_1)f(v_2),\quad v_1\neq v_2, \tag{27}$$

is the single-particle where $f(v_1)$ and $f(v_2)$ are single-particle distribution functions (19). Hence it can be seen that the pairwise correlation function in our case has in fact the nature of a δ correlation and is completely determined by the single-particle distribution. As can be seen from relation (27), the spatial correlation functions are proportional to $(1/t)\,\delta(s/t-(v_1-v_2))$ and, consequently, for any finite s decay with time as $1/t$.

Note that in the derivation of relation (27) it was assumed that the initial function $u_0(x)$ is aperiodic. It can be easily seen that in the case of a periodic function we obtain analogous relations, but expression (27) will contain a sum of terms $\delta(s-(v_2-v_1)t+X_0n)$, where n is some integer and X_0 is the period. Analogous relations also obtain for the higher correlation functions F_n, which are described by the equation

$$\frac{\partial F_n}{\partial t}+v_1\frac{\partial F_n}{\partial x_1}+\ldots+v_n\frac{\partial F_2}{\partial x_n}=0.$$

Thus, in the limit $t\to\infty$ there exists only one nonzero function $f(u)$. This implies that the initial distribution $u_0(x)$ becomes completely washed out. The limiting mixed state (16), (19) is thus completely uncorrelated.

Let us consider some examples of the mixing process.

Example 1. Initial data in the form of a periodic function.

We assign the initial function in the form

$$u_0 = a\sin x. \tag{28}$$

It follows from Eqs. (28) and (19) that $x=\arcsin(u/a)$ and

$$f(u)=\frac{C}{a\,|\cos x(u)|}=\frac{1}{\pi\sqrt{a^2-u^2}}. \tag{29}$$

The normalization constant in formula (29) is chosen from the condition

$$\int_{-a}^{a} f(u)\,du=1.$$

We emphasize that although the probability $f(u)$ defined by formula (29) for the periodic boundary condition $u_0(x+X_0)=u_0(x)$ has a completely real meaning, total stochastic mixing in this case does not occur: for example, the pairwise correlation function for $v_1=v_2$ and all higher correlation functions have, to first order in $1/t$, infinite peaks for $x_k-x_m=nX_0$ (26).

Example 2. Quasiperiodic initial conditions.

Let the initial velocity be prescribed in the form

$$u_0(x)=(a_0+a_1\sin(kx))\sin x. \tag{30}$$

where k is an arbitrary irrational number, $k\ll 1$. Separating the fast and slow variables, we represent the function $u_0(x)$ as

$$u_0(x)=a(X)\sin x,$$

where

$$X=k\,x,\quad a(X)=a_0+a_1\sin X.$$

Thus, to first order in $O(k)$ we obtain for the function $f_\infty(u)$

$$f(u)=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{dX}{\sqrt{a^2(X)-u^2}}. \tag{31}$$

This function can be expressed in terms of elliptic integrals.

Example 3. Infinite sequence of isolated pulses.

We consider the initial dynamical problem $u(x,0)$ in the form of an infinite sequence of pulses of identical shape F:

$$u(x,0)=u_0(x)=\sum_i F\left(\frac{x-x_{0i}}{\theta}\right). \tag{32}$$

Here θ is the pulse width and the point x_{0i} is the position of the maximum of the ith pulse. We assume that $F(0)=1$ holds at the maximum, thus $0\leqslant F\leqslant 1$. The mean distance between neighboring points x_i is equal to l, and the ratio of the pulse width θ to the distance between pulses is

$$\gamma=\frac{\theta}{l}. \tag{33}$$

The quantity γ is a characteristic parameter of the initial-value problem under consideration. With the help of Eq. (19), we find the distribution function in the form

$$f(u)=(1-a^{(1)}\gamma)\,\delta(u)+\gamma f_1(u), \tag{34}$$

where the function $f_1(u)$ has the usual form:

$$\gamma f_1(u)=\left\langle\left|\frac{du}{dx}\right|^{-1}\right\rangle_{x_i}=\frac{2}{l}\left|\frac{du}{dx}\right|^{-1}; \tag{35}$$

the $x_i(u)$ are the roots of Eq. (19), determined by the shape of the initial pulse $F(\xi)$. The average in Eq. (35) is calculated on a scale exceeding the width of the initial pulse θ. The presence of the δ function in the solution is completely understandable—the initial distribution (32) has as its most probable value $u=0$. The constant $a^{(1)}$ is determined by the normalization condition.

For example, if the pulse shape has the form

$$F(\xi)=1-\frac{\xi^2}{(\theta/2)^2},\quad u_0(x)=1-\frac{x^2}{(\theta/2)^2}, \tag{36}$$

Then from Eq. (35) we find that

$$x(u)=\pm\frac{\theta}{2}\sqrt{1-u},\quad \left|\frac{du}{dx}\right|=\frac{4}{\theta}\sqrt{1-u}$$

and correspondingly

$$f_1(u)=\frac{1}{2\sqrt{1-u}},\quad a^{(1)}=\int_0^1 f_1(u)\,du=1. \tag{37}$$

Example 4. Random initial function.

Above we considered a purely dynamical initial model. The mixing process takes place in a completely analogous way when the initial condition is prescribed by a random function. For example, if the initial function is periodic with random amplitude a, then, as can be easily seen, for the distribution function that sets up after mixing, instead of the dependence (29) we have

$$f(u)=\int\psi(a)\frac{da}{\pi\sqrt{a^2-u^2}}, \tag{38}$$

where $\psi(a)$ is the prescribed distribution of the random variable a. Averaging over the parameters a_0, a_1, and k can be done analogously for a quasiperiodic initial function.

3. STOCHASTIC SOLUTIONS OF THE KdV EQUATION

Let us turn now to our main problem, the asymptotic properties of the solution of the KdV equation (1) for the case of aperiodic oscillating initial conditions (2)–(5). Since we will make extensive use here of properties of the so-called multiphase finite-band solutions of the KdV equation, it is only natural to first recall the structure of these solutions.

3.1. Multiphase KdV solutions

The structure of the initial potential $-u_0(x)$ plays an important role in the integration of the KdV equation by the inverse-scattering method.

In the case in which the initial potential is a periodic function of a special class, the spectrum of the scattering problem consists of a finite number of allowed bands (see Refs. 5–7). The solutions of the KdV equation constructed for such initial conditions are multiphase (g-phase). They are found in the theory of finite-band integration with the help of the so-called trace formula (see Refs. 17 and 1)

$$u_g(x,t)=r_1-2\sum_{j=1}^{g}\left(\mu_j(x,t)-\frac{r_{2j}+r_{2j+1}}{2}\right), \tag{39}$$

where the constants $r_j>0$, $j=1,\ldots,2g+1$, are the branch points of the spectral surface of the complex variable λ in the finite-band scattering problem. This surface is prescribed by the algebraic equation

$$y^2=R_g(\lambda)=-\prod_{j=1}^{2g+1}(\lambda-r_j), \tag{40}$$

$$r_1<r_2<\ldots<r_{2g+1},$$

and is a two-sheet Riemann surface of genus g with cuts along the allowed bands $[-\infty,r_1]$, $[r_2,r_3]$, ..., $[r_{2g},r_{2g+1}]$.

The functions $\mu_j(x,t)$ (the so-called auxiliary spectrum of the finite-band scattering problem) are described by the system of Dubrovin ordinary differential equations

$$\varepsilon\frac{\partial\mu_j}{\partial x}=\frac{2i\sigma_jR_g^{1/2}(\mu_j)}{\Pi_{j\neq k}^{g}(\mu_j-\mu_k)},\quad j=1,2,\ldots,g. \tag{41}$$

Here $\sigma_j=\pm1$ are the signs of the square root of the function $R_g(\mu_j)$. Each μ_j lies in its own allowed band $[r_{2j-1},r_{2j}]$ and oscillates with variation of x, where each time μ_j reaches the edge of the band, σ_j changes sign and the motion advances to the next sheet of the Riemann surface. The characteristic period of the oscillations of the functions μ_j, as is clear from Eqs. (41), is of order ε.

Equations (41) in general describe a system of coupled nonlinear oscillators. For $g=1$ there is only one oscillator and Eqs. (39) and (41) yield a single equation whose solution is expressed in terms of the Jacobian elliptic function:[18]

$$u_1(x)=2(r_2-r_1)\mathrm{cn}^2\left((r_3-r_1)^{1/2}\frac{x}{\varepsilon}+\varphi_0\middle|m\right)+r_1+r_3-r_2, \tag{42}$$

where $m=(r_2-r_1)/(r_3-r_1)$ is the elliptic function parameter and φ_0 is an arbitrary initial phase (integration constant). The solution $u_1(x)$ is a periodic function of x with amplitude a and period T:

$$a=2(r_2-r_1),\quad T=2^{1/2}\varepsilon(r_3-r_1)^{-1/2}K(m), \tag{43}$$

where $K(m)$ is the complete elliptic integral of the first kind. As $m\to1$ the oscillations degenerate into a chain of isolated solitons of height a, where the distance between solitons grows as the distance r_3-r_2, i.e., the width of the allowed band, decreases:

$$T=2^{-1/2}\varepsilon\ln\left(16\frac{r_3-r_1}{r_3-r_2}\right). \tag{44}$$

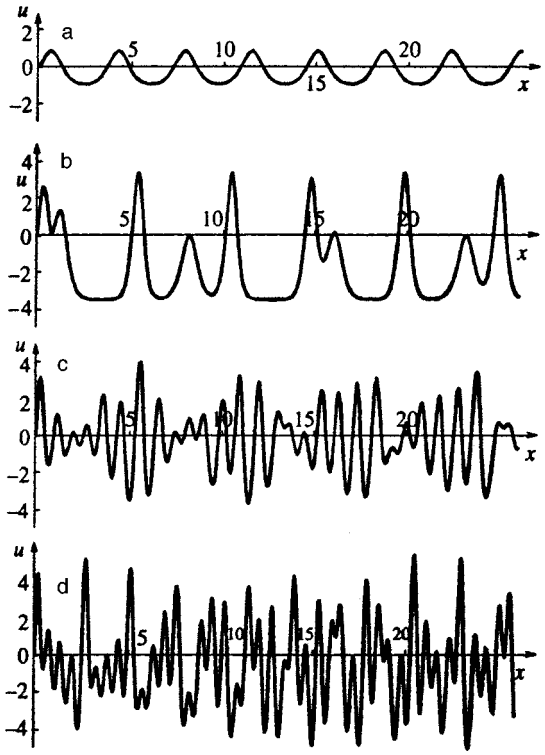


FIG. 2. Multiphase solutions of the Korteweg–de Vries equation. a — $r_1=0$, $r_2=0.2$. b — Two-phase: $r_1=0$, $r_2=0.001$, $r_3=3.5$, $r_4=3.5005$, $r_5=7$. c — Five-phase: $r_1=0$, $r_2=1$, $r_3=2$, $r_4=3$, $r_5=4$, $r_6=5$, $r_7=6$, $r_8=7$, $r_9=10$, $r_{10}=11$, $r_{11}=14$. d — Six-phase: $r_1=0$, $r_2=0.7$, $r_3=4$, $r_4=5.6$, $r_5=5.7$, $r_6=5.8$, $r_7=8.6$, $r_8=13.2$, $r_9=15.6$, $r_{10}=22$, $r_{11}=23.1$, $r_{12}=23.11$, $r_{13}=27.4$.

For $g=2$, the system (41) describes two coupled oscillators, in particular, the motion of a heavy top (the Kovalevskaya case) (see, e.g., Ref. 19).

Results of a numerical integration of system (41) for various values of g are plotted in Fig. 2. It can be seen that with increasing g the shape of the oscillations (39) becomes increasingly more complex, and for large enough g they outwardly appear to be disordered. This is understanda ble: if the single-band solution $g=1$ contains only one period and its nonlinear harmonics, then the g-phase solution is a quasiperiodic function and contains g independent periods, and all their nonlinear harmonics.

At the same time, it should be emphasized that in the multiphase solution, as in case (42) $g=1$, narrowing of the jth allowed band, $r_{2j+1}-r_{2j}\to0$, leads to the appearance of special oscillations in the form of the solitons corresponding to this band, separated by a large interval (44):

$$T_j\propto\varepsilon\ln(1/(r_{2j+1}-r_{2j})). \tag{45}$$

On the other hand, in the case when all the allowed bands are narrow, $r_{2j+1}-r_{2j}\to0$, $j=1,2,\ldots,g$, the entire g-phase solution decays into solitons of different amplitudes [Fig. 2b]. For $\varepsilon g\ll T_j$ these solitons are free to first order in the parameter $\varepsilon g/T$.

The temporal evolution of the functions $\mu_j(x,t)$ is described by a weakly nonlinear system of differential equations

$$\frac{\partial \mu_j}{\partial t}=2(-u_g(x,t)+2\mu_j)\frac{\partial \mu_j}{\partial x} \tag{46}$$

with initial data which are found by integrating the steady-state system (41).

Also note that the finite-band KdV solution (39)–(41), (46) can be expressed in terms of an algebraic function on a g-dimensional torus—the theta function of the Riemann surface (40) (see Ref. 19).

3.2. The stochastization process

We will start off with a brief description of the evolution of a single localized perturbation

$$u(x,0)=u_0(x), \quad u_0(+\infty)=u_0(-\infty)=0. \tag{47}$$

The development with time of the smooth perturbation $u(x,t)$ is described by the Hopf equation (2) at the outset, since the term with the higher derivative in the weak dispersion limit $\varepsilon\to 0$ is unimportant to start with. After passage at the time t_k through the critical reversal point x_k, where $|\partial u/\partial x|\to\infty$, three streams (7) and (12) appear in the solution of the Hopf equation in the vicinity of the point x_k, i.e., a multistream flow arises for $t>t_k$ (Fig. 1). The dispersion term with the higher derivative $\varepsilon^2\partial^3 u/\partial x^3$ begins to play a dominant role in the solution of the KdV equation. For this reason the solution of the KdV equation remains always single-valued, i.e., single-stream, but as if to make up for it, small-scale (with period $\sim\varepsilon$) oscillations are excited in the vicinity of the reversal point. It is important that these small-scale oscillations can be represented in the form of a single-phase solution of the KdV equation (42), (43) with slowly varying (in space x and in time t) branch points on the Riemann surface—the parameters r_1,r_2,r_3. The variation of the parameters $r_i(x,t)$ is described by the system of Whitham modulation equations:[20]

$$\frac{\partial r_i}{\partial t}+V_i(r_1,..r_3)\frac{\partial r_i}{\partial x}=0, \quad i=1,2,3, \tag{48}$$

$$r_3\geqslant r_2\geqslant r_1\geqslant 0, \quad V_3\geqslant V_2\geqslant V_1\geqslant 0.$$

We see that the system of equations for the Riemann invariants r_1,r_2,r_3 in the region of the oscillations is entirely analogous to the Hopf equations for three streams arising after reversal (7). This analogy extends further in the character of the matching with the nonoscillatory part of the solution:[4] the matching takes place at a caustic point $x^-(t)$, where not only the invariants r_1 and r_2, but also the velocities V_1 and V_2 coalesce; an analogous coalescence of the invariants r_2 and r_3 and the velocities V_2 and V_3 takes place at the point $x^+(t)$ [see Fig. 3a]. The difference from the three-stream Hopf equations (7) is that the group velocities V_i now depend on all three invariants, not just r_i. However, fulfillment of conditions (48) and the matching conditions for the invariants shows that in a qualitative sense the dynamics of the Riemann invariants is entirely similar to the development of three-stream flow.

This is fully confirmed by the solutions of the corresponding problems obtaining for different initial conditions (see Refs. 4, 21–26).

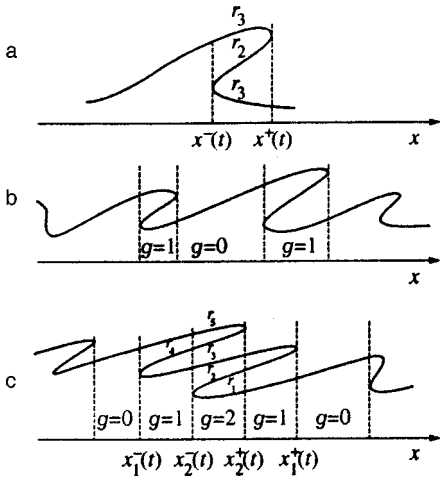


FIG. 3. Appearance of a multiphase structure.

We turn now to the oscillating initial condition of interest to us (2)–(5). In this case, reversal occurs at each inflection point of the initial curve $u_0(x)$ [Fig. 3b]. Here bands of single-phase oscillations develop. Then, with the passage of time the bands of single-phase oscillations begin to overlap. In this way, two-phase, and then three-phase oscillations, etc., develop [Fig. 3c]. They are described, respectively, by five, seven, etc. Riemann invariants. The variation of the Riemann invariants r_i for a g-phase structure is described by a generalized system of Whitham equations (Flaschka, Krichever *et al.*[27,3,28,29]):

$$\frac{\partial r_i}{\partial t}+V_i(r_1,\ldots,r_{2g+1})\frac{\partial r_i}{\partial x}=0, \quad i=1,\ldots,2g+1, \tag{49}$$

$$r_{2g+1}\geqslant r_{2g}\geqslant\ldots\geqslant r_1\geqslant 0,$$

$$V_{2g+1}\geqslant V_{2g}\geqslant\ldots\geqslant V_1\geqslant 0.$$

Equations (49) are written in a coordinate system moving with velocity $u_{0\min}$. In this coordinate system, the transition from a g-phase structure to a $(g+1)$- or $(g-1)$-phase structure is completed at the caustic points $x_g^{\pm}$, where, according to the rule described above for a single-phase system, the invariants coalesce.[30] For the initial-value problem (2)–(5), with the passage of time the number of phases at each point grows, $g\propto t$, but since the region of definition of the parameters r_i is prescribed by the initial conditions (4) in the bounded scale

$$u_{0\max}-u_{0\min}\geqslant r\geqslant 0, \tag{50}$$

this means that the mean distance between the invariants $\langle\Delta r\rangle=\langle|r_{i+1}-r_i|\rangle$ decreases with time as t:

$$\langle\Delta r\rangle\propto 1/t.$$

The above-described process is completely similar to the multiplication of streams in the solution of the Hopf equation and takes place uniformly over the entire x axis. Therefore, in analogy with the multistream solution of the Hopf equation investigated in Sec. 2, a steady-state distribution $F(r)$ over the invariants r that is completely determined to first order in $O(1/t)$ is asymptotically set up as $t\to\infty$. The distri-

bution function $F(r)$ describes the relative density of the invariants r_i in the interval from r to $r+dr$. A natural normalization rule is

$$\int_{r_{\min}}^{r_{\max}} F(r)\,dr=1. \tag{51}$$

The values $r_{\min}$ and $r_{\max}$ are defined in accordance with condition (50). The phase relations of the various modes of the oscillations in this case are unimportant, so the transition as $t\to\infty$ to a continuous distribution $F(r)$ already constitutes a stochastic description of the system. The continuum limit of the theta function introduced by Venakides[31] may serve as a definite analogy of such a description.

3.5. Stochastic multisoliton structure

Let us pause to note one important difference between the structure of the Riemann invariants of the KdV solution for the initial-value problem (2)–(5) and the structure of a multistream Hopf flow. Toward this end, we consider an analysis of a single localized perturbation. For $t\gg 1$ the asymptotic behavior of the solution of the initial-value problem (47) takes the form of a so-called soliton wave,[23,24] a set of solitons aligned in amplitude: the lead soliton has the greatest amplitude, followed by a soliton with lesser amplitude, etc. The distance between solitons grows linearly with time, $T_s\propto t$, which is a natural consequence of the linear dependence of the soliton velocity on the amplitude.[2]

But, as follows from Eq. (44), this means that the width of the allowed band r_3-r_2 narrows exponentially with time:

$$r_3-r_2=r_3\exp\left(-\frac{t}{\tau(r_3)}\right),\quad r_1\to 0,\quad t\gg 1. \tag{52}$$

An analogous situation obtains for a two-phase, three-phase, and indeed any multiphase system, as $t\to\infty$. Solitons of different amplitude a_{sj}, belonging to any given phase j, move with a different velocity $V(a_{sj})$. Consequently, the period T_j increases with time and consequently, according to relation (45), the width of the jth allowed band decreases exponentially with time:

$$r_{2j+1}-r_{2j}\propto\exp(-t/\tau_j). \tag{53}$$

Thus, in the multiphase structure under consideration, as $t\to\infty$ the widths of the allowed bands at any point x narrow down exponentiall y, and the widths of the forbidden bands (r_{2j},r_{2j+1}), (r_{2j+2},r_{2j+3}), etc., decrease with time as $1/t$. This peculiarity distinguishes the structure of the distribution of invariants in the KdV equation from the structure of multistream flows of the Hopf equation, where all $2j+1$ streams are equivalent. At the same time, Eq. (53) yields an important consequence for the chaotic structure of the KdV equation. First, it follows from fulfillment of condition (53) for each phase that the structure developing as $t\to\infty$ from the initial condition (2)–(5) is a system of interacting solitons [see relation (45)]. Further, it follows from condition (53) that the solitons belonging to a given mode, i.e., related by their common origin from one hump of the initial function $u_0(x)$ (2) and for this reason connected from the outset by a definite phase relation, are found at a large and ever growing distance T_j from each other: $T_j\propto t$. But this means that as a consequenc e of the general law of conservation of mean density, the space between them is filled with solitons belonging to other modes, i.e., solitons not related by a common origin and consequently possessing random phases. As a result, as $t\to\infty$ a random distribution of solitons with uniformly distributed phases sets up over all space. Each soliton can therefore be characterized by just its amplitude a. The maximum value of the amplitude is bounded by the prescribed initial condition $u_0(x)$ [Eqs. (2), (43), and (50)]:

$$a_m=2(u_{0\max}-u_{0\min}).$$

In this case, for the aperiodic oscillating initial-value problem (2)–(5) the set of soliton amplitudes is countable and hence dense on the interval $0\geqslant a\geqslant a_m$. The arising chaotic structure can therefore be described by the amplitude distribution function

$$f(a),\quad 0\leqslant a\leqslant a_m. \tag{54}$$

In what follows we will take $u_{0\min}=0$ and $u_{0\max}=1$ and consequently $a_m=2$.

The existence as $t\to\infty$ of the x-independent limit $f(a)$ is a consequence of the ergodicity of the initial function $u_0(x)$. The soliton amplitude distribution function $f(a)$ is determined directly by the form of the initial function $u_0(x)$, Eq. (2). Indeed, according to the inverse scattering problem, the solitons correspond to levels of the discrete spectrum of the Schrödinger equation with the potential $-u_0(x)$. By virtue of the smallness of the parameter ε it is possible to use the semiclassical approximation in the solution of the Schrödinger equation. Thus, the soliton amplitude distribution $f(a)$ describing the number of solitons with amplitudes in the interval from a to $a+da$,

$$dW=f(a)\,da, \tag{55}$$

is determined by the shape of each individual potential well of the initial function. Specifically,[23]

$$f(a)=-\frac{1}{4\pi\varepsilon}\int_{a/2}^{1}\frac{D'(z)}{\sqrt{z-a/2}}dz. \tag{56}$$

Here $D(z)$ is the width of the initial well at the level $u=z$. The derivative $D'(z)$ is related to the initial shape of the well $F(\xi)$ by the obvious relation

$$D'(z)=-\left(\frac{\theta}{|dF/d\xi|_{\xi_1(z)}}+\frac{\theta}{|dF/d\xi|_{\xi_2(z)}}\right), \tag{57}$$

where $\xi_1(z)$ and $\xi_2(z)$ are the values of the two-valued inverse function of $F(\xi)$ at the point $F(\xi)=z$. By virtue of the smallness of the parameter ε the system of levels can be considered separately in each well. As a result we obtain

$$f(a)=\frac{1}{m}\sum_{k=1}^{m} f_k(a), \tag{58}$$

where f_k is the distribution function in the kth well.

Thus the function $f(a)$ is normalized to the mean number of solitons per well, N. According to condition (4), it is possible to limit the number of wells in this sum to a finite value m on a sufficiently large interval L of the x axis. In this

case, according to the initial-value problem (2)–(5), the distribution function $f(a)$ to first order in $O(\theta/L)$ does not depend on which segment L of the initial function it is calculated on. We give examples.

Example 1. Let $u_0(x)$ be a sequence of isolated pulses of the form (38). Then, according to formula (56) we have

$$f(a)=\frac{\theta}{8\pi\varepsilon}\int_{a/2}^{1}\frac{dz}{\sqrt{1-z}\sqrt{z-a/2}}=\frac{\theta}{8\varepsilon}. \tag{59}$$

We assume that the width of the pulses is modulated:

$$\theta=\theta_0(1+b\cos^2(kx)),\quad k\ll 1.$$

In this case, averaging on a scale $L\gg 1/k$, we obtain the continuous soliton distribution

$$f(a)=\frac{\theta_0}{8\varepsilon}\left(1+\frac{b}{2}\right).$$

Example 2. If $u_0(x)=b(1-\cos x)$, then $D(z)=2\arccos(1-z/b)$ and from formula (56) we obtain

$$f(a)=\frac{1}{2\pi\varepsilon\sqrt{b}}\int_{\arccos(1-a/2b)}^{0}\frac{dt}{\sqrt{1-a/2b-\cos t}}=\frac{1}{\pi\varepsilon}\sqrt{\frac{2}{ab}}\,F\left(\arcsin\left(\sqrt{\frac{a}{8b}}\right);\sqrt{\frac{8b}{a}}\right), \tag{60}$$

where $F(\phi;m)$ is the incomplete elliptical integral of the first kind.

Next, considering the slow variation of the amplitude of the oscillations of an initial function $u_0(x)$ of the form $b=1+b_1\sin(kx)$ ($b_1<1$; $k\ll 1$, where k is irrational), it is also possible here to obtain a continuous distribution $f(a)$ by averaging expression (60) on a scale $L\gg 1/k$:

$$f(a)=\frac{1}{\sqrt{2a}\,\pi^2\varepsilon}\int_{-\pi}^{\pi}\frac{dz}{\sqrt{1+b_1\sin z}}\times F\left(\arcsin\sqrt{\frac{a}{8(1+b_1\sin z)}};\sqrt{\frac{8(1+b_1\sin z)}{a}}\right).$$

Another important parameter is the mean spatial density of the solitons C characterizing their interaction. The mean spatial density of the solitons C is also determined directly by the form of the initial function $u_0(x)$. Indeed, the total number of solitons in one individual well, according to formula (56) is

$$N_i=\int f_i(a)\,da. \tag{61}$$

Determining the total number of solitons in all the wells on a ''representative'' scale L, we find the mean soliton density corresponding to the given initial-value problem:

$$C=\frac{1}{L}\sum_{i=1}^{m}N_i. \tag{62}$$

According to conditions (3)–(5) the density C does not depend on the choice of segment of the initial distribution. For example, for the soliton density in the case (59) we have

$$C=\frac{\gamma}{4\varepsilon}\left(1+\frac{b}{2}\right),$$

where according to Eq. (33) $\gamma=\theta_0/l$ (l is the mean distance between pulses). We emphasize that the density is bounded—it does not exceed a quantity of order $1/\varepsilon$. For example, in the case of a single-phase wave (42) the minimum value of the period of the oscillations $T_{\min}$ is reached as $m\to 0$ and according to formula (43) is equal to

$$T_{\min}=\frac{\pi\varepsilon}{a^{1/2}}.$$

Consequently, the maximum value of the density is equal to

$$C_m=\frac{1}{T_{\min}}=\frac{a^{1/2}}{\pi\varepsilon}.$$

Hence it is clear that C_m decreases as the soliton amplitude decreases. Note that the soliton distribution function normalized to the spatial density of the solitons is $f_C(a)=Cf(a)$.

We emphasize that the possibility of describing the main properties of the structure under consideration with the help of a chaotic system of solitons characterized only by the amplitude distribution and the mean density is determined by the form of the initial function $u_0(x)$.

An important role is played here by the condition of spatial homogeneity of the initial function $u_0(x)$ and condition (5), which exclude the effect of a singular point at infinity ($x\to\pm\infty$).

Despite the fact, as was indicated above, that the continuum limit of the theta function[31] is a definite analog of the system considered here, it is necessary to also turn our attention to an important difference between them. The point is that Ref. 31 considers the limit for a *g*-band structure, arising, for example, in the case of a periodic initial-value problem. This structure is formed by a large number $g\sim N\sim 1/\varepsilon$ of soliton lattices,[29] each of which corresponds to a definite band. The width of the allowed bands describing the lattices in this case is exponentially small in comparison with the width of the forbidden bands describing the difference in the levels of the solitons belonging to neighboring lattices. In other words, the soliton amplitudes in a mixed system are not arbitrary, but vary by small ($\sim\varepsilon$) jumps.

In the system under consideration, in the asymptotic limit as $t\to\infty$ no distinct soliton lattices arise. The quantum levels of the initial distribution form a continuous system. Correspondingly, a continuous distribution of soliton amplitudes (56)–(58) arises as $t\to\infty$. We emphasize that although in the semiclassical limit considered in Ref. 31 the continuumized description of the discrete spectrum is identical to a description of the continuous soliton distribution, upon closer analysis, e.g., in a consideration of the correlation functions, fundamental differences between them can arise, as is already evident in the case of the Hopf equation. The paper by Krylov *et al.*[32] is dedicated to a more detailed treatment of the corresponding mathematical questions.

4. SEQUENCE OF ISOLATED PULSES IN THE KdV EQUATION

We consider for the KdV equation with weak dispersion $\varepsilon \ll 1$ initial data in the form of an infinite sequence of pulses of identical shape (32). We assume the pulses to be well separated:

$$\gamma = \frac{\theta}{l} \ll 1. \tag{63}$$

Over a time $t \leqslant 1/\gamma \gg 1$ each pulse can be treated separately. After this time the pulse transforms into a soliton wave, i.e., it decays into a large number of solitons $N \sim \theta/\varepsilon$, aligned in amplitude.[23,24] The total number of solitons arising as a result of the decay of one pulse is equal to

$$N = \int_0^2 f(a)\, da = \frac{\theta}{\pi\varepsilon} \int_{-\infty}^{\infty} \sqrt{F(\xi)}\, d\xi. \tag{64}$$

Each soliton moves with its own velocity v_s, which is proportional to its amplitude:

$$v_s = 2a. \tag{65}$$

We now consider the asymptotic behavior of the solution of the KdV equation with the initial condition (2) with condition (63) taken into account in the limit $t \to \infty$. As was shown in Sec. 3, as a consequence of ergodicity of the problem the system passes into a uniformly mixed (on the x axis) stochastic state. The process of multiplication of modes investigated in Sec. 3 is equivalent in our case to a gradual superposition of the soliton waves arising as a consequence of the differences in the velocities of the solitons (65). The considered mixing process is more obvious by virtue of the fact that in the linear-in-γ approximation the interaction of the solitons can be neglected. We are thus talking about the mixing in one-dimensional space of an ideal gas of noninteracting particles (solitons) with velocity spectrum $f(v_s)$. Note that the kinetics of a gas of solitons with their interactions taken into account was considered by Zakharov.[33]

For the problem of interest to us, it is important that each soliton has a completely determined dependence of its shape u on x and t:

$$u = u_s(x) = a \cosh^{-2}\left(\frac{x - v_s t}{\varepsilon} \left(\frac{a}{2} \right)^{1/2} \right). \tag{66}$$

Therefore, following the results of Sec. 2, in order to find the velocity distribution function $f_\infty(u)$ in the stochastically mixed state in the leading approximation, it is necessary first to consider the probability density distribution function in a soliton wave of fixed amplitude,

$$f_a(u) = \langle \delta(u - u_s(x)) \rangle_{x_i} = \frac{2}{l} \left| \frac{du_s}{dx} \right|^{-1},$$

and then average it over the soliton amplitudes:

$$f_\infty(u) = \frac{2}{l} \left\langle \frac{1}{|du_s/dx|} \right\rangle_a = \frac{2}{l} \int_u^2 f(a) \left| \frac{du_s}{dx} \right|^{-1} da. \tag{67}$$

Here we have taken into account that the soliton amplitudes are distributed according to the rule (55), (56) and that only solitons with amplitude greater than u contribute to $f_\infty(u)$. We now take into account that the quantity $|du/dx|$ for solitons with the shape (66) has the universal form

$$\left| \frac{du_s}{dx} \right| = \frac{\sqrt{2}}{\varepsilon} u \sqrt{a - u}. \tag{68}$$

Substituting expression (68) into formula (56) and integrating, we arrive at a universal expression for the steady-state distribution function:

$$f_\infty(u) = \gamma \frac{C_1}{u} D\left(\frac{u}{2} \right). \tag{69}$$

Here C_1 is a normalization constant.

It is clear from Eq. (69) that the distribution function $f_\infty(u)$ grows without bound like $1/u$ as $u \to 0$, so that the distribution (69) is not normalizable. It should, however, be borne in mind that for very small u the implemented approximation of noninteracting solitons breaks down.

To calculate the interaction between the solitons (see Ref. 1), note that as $\gamma \to 0$ the solitons overlap mainly in the region of their exponential tails.

Therefore the velocity $u(x)$ in the interaction region can be represented in the form of a superposition of two exponential soliton ''tails'':

$$u(x) = a \exp\left(-\frac{x}{\varepsilon} \sqrt{\frac{a}{2}} \right) + a_2 \exp\left(\frac{x - x_0}{\varepsilon} \sqrt{\frac{a_2}{2}} \right). \tag{70}$$

As follows from Eq. (67), we must calculate $x'(u)$ as a function of u. Differentiating expression (70) and regrouping terms, we obtain a relation between $u'(x)$ and u:

$$\ln[(\sqrt{a}u + \varepsilon\sqrt{2}u')(u\sqrt{a_2} - \varepsilon\sqrt{2}u')^{\sqrt{a_2/a}}]$$
$$= -\frac{x_0}{\varepsilon}\sqrt{a_2/2} + \ln[(a^{3/2} + a\sqrt{a_2})^{\sqrt{a_2/a}}(a_2\sqrt{a} + a_2^{3/2})]. \tag{71}$$

By virtue of the smallness of the parameter γ the following condition is fulfilled:

$$\frac{x_0}{\varepsilon}\sqrt{\frac{a_2}{2}} \gg \ln[(a^{3/2} + a\sqrt{a_2})^{\sqrt{a_2/a}}(a_2\sqrt{a} + a_2^{3/2})].$$

Therefore the solution of Eq. (71) can be represented in the form

$$(\sqrt{a}u + \varepsilon\sqrt{3}u')(u\sqrt{a_2} - \varepsilon\sqrt{2}u')^{\sqrt{a_2/a}}$$
$$= \exp\left\{ -\frac{x_0}{\varepsilon}\sqrt{\frac{a_2}{2}} \right\}. \tag{72}$$

Analysis of expression (72) shows that there exists a minimum value $u_{\min}$ such that there is no solution for $u < u_{\min}$, where

$$u_{\min} = \frac{1}{\sqrt{a}} \left(\sqrt{\frac{a_2}{a}} \right)^{-\sqrt{a_2/a}/(1+\sqrt{a_2/a})}$$
$$\times \exp\left\{ -\frac{x_0}{\varepsilon} \frac{\sqrt{a_2/2}}{1 + \sqrt{a_2/a}} \right\}.$$

In this case, for dx/du we have the following relation:

$$\left|\frac{dx}{du}\right|_{u\to u_{\min}} \propto \frac{1}{\sqrt{u-u_{\min}}}, \quad u \geqslant u_{\min}.$$

Thus we see that the $1/u$ divergence obtained earlier in expression (69) for $f_\infty(u)$ is cut off at small u at values $u=u_{\min}$. Therefore, to eliminate the divergence in expression (69) it is necessary when averaging to introduce a cutoff at $u=u_{\min}$. Note that exact calculation shows that the result does not depend on how the cutoff is effected. Therefore it is simplest to make the cutoff by introducing the Heaviside step function $H(x)$, which, as usual, is equal to zero for $x<0$ and one for $x>0$. This function must be averaged over the position of the second soliton x_0 and over the amplitudes a, a_2. It can be shown that the probability of the relative arrangement of the solitons in the mixed stochastic state obeys Poisson statistics regardless of the form of the initial distribution. The Poisson distribution is a spatial analog of the uniform distribution of the phases over their periods. In other words, the probability that the distance between the solitons will take the value x_0 is

$$P(x_0)=\frac{4\varepsilon}{\gamma}\exp\left(-x_0\frac{\gamma}{4\varepsilon}\right). \tag{73}$$

Here we have taken into account that the mean distance between the solitons in our case (32), (62), (63) is equal to $4\varepsilon/\gamma$. Averaging over the Poisson distribution (73), we obtain

$$\langle H(u)\rangle_{x_0}=\int_0^\infty P(x_0)H(u-u_{\min}(x_0,a_2,a))\,dx_0$$

$$=u^{\gamma/2\sqrt{2}}\left(\frac{1}{\sqrt{a_2}}+\frac{1}{\sqrt{a}}\right). \tag{74}$$

Next it is necessary to average expression (74) over the soliton amplitudes a and a_2. This averaging is performed using the soliton amplitude distribution function. As a result, instead of formula (56) for $f_\infty(u)$, we obtain the normalized distribution

$$f_\infty(u)=\frac{\gamma}{2}u^{-1+\gamma/2}D\left(\frac{u}{2}\right). \tag{75}$$

The normalization constant in formula (75), as usual, is found from the equation

$$\int_0^\infty f_\infty(u)\,du=1.$$

Thus, for example, for the case when the initial function is prescribed by a set of parabolic pulses (36), the distribution function for $\gamma\ll 1$ has the form

$$f_\infty(u)=\frac{\gamma}{2}u^{-1+\gamma/2}\sqrt{1-\frac{u}{2}}. \tag{76}$$

The distribution function (76) is plotted in Fig. 4.

Note that the divergence of the distribution functions (75) and (76) as $u\to 0$ is a consequence of working to first order in γ. This approximation does not take account of the special role of the small-amplitude solitons. Indeed, as is clear from formula (66), the width of the soliton Δx_s grows as its amplitude decreases:

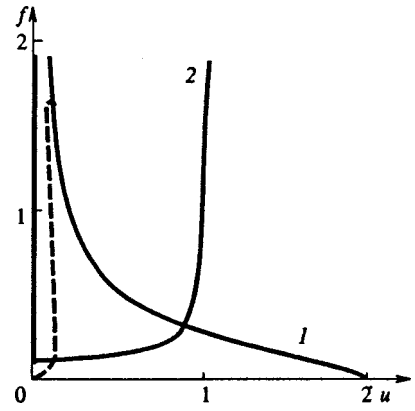


FIG. 4. Velocity distribution function $f(u)$ for the initial function $u(x)$, defined by Eqs. (32) and (36); *1* — Korteweg–de Vries equation (76). Dashed line — corrections $\propto\gamma^2$; *2* — dispersionless limit (34), (37) (Hopf equation).

$$\Delta x_s \approx \frac{\varepsilon}{\sqrt{a}}.$$

The mean distance between solitons is equal to

$$\langle\Delta x\rangle=\frac{l}{N}\sim\varepsilon\frac{l}{\theta}\sim\frac{\varepsilon}{\gamma}.$$

Therefore, for $a\sim\gamma^2$ the solitons overlap substantially and cannot be treated as weakly interacting. Thus, the distribution function (75) is valid only up to values $u\sim\gamma^2$, i.e., up to values $f_\infty\sim 1/\gamma$. For values $u<\gamma^2$ its growth is truncated due to the strong interaction of the small-amplitude solitons.

This is also why the function $f_\infty(u)$ takes the form depicted qualitatively by the dashed line in Fig. 4. Its exact calculation requires a complete account of effects $\sim\gamma^2$, which goes beyond the scope of the present work.

We note in conclusion that for the same initial conditions (32) there exists a structural similarity between the distribution functions in the solution of the KdV equation, $f_\infty(u)$ given by Eq. (76) and the Hopf equation $f(u)$ given by (34): both have a characteristic value of order γ for $u\sim 1$ and a large maximum in the limit $u\to 0$ (see Fig. 4). These features reflect the properties of the initial distribution $u_0(x)$.

5. THE CORRELATION FUNCTION

Let us turn now to the correlation function in the mixed state. As usual, we define it by the formula

$$\mathbf{K}(s)=\langle u(x+s)u(x)\rangle-\langle u(x)\rangle^2. \tag{77}$$

Here the angle brackets $\langle\ldots\rangle$ denote averaging over space x. According to our earlier analysis of the probability function $f_\infty(u)$ in the semiclassical problem ($\varepsilon\to 0$), in the case of a well-separated sequence of initial pulses ($\gamma\to 0$) two regions stand out. The first of these is the main region of values $u\sim 1$, where the solution decays into a system of

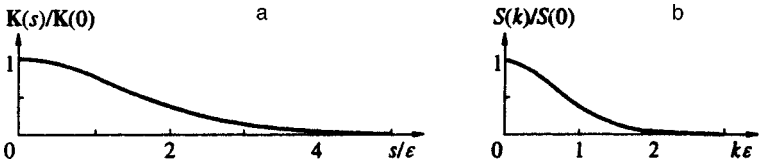


FIG. 5. a — Correlation function $\mathbf{K}(s)$; b — Fourier spectrum $S(k)$.

well-separated solitons such that to first order in the parameter γ they can be taken to be noninteracting. The correlations here are short-range and are determined only by the soliton scale ε. The second region is the region of very small values of $u<\gamma^2$, where the small-amplitude solitons $a\sim\gamma^2$ play the main role. In this case, the interaction between the solitons turns out to be not only important, but in fact predominant. In this region long-range correlations should naturally arise.

We consider here the first approximation in the parameter γ and, consequently, only the region of short-range correlations. In this approximation only the first term in expression (77) makes a contribution of order γ to the correlation function. It can be rewritten in the form

$$\mathbf{K}(s)=\int_0^2 f(a)K_a(s)\,da, \tag{78}$$

where $f(a)$ is the soliton amplitude distribution function (56), and $K_a(s)$ is the correlation function of solitons of a given amplitude a:

$$\begin{aligned}K_a(s)&=a^2\lim_{L\to\infty}\frac{1}{2L}\\&\quad\times\int_{-L}^{L}\cosh^{-2}\left(\frac{x+s}{\varepsilon}\left(\frac{a}{2}\right)^{1/2}\right)\cosh^{-2}\left(\frac{x}{\varepsilon}\left(\frac{a}{2}\right)^{1/2}\right)dx\\&=\frac{\varepsilon a^{3/2}}{\sqrt{2}l}\int_{-\infty}^{+\infty}\cosh^{-2}(x_1+s_1)\cosh^{-2}(x_1)\,dx_1,\\s_1&=\frac{s}{\varepsilon}\left(\frac{a}{2}\right)^{1/2}.\end{aligned} \tag{79}$$

Here l is the mean distance between pulses. After integration, formula (78) yields

$$K_a(s_1)=\frac{\sqrt{2}\varepsilon a^{3/2}}{l}\left[\frac{\cosh s_1}{(\sinh s_1)^3}s_1-\frac{1}{\sinh^2 s_1}\right]. \tag{80}$$

The function $K_a(s)$ takes its maximum value at $s=0$. Near the maximum for $s\ll\varepsilon/a^{1/2}$ it can be represented in the form

$$K_a(s)=\frac{\sqrt{2}}{3}\varepsilon\frac{a^{3/2}}{l}\left(1-\frac{a}{5\varepsilon^2}s^2\right).$$

For $s\gg\varepsilon/a^{1/2}$ the correlation function decreases exponentially:

$$K_a(s)_{s\gg\varepsilon/a^{1/2}}=4s\frac{a^2}{l}\exp\left(-\sqrt{2}\frac{sa^{1/2}}{\varepsilon}\right).$$

The total correlation function depends on the soliton amplitude distribution. Let us consider the specific example (36). In this case

$$D'(z)=\frac{2}{|u'(z)|}=\frac{\theta}{2\sqrt{1-z}}.$$

Therefore

$$f(a)=\frac{\theta}{8\pi\varepsilon}\int_{a/2}^{1}\frac{dz}{\sqrt{(1-z)(z-a/2)}}=\begin{cases}\dfrac{\theta}{8\varepsilon}, & 2\geqslant a\geqslant 0,\\ 0, & a>2.\end{cases} \tag{81}$$

It follows from Eqs. (78), (80), and (81) that

$$\mathbf{K}(s)=\frac{\sqrt{2}\theta}{8l}\int_0^2\left[\frac{\cosh s_1}{(\sinh s_1)^3}s_1-\frac{1}{\sinh^2 s_1}\right]a^{3/2}\,da. \tag{82}$$

From this expression we find the asymptotic forms of the correlation function:

$$\mathbf{K}(s)_{s\ll\varepsilon}=\frac{2}{15}\gamma\left(1-\frac{2s^2}{7\varepsilon^2}\right), \tag{83}$$

$$\mathbf{K}(s)_{s\gg\varepsilon}=3\gamma\left(\frac{\varepsilon}{s}\right)^5\left[1-\frac{2}{3}\left(\frac{s}{\varepsilon}\right)^5\exp\left(-\frac{2s}{\varepsilon}\right)\right].$$

Hence it is clear that the maximum value of the correlator is $\mathbf{K}(0)=(2/15)\gamma$, and for large values of the parameter s it falls off as s^{-5}. The total integral of the correlation function is of course proportional to the small parameter $\gamma\varepsilon$:

$$J=\int_{-\infty}^{\infty}\mathbf{K}(s)\,ds=\sqrt{2}\gamma\varepsilon.$$

The form of function (82) is depicted in Fig. 5a.

The Fourier transform of the correlation function—the spectral power of the process[34]—is usually also of significant interest:

$$S(k)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{K}(s)e^{iks}ds. \tag{84}$$

In the case under consideration of a pulsed process with nonoverlapping pulses

$$S(k)=\frac{1}{l}\int_0^2\langle|u_{sa}(k)|^2\rangle f(a)\,da,$$

where

$$u_{sa}(k)=\frac{1}{2\pi}\int_{-\infty}^{\infty}u_{sa}(x)e^{-ikx}dx=\frac{\varepsilon^2 k}{2\pi\sinh(k\varepsilon(2/a)^{1/2})}$$

is the Fourier transform of an isolated soliton.

In the example (36) considered above the integrated Fourier spectrum has the form

$$S(k)=\frac{\theta}{8l\varepsilon}\int_0^2\frac{k^2\varepsilon^4}{4\pi^2\sinh^2(\sqrt{2}k\varepsilon/a^{1/2})}da$$

$$=\frac{1}{8\pi^2}\gamma\varepsilon(k\varepsilon)^4\int_{k\varepsilon}^{\infty}\frac{dt}{t^3\sinh^2 t}. \quad (85)$$

Expression (85) has the following asymptotic limits. For $k\varepsilon\ll 1$ the spectrum tends toward a constant:

$$S(0)=\frac{\gamma\varepsilon}{32\pi^2}.$$

And for $k\varepsilon\gg 1$ the spectrum decreases exponentially:

$$S(k)=\frac{1}{4\pi^2}(\gamma\varepsilon)(k\varepsilon)^2e^{-2k\varepsilon}.$$

The general form of the Fourier spectrum (85) is shown in Fig. 5b.

6. CONCLUSION

The process examined in the present paper can be regarded as the generation of soliton turbulence in the dynamics of a continuous, one-dimensional nondissipative medium described by the integrable KdV equation. We recapitulate the main results of the theory.

1. For initial conditions having the form of large-scale oscillations $u_0(x)$ prescribed on the entire x axis, as a consequence of the dynamics of the KdV equation stochastic small-scale oscillations set up as $t\to\infty$, which can be described in terms of a continuous random process: the probability density $f(u)$, the correlation function $K(s)$, and other higher correlations.

2. The statistical characteristics $f(u)$, $K(s)$, and others are uniquely determined by the form of the initial function $u_0(x)$ and can be calculated. They are stable, i.e., they remain invariant for arbitrary finite perturbations J of the initial function $u_0(x)$:

$$J=\int_{-\infty}^{\infty}|\delta u(x)|dx<\infty,\quad \delta u=\tilde{u}_0(x)-u_0(x). \quad (86)$$

A general proof of the last statement proceeds as follows: a finite perturbation of the initial condition (86) corresponds to a finite perturbation of the initial soliton structure and therefore cannot alter the distribution function $f(a)$, which is formed by an infinite number of solitons.

3. All the dynamical properties of the system are described by corrections that are $O(1/t)$. As $t\to\infty$ they disappear asymptotic ally and only a statistical description remains possible.

Of course, an exact proof of these statements would require a detailed mathematical study, which would go beyond the scope of the present work. We also emphasize that the final expressions obtained here for the simple specific cases given can be generalized using the analytical methods developed in Refs. 3, 31, and 25 for minimizing the multisolito n interaction, or by numerical methods. It should also be noted that the proposed theory can be generalized to other completely integrable systems.

In conclusion, we make a few remarks about the connection between the process considered here and other known models of the appearance of chaotic motion.

1. It is possible to discern an analogy between the Landau–Hopf mechanism of successive bifurcations in the development of an instability in the flow of a dissipative continuous medium[13] and the successive increase in the number of modes of a multiband structure in the theory considered here. But one should also note the substantial difference between them. While Landau–Hopf bifurcations develop as linear streaming instabilities, in our case singular points in x-space play the defining role. It is precisely the appearance of singularities as a consequence of the successive reversal of the velocity profile u that leads to the generation of new modes which immediately develop as a strongly nonlinear process in the local vicinity of the given singular point x_k. Note that the contemporary theory of hydrodynamic turbulence (see, e.g., Ref. 35) also points to the important role of singularities in the higher correlations.

2. Since $u(x,t)$ is the velocity, the trajectory of any given point x in the coordinate system moving with constant mean velocity $\bar{u}=\langle u_0(x)\rangle$, is described by the equation

$$\frac{dx}{dt}=u(x,t)-\bar{u}. \quad (87)$$

In our case all the trajectories $x(t)$ begin as large-scale motion defined by Eq. (87) with initial condition (2)–(5), but then with the passage of time contracts to small-scale oscillations $x(t)$ with characteristic scale of order ε. This process of compactification of the trajectories $x(t)$ has a possible analogy with the attractors considered in the theory of dissipative turbulence.

3. According to the theory of integrability of the KdV equation, an arbitrary large-scale, finite function $u_0(x)>0$ can be approximated with a high degree of accuracy [of order $\exp(-1/\varepsilon)$] by a multisoliton solution with $N\sim 1/\varepsilon$ (Ref. 3). It is important that each soliton has a definite amplitude a_i, but the position of its center x_i in this case is rigidly fixed. In regard to the oscillating initial function $u_0(x)$ (2)–(5) it can be said that the position of the soliton inside each large-scale oscillation has a fixed phase $\varphi_i(x_i)$. Thus, at the initial time $t=0$ there is a rigid connection between the amplitudes a_i and phases φ_i of the solitons. Their motion in space x takes place with different velocities depending on their amplitude. As a result, the rigid initial connection between their amplitudes a_i and phases φ_i breaks down. We stress that the solitons do not move freely, but interact with one another, and the more strongly, the larger is their mean density. Despite this fact, solitons of greater amplitude move on average faster than those with smaller amplitude, which follows directly from relations (48) and (52). As a result, as $t\to\infty$ complete phase mixing takes place. Any value of the phase φ between 0 and 2π becomes possible for any soliton. Such a situation corresponds to the onset of soliton turbulence.

From this standpoint the process of turbulization of a one-dimensional continuous medium described by the KdV equation can be considered as ergodic. This fact is under-

scored by the analogy with the solution of the Hopf equation (2) indicated in Sec. 3.2. On the other hand, it cannot go without mention that phase shifts also arise as a consequence of interactions between the solitons.[1,33] This process amplifies the phase mixing.

4. We have not touched here on the huge literature devoted to plasma turbulence. The main difference in our work consists in an analysis of the dynamics of an exactly integrable system.

We are deeply indebted to A. L. Krylov; without the many valuable discussions we had with him this work would never have been finished. We are also grateful to S. Venakides, N. G. Mazur, V. V. Geordz haev, and M. O. Ptitsyn for helpful discussions.

This work was carried out with the financial support of the U.S. Civilian and Defense Research Foundation (Grant No. RM1-145). One of us (G. A. É.) also thanks the Russian Fund for Fundamental Research for partial financial support (Grant No. 96-01-01453).

*)E-mail: alex@td.lpi.ac.ru

[1] S. P. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method* (Consultants Bureau, New York, 1984) [Russ. original, Nauka, Moscow, 1980].
[2] V. I. Karpman, *Non-Linear Waves in Dispersive Media* (Pergamon Press, Oxford, 1975)
[3] P. D. Lax and C. D. Levermore, Commun. Pure Appl. Math. **36**, 253, 571, 809 (1983).
[4] A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **65**, 590 (1973) [Sov. Phys. JETP **38**, 291 (1974)].
[5] S. P. Novikov, Funk. An. Pril. **8**, 54 (1974).
[6] P. D. Lax, Commun. Pure Appl. Math. **26**, 141 (1975).
[7] B. A. Dubrovin, V. B. Matveev, and S. P. Novikov, Usp. Mat. Nauk **31**, 59 (1976).
[8] S. Venakides, Am. Math. Soc. Trans. **301**, 189 (1987).
[9] J. N. Elgin, Phys. Lett. A **110**, 441 (1985).
[10] F. Kh. Abdullaev, Phys. Rep. **179**, 1 (1989).
[11] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1988) [Russ. original, Nauka, Moscow, 1982].
[12] L. A. Pastur and A. L. Figotin, *Spectra of Random and Almost-Periodic Operators* (Springer-Verlag, New York, 1992) [Russ. original Nauka, Moscow, 1991].
[13] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, Oxford, 1987).
[14] A. V. Gurevich and K. P. Zybkin, Zh. Éksp. Teor. Fiz. **94**, 3 (1988) [Sov. Phys. JETP **67**, 1 (1988)]; Usp. Fiz. Nauk **165**, 723 (1995).
[15] E. Weinan, K. Khanin, A. Mazel, and Ya. Sinai, Phys. Rev. Lett. **78**, 1904 (1997).
[16] I.P. Cornfeld, S.V. Fomin, Ya.G. Sinai, *Ergodic theory* (Springer-Verlag, New York, 1982) [Russ. original Nauka, Moscow, 1980].
[17] B. M. Levitan, *Inverse Sturm–Liouville Problems* (VNU Science Press, Utrecht, 1987).
[18] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[19] B. A. Dubrovin, Usp. Mat. Nauk **36**, 215 (1981).
[20] G. B. Whitham, Proc. R. Soc. London, Ser. A **283**, 238 (1965).
[21] A. V. Gurevich, A. L. Krylov, and G. A. El', Zh. Éksp. Teor. Fiz. **101**, 1797 (1992) [Sov. Phys. JETP **74**, 957 (1992)].
[22] F. R. Tian, Commun. Pure Appl. Math. **46**, 1093 (1993).
[23] A. V. Gurevich, A. L. Krylov, and N. G. Mazur, Zh. Éksp. Teor. Fiz. **95**, 1674 (1989) [Sov. Phys. JETP **68**, 966 (1989)].
[24] A. V. Gurevich, A. L. Krylov, N. G. Mazur, and G. A. Él', Dokl. Akad. Nauk SSSR **323**, 876 (1992) [Sov. Phys. Dokl. **37**, 198 (1992)].
[25] N. G. Mazur, Teor. Mat. Fiz. **106**, 44 (1996).
[26] P. Deift, S. Venakides, and X. Zhou, Int'l. Math. Res. J., No.4, 285 (1997).
[27] H. Flaschka, G. Forest, and D. W. McLaughlin, Commun. Pure Appl. Math. **33**, 739 (1979).
[28] I. M. Krichever, Funk. An. Pril. **22**, 37 (1988).
[29] B. A. Dubrovin and S. P. Novikov, Usp. Mat. Nauk **44**, 35 (1989).
[30] G. A. El, Phys. Lett. A **222**, 393 (1996).
[31] S. Venakides, Commun. Pure Appl. Math. **42**, 711 (1989).
[32] A. L. Krylov, G. A. Él', and A. V. Gurevich, in Proc. Conf. in honor of Prof. S. Novikov, Moscow (1998), p. 17.
[33] V. E. Zakharov, Zh. Éksp. Teor. Fiz. **60**, 1012 (1971) [Sov. Phys. JETP **33**, 548 (1971)].
[34] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskiĭ, *Principles of Statistical Radiophysics*, Vol. 1 (Springer-Verlag, Berlin, 1987).
[35] L. Sirovich, L. Smith, and V. Yahot, Phys. Rev. Lett. **72**, 344 (1994).

Translated by Paul F. Schippnick

Dynamics of an anharmonic oscillator with a periodic perturbation

Yu. L. Bolotin, V. Yu. Gonchar,*) M. Ya. Granovskiĭ, and A. V. Chechkin

Kharkov Institute of Physics and Technology, National Scientific Center, 310108 Kharkov, Ukraine
(Submitted 26 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 361–377 (January 1999)

We study the features of the stochastic dynamics of a Hamiltonian system with the potential x^{2n} subjected to an external monochromatic perturbation. Three regimes of stochastic diffusion, which differ in the value of the amplitude of the external perturbation, are detected. We demonstrate the possibility of chaotic regimes manifesting themselves in pendulum vibrations of the well of a water-moderated and -cooled nuclear power reactor as an application of the model being investigated. Finally, we propose a method of simple proportional control, which makes it possible to control the chaotic vibrations of the anharmonic oscillator.
 [S1063-7761(99)02801-2]

1. INTRODUCTION

The revival of interest in the classical problem of the dynamics of a linear system subjected to a periodic perturbation is due to two facts. On the one hand, the interest stems from the need to describe the rapidly growing number of experiments dealing with the interaction of periodic fields and nonlinear systems, and on the other, the new ideas about the dynamics of nonlinear systems, developed in the last decades,[1–4] require confirmation by models. The essence of these new ideas is that under certain conditions the motion of a strictly deterministic system acquires all the features of random motion, although no visible source of stochasticity is present. Examples of dynamical chaos have been discovered in essentially all areas of physics, and their number is still increasing.

A one-dimensional Hamiltonian system with a time-dependent interaction is the simplest dynamical system that allows for chaotic behavior. The case of a monochromatic perturbation is convenient for analysis and reflects the experimental situation fairly accurately. The reason is that in many cases the motion of three-dimensional objects can be approximately described by one effective degree of freedom. Such a situation occurs, for instance, in the case of Rydberg states in a microwave field[5,6] or surface states of an electron above liquid helium.[7] These systems are also interesting conceptually from the viewpoint of the problem of quantum chaos.[8,9]

A general study of the dynamics of any nonlinear system in which chaotic behavior is possible consists of the following stages.

1. The numerical study of the structure of classical phase space.

2. Analytical estimates of the critical parameters of the regularity-to-chaos transition.

3. The study of the dynamics of a system in chaotic regimes and the development of methods for describing such behavior.

Our previous work[10,11] was devoted to solving the first two problems for an anharmonic oscillator under a periodic perturbation. The aim of the present work is to study the chaotic regimes in this model. In Sec. 2 we briefly describe the structure of phase space and give an expression for the amplitude of the periodic perturbation, which makes it possible to determine the intervals of regular and chaotic motion. In Sec. 3 we use the quasilinear approximation to derive an expression for the local diffusion coefficient and study the various diffusion regimes. In Sec. 4 we employ the model of an anharmonic oscillator with a periodic perturbation to describe the unidirectional vibrations of the well of a nuclear power reactor. We show that for the parameters of working water-moderated and -cooled (water–water) reactors the vibrations of the reactor well can become stochastic. Finally, in Sec. 5 we demonstrate the possibility of converting the chaotic vibrations of an anharmonic oscillator into regular vibration by introducing a small programmed perturbation into the system. We study the stability of the proposed algorithm of chaos control against additive Gaussian noise.

2. THE STRUCTURE OF PHASE SPACE

In this paper we study the classical particle dynamics generated by the Hamiltonian

$$H(p,x,t)=H_0(p,x)+Fx\cos\Omega t, \tag{1}$$

where the unperturbed Hamiltonian is

$$H_0(p,x)=\frac{p^2}{2m}+Ax^n=E \quad (n=2l, \quad l>1). \tag{2}$$

Here p, x, and m are the particle momentum, coordinate, and mass, and F and Ω are the amplitude and frequency of the external field.

The anharmonic oscillator described by the Hamiltonian (2) fills the gap between two important physical models, the harmonic oscillator ($n=2$) and the infinitely high rectangular well ($n=\infty$). Research in both classical and quantum dynamics of the anharmonic oscillator (2) has a rich history. (A detailed list of papers on the subject can be found in Ref. 12.)

The Hamiltonian (1) can be written in terms of dimensionless variables if we introduce arbitrary units of mass m_0, length l_0, and time t_0. The new parameters (m', A', F', and Ω') are linked to the old ones through the relationship

$$A'=\frac{Fl_0^{n-2}t_0^2}{m_0}, \quad \Omega'=\Omega t_0,$$

$$F'=\frac{FT_0^2}{m_0 l_0}, \quad m'=\frac{m}{m_0}. \tag{3}$$

By proper selection of the units m_0, l_0, and t_0 we can use (3) to obtain ''convenient'' values of the parameters m', A', F', and Ω'. Of the four parameters three can be fixed. For instance, by selecting $A'=m'=\Omega'=1$ we determine the following set of the basic units:

$$m_0=m, \quad t_0=\frac{1}{\Omega}, \quad l_0=\left(\frac{m\Omega^2}{A}\right)^{1/(n-2)}. \tag{4}$$

Here the only parameter F that enters into the Hamiltonian

$$H=\frac{p^2}{2}+x^n+Fx\cos t \tag{5}$$

(below we drop the prime wherever this does not lead to a misunderstanding) can be expressed in terms of the initial physical parameters:

$$F'=F\left(\frac{1}{m\Omega^2}\right)\left(\frac{A}{m\Omega^2}\right)^{1/(n-2)}=\frac{F}{m\Omega^2 l_0}. \tag{6}$$

The scaled Hamiltonian $H_0(p,x)$ can be expressed in terms of the action and angle variables (I and θ) as follows:[11,13]

$$H_0(I)=\left[\frac{2\pi}{\alpha G(n) I}\right]^{\alpha}, \tag{7}$$

where

$$G(n)=2(2\pi)^{1/2}\frac{\Gamma(1+1/n)}{\Gamma(1/2+1/n)}, \quad \alpha=\frac{2n}{n+2}. \tag{8}$$

The resonant values of the action, I_k, which are determined from the condition

$$k\omega(I)=1, \quad \omega(I)=\frac{\partial H_0}{\partial I}, \quad k=1,2,\ldots, \tag{9}$$

are

$$I_k=\alpha\left[\frac{G(n)}{2\pi}\right]^{2n/(n-2)}k^{(2+n)/(2-n)}. \tag{10}$$

If we expand the perturbation in a Fourier series in θ, the total system Hamiltonian can be written

$$H(I,\theta,t)=H_0(t)+\sum_{k=-\infty}^{\infty}x_k(I)\cos(k\theta-t), \tag{11}$$

where the Fourier coefficients x_k are given by the integrals

$$x_k(I)=\int\frac{d\theta}{2\pi}x(I,\theta)\exp(ik\theta). \tag{12}$$

An analysis based on the Chirikov resonance overlap criterion[14] leads to an expression for the critical external-field amplitude that ensures the overlap of neighboring odd resonances[11] (to first order in the amplitude F of the external field the widths of even resonances for the potential considered here are zero):

$$F_k^{\mathrm{cr}}=2^{(2-3n)/(n-2)}\frac{n(n-2)}{(n+2)^2x_k}\left[\frac{G(n)}{\pi}\right]^{2n/(n-2)}k^{4/(n-2)}$$
$$\times[k^{(2+n)/(2-n)}-(k+1)^{(2+n)/(2-n)}]^2. \tag{13}$$

This expression solves the problem of reconstructing the structure of the phase space of the Hamiltonian (1) for arbitrary values of the parameters. The analysis of (13) reveals the following main features of this structure:

1. For a rectangular well ($n=\infty$), for every value of F there exists an energy (or number of the resonance) above or below of which the motion is regular or chaotic, respectively.

2. For every $n<\infty$ there is always a value $F_0(n)$ of the external perturbation at which for all $F<F_0(n)$ the motion is regular at all energies.

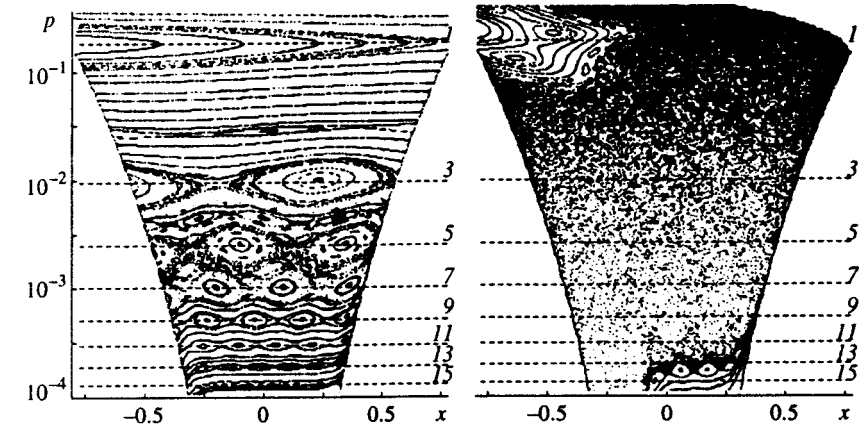
3. For every $F>F_0(n)$ the anharmonic oscillator undergoes a regularity-to-chaos-to-regularity transition, i.e., in this case we can always specify an energy interval within which the motion is chaotic and outside of which the motion is regular.

The reason for a regularity-to-chaos-to-regularity transition is the following.[10,11] For high-lying resonances (small k) and fixed F, the resonance widths decrease as a function of the order more slowly than the distances between the resonances. This ensures the existence, as the perturbation increases in strength, of the well-known (normal) regularity-to-chaos transition. However, this is true only within a limited energy interval. The increase of the relative rate of decrease of the resonance width in comparison with the rate at which the distance between resonances decreases leads to a situation in which the conditions for resonance overlap cease to be met, which results in a new (anomalous) chaos-to-regularity transition.

The structure of the phase space varies smoothly with the degree of anharmonicity n. Hence for a typical example we take the numerical results for $n=8$. Figure 1 depicts a phase diagram that makes it possible, for a fixed level of external perturbation, to determine the energy intervals of regular and chaotic motions. The synoptic dependence $E(x,t_n=2\pi n)$ supports the hypothesis that there is an anomalous chaos-to-regularity transition. Figure 1 clearly shows that for $k>13$ there are still isolated resonances corresponding to regular motion.

3. STOCHASTIC DIFFUSION

When the amplitude F of the external perturbation is large, there is a strong overlap of resonances, and the evolution of the system can be described by a distribution function $f(I,t)$ satisfying the Fokker–Planck equation (the quasilinear approximation):



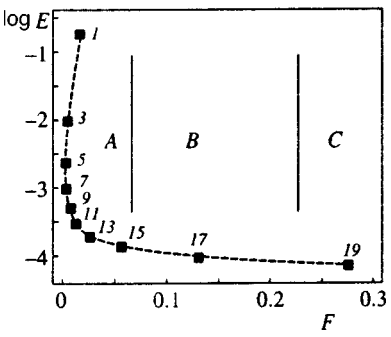


FIG. 1. Above: stroboscopic images of the phase space for $n=8$; to the right, $F=0.01$; to the left, $F=0.2$. Below: the phase diagram for $n=8$. The dashed curve separates the regular region (to the left of the curve) from the chaotic region (to the right of the curve). The black squares show the stability boundary for the resonances whose numbers stand at the squares, and A, B, and C are the three regions in the phase diagram corresponding to different stochastic diffusion regimes.

$$\frac{\partial f}{\partial t}=\frac{\partial}{\partial I}D_{\rm ql}(I)\frac{\partial f}{\partial I}, \tag{14}$$

where $D_{\rm ql}(I)$ is the quasilinear diffusion coefficient. A consistent derivation of Eq. (14) and of an expression for the quasilinear diffusion coefficient presupposes the use of the Liouville equation as the starting equation for the distribution function in the (I,θ) space with subsequent averaging of the distribution function over the initial phases and with analysis of the conditions for the smallness of the terms discarded in the equation for the averaged distribution function (see, e.g., Ref. 1). In this paper we limit ourselves to the derivation of a formula for $D_{\rm ql}$. For the starting expression we use the formula for the diffusion coefficient for a test particle,

$$D_{\rm ql}(I)=\lim_{t\to\infty}\frac{1}{2t}\int_0^t dt_1\int_0^t dt_2\,\overline{(\dot I(t_1)\dot I(t_2))}\,, \tag{15}$$

where $\dot I\equiv dI/dt$, and the horizontal bar indicates averaging over the initial phases. Below (in the present section) we distinguish between dimensional (unprimed) and dimensionless (primed) quantities. The correlator in (15) can be calculated via the equations of motion in the (I,θ) space:

$$\begin{aligned}\frac{dI}{dt}&=F\sum_{k=-\infty}^{\infty}kx_k(I)\sin(k\theta-\Omega t),\\ \frac{d\theta}{dt}&=\omega(t)+F\sum_{k=-\infty}^{\infty}\frac{dx_k(I)}{dI}\cos(k\theta-\Omega t).\end{aligned} \tag{16}$$

What is important is that the integration of the second equation in (16) in the quasilinear approximation is done along the unperturbed path,

$$\theta(t)\approx\theta_0+\omega\tau. \tag{17}$$

Plugging (17) into the first equation in (16) and then into (15), averaging over the initial phases, and using the new variables t_1-t_2 and $(t_1+t_2)/2$, in the $t\to\infty$ limit we arrive at an expression for the quasilinear diffusion coefficient:

$$D_{\rm ql}\frac{\pi F^2}{2}\sum_{k=-\infty}^{\infty}k^2x_k^2\delta(k\omega(I)-\Omega). \tag{18}$$

We can use this formula to estimate the local diffusion coefficient at $I\approx I_k$. To this end we could use the explicit expression for the Fourier coefficients, which are defined by (12). This expression is too cumbersome, however. It was found that to make simple estimates we can approximate $x_k'=x(A/E)^{1/n}$ by the formula

$$x_k'\approx 0.47\,\exp\left(-\frac{3k}{4}\right) \tag{19}$$

at $n=8$. Figure 2 illustrates the fact that (19) is a good approximation of the x_k' vs. k dependence for $k\geqslant 3$. Then, combining (18) and (19), we arrive at an estimate (to within a numerical factor of order unity) for the local quasilinear diffusion coefficient with $n=8$:

$$D_{\rm ql}(I_k)\approx F^2\left(\frac{E_k}{A}\right)^{1/4}\frac{\Omega^2}{\omega^3(I_k)}\exp\left\{-\frac{3\Omega}{2\omega(I_k)}\right\}. \tag{20}$$

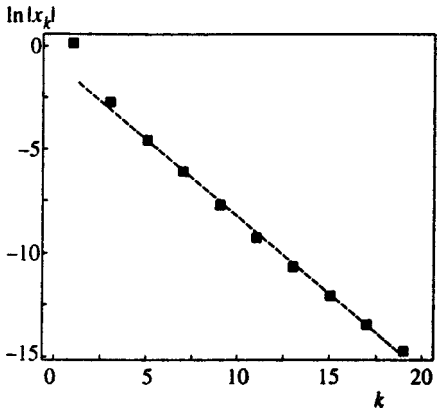


FIG. 2. Values of the logarithm of the dimensionless Fourier coefficient x_k for k odd (the black squares). The dashed line represents Eq. (19).

The results of a numerical investigation of stochastic diffusion in our model ($n=8$) are depicted in Fig. 3. Numerical analysis revealed the presence of three regions in the bifurcation diagram in the dimensionless variables (E,F) characterized by the different diffusion regimes A, B, and C (see Fig. 1). Below we give a brief qualitative characteristic for each regime.

Region A, in which the amplitude F of the external perturbation is only slightly larger than the minimum critical value $F_0(8)$. Here the primary resonances overlap only slightly, and diffusion involves secondary resonances. Figure 3a depicts a chaotic path of the dimensionless variables t_n and I, with $t_n=2\pi n$, which first wanders around the fifth primary resonance and then jumps to the seventh primary resonance. Figure 3c depicts the dependence of $\langle I^2\rangle$ on t_n at $F=0.01$, which is of a linear nature, with $\langle\ \rangle$ standing for the probability average. Note that the diffusion coefficient $D=\lim_{t\to\infty}\langle I^2\rangle/t$, defined in this diagram as the slope of the line to the time axis, is different for the random walks of the particle near the fifth and seventh primary resonances.

Region B corresponds to moderate external-perturbation amplitudes. Here the resonances effectively overlap and hence the conditions for the quasilinear approximation are met. The chaotic path $I(n)$ at $F=0.2$ is depicted in Fig. 3b, and the dependence of $\langle I^2\rangle$ on n is shown in Fig. 3d. The linear nature of this dependence is evident.

Region C corresponds to large external-perturbation amplitudes. Figure 4 illustrates the specific features of diffusion in this region. Clearly, the dependence of the diffusion coefficient on the square of F, which is linear in B [this corresponds to the quasilinear approximation; see Eq. (18)], in C becomes less slanted. This may be explained as follows. In deriving the expression for the quasilinear diffusion coefficient we integrated with respect to the angle along the unperturbed path [see Eq. (17)]. As F increases, we must allow for its effect on the path. In the theory of resonance broadening,[15] this effect is taken into account by replacing the delta function in the quasilinear diffusion coefficient with a resonance curve of a finite width, which increases with F. We allow for this feature by replacing the delta function in (18) with an exponential:

$$\delta(k\omega-\Omega)\to\frac{1}{\sqrt{\pi}\,\gamma}\exp\left\{-\frac{(k\omega-\Omega)^2}{\gamma^2}\right\}, \tag{21}$$

where $\gamma(F)$ is the exponential-curve width, which increases with F. Then for $\gamma\ll\omega,\Omega$ we have an estimate for the quasilinear diffusion coefficient:

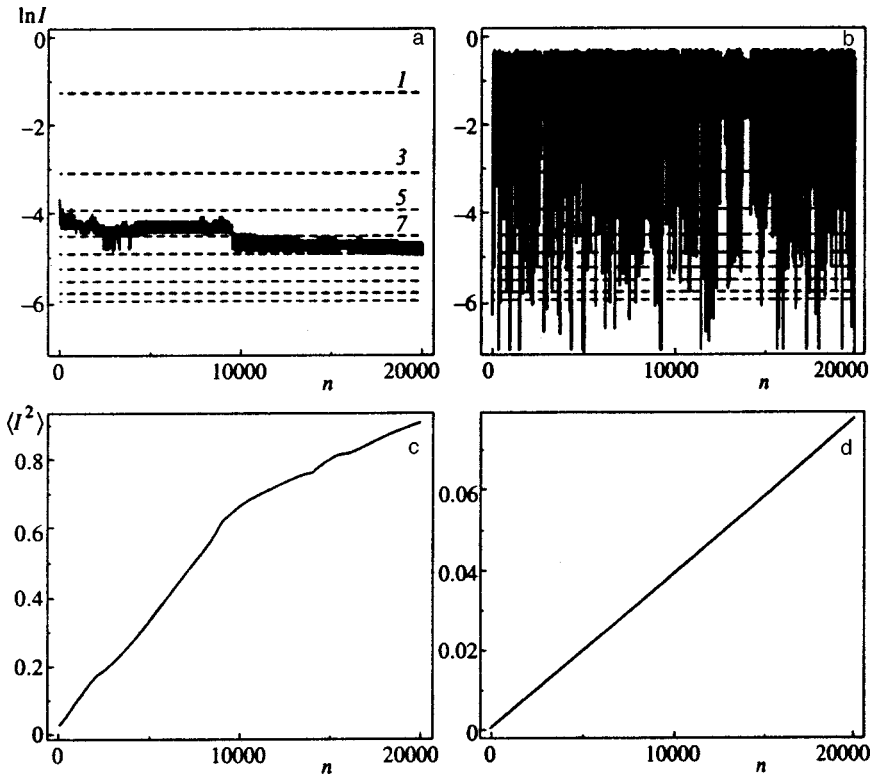


FIG. 3. Upper row: a chaotic path in the $(\ln I,n)$ space. Lower row: the dependence of $\langle I^2\rangle$ on n: (a) and (c), at $F=0.01$ (region A); and (b) and (c), at $F=0.2$ (region B).

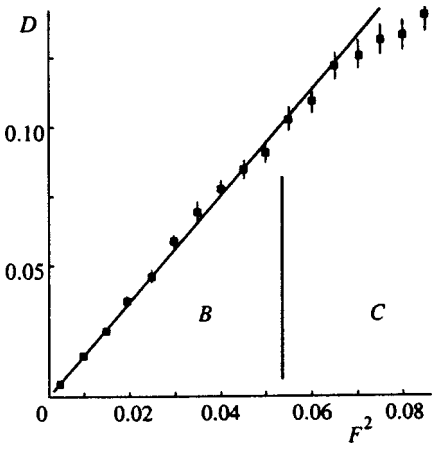


FIG. 4. Diffusion coefficient $D=\langle I^2\rangle/n$ as a function of the square of the external-perturbation amplitude for regions *B* and *C*.

$$D_{\mathrm{ql}}(I)\to\tilde{D}_{\mathrm{ql}}(I)=D_{\mathrm{ql}}(I)(1-\Delta),\quad \Delta\propto\frac{\gamma^2}{\omega\Omega}. \tag{22}$$

For $\gamma=0$ (22) leads to the previous estimate (20) and hence to a linear dependence of the diffusion coefficient D on F^2. The estimate (22) clearly shows that an increase in F causes an increase in γ, so that the dependence of the diffusion coefficient on F^2 becomes less slanted.

4. THE ANHARMONIC OSCILLATOR UNDER A PERIODIC PERTURBATION AS A MODEL OF REACTOR-WELL VIBRATIONS

As a possible application of the above results, we study the possibility of chaotic regimes setting in in reactors of the water-moderated water-cooled power type (VVÉR in Russian), or simply a water–water reactor. In the course of operation, the reactor well experiences stresses due to the powerful fluxes of the heat-transfer agent from the cold legs of the circulation loops. In view of this, the well vibrates in relation to the reactor vessel, and these vibrations are registered and analyzed by neutron-sensing elements placed outside the well but inside the vessel.[16] The monitoring of the movements of the well and the inner components during operation is an important problem of reactor noise diagnostics, which is being actively developed in countries that have nuclear reactors (see the materials[17] of the Specialists Meeting on Reactor Noises, SMORN VII). What is especially worth mentioning is the possibility that (due to wear and tear of some of the fastening elements) (quasi)pendulum vibrations of the well as a beam with a fixed upper end can develop, vibrations involving contact with the vessel in the lower part of the well.[18] Identifying, describing, and monitoring changes in the characteristics of these vibrations are important in the early stages of diagnosing construction defects acquired in the process of operation.

Monitoring the reactor well amounts to measuring and processing the signals from noise sensors and analysis of these signals via theoretical models. In interpreting the data on the vibrations of the well (and other construction elements of the reactor), models of linear oscillators (or coupled linear oscillators) are used. In these models the values of the masses and the stiffness and friction coefficients are obtained by combining the results of theoretical calculations, bench tests, computer modeling, and pre-operational tests.[19] Generally, the external force (perturbation) in these models consists of two components: the random component, which results from local turbulent pulsations, and the regular component, with several high peaks in the spectrum (for more details concerning the origin of forces acting on the reactor well see the review in Ref. 19).

From the variations from measurement to measurement of, say, the spectral characteristics (auto power spectra, coherence functions, and the phase) of the noise signals, which register well vibrations, we can draw conclusions concerning the state of the devices inside the reactor vessel and of fastening elements. The possibility of this follows from the fact that different types of possible defects lead to variations in different parts of the spectrum.[20] However, when interpreting the data of measurements one must bear in mind that due to contact with the reactor vessel the (quasi)pendulum vibrations of the well are essentially nonlinear,[18] so that a periodic external force may make these vibrations chaotic. The properties of chaotic motion cannot be taken into account by linear models, with the result that such models must be augmented by an analysis of possible nonlinear effects.

The first to study nonlinear effects in (quasi)pendulum vibrations of a reactor well was Vavrin,[21] who, however, did not consider the possibility of stochastization of these vibrations by an external periodic force. According to a recent report,[22] theoretical and experimental investigations of the nonlinear effects in the vibrations of other elements of the intravessel devices, such as control rods and clusters of fuel elements, have just started.

Only estimates done with appropriate models can answer the question of whether chaotic regimes occur at values of parameters corresponding to operating reactors. As one of the simplest models describing unidirectional quasipendulum vibrations of a reactor well we can take the above model of an anharmonic oscillator with a periodic perturbation. The oscillator's potential energy models the elastic interaction that emerges when the reactor well contacts the reactor vessel, while the periodic perturbation models the force at the frequency of rotation of the main circulating pumps.

To determine the type of motion (chaotic or regular) occurring in quasipendulum vibrations we must connect the parameters of the model with the parameters of the reactor. Let Δ be the amplitude of reactor-well vibrations at a fixed energy,

$$E=A\Delta^n. \tag{23}$$

Then for the stiffness coefficient A expressed in terms of the amplitude Δ we have

$$A=\frac{2}{\pi}m\omega^2\Gamma_n^2\Delta^{2-n},\quad \Gamma_n\equiv\frac{\Gamma(1+1/n)}{\Gamma(1/2+1/n)}. \tag{24}$$

In terms of Δ, the unit of length, l_0, is

$$l_0=\left[\left(\frac{2}{\pi}\right)^{1/2}\Gamma_n\omega\right]^{2/(2-n)}\Delta, \tag{25}$$

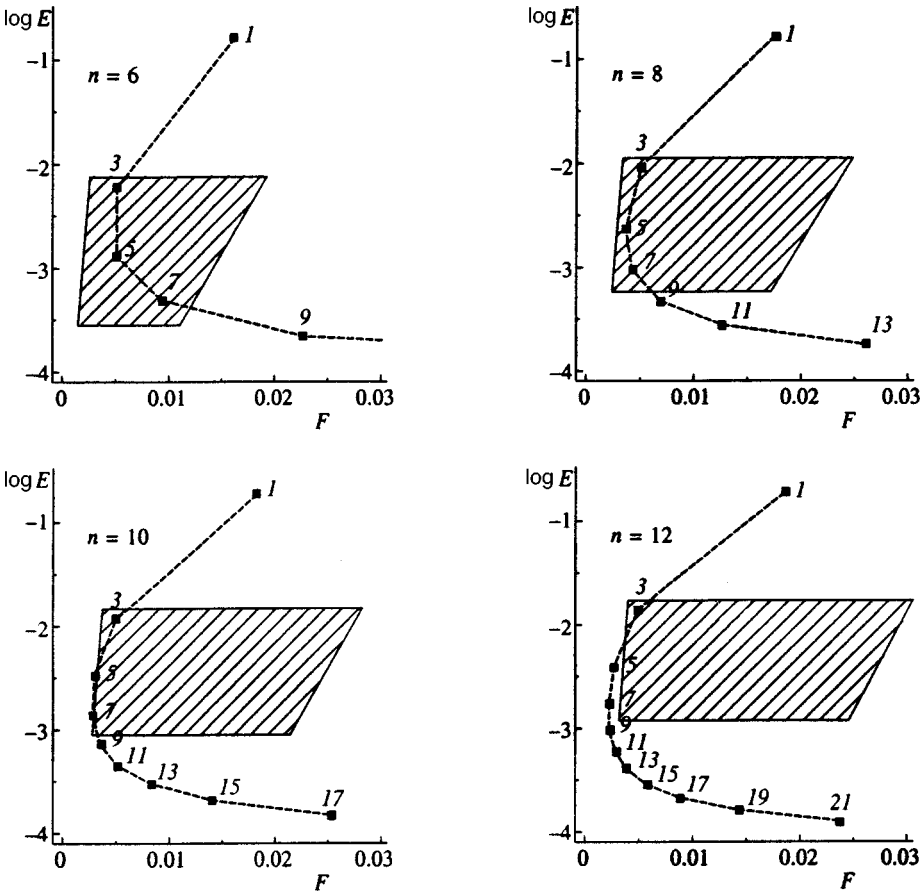


FIG. 5. Phase diagrams for different n. The hatched areas correspond to the parameters of the VVÉR-1000 reactor.

and hence the dimensionless energy is

$$E' = \frac{E}{m\Omega^2 l_0^2} = \left[\frac{2}{\pi}\frac{\Gamma_n \omega}{\Omega}\right]^{2n/(2-n)}. \tag{26}$$

The law of variation of the momentum of the coolant flow acting on the reactor well yields the force exerted by the cold leg of a circulation loop:

$$F = \rho \nu^2 \pi r^2, \tag{27}$$

where ρ is the density of the heat-transfer agent, ν is the average velocity of the heat-transfer agent in the cold leg, and r is the radius of the pipe. The dimensionless resulting force exerted by two opposite circulation loops is

$$F' = \frac{F}{m\omega^2 l_0} = \frac{\alpha}{\Delta}\frac{\rho \nu^2 \pi r^2}{m\omega^2}\left[\left(\frac{2}{\pi}\right)^{1/2}\frac{\Gamma_n \omega}{\Delta}\right]^{2/(n-2)}, \tag{28}$$

where the coefficient α characterizes the imbalance of the opposing pumps.[23]

Equations (26) and (28) make it possible to determine where in the diagram in Fig. 1 the region corresponding to the characteristic values of the parameters for quasipendulum vibrations of the well of a VVÉR-1000 reactor is situated. The rated values of the parameters are (1) $m = 2.20 \times 10^5$ kg, (2) $W = 2\pi 16.6$ Hz, (3) $r = 0.425$ m, and (4) $\Delta = 0.06$ mm.

In Fig. 5, the hatched areas of the phase diagrams for different values of n correspond to the following parameter intervals: $\omega = 2\pi \times (2-6)$, and $\alpha = 0.02-0.15$ (see Ref. 24; as in the previous figures, we have discarded the primes on dimensionless quantities). We see that for a ''typical'' set of parameters chaotic modes may set in. Stochastization of the vibrations of the reactor well occurs in the vicinity of the resonances that overlap for the given level of periodic perturbation. Here the parameter n determines the ''relative'' weight of the region with regular and chaotic motion within the range of admissible values of the parameters.

5. CONTROL OF THE CHAOTIC VIBRATIONS OF AN ANHARMONIC OSCILLATOR

An important feature of the dynamics of a nonlinear system interacting with a periodic external field is the possibility of nonresonant (diffuse) absorption of energy. As applied to the problem being discussed, this process may cause the amplitude of the reactor well vibrations to build up dangerously due to absorption of the energy of the heat-transfer agent. Chaos plays a useful role since, by stimulating mixing, it affords a powerful heat and mass transfer mechanism. However, in many cases (in particular, for vibrations of the reactor well) chaos is undesirable, since it leads to additional mechanical fatigue of the fastening elements, has an adverse effect on the operation of measuring devices, and hinders the analysis of spectra. For this reason there have been significant recent attempts to control chaos, i.e., to convert it into a periodic process. To this end a small programmed variation of one (or several) parameters of the system is enforced[25] or a specially constructed external perturbation is used.[26]

As is well known, the extremely high sensitivity to small perturbations is the key feature of chaotic systems. In 1990, the OGY group (Ott, Grebogi and Yorke[25]) proposed using this feature to control dissipative dynamical systems whose phase space contains a small attractor. OGY suggested the following procedure to stabilize any unstable periodic orbit lying on the attractor. In the preliminary stage, the equations of motion or time series of the measured characteristics of the system[27] are used to construct the Poincaré section $\mathbf{Z}_{n+1}=\mathbf{F}(\mathbf{Z}_n,p_0)$ for a certain fixed value of the parameter $p=p_0$ of the system available for control. When a point $\mathbf{Z}_n$ on the chaotic path of the system in the Poincaré section is found to be close to a point $\mathbf{Z}^*(p_0)$ of the objective unstable orbit of period $k(\mathbf{Z}^*(p_0)=\mathbf{F}^k(\mathbf{Z}^*(p_0),p_0))$, we slightly change the value of the control parameter p $(p_n=p_0+\delta p_n)$ so that in the next iteration the point $\mathbf{Z}_{n+1}=\mathbf{F}(\mathbf{Z}_n,p_n)$ lands in the stable set of the saddle point $\mathbf{Z}^*(p_0)$. This condition leads to the main formula of the OGY control method:

$$\delta p_n=\frac{\lambda_u \mathbf{f}_u\cdot\delta\mathbf{Z}_n}{(\lambda_n-1)\mathbf{f}_u\cdot\mathbf{w}}, \tag{29}$$

where

$$\delta\mathbf{Z}_n=\mathbf{Z}_n-\mathbf{Z}^*(p_0),\quad \mathbf{w}=\left.\frac{\partial\mathbf{F}(\mathbf{Z},p)}{\partial p}\right|_{\mathbf{Z}^*,p_0},$$

λ_u is an unstable eigenvalue of the Jacobi matrix describing in the linear approximation the dynamics in the vicinity of $\mathbf{Z}^*(p_0)$, and $\mathbf{f}_u$ is a contravariant unstable eigenvector of the Jacobi matrix. Formula (29) can be written

$$\delta p_n=C\mathbf{f}_u\cdot\delta\mathbf{Z}_n, \tag{30}$$

i.e., the deviation of the parameter from the rated value needed for control is proportional to the vector $\delta\mathbf{Z}_n$ on the unstable direction $\mathbf{f}_u$, and the constant C is calculated from the projection on the same direction of the system's reaction $\mathbf{w}$ to the perturbation of parameter p. The effectiveness of the OGY control method and its numerous generalizations has been demonstrated in all areas of science (physics, chemistry) and in medicine (see Ref. 28).

The vast majority of examples of chaos control achieved to date refer to motion on a strange attractor. The complexity of chaos control in Hamiltonian systems is due to two facts.[29]

1. In view of conservation of phase volume, for some points of a periodic path may have complex eigenvalues lying on the unit circle. Hence in this case the OGY formula (29) cannot be used directly. Of course, we can, using the Jacobi matrix for the orbit of period k, obtain real eigenvalues via formula (29). In this case, however, using the control method at each kth step, we may either fail to achieve control or may lose control very fast even in the event of a weak noise.

2. The second reason is related to the duration τ of the transient period just before control locking. In dissipative chaotic systems, for randomly selected initial conditions, τ has an exponential probability distribution,[30]

$$P(\tau)\propto\exp\left(-\frac{\tau}{\langle\tau\rangle}\right) \tag{31}$$

for τ large, where $\langle\tau\rangle$ is the average control locking time, which is a power-law function of the range of admissible values of the parameter,

$$\langle\tau\rangle\propto(\delta p)_{\max}^{-D}, \tag{32}$$

where D is the information size of the attractor. Hence in this case $\langle\tau\rangle$ is always finite. In Hamiltonian systems,[31]

$$P(\tau)\propto\tau^{-\alpha}, \tag{33}$$

where $1<\alpha<2$. In this case the average control locking time $\langle\tau\rangle$ is infinite.

In view of these difficulties, to stabilize the chaotic vibrations of an anharmonic oscillator, we use an alternative approximation, which we call simple proportional control. In order to formulate the procedures of this approximation, we note that the control process can be broken down into several stages.

1. The system approaches the objective periodic orbit by ergodic random walks. Special methods are sometimes used to shorten this stage.

2. Control locking. Here maximum control is exerted on the system.

3. Exponential approach to the objective (γ is the exponent). The numerical value of γ can be interpreted as the measure of effectiveness of control and can be used as the objective function to optimize control.

4. The stabilization stage with a small chaotic component, whose value is determined either by noise or by fluctuations due to the finite accuracy with which numbers are represented in a computer.

5. Recreation of chaos after control has ceased. The stage is characters by exponentially rapid departure from the objective orbit and by chaos buildup.

Note that the classification is universal both for discrete maps and for fluxes described by ordinary differential equations.

We consider a system whose dynamics is described by a two-dimensional map (for continuous systems we assume that this map is generated by the corresponding Poincaré section)

$$\mathbf{Z}_{n+1}=\mathbf{F}(\mathbf{Z}_n). \tag{34}$$

Then we call the map

$$\mathbf{Z}^c_{n+1}=\mathbf{F}(\mathbf{Z}_n)+\Delta\mathbf{F}(\mathbf{Z}_{n+1}) \tag{35}$$

a totally controlled map, where $\Delta\mathbf{F}(\mathbf{Z}_{n+1})=\hat{\mathbf{C}}(\mathbf{Z}_{n+1}-\mathbf{Z}^*_{n+1})$, with

$$\hat{\mathbf{C}}\equiv\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}$$

the matrix of control coefficients, and $\mathbf{Z}^*_{n+1}$ an unstable fixed point or one of the points of the unstable period orbit of the map. Note that in the case where the additive control parameter enters only into one map row (e.g., the upper row), the

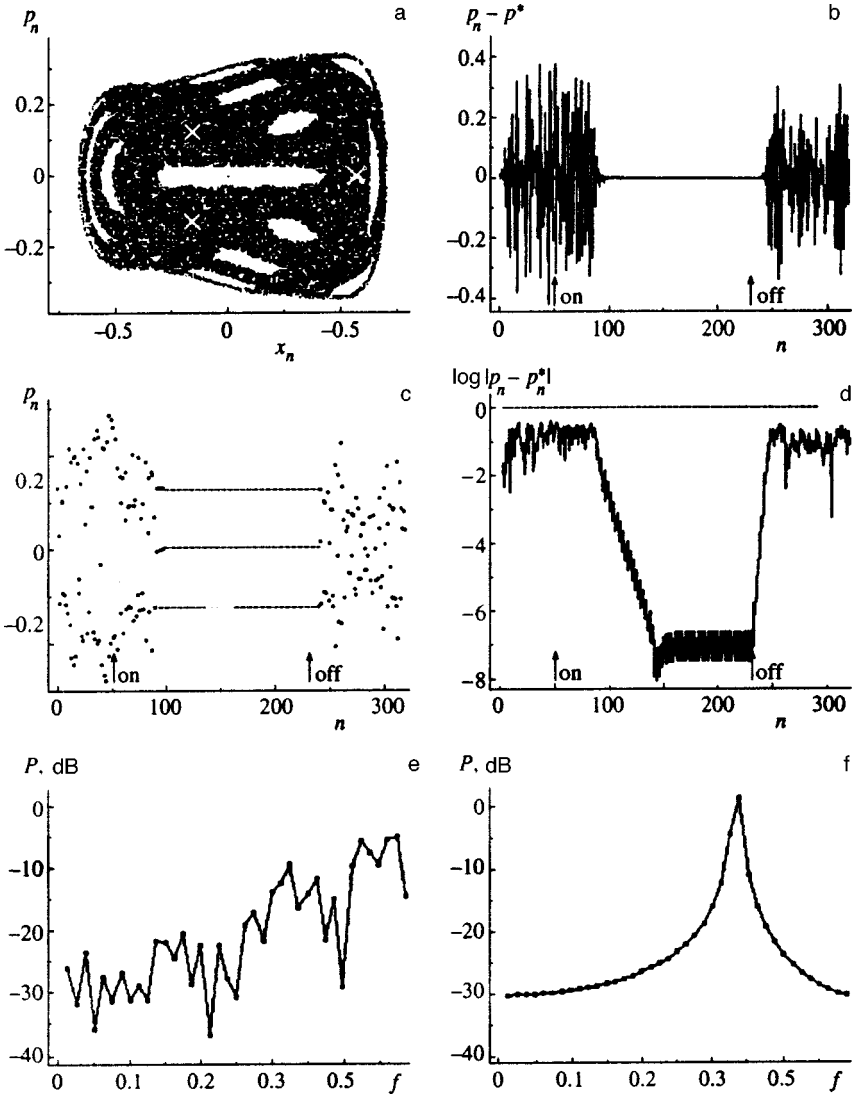


FIG. 6. Stabilization of the chaotic vibrations of an anharmonic oscillator by simple proportional control. The unstable orbit with period 3 is the objective orbit.

map (35) reduces to a controlled OGY map: the coefficients C_{11} and C_{12} are determined by OGY formulas, and $C_{21}=C_{22}=0$.

To find the coefficients C_{ij} we can use one of the usual optimization methods (e.g., the gradient method, the method of conjugate directions, the random search method), taking the exponent γ as the objective function.

Figure 6 depicts the results of stabilization by controlling the unstable orbit with period 3 of the anharmonic oscillator (5), with the matrix $\hat{\mathbf{C}}$ found by random search. The admissible perturbation of the system in the process of control is at most 0.03. Figure 6a depicts the synoptic Poincaré section of the classical phase space $\{p(t_n),x(t_n)\}$ at times $t_n=2\pi n(n=0,1,2,\ldots)$. All points in the chaotic region were obtained via a single path; the light $\times$'s denote the objective unstable orbit with period 3. Figure 6b depicts the difference $p(t_n)-p_F(i)$ $(i=1,2,3)$ between the running synoptic component $p(t_n)$ and the p-coordinateof the unstable periodic orbit.

The control method was initiated at the 50th iteration and was discontinued at the 230th iteration (the small vertical arrows indicate these moments). We see that the the stage of ergodic random walks, which precedes control locking, lasts 30 to 40 iterations. In Fig. 6c the control process is illustrated by the time series $p(t_n)$. The control process depicted in Fig. 6d on the log-linear scale makes it possible to follow all the control stages mentioned earlier. Finally, in Figs. 6e and 6f we depict the power spectra $P(f)$ of the system without the control method (e) and with the control method (f) $(n=150–230)$. In the latter case there is a distinct peak, corresponding to the frequency of the objective periodic orbit $(f=1/3)$.

Figure 7 illustrates the stability of the adopted algorithm of chaos control against additive Gaussian noise. It depicts the same characteristics of the system as Fig. 6, with the only difference that $\varepsilon\delta_{xn}$ and $\varepsilon\delta_{pn}$ have been added to the right-hand sides of the equations. The independent random variables δ_{xn} and δ_{pn} have a Gaussian probability distribution, a zero mean value, and a unit variance. The noise amplitude ε was assumed equal to 0.03. It is possible to retain an adequate quality of control (Figs. 7b, 7c, and 7d) and fairly good spectral purity (Fig. 7f) provided that the perturbation level grows only up to 0.06.

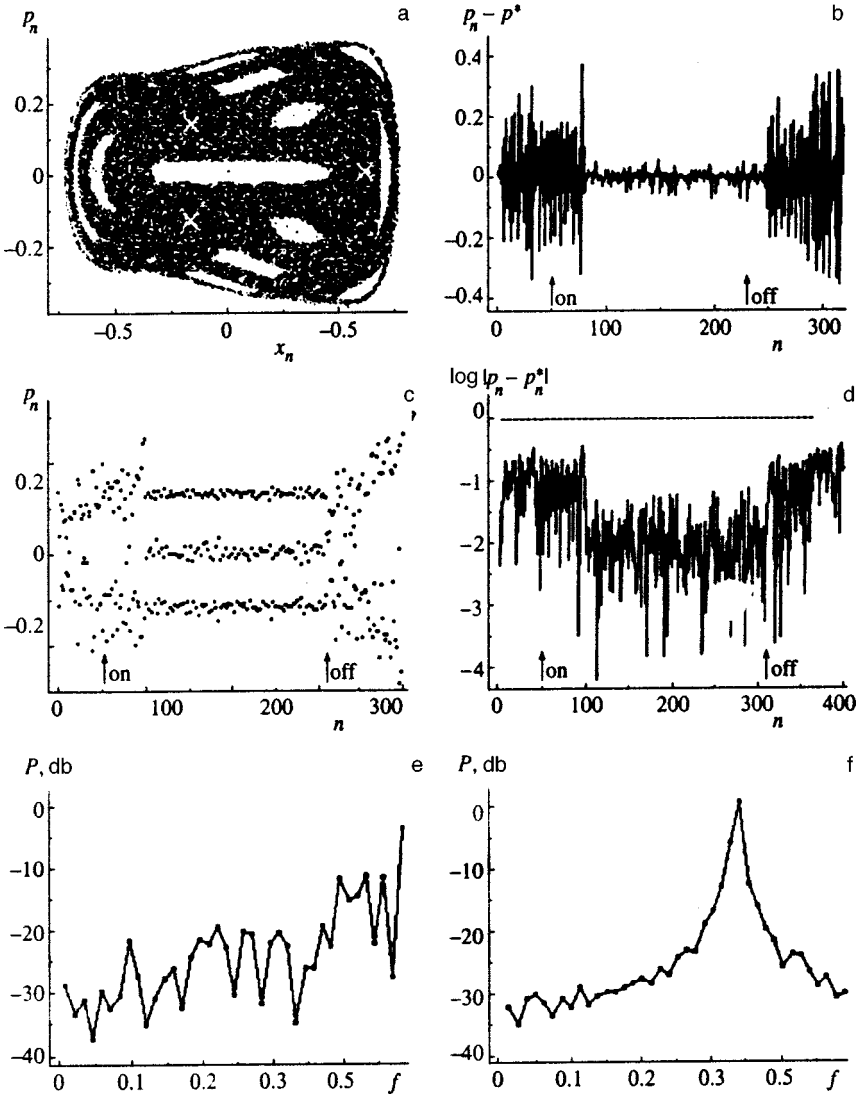


FIG. 7. The same as in Fig. 6 with external noise.

6. CONCLUSION

In this paper we have studied the dynamics of an anharmonic oscillator subjected to a periodic monochromatic perturbation in relation to the stochastization of the vibrations when the system parameters are varied.

A numerical analysis of the equations of motion made it to distinguish three regions representing different diffusion regimes in the (E,F) plane. There is the region of small amplitudes, where diffusion primarily involves secondary resonance, and only occasionally do transitions between neighboring primary resonances occur. There is also the region of moderate amplitudes, where the conditions for the quasilinear approximation are met. The $\langle I^2 \rangle$ vs. t_n dependence in this region is clearly linear. Finally, there is the region of large amplitudes, in which deviations from the quasilinear approximation occur. Allowance for the effect of an external perturbation on the unperturbed path used in deriving the quasilinear diffusion coefficient provided a qualitative explanation for the observed deviation.

As a possible application of our results we have examined the possibility of chaotic regimes developing in quasipendulum vibration of the well of a reactor of the water–water type. We have found that for values of the parameters of operating reactors of this type the well vibrations may be either regular or chaotic, depending on the well–vessel gap and the resulting force with which the heat-transfer agent flux acts on the reactor well. Allowance for the possibility of stochastization of well vibrations may help to interpret the noise spectra and hence to increase the reliability of conclusions about the state of intravessel devices.

We have proposed an algorithm of simple proportional control, which converts the chaotic vibrations of an anharmonic oscillator into periodic vibrations, thus preventing a dangerous buildup of the vibration amplitude (in the case of chaotic vibrations of the reactor well) due to nonresonant energy absorption.

The authors would like to express their deep gratitude to V. V. Bulavin (Russian Science Center ''Kurchatov Institute,'' Russia) and I. Pazsit (Chalmers University of Science, Sweden) for discussing the problems related to applying the nonlinear oscillator model to reactor well vibrations of water–water reactors. The work was made possible by a grant from the Fund for Basic Research of the Ukrainian

Ministry of Science (Project 2.4/342) and by information support from the INTAS Project (LA-96-09).

*)E-mail: vgonchar@kipt.kharkov.ua

[1] G. M. Zaslavsky, *Chaos in Dynamical Systems*, Harwood Academic, New York (1985).
[2] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, Springer, Berlin (1992).
[3] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics*, Wiley, New York (1989).
[4] L. E. Reichl, *The Transition to Chaos*, Springer, Berlin (1992).
[5] N. B. Delone, V. P. Kraĭnov, and D. L. Shepelyanskiĭ, Usp. Fiz. Nauk **140**, 355 (1983) [Sov. Phys. Usp. **26**, 551 (1983)].
[6] G. Casati *et al.*, Phys. Rep. **154**, 77 (1987).
[7] R. V. Jensen, Phys. Rev. A **30**, 386 (1984).
[8] M. Guttzviller, *Chaos in Classical and Quantum Mechanics*, Springer, Berlin (1991).
[9] P. V. Elyutin, Usp. Fiz. Nauk **155**, 397 (1988) [Sov. Phys. Usp. **31**, 597 (1988)].
[10] Yu. L. Bolotin, V. Yu. Gonchar, and M. Ya. Granovskiĭ, JETP Lett. **59**, 651 (1994).
[11] Yu. L. Bolotin, V. Yu. Gonchar, and M. Ya. Granovsky, Physica D **86**, 500 (1995).
[12] H. Breuer and M. Holthaus, Ann. Phys. (N.Y.) **211**, 249 (1991).
[13] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).
[14] B. V. Chirikov, At. Energ. **6**, 630 (1959).
[15] T. H. Dupree, Phys. Fluids **9**, 1773 (1966).
[16] I. F. Akerhielm, R. Espefalt, and J. Lorenzen, Prog. Nucl. Energy **9**, 453 (1982).
[17] *Proc. Symposium on Nuclear Reactor Surveillance and Diagnostics (SMORN VII)*, 19–23June 1995, Avignon, France (1995).
[18] V. V. Bulavin, V. I. Pavelko, and D. F. Gutsev, At. Energy **79**, 343 (1995).
[19] E. Alstadt, M. Scheffler, and F.-P. Weiss, Prog. Nucl. Energy **29**, 129 (1995).
[20] J. A. Thie, Nucl. Technol. **45**, 5 (1979).
[21] J. Vavrin, in *Proc. Symposium on Nuclear Reactor Surveillance and Diagnostics (SMORN VI)*, Gatlinburg, Tennessee, USA (1991).
[22] O. Thomson, N. S. Garis, and I. Pazsit, submitted to Nucl. Technol.
[23] J. Runkel, D. Stegemann, J. Fiedler *et al.*, in *Proc. Symposium on Nuclear Reactor Surveillance and Diagnostics (SMORN VII)*, Avignon, France (1995).
[24] G. Por, K. A. Kantor, and L. A. Sokolov, in *Proc. Symposium on Nuclear Reactor Surveillance and Diagnostics (SMORN VII)*, Avignon, France (1995).
[25] E. Ott, C. Grebogi, and J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).
[26] K. Pyragas, Phys. Lett. A **170**, 421 (1992).
[27] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[28] T. Shinbrot, Adv. Phys. **44**, 71 (1995).
[29] Y.-C. Lai, M. Ding, and C. Grebogi, Phys. Rev. E **47**, 86 (1993).
[30] C. Grebogi, E. Ott, and J. A. Yorke, Phys. Rev. Lett. **57**, 1294 (1986).
[31] C. F. F. Karney, Physica D **8**, 360 (1983).

Translated by Eugene Yankovsky

ERRATA

Erratum: Hexagonal optical structures in photorefractive crystals with a feedback mirror [JETP 86, 614–627 (March 1998)]

P. M. Lushnikov

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
Zh. Éksp. Teor. Fiz. **115**, 378 (January 1999)

[S1063-7761(99)02901-7]

1. The unnumbered equation following Eq. (42) on page 623 should read

$$\delta\Psi_k = \sum_{\mathbf{k}_1+\mathbf{k}_2=\mathbf{k}} A_{k_1} A_{k_2} \psi_{k_1} \psi_{k_2}.$$

2. At the beginning of the second paragraph on page 623 ''Substituting Eqs. (42) and (43) into ....'' should be replaced by ''Substituting Eq. (42) into ....''

3. Equation (43) on page 623 should be replaced by

$$\frac{\partial A_k}{\partial t} = \nu_k A_k + \frac{U}{2} \sum_{\mathbf{k}_1+\mathbf{k}_2=\mathbf{k}} A_{k_1} A_{k_2} - \sum_{\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=\mathbf{k}} \{[-\langle \psi_{\mathbf{k}}^{c(0)} | \eta(\psi_{k_1}^{(0)}, \psi_{k_2}^{(0)} \psi_{k_3}^{(0)}) + \eta(\psi_{k_2}^{(0)} \psi_{k_3}^{(0)}, \psi_{k_1}^{(0)}) \rangle$$
$$- \langle \psi_k^{c(0)} | \Theta(\psi_{k_1}^{(0)}, \psi_{k_2}^{(0)}, \psi_{k_3}^{(0)}) \rangle] / \langle \psi_k^{c(0)} | \mathbf{J} \psi_k^{(0)} \rangle \} A_{k_1} A_{k_2} A_{k_3}, \tag{43}$$

Translated by M. E. Alferieff

Erratum: Symmetries and causes of the coincidence of the emission spectra of mirrors and charges in 1+1 and 3+1 spaces [JETP 87, 25–34 (July 1998)]

V. I. Ritus

P. N. Lebedev Physical Institute, Russian Academy of Sciences, 117924 Moscow, Russia
Zh. Éksp. Teor. Fiz. **115**, 378 (January 1999)

[S1063-7761(99)03001-2]

The abstract should read as follows:

This paper discusses the symmetry of the wave field that lies to the right and left of a two-sided accelerated mirror in 1+1 space and satisfies a single condition on it. The symmetry is accumulated in the Bogolyubov matrix coefficients α and β that connect the two complete sets of solutions of the wave equations. The amplitudes of the quantum processes in the right and left half-spaces are expressed in terms of α and β and are related to each other by transformation (12). Coefficient $\beta^*_{\omega'\omega}$ plays the role of the source amplitude of a pair of particles that are directed to opposite sides with frequencies ω and ω' but that are in either the left or the right half-space as a consequence of the reflection of one of them. Such an interpretation makes $\beta^*_{\omega'\omega}$ observable and explains the fact, given by Eq. (1) and found earlier by Nikishov and Ritus [JETP **81**, 615 (1995)] and by Ritus [JETP **83**, 282 (1996)], that the emission spectra of a mirror in 1+1 space coincide with those of charges in 3+1 space by the fact that the moment of the pair emitted by the mirror coincide with the moment of the single particle emitted by the charge.

The quantity 1 on the left-hand side of the first member of Eq. (71) should be replaced by 0, and the phrase in front of this formula should be replaced by the following phrase: ''In these cases the rms fluctuation of the number of bosonic pairs is always greater than $\overline{n^B}$ and the rms fluctuation of the number of fermionic pairs is always less than $\overline{n^F}$, being equal to $\overline{n}(1\pm\overline{n})$, where,'' etc.

Erratum: Nonlinear waves in zinc [JETP 87, 396–400 (August 1998)]

V. G. Skobov and A. S. Chernov

Moscow State Institute of Engineering Physics, 115409 Moscow, Russia
Zh. Éksp. Teor. Fiz. **115**, 379 (January 1999)

[S1063-7761(99)03101-7]

In the article, the following sentence was omitted: ''This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 98-2-18393).''

Translated by M. E. Alferieff

Long-wavelength structure on a charged liquid surface

G. V. Kolmakov and E. V. Lebedeva*)

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 5 May 1998)
Zh. Éksp. Teor. Fiz. **115**, 43–49 (January 1999)

The problem of the equilibrium form of a charged surface of a dielectric liquid in a strong electric field, such that a flat surface becomes unstable, is studied. A periodic long-wavelength structure with a small amplitude can arise when the gap between the surface and a charged electrode is small compared with the capillary length and the charge completely screens the electric field. The equilibrium form of the surface is calculated assuming that the resulting wave is one-dimensional. The effect of the boundary conditions at the vessel walls on the dependence of the amplitude of the standing wave on the applied voltage is estimated. It is shown that this dependence is very sensitive to the conditions of contact between the vessel walls and the liquid. The possibility is discussed of using the theory developed in this paper to explain the experimental results obtained with a charged liquid-hydrogen surface. © *1999 American Institute of Physics.* [S1063-7761(99)00401-1]

1. INTRODUCTION

In recent experiments on the reconstruction of a charged liquid-hydrogen surface,[1,2] a stationary structure was observed to appear on the surface of the liquid. This was a soliton-like hump with amplitude much less than the capillary length $a=\sqrt{\sigma/\rho g}$, where σ and ρ are, respectively, the surface tension and density of the liquid. Such a phenomenon cannot be explained on the basis of the existing theory,[3–6] which was developed for electrons localized above the surface of bulk helium and predicted a ''hard'' transition into a reconstructed state with a structural amplitude of order a.

In the present paper the equilibrium form of the surface is calculated for the experimental conditions of Ref. 1, i.e., when the distance d between the controlling electrode and a charged liquid-hydrogen surface is much smaller than the characteristic horizontal size of the deformation wave appearing on the surface and the surface charge completely screens the external electric field (the charged clusters form a quasi-two-dimensional layer beneath the surface of the liquid). The calculation shows that a stationary periodic wave with amplitude much less and period much greater than the capillary length can form in fields exceeding the critical value $V_0=\sqrt{4\pi\rho g d^3}$. The hump observed in the experiments of Ref. 1 corresponds to one period of a wave with a maximum inside the experimental cell. Numerical calculations of the wave amplitude as a function of the voltage applied to the diode faces were performed, taking account of the contact conditions between the cell walls and the liquid.

The existence of a small parameter, the slope angle of the surface, makes it possible to simplify the calculation substantially and to write down a local equation for the form of the surface, identical to the equation for a periodic nonlinear traveling wave on the surface of a liquid in a channel.[7] The approach based on an expansion in terms of a small parameter was used in Ref. 8 to calculate the stationary form of a charged liquid-helium surface.

2. EQUILIBRIUM FORM OF THE SURFACE

The equilibrium form of the surface corresponds to the minimum of the effective energy $\mathscr{E}$ of the liquid in an electric force field with the potential V prescribed on the surface and in the volume of the liquid,[9]

$$\mathscr{E}=\mathscr{E}_{\rm mech}-\frac{1}{8\pi}\int d\mathbf{r}\int_{\zeta(\mathbf{r})}^{d} dz\,E^2, \tag{1}$$

where $\mathscr{E}_{\rm mech}$ is the total mechanical energy of the liquid, equal to the sum of the energy in the gravitational field and the surface energy. The second term in Eq. (1) is the energy stored in the electric field (the integration extends over the volume between the surface and the controlling electrode, located at a distance d above the liquid). The z axis is oriented vertically in a direction opposite to the force of gravity, $\mathbf{r}$ is a vector in the (x,y) plane, and $z=\zeta(\mathbf{r})$ is the equation of the surface.

As the subsequent calculation will show, there exists an interval of voltages V where the spatial derivative satisfies $|\nabla\zeta(\mathbf{r})|\ll 1$. When this inequality holds, perturbation theory can be used, treating $\nabla\zeta(\mathbf{r})$ as a small parameter, to calculate the electrostatic energy (1). To simplify the calculations we shall also assume that, together with the derivative, the deformation of the surface itself is small, $\zeta\ll d$ (the latter assumption is not fundamental).

In the case at hand ($d\ll a$) the expression (1) for the energy of the liquid has the form, up to terms cubic in ζ/d,

$$\mathscr{E}=\frac{\sigma}{2}\int d\mathbf{r}\,(\nabla\zeta)^2-\frac{\rho g\gamma}{2}\int d\mathbf{r}\,\zeta^2-\frac{V^2}{8\pi d^4}\int d\mathbf{r}\,\zeta^3. \tag{2}$$

Here $\gamma=V^2/V_0^2-1, V_0=\sqrt{4\pi\rho g d^3}$. We note that the quantity $\nabla\zeta$ arises in the expression (2) only from the surface energy of the liquid [i.e., from the first term in Eq. (1)].

For voltages $V>V_0$ (i.e., for $\gamma>0$) the squared frequency of surface waves with small wave numbers $\omega^2<0$ and a flat liquid surface is unstable.[10] For $\gamma>0$ the equation describing the equilibrium form $\zeta(\mathbf{r})$ of an incompressible liquid, taking account of the fact that the total volume of the liquid remains constant when the shape of the volume changes, can be obtained by setting to zero the variation of the sum of the energy $\mathscr{E}$ of the liquid and the volume with an undetermined Lagrange multiplier p:[11]

$$\delta\left(\mathscr{E}-p\int dx\,\zeta\right)=0. \tag{3}$$

The quantity p is a pressure acting on the surface of the liquid. It follows from Eqs. (2) and (3) that the equilibrium form of the liquid surface is described by the equation

$$\zeta''+\frac{D-1}{x}\zeta'+\gamma\zeta+\frac{3V^2}{2}\zeta^2+p=0. \tag{4}$$

Here D is the dimension of the space, and Eq. (4) was written in dimensionless variables: distances in the horizontal plane xy are measured in units of the capillary length, voltage is measured in units of V_0, and the deformation $\zeta(\mathbf{r})$ is measured in units of d.

In what follows we shall consider the one-dimensional case ($D=1$), where the deformation of the surface depends only on the coordinate x. This corresponds to a system of ''rolls'' on the surface of the liquid. Then Eq. (4) has a first integral of the form

$$\zeta'^2/2+U(\zeta)=C, \tag{5}$$

where $U(\zeta)=\gamma\zeta^2/2+V^2\zeta^3/2+p\zeta$. The form $\zeta(x)$ of the surface can be obtained by integrating Eq. (5) and then calculating p from the condition

$$\int dx\,\zeta=0. \tag{6}$$

(The level of the liquid is measured from its average value.)

The equation describing the equilibrium form of the surface (4) possesses periodic solutions with period X_0 depending on the voltage V and the amplitude of the wave. The following mechanical analogy is helpful in describing the properties of such solutions. For fixed p Eq. (4) can be interpreted as the equation of motion of a fictitious point particle with mass $m=1$ in an external field with potential energy $U(\zeta)$. Here ζ plays the role of the coordinate of the particle and x plays the role of time. Then the functional (2) is the Lagrangian of the particle. The equation (5) expresses the law of conservation of energy; the first term on the left-hand side is the ''kinetic energy'' of the particle and C is the ''total energy'' of the particle. A plot of the potential energy for $\gamma>0$ is displayed in Fig. 1.

The equation $U(\zeta)=C$ has three solutions: $\zeta=\alpha_i$, $i=1,2,3$ (we assume $\alpha_1>\alpha_2>\alpha_3$). The periodic deformation wave of interest to us corresponds to the motion of a particle in the region $\alpha_2\leqslant\zeta\leqslant\alpha_1$, where $\zeta'^2=C-U(\zeta)>0$.

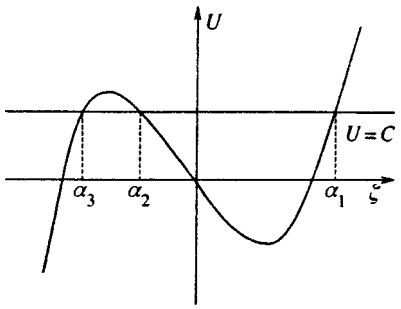


FIG. 1. Plot of the potential energy for the motion of a fictitious particle.

Let the function ζ assume its minimum value α_2 at a certain point $x=x_0$. This corresponds to zero initial velocity ζ' of the fictitious particle. Then for $0\leqslant x-x_0\leqslant X_0/2$ the form of the surface can be found from the equation

$$\int_{\alpha_2}^{\zeta(x)}\frac{dz}{\sqrt{G(z)}}=x-x_0, \tag{7}$$

where $G(\zeta)=V^2(\alpha_1-\zeta)(\zeta-\alpha_2)(\zeta-\alpha_3)$. The integral in Eq. (7) can be expressed in terms of an incomplete elliptic integral of the first kind. A calculation using Eq. (7) gives the following expression for the equilibrium form of the surface:

$$\zeta(x)=\alpha_3+\frac{\alpha_{32}}{\mathrm{dn}^2 u}, \tag{8}$$

where $u=\sqrt{\alpha_{31}}V(x-x_0)/2$, dn is an elliptic function with modulus $k=\sqrt{\alpha_{21}/\alpha_{31}}$, and $\alpha_{ik}=\alpha_k-\alpha_i$.

For real u the period of the function dn u is $2K(k)$, where $K(k)$ is a complete elliptic integral of the first kind. Therefore the period of the stationary wave (8) is

$$X_0=\frac{4K(k)}{\sqrt{\alpha_{31}}V}. \tag{9}$$

The deformation of the surface assumes the value

$$\zeta=\begin{cases}\alpha_1>0, & x-x_0=(n+1/2)X_0,\\ \alpha_2<0, & x-x_0=nX_0,\end{cases}$$

where $n=0,\pm1,\ldots$. The presence of the parameter x_0 in Eq. (8) expresses the translational invariance of Eq. (4).

Calculation of the integral (6) with the function $\zeta(x)$ from Eq. (8) gives the following equation that defines p implicitly as a function of the energy C:

$$E(k)+\frac{\alpha_3}{\alpha_{31}}K(k)=0. \tag{10}$$

We first consider a small-amplitude wave ($C\to0$). The modulus k is close to zero, and the roots α_i can be found using the well-known asymptotic expressions for the integrals K and E.[12] It follows from Eqs. (8) and (10) that we have $\alpha_2=-\alpha_1$ and

$$\zeta=\zeta_0\cos qx. \tag{11}$$

The wave number $q=\sqrt{\gamma}$ does not depend on C, $\zeta_0=\zeta_0(C)$ is the wave amplitude, and $p=0$. This solution

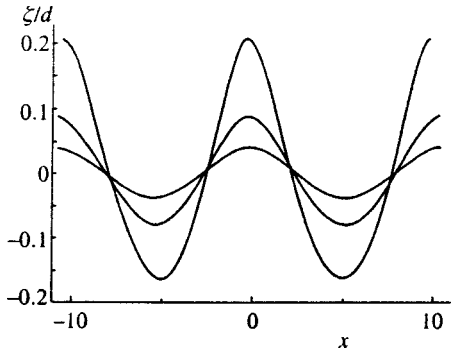


FIG. 2. Form of the liquid surface for fixed voltage and various wave amplitudes ($\gamma=0.3$).

corresponds to oscillations of a fictitious particle near the minimum $\zeta=0$ of the potential $U(\zeta)$, neglecting nonlinear terms (i.e., $V^2\zeta_0 \ll \gamma$).

For finite $C>0$ the roots α_i required to calculate the form of the surface using Eq. (8) can be found numerically. The quantity p determined from Eq. (10) is negative. A plot of the function (8) for various wave amplitudes is presented in Fig. 2.

The numerical calculations of the amplitude ζ_0 of a one-dimensional wave as a function of the period X_0 are displayed in Fig. 3. It follows from the figure that for fixed V the period of the wave decreases with increasing amplitude.

For arbitrary values of the roots α_i, Eq. (8) possesses a solution in the form of a solitary hump, which can be obtained by passing to the limit $X_0 \to \infty$. However, the formation of such a wave is forbidden by the condition that the total volume of the liquid is constant, as expressed by Eq. (10). Indeed, it follows from Eq. (9) that such a limit solution arises when the roots α_2 and α_3 of the polynomial $G(\zeta)$ merge. Then $k \to 1$. In addition, Eq. (10) has no solutions for k close to 1; this can be shown by expanding the elliptic integrals in terms of the small quantity $k'^2=1-k^2$.

It follows also from the absence of solutions with $\alpha_2=\alpha_3$ that for $C>0$ the equation $U(\zeta)-C=0$ always possesses three roots. Therefore the solution (8) of Eq. (4) is a periodic wave for any energy $C>0$ (i.e., for arbitrary amplitude of the wave).

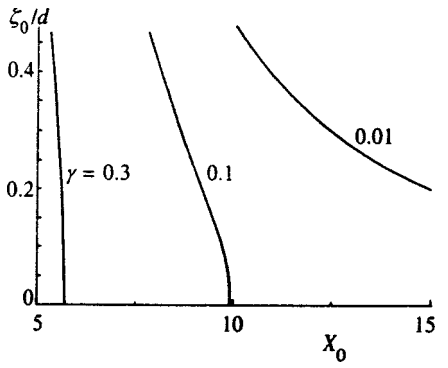


FIG. 3. The amplitude of a wave on an infinite liquid surface versus the period of the wave for various voltages.

This calculation shows that for a one-dimensional wave arising on a horizontally unbounded surface and fixed voltage $V>V_0$ the shape of the surface can be described by a one-parameter family of curves of the form (8). The wave amplitude [or the constant $C>0$ related to it by (10)] is a parameter. Therefore, in contrast to the well-known problem of the equilibrium form of the surface of bulk liquid helium with a fixed number of charges localized at the surface, in the case at hand ($d \ll a$) it is impossible to determine unequivocally the equilibrium form of the surface without taking account of the boundary conditions at the vessel walls.

Let us now consider the effect of the boundary conditions on the possible stationary form of the surface. The simplest boundary conditions correspond to setting the contact angle β between the liquid and the vertical walls of the vessel (whose width we denote by L):

$$\zeta'(-L/2)=-\theta_0, \quad \zeta'(L/2)=\theta_0, \tag{12}$$

where $\theta=\theta(x_0)=\cot\beta$. We assume that $\theta_0 \ll 1$; otherwise, the condition that ζ' is small, which is fundamental for the derivation of Eq. (4), is not satisfied.

The conditions (12) limit the possible values of C to a discrete sequence C_n ($n=0,1,2,\ \ldots$). The number n is the number of maxima of the function $\zeta(x)$ in the interval $-L/2<x<L/2$.

We first consider the case $\theta_0=0$. The dependence of the wave amplitude on the applied voltage can be determined from the function $\zeta_0(X_0)$ calculated for an infinite surface. The amplitude of a wave with a fixed number n of maxima is the ordinate of the point of intersection of the curve $\zeta_0(X_0)$ and the vertical straight line $X_0=L/n$. It follows from Fig. 3 that ζ_0 decreases monotonically with increasing voltage V.

The voltage dependence of ζ_0 for $\theta_0 \neq 0$ can be calculated similarly. This requires studying the points of intersection of the vertical straight line $x_1=L/n$ and the plot of $\zeta_0(x_1)$, describing the amplitude as a function of the x coordinate of the point (designated as x_1) where the condition (12) holds. A numerical calculation for the case $\theta_0>0$ gives a double-valued function $\zeta_0(x_0)$ for fixed γ. The result is that the function $\zeta_0(\gamma)$ possesses two branches, one of which decreases monotonically and the other possesses a minimum. At some $\gamma=\gamma_1$ one branch passes into the other branch. If $\gamma>\gamma_1$ holds, a stationary solution with fixed n does not exist, since it is impossible to satisfy the boundary conditions (12). Figure 4 shows the amplitude ζ_0 of a wave with one hump versus γ for two values of the vessel width L.

In summary, it is impossible to pass continuously between states with a different number n of maxima, and a hump forms on the surface with a jump in the wave amplitude. For each n there exists a minimum possible value ζ_m of the amplitude of the wave formed. For small θ_0, we have $\zeta_m \sim \theta_0 d$ in order of magnitude.

3. CONCLUSIONS

It follows from the calculations performed in this work that for $d \ll a$ a long-wavelength stationary wave of the form (8) with a small amplitude (''soft'' reconstruction of the surface) can form on a charged liquid surface at voltages above

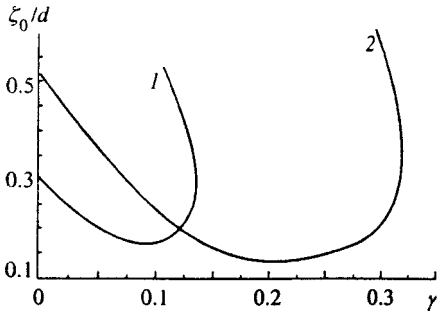


FIG. 4. Voltage dependence of the amplitude of a wave on a liquid surface in a vessel with two unequal dimensions: *1* — $x_0=14$; *2* — $x_0=10$. Here $\theta_0^2=0.001$.

a critical value V_0 (for which a flat surface becomes unstable). For an infinite surface, the form and period of the wave with fixed voltage V depend on the positive parameter C and cannot be uniquely determined without taking account of the boundary conditions at the vessel walls. Taking account of the contact conditions at the boundary of a finite-size vessel distinguishes a discrete spectrum of possible values C_n, where n is the number of maxima in the profile of a wave inside the finite vessel. The character of the dependence of the wave amplitude on the applied voltage is different for zero and finite contact angles.

In the first case the wave amplitude decreases with increasing voltage. For $\theta_0>0$ the amplitude of the structure arising on the surface can depend nonmonotonically on the voltage. A transition of the initial state (prior to reconstruction) into a state with one or several maxima occurs with a jump $\sim\theta_0 d$ in the amplitude, i.e., the jump is small within the range of applicability of the theory (for $\theta_0\ll 1$). At the voltage $V_2=V_0\sqrt{1+\gamma_1}$, which depends on the dimensions of the vessel, the reconstructed surface becomes unstable.

This calculation indicates that the amplitude dependence of the wave formed in a finite vessel is extremely sensitive to the form of the boundary conditions at the vessel walls.

We are deeply grateful to L. P. Mezhov-Deglin, A. A. Levchenko, and A. B. Trusov for providing the experimental data and for many discussions of the computational results. This work was supported in part by grant PKA-NASA TM-17 and a grant from the Russian Fund for Fundamental Research (No. 980216640).

*)E-mail: lebedeva@issp.ac.ru

[1] A. A. Levchenko, E. Teske, G. V. Kolmakov, P. Leĭderer, L. P. Mezhov-Deglin, and V. B. Shikin, JETP Lett. **65**, 572 (1997).
[2] G. V. Kolmakov, A. A. Levchenko, L. P. Mezhiv-Deglin, and A. B. Trusov, Fiz. Nizk. Temp. **24**, 158 (1998) [Low Temp. Phys. **24**, 114 (1998)].
[3] V. B. Shikin and Yu. P. Monarkha, *Two-Dimensional Charge Systems in Helium* (Nauka, Moscow, 1989).
[4] V. B. Shikin and P. Leiderer, JETP Lett. **32**, 416 (1980).
[5] D. M. Chernikova, Fiz. Nizk. Temp. **6**, 1513 (1980) [Sov. J. Low Temp. Phys. **6**, 737 (1980)].
[6] V. I. Mel'nikov and S. V. Meshkov, Zh. Éksp. Teor. Fiz. **81**, 951 (1981) [Sov. Phys. JETP **54**, 505 (1981)].
[7] S. P. Novikov, S. V. Manakov, L. P. Pitaevskiĭ, and V. E. Zakharov, *Theory of Solitons Theory of Solitons* (Consultants Bureau, New York, 1984) [Russian orig., Nauka, Moscow, 1980)].
[8] V. Shikin and P. Leĭderer, Fiz. Nizk. Temp. **23**, 624 (1997) [Low Temp. Phys. **23**, 468 (1996)].
[9] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, New York) [Russian original, Nauka, Moscow, 1982].
[10] D. M. Chernikova, Zh. Éksp. Teor. Fiz. **68**, 249 (1975) [Sov. Phys. JETP **41**, 121 (1975)].
[11] Ya. I. Frenkel', Zh. Éksp. Teor. Fiz. **18**, 659 (1948).
[12] I. S. Gradshteĭn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, transl. of 4th Russ. ed. (Academic, New York, 1980) [Russ. original, 5th ed., Nauka, Moscow, 1971].

Translated by M. E. Alferieff

Computer simulation of the structure of liquid cesium and determination of the pair potentials over a wide temperature range

D. K. Belashchenko*) and A. S. Ginzburg†)

Moscow Institute of Steels and Alloys (Technological University), 117936 Moscow, Russia
(Submitted 5 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 50–60 (January 1999)

Using diffraction data for liquid cesium structure over a wide temperature range, models of liquid cesium are constructed and the effective pair potentials are extracted using the theory of liquids. The iterative procedure proposed by L. Reatto is used. In the range 323–1923 K the pair potentials are weakly temperature-dependent. The potentials extracted from the diffraction data differ from the potentials calculated using the Animalu–Heine pseudopotential. The self-diffusion coefficients in liquid cesium are determined. Their temperature dependence is described satisfactorily by a power-law function. © *1999 American Institute of Physics.* [S1063-7761(99)00501-6]

1. INTRODUCTION

A new and interesting direction in the physics of noncrystalline condensed systems is unedr active study—the construction of atomic models of one- and two-component liquids or amorphous substances on the basis of existing diffraction structural data (specifically, on the basis of known structure factors or pair correlation functions). This problem can be formulated as a purely geometric one. In this case, points representing the atoms of the components must be arranged in space so that the pair correlation function of a one-component system or three independent partial pair correlation functions of a two-component system for pairs 11, 12, and 22 are the same (to within reasonable accuracy) as the corresponding functions obtained for a real substance by diffraction methods (x-ray or neutron scattering). If the model constructed in this manner is adequate, then investigation of the model will yield additional information about the structure of the system of interest. Doubtless this method of analyzing noncrystalline structures has a great future.

The problem of well-posedness is directly related to the question of the uniqueness of the construction of the model. Here uniqueness is taken to mean that the main structural characteristics are the same as the corresponding characteristics of a real body, provided that the pair correlation functions are the same. Existing data show that in principle the problem can be well-posed (to within the limits of accuracy of the initial data) for ''dense'' structures, such as simple liquids and liquid and amorphous metals. However, for loose systems with low coordination numbers (such as liquid silicon, silica, and so on) agreement between the pair correlation functions in the model and in the real body does not guarantee that the three-particle correlation functions, distributions of the azimuthal angles and Voronoĭ polyhedra, and so on will agree.

However, the problem at hand is not purely geometric, since the equilibrium arrangement of the atoms at fixed temperature and density is due to the form of the interparticle interaction potentials. Additional considerations of a physical character can therefore be used to construct the models. The most common one is the assumption that the interaction responsible for the structure is pairwise, i.e., in constructing models the three-particle and cooperative contributions to the energy can be neglected and effective pair interatomic interaction potentials with a fixed volume of the system can be introduced. This approach is used, for example, for liquid and amorphous metals.

Including interaction potentials in the model makes it possible to formulate the inverse problem also, specifically, the problem of finding an interparticle potential such that the model structure constructed for a liquid or amorphous material is quite close to the structure of the real body. This problem is solved in the present paper for liquid cesium. In contrast to previous work, where individual states of a liquid are studied (see below), we have constructed a series of models of liquid cesium on the basis of diffraction data obtained in a very wide range of temperatures, right up to 1923 K. The existence of a definite temperature dependence of the extracted pair potentials, which is due to an appreciable decrease in the density with increasing temperature, could serve as an additional adequacy criterion for the models constructed.

2. METHODS FOR CONSTRUCTING MODELS USING DIFFRACTION DATA

Several methods have been proposed for calculating interparticle potentials from existing diffraction structural data for a liquid assuming a pair interaction (which may not be valid in a real liquid).[1–10] This problem was first formulated in Ref. 1, where it was solved on the basis of the approximate Born–Green–Kirkwood and the Percus–Yevick equations. Schommers proposed a different approach.[2,3] He found a pair potential by an iterative procedure employing a step in which a molecular-dynamics model of the liquid is constructed using a trial potential and corrections are then introduced into this potential. The correction was calculated using a special algorithm that takes account of the discrepancy be-

tween the pair correlation function of the model and the prescribed pair correlation function of the real (''target'') liquid. Schommers used his method to construct a model of liquid Ga and to find the pair interparticle potential at the same time. The procedure required a series of iterations in which the potential and the pair correlation function $g(r)$ converged asymptotically to the desired solution.

This procedure proved to be inadequate. While the model function converges well to the target function, the reconstructed potential $u(r)$ can approach a solution that is different from the target solution. This was pointed out by Reatto.[4,5] To calculate the corrections to the trial potential, he proposed using the complete equation, containing a so-called bridge function $B(r)$, from the theory of liquids:

$$\frac{u(r)}{kT}=g(r)-1-c(r)-\ln g(r)+B(r). \tag{1}$$

Here $u(r)$ is the pair potential, $g(r)$ is the pair correlation function of the liquid, and $c(r)$ is the direct correlation function. The form of the function $B(r)$ for an arbitrary potential is unknown. The functions $g(r)$ and $c(r)$ are related by the Ornstein–Zernicke equation

$$h(r)=g(r)-1=c(r)+\frac{2\pi N}{Vr}\int_0^\infty c(s)s\,ds\int_{|r-s|}^{r+s}h(t)t\,dt, \tag{2}$$

where $n=N/V$ is the number of particles per unit volume. If it is assumed that the function $B(r)$ is insensitive to the form of the potential $u(r)$, then two successive approximations of the potentials u_1 and u_2 satisfy

$$\frac{u_2(r)}{kT}=\frac{u_1(r)}{kT}+g(r)-g_0(r)-\ln\frac{g(r)}{g_0(r)}+c(r)-c_0(r). \tag{3}$$

Here $g(r)$ and $c(r)$ are the correlation functions of the liquid with the potential $u_1(r)$; $u_2(r)$ is the next approximation for the potential; and $g_0(r)$ and $c_0(r)$ are the correlation functions of the target model. In Refs. 4 and 5 this procedure gave convergence of the potential to the Lennard–Jones potential

$$u_0(r)=4\varepsilon[(\sigma/r)^{12}-(\sigma/r)^6], \tag{4}$$

used to construct the target model.

A different iteration scheme was proposed in Ref. 6. This scheme used a ''hybrid algorithm'' to reconstruct the interparticle potential. The algorithm can operate only at absolute zero temperature, and it is therefore suitable for amorphous systems. The hybrid algorithm employed the condition of mechanical equilibrium for all particles and was found to be suitable for amorphous iron.

Finally, an iteration algorithm that employs the complete Born–Green–Bogolyubov equation, without the superpositional Kirkwood approximation, to extract the potential is proposed in Ref. 7. In this method the three-particle correlation function is calculated for a trial model of the liquid. Given this function and the pair correlation function of the model, a trial pair potential can be calculated and this potential can then be used for molecular-dynamics construction of the next model, and so on.

The methods of Refs. 2, 6, and 7 have also been extended to binary systems. In Ref. 8 Schommer's method was used to calculate the pair potentials in the liquid alloy Ag–Ge. In Refs. 9 and 10 the Born–Green–Bogolyubov equation method was used for the liquid alloys Ag–Ge and Fe_2Tb.

In Ref. 11 Reatto's iterative method was used for a liquid with the pair potential (4) under conditions different from those of Refs. 4 and 5. The pair correlation function for this state was calculated earlier by Verlet.[12] The standard deviation

$$R_f=\left\{\frac{1}{n_2-n_1+1}\sum_{n_1}^{n_2}[f_2(r_i)-f_1(r_i)]^2\right\}^{1/2} \tag{5}$$

can be used as a measure of the closeness of two functions $f_1(r)$ and $f_2(r)$ in tabular form. Here n_1 and n_2 are the

TABLE I. Iteration procedure for constructing models of liquid cesium at 573 K.

N	R_g	R_c	R_a	$R_u\cdot 1000$	r_m, Å	U_m, eV	$r_{\min}$, Å	r_1, Å	$G(r_1)$	C_0	H_0
0	0	0.8781	0.5520	49.4	5.7	−0.074	3.5	5.5	1.978	0	0
2	0.0596	0.1279	0.0619	0.8765	5.7	−0.073	3.3	5.5	2.161	−19.14	−0.8804
4	0.0458	0.0936	0.0501	0.5818	5.7	−0.072	3.3	5.5	2.118	−18.16	−0.8806
5	0.0378	0.0805	0.0394	0.4787	5.8	−0.072	3.3	5.5	2.082	−17.74	−0.8718
6	0.0354	0.0796	0.0376	0.4679	5.8	−0.071	3.3	5.5	2.090	−17.60	−0.8743
7	0.0303	0.0754	0.0320	0.3983	5.8	−0.071	3.3	5.5	2.055	−17.35	−0.8600
8	0.0258	0.0705	0.0260	0.3562	5.8	−0.071	3.3	5.5	2.059	−17.09	−0.8730
9	0.0230	0.0680	0.0231	0.3707	5.8	−0.071	3.3	5.5	2.035	−16.85	−0.8807
10	0.0187	0.0507	0.0196	0.2544	5.8	−0.071	3.3	5.5	2.038	−16.72	−0.8671
11	0.0151	0.0645	0.0143	0.2903	5.8	−0.071	3.3	5.5	2.027	−16.55	−0.8682
12	0.0138	0.0435	0.0119	0.2155	5.8	−0.071	3.3	5.5	2.010	−16.50	−0.8842
20	0.0081	0.0372	0.0069	0.1776	5.8	−0.072	3.3	5.5	2.002	−16.15	−0.8585

Note: N — number of iterations; R_g, R_c, and R_a — standard deviations for the pair correlation function, $C(r)$, and the structure factor, respectively; R_u — standard deviation of the pair potentials at the preceding and given iterations; r_m — coordinate of the potential minimum; U_m — value of the pair potential at the minimum, $r_{\min}$ — coordinate of the point where the pair correlation function is zero; r_1 — coordinate of the first peak of this function; $G(r_1)$ — height of this peak; C_0 — value of the direct correlation function at $r=0$; H_0 — value of the function $h(K)=a(K)-1$ in the limit $K\to 0$.

TABLE II. Values of the bridge function at iterations 21–23 in the simulation of cesium at 1923 K.

Distance, Å	3.0	3.2	3.4	3.6	3.8	4.0	4.2	4.4	4.6
$i=21$	0.528	0.242	0.148	0.092	0.063	0.063	0.052	0.034	0.028
$i=22$	0.458	0.239	0.124	0.088	0.066	0.058	0.047	0.049	0.048
$i=23$	0.487	0.207	0.109	0.076	0.025	0.034	0.039	0.016	0.012

summation limits of the tabular data. The deviation R_f of the pair correlation function of the model $M1$ from the pair correlation function of the Verlet model is 0.0074 (on the section $0.88 \leqslant r \leqslant 2.40$). Such a value indicates that the functions are essentially identical. The values of the potential energy U and the factor pV/NkT (with a correction for cutoff of the potential; p is the pressure) also were in good agreement. However, the target potential (4) in the region of attraction could not be extracted using Eq. (3) even with a large number of iterations (greater than 20), although the repulsive branch was extracted satisfactorily. The standard deviation R_f can be decreased to extremely low values (0.005–0.008 for the pair correlation function and 0.001–0.006 for the structure factor) by increasing the molecular-dynamics run lengths up to 10000–15000 steps, but the agreement between the extracted and target potentials does not improve in the process. Moreover, asymptotic convergence to a definite function $u(r)$ was not observed with the Reatto procedure as implemented in Ref. 11, since even in states with the lowest values of R_f the potential continued to vary from one iteration to the next. This is evidently due to the natural fluctuations of the pair correlation function and the exceedingly high response of the direct correlation function to this near the first peaks in the pair correlation function.

It is nonetheless evident from Refs. 4, 5, and 11 that for liquids Reatto's procedure makes it possible to construct an almost ideal model in the sense that its pair correlation function and the structure factor of the model agree with their target analogs. In this connection, in the present work this procedure was used to construct models and to extract interparticle potentials for liquid cesium, for which diffraction structural data at temperatures 323–1923 K are available.[13]

3. ANALYSIS OF THE DIFFRACTION DATA

Professor F. Hensel (Marburg, Germany) kindly provided us with the initial structure factors of liquid cesium.[13] The pair correlation function of liquid cesium at various densities and pressures were obtained using Filon's implementation of the Fourier transform. In the process spurious oscillations of the pair correlation function are ordinarily observed at short distances. A procedure proposed in Ref. 14 for correcting the initial structure factor was used to improve the quality of the Fourier transform. The least-squares method was used to find minimal corrections to the structure factor that at the same time minimize the amplitude of the spurious oscillations. For liquid cesium, this amplitude was decreased by approximately an order of magnitude by introducing corrections of order 0.01 to the structure factor. The pair correlation functions obtained in the process were used in Reatto's procedure.

4. CONSTRUCTION OF MODELS OF LIQUID CESIUM AND EXTRACTION OF INTERPARTICLE POTENTIALS

In all models, the main cube contained 1000 particles. The initial potential for the Reatto procedure was calculated by the method of Refs. 4 and 5. The molecular-dynamics runs ordinarily consisted of 5000 steps at the initial iterations and 10000 steps at the last stage. The NVT-ensemble method was used, and the size of the main cube was set according to the actual density of the metal. The pair correlation functions

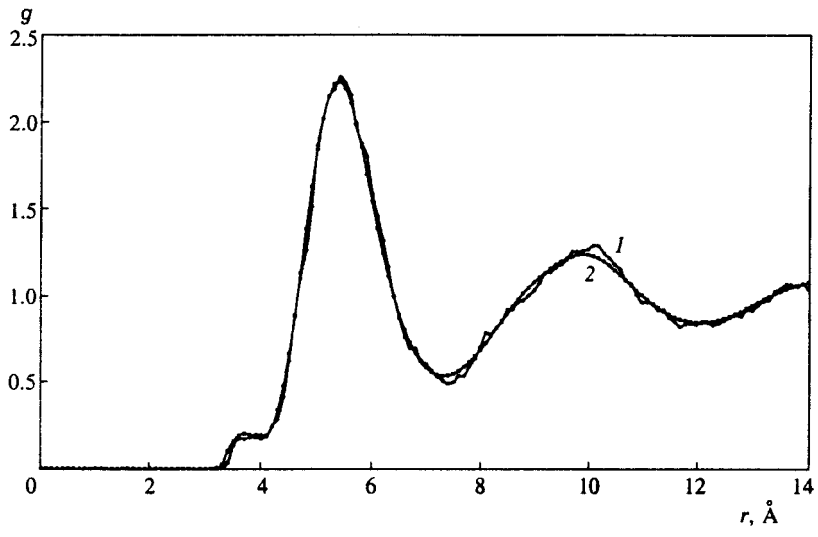


FIG. 1. Pair correlation functions in liquid cesium at 323 K: *1* — model pair correlation function, *2* — diffraction pair correlation function.

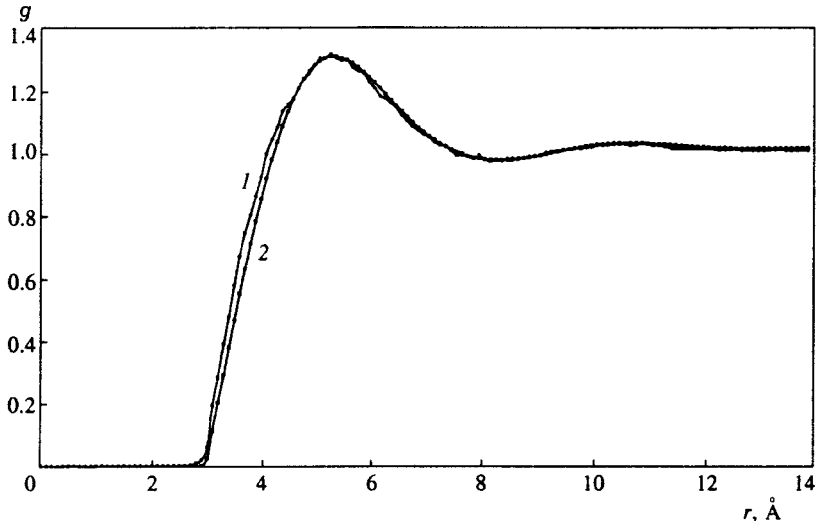


FIG. 2. Pair correlation functions in liquid cesium at 1923 K: *1* — model pair correlation function, *2* — diffraction pair correlation function.

were calculated in a molecular-dynamics run up to distances $\sim L/3$, where L is the edge length of the main cube, and then the lengths were extended up to 60 Å using the Ornstein–Zernicke equation analogously to Refs. 4, 5, and 11. To extend the pair correlation function to distances exceeding $L/3$, the potential (and therefore the direct correlation function, in accordance with the Percus–Yevick equation) was assumed to be zero.

As an example, Table I gives data for the iteration procedure used to construct models of liquid cesium at 573 K. After 20 iterations, quite small values were obtained for R_g (about 0.008) for the pair correlation function and for R_a (less than 0.007) for the structure factor and somewhat larger values were obtained for R_c (about 0.037) for the direct correlation function. At higher temperatures, in general, fewer iterations were required to achieve an acceptable value of R_g for the pair correlation function. As shown in Ref. 11, in simulations with runs of 10000 steps the natural limit of accuracy (because of fluctuations of the function itself) is $R_g = 0.003 - 0.004$, and for runs with 15000 steps $R_g \simeq 0.0025$. Hence to extract a potential from the diffraction data a logical rule for terminating the iteration process (provided it converges) would be a deviation of the model function from the target function of about 0.003–0.005. Approximately such values were in fact attained for models of cesium at temperatures above 573 K.

Just as in Ref. 11, the direct correlation function was found to be very sensitive to small differences between two pair correlation functions. The assumption that bridge function varies little while a potential is being extracted is also unjustified. As an example, the values of the bridge function $B(r)$ in 21–23 iterations in the simulation of cesium at 1923 K are given in Table II. It is evident from these data that the bridge function fluctuates from one iteration to another within several hundredths, i.e., more strongly than the pair correlation itself changes. In accordance with Eq. (1), this leads to fluctuations of the computed potential of order $0.01kT$, i.e., by several meV with the minimum of the potential being only ~ 56 meV.

Hence it follows that Eq. (3) is inaccurate and the extracted pair potential need not approach a definite limit function. Indeed, the form of the extracted pair potential changes continuously during the iteration process. An appreciable decrease in the rate of change of the depth of the minimum of the computed potential is noticeable only at 1923 K, where 24 iterations were performed. In all other cases this does not

TABLE III. Characteristics of liquid cesium models constructed from the diffraction data. Model size — 1000 particles. Number of iterations — from 8 to 20.

T, K	V/N, $\text{cm}^3\,\text{mole}^{13}$	L, Å	R_g	R_c	R_s	$a(0)$		P, MPa		E, kJ/mole	U, kJ/mole	$D\cdot 10^5$, cm^2/s
						Calc.	Exp.[13]	Calc.	Exp.[13]			
323	72.6	49.406	0.0188	0.1177	0.0498	0.087	0.015	98.3	0.3	−16.9	−20.9	2.84
573	79.1	50.835	0.0081	0.0372	0.0069	0.142	0.048	262	0.3	−11.7	−18.8	9.58
773	84.6	51.996	0.0055	0.0365	0.0075	0.102	0.08	461	0.3	16.1	6.38	18.6
1073	95.6	54.149	0.0068	0.0253	0.0078	0.194	0.18	333	0.6	12.8	−0.534	31.2
1173	99.9	54.952	0.0063	0.0157	0.0139	0.372	0.29	146	0.3	3.70	−10.9	42.7
1373	109.9	56.727	0.0064	0.0203	0.0058	0.434	0.52	173	2.0	5.98	−11.1	54.6
1673	139.0	61.345	0.0057	0.0122	0.0638	0.814	1.10	105	5.3	9.20	−11.7	96.6
1923	225.3	72.052	0.0140	0.0164	0.0209	2.28	2.54	21.3	96	15.0	−8.96	196

Note: V/N — molar volume, L — edge length of the main cube, $a(0)$ — zero limit of the structure factor, E — total energy, U — potential energy, D — self-diffusion coefficient.

happen, even though the values of R_g were already in the limit of stable molecular-dynamics determination in runs with 10000 steps. In essence, the iteration process could be continued further, even with an indefinite result for the potential. True, the changes in the depths of the minimum of the potential from one iteration to another are small. Therefore the problem of extracting a potential on the basis of the structure of the liquid cannot be solved uniquely if a certain lower level of the standard deviation is fixed for the pair correlation function (determined by the size of the models and the length of the molecular-dynamics runs). This result is at variance with Refs. 4 and 5, where an asymptotic solution for the potential was obtained with a finite number of Reatto iterations.

The pair correlation functions of our models are displayed in Figs. 1 and 2. For the values of R_g attained in this work, the difference between the ''target'' and model functions is very small. The other characteristics of the models constructed are given in Table III. For not very high temperatures, the zero limit of the structure factor $a(0)$, obtained by extending the pair correlation function using the Ornstein–Zernicke equation, turned out to be greater than the actual value[13] (see Table III). From the relation between $a(0)$ and the isothermal compressibility β and the density of the liquid (N is the number of atoms in a volume V),

$$a(0)=\frac{N}{V}\beta kT, \tag{6}$$

it can be concluded that far from the critical point the compressibility of real cesium is 1.3–5.8 times smaller than that of the model metal with a pair interparticle interaction. The discrepancy decreases with increasing temperature. This effect is evidently due to the negative contribution of the electron gas to the compressibility and has been discussed theoretically in previous work.[15] Using Eq. (6), the bulk modulus $K=1/\beta$ can be calculated. The actual moduli K_{exp} of liquid cesium and the computed moduli K_{theor} due to the pair interaction are compared in Table IV; the moduli K_{theor} were calculated taking account of the zero limits $a(0)$ given in Table III. The difference $\Delta K=K_{\text{exp}}-K_{\text{theor}}$ is due to the contribution of the electron gas. Even though the accuracy of the calculations is not very high, it is evident that ΔK decreases rapidly (approximately as V^{-12}) as the metal expands. The free-electron model at $T=0$ gives a much weaker dependence $\Delta K\sim V^{-5/3}$. At 1923 K temperature the difference between K_{theor} and K_{exp} becomes less than the error in determining K_{exp}, so that at this temperature the electron gas no longer makes an appreciable contribution to the compressibility. The electrical conductivity also decreases rapidly as the metal expands: At 1900 K and a pressure of 86 bar it is 0.016 times the conductivity at the melting temperature.[16]

TABLE IV. Comparison of the actual moduli K_{exp} of liquid cesium with the computed moduli K_{theor} determined by the pair interaction.

T, K	323	573	773	1073	1173	1923
V/N, cm^3/mole	72.6	79.1	84.6	95.6	99.9	225.3
$10^{-8}\cdot K_{\text{exp}}$, Pa	24.6	12.5	9.49	5.18	3.36	0.279
$10^{-8}\cdot K_{\text{theor}}$, Pa	4.25	4.24	7.44	4.81	2.62	0.311
$10^{-8}\cdot \Delta K$, Pa	20.4	8.26	2.05	0.37	0.74	~ 0

Various characteristics of cesium models are given in Table III: the pressure as well as the total energy E and the potential energy U due to the effective pair interaction. Since they were calculated neglecting the volume electronic contributions to the energy, their values cannot be compared directly with the actual values.

The computed pair interparticle potentials in cesium at various temperatures are shown in Figs. 3 and 4. They decrease rapidly in absolute magnitude with increasing distance; for $r>8$ Å they do not exceed a few meV, and for $r>10$ Å they are less than 1 meV. Monotonic variation of the potentials with increasing temperature is not observed. In the repulsion region the potential increases on heating up to 773 K and once again decreases as temperature increases further. This is evidently because the calculated potentials are highly sensitive to the form of the pair correlation function; comparatively small errors in the computed functions lead to changes in the potentials that are much greater than the regular changes occurring with increasing temperature.

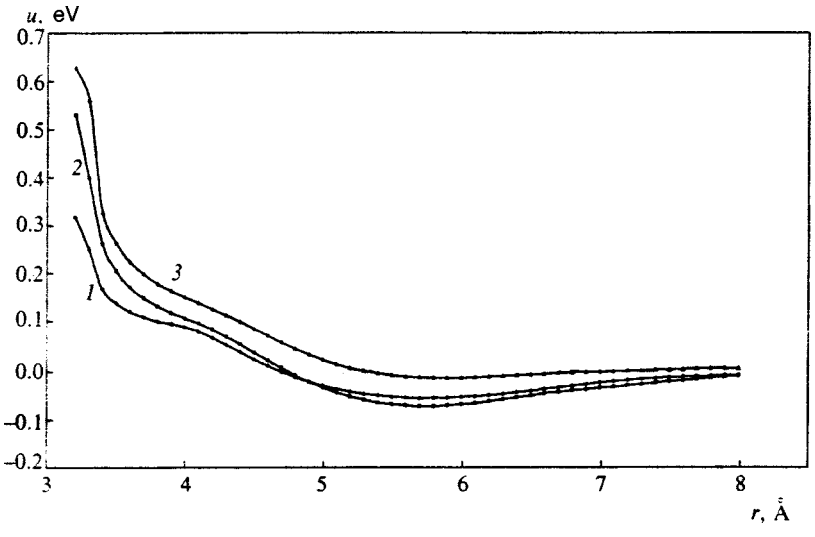


FIG. 3. Interparticle potentials in cesium at various temperatures: *1* — 323 K, *2* — 573 K, *3* — 773 K.

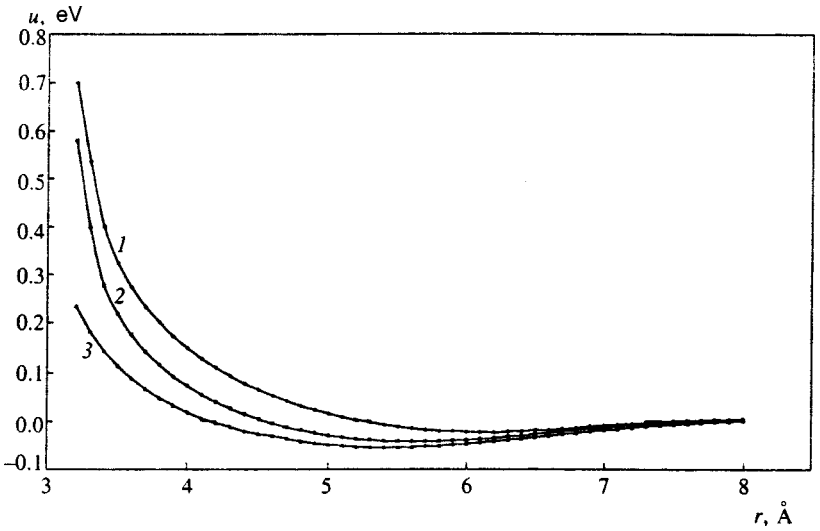


FIG. 4. Interparticle potentials in cesium at various temperatures: *1* — 1073 K, *2* — 1173 K, *3* — 1973 K.

The effective interionic potential in liquid cesium near the melting temperature was calculated in Refs. 17 and 18 using the Animalu–Heine pseudopotential and Geldart–Vosko screening. It differs from the potential found from the diffraction data. For example, the first node of our potential occurs at 4.7 Å and the first node of the pseudopotential occurs at 4.9 Å ; the second nodes are located at 9.1 and 7.4 Å , respectively. The diffraction potential increases at small distances less rapidly than the pseudopotential (for example, at 4.5 Å the diffraction potential is ~ 24 meV, while the theoretical potential is 53 meV). The depth of the minimum of the diffraction potential at 5.7 Å is -55.6 meV, while for the computed potential the minimum lies at the same location but its depth is -26 meV.

The self-diffusion coefficients found from the dependence of the mean-square displacement of cesium atoms on the molecular-dynamics relaxation time are shown in the last column of Table III. The self-diffusion coefficient increases by the factor ~ 70 in the temperature range 323–1923 K. The temperature dependence is described satisfactorily by the power-law expression

$$D = 6.69 \times 10^{-11} T^{2.228} \text{ cm}^2/\text{s}. \quad (7)$$

The Arrhenius-type exponential formula $D = 3.35 \times 10^{-3} \exp(-12636 \text{ J}/RT)$ gives a much worse description of the results. Power-law expressions for $D(T)$ have been proposed in previous work,[19,20] taking account of the fact that the diffusion mechanism in a liquid metal is not activational.

5. DISCUSSION

Judging from our data, an entire family of computed pair potentials giving good agreement with the target pair correlation function with a reasonable standard deviation can exist at fixed temperature and pressure. Accordingly, the procedure for extracting the potential using the equations from the theory of liquids can, in principle, be unstable, so that for a sufficiently large number of iterations different terms of this family will appear in succession. The successful extraction in Refs. 4 and 5 of the Lennard–Jones potential could be due to a felicitous choice of the initial conditions and the initial potential. However, there is no guarantee that a transition to a different potential will not occur if the iteration process is extended further.

This result is unpleasant from the standpoint of extracting the potential of real liquids, where the optimal initial conditions are unknown. As for the adequacy of the extraction of the structure of a liquid, for dense systems good agreement between the actual and model pair correlation functions evidently guarantees agreement between other structural characteristics, specifically, the angular correlations, distributions of the Voronoĭ polyhedra, and so on.[10,21] The obtained pair potential can be used to calculate properties of a liquid, such as the vibrational spectrum, the self-diffusion coefficient, and the viscosity. This is no longer true for topologically loose systems with low coordination numbers (such as liquid silicon), since ideal agreement between the pair correlation functions of two states with completely different three-particle correlation functions can be obtained.[22]

*)E-mail: dkbel@bel.misa.ac.ru
†)E-mail: postmaster@phch.misa.ac.ru

[1] M. D. Johnson, P. Hutchinson, and N. H. March, Proc. R. Soc. London, Ser. A **282**, 283 (1964).
[2] W. Schommers, Phys. Rev. A **28**, 3599 (1983).
[3] W. Schommers, J. Non-Cryst. Solids **61-62** (Part 1), 571 (1984).
[4] L. Reatto, D. Levesque, and J. J. Weis, Phys. Rev. A **33**, 3451 (1986).
[5] L. Reatto, Phil. Mag. **58**(1), 37 (1988).
[6] M. I. Mendelev and D. K. Belashchenko, Rasplavy, No. 4, 60 (1992).
[7] D. K. Belashchenko and M. I. Mendelev, Metally, No. 5, 80 (1993).
[8] D. K. Belashchenko and M. P. Momchev, Izv. Vyssh. Uchben. Zaved., Chernaya Metallurgiya, No. 7, 72 (1992).
[9] M. I. Mendelev and D. K. Belashchenko, N. Engl. J. Med. **31**(2), 215 (1995).
[10] M. I. Mendelev and D. K. Belashchenko, N. Engl. J. Med. **30**(3), 379 (1994).
[11] D. K. Belshenko, Metally, No. 4, 101 (1998).
[12] L. Verlet, Phys. Rev. **165**, 201 (1968).
[13] R. Winter, F. Hensel, T. Bodensteiner, and W. Glaser, Ber. Bunsenges. Phys. Chem. **91**, 1327 (1987).
[14] D. K. Belshenko, Kristallografiya **43**(3), 400 (1998) [Crystallogr. Rep. **43**, 362 (1998)].
[15] M. Hasegava and W. H. Young, J. Phys. F: Metal Phys. **8**(4), L81 (1978).

[16] F. Hensel, S. Jungst, F. Noll, and R. Winter, in *Localization and Metal-Insulator Transition*, edited by D. Adler and H. Fritzsche (Plenum Press, New York, 1985), p. 109.
[17] A. M. Bratkovskiĭ, V. G. Vaks, S. P. Kravchuk, and A. V. Trefilov, Preprint IAÉ-3392/9 (I. V. Kurchatov Institute of Atomic Energy, Moscow, 1981).
[18] A. M. Bratkovsky, V. G. Vaks, S. P. Kravchuk, and A. V. Trefilov, J. Phys. F: Metal Phys. **12**(7), 1293 (1982).
[19] R. A. Swalin, Acta Metall. **7**, 736 (1959).
[20] R. A. Swalin, Acta Metall. **9**, 379 (1961).
[21] D. K. Belashchenko and M. I. Mendelev, Metally, No. 5, 80 (1993).
[22] V. V. Alenkov, D. K. Belashchenko, and G. D. Kuznetsov, Rasplavy, No. 4, 65 (1989).

Translated by M. E. Alferieff

Discrete thinning of free-standing smectic films in the de Gennes ''pre-smectic liquid'' model

E. E. Gorodetskiĭ, E. S. Pikina, and V. É. Podnek*)

Petroleum and Gas Institute, Russian Academy of Sciences, Moscow, Russia
(Submitted 9 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 61–69 (January 1999)

It is shown that the successive discrete thinning of free-standing smectic films (FSSFs), which is observed when the films are heated above the temperature of the smectic A–nematic bulk phase transition, has a natural explanation in terms of the de Gennes ''pre-smectic liquid'' model, provided that a sufficiently large external compressive force is applied to the free surfaces of the FSSF. In a real situation this force stems from the curvature of the surrounding miniscus, which plays the role of a volume reservoir. In this model a superheated FSSF is stabilized by balancing the external compressive and elastic forces. When heating takes place the bulk modulus of the pre-smectic lattice decreases, and when the superheating reaches a critical value, the FSSF is subject to a long-wavelength instability in thickness beause the external compressive and elastic forces can no longer be balanced for a fixed number of smectic layers. If a superheated FSSF possesses adequate stability against disruption, the balance of forces, which was disrupted, and hence the stability of the FSSF can be restored as a result of spontaneous thinning of the film to a thickness corresponding to a smaller number of smectic layers. In general, heating of a superheated FSSF is accompanied by a series of such thinning transitions. Near the critical points where the balance of the forces breaks down, the dislocation mechanism of spontaneous thinning, which could be responsible for the stratified nature of the progressive discrete thinning of real FSSFs, can become dangerous.
 [S1063-7761(99)00601-0]

1. Discrete thinning of free-standing smectic films (FSSFs), which is observed when the films are heated above the temperature of the smectic A–isotropic liquid[1] and smectic A–nematic (NA)[2,3] bulk phase transitions, remains one of the unsolved problems of the physics of liquid crystals. It is well known that FSSFs prepared below the temperature at which smectic order is destroyed in the bulk can be easily superheated even above the indicated temperature. Such superheating is made possible by the ''attracting'' action of the free surfaces and the remarkable stability of smectic films against the formation of perforations.[4] When heated, superheated FSSFs undergo a series of spontaneous thinning transitions at various temperatures that terminates with the films rupturing.[1–3] The maximum possible superheating of FSSFs is a power-law function of the film thickness,[1–3] the thermal expansion coefficient of a superheated FSSF is negative,[5] and the process of discrete thinning itself is of a pronounced monotropic character.[1–3] It is shown in the present paper that the entire spectrum of phenomena enumerated above has a natural explanation in the phenomenological pre-smectic liquid model of de Gennes,[6] provided that a sufficiently large external compressive force is applied to the free surfaces of the FSSF. The remarkable fact is that in a real situation this force is due to the curvature of the surrounding miniscus, which for a superheated FSSF acts like a volume reservoir (see Ref. 7).

2. We recall that the de Gennes pre-smectic liquid model[6] is a simplified version of the phenomenological model of a NA transition,[4,8] describing short-range smectic order effects above the bulk NA transition temperature. In this model a superheated FSSF is a thin layer of a pre-smectic (nematic) liquid which is bounded by two parallel free surfaces and is connected at the periphery with the surrounding volume reservoir. A pre-smectic wave of mass density is induced in the space between the free surfaces. This makes it possible to speak of the indicated system as a ''smectic'' film.

For definiteness, let the free surfaces bounding a superheated FSSF of thickness L be located at $z=\pm L/2$, and let the pre-smectic wave of mass density be parametrized as

$$\delta\varrho(z)=\sqrt{2}\,\varrho_0(z)\cos[\,q_0(z+u(z))\,]\ , \tag{1}$$

where $\varrho_0(z)$ is the modulation amplitude of the mass density, $q_0=2\pi/d_0$ is the optimal wave number, d_0 is the period of the volume smectic lattice, and $u(z)$ is a long-wavelength variable describing the elastic displacement of the smectic layers.

In the harmonic approximation the free-energy functional of a superheated FSSF can be written in the simple form (compare with Ref. 6)

$$F_{Sm}[\delta\varrho]=S\frac{\alpha}{2}\int_{-L/2}^{L/2}dz\Bigg[\tau\delta\varrho^2+\frac{\xi_{\|0}^2}{4q_0^2}[(\nabla_z^2+q_0^2)\delta\varrho]^2\Bigg], \tag{2}$$

where $\tau=(T-T_{NA})/T_{NA}$ is the dimensionless superheating of the FSSF relative to the critical temperature of the bulk NA transition ($\tau>0$), $\xi_{\|0}$ is the direct longitudinal smectic correlation length, α is a dimensionless phenomenological constant, and S is the area of the FSSF.

Substituting (1) into Eq. (2) reveals the nontrivial character of the de Gennes model and yields the following expression for the free-energy functional of a superheated FSSF:[6]

$$F_{Sm}[\varrho_0,u]=S\frac{\alpha\tau}{2}\int_{-L/2}^{L/2}dz[\varrho_0^2+\xi_C^2\{(\nabla_z\varrho_0)^2+q_0^2\varrho_0^2(\nabla_z u)^2\}], \tag{3}$$

where $\xi_C=\xi_{\|0}\tau^{-1/2}$ is the bulk value of the longitudinal smectic correlation length. Here and below we have

$$L\gg\xi_C\gg q_0^{-1}, \tag{4}$$

which is the condition for simultaneous applicability of the long-wavelength approximation and ''phonon'' parametrization of the pre-smectic wave of mass density.

3. In terms of the functional (3), the spatial distribution of the amplitude $\varrho_0(z)$ and the displacement $u(z)$ over the thickness of the FSSF is given by the following system of Euler equations:[6]

$$\nabla_z[\varrho_0^2\nabla_z u]=0, \tag{5}$$

$$\varrho_0[1+q_0^2\xi_C^2(\nabla_z u)^2]=\xi_C^2\nabla_z^2\varrho_0 . \tag{6}$$

Following de Gennes,[6] it is natural to assume that the positions of the maxima of the pre-smectic wave of mass density are rigidly fixed on the free surfaces of the FSSF. Since the functional (3) is invariant under the substitution $-\varrho_0\to\varrho_0$, this means that

$$q_0\left[L+2u\left(\frac{L}{2}\right)\right]=2\pi N, \tag{7}$$

where N is the integer, equal to the number of smectic layers, closest to L/d_0.

The solution of the system (5) and (6) with the boundary condition (7) is given in Ref. 6. The first derivation of an expression for the free energy of a pre-smectic wave of mass density induced in the space between the two parallel ''attracting'' surfaces is also given there. For a fixed surface value of the squared amplitude $[\varrho_0(\pm L/2)]^2\equiv\varrho_{0s}^2$, the result of de Gennes for the free energy of a pre-smectic wave of mass density has the form[9]

$$F_{Sm}(L)=S\alpha\tau\xi_C\varrho_{0s}^2\left[\tanh\left(\frac{L}{2\xi_C}\right)+\frac{1-\cos[\phi]_N}{\sinh(L/\xi_C)}\right], \tag{8}$$

where $[\phi]_N=q_0(L-Nd_0)$. The first term in Eq. (8) describes the usual (harmonic theory) contribution due to the spatial nonuniformity of the amplitude $\varrho_0(z)$ to the free energy of the FSSF, while the second term describes the elastic deformation energy unavoidably arising for thicknesses L different from the ''integer'' values $L_N=Nd_0$.

Subtracting from Eq. (8) the L-independent purely surface part of the free energy, we obtain the dimensional component of the free energy of a superheated FSSF:

$$\Delta F_{Sm}(L)=S\frac{\alpha\tau\xi_C\varrho_{0s}^2}{\sinh(L/\xi_C)}\left[\exp\left(-\frac{L}{\xi_C}\right)-\cos[\phi]_N\right]. \tag{9}$$

It follows from Eq. (9) that the free energy of a superheated FSSF has an infinite sequence of local minima for $L\approx L_N$, the depth of these minima increasing with decreasing N.[6] The latter signifies that superheated FSSFs are systems in a metastable thermodynamic state.

4. It is obvious that as a result of the existence of a volume reservoir, the thickness L of a superheated FSSF is a free thermodynamic parameter. The oscillator L-dependence of the energy (9) has the effect that the range of values of L is divided into sequentially alternating ''allowed'' ($[\Delta F_{Sm}]''_{LL}\geqslant 0$) and ''forbidden'' ($[\Delta F_{Sm}]''_{LL}<0$) zones differing by the sign of the effective bulk modulus of the pre-smectic lattice. When the inequality (4) is satisfied, the widths of the alternating zones are close to $d_0/2$, and the allowed zones (regions of positive elasticity of the pre-smectic lattice) are centered near ''integer'' values of L. We note that the positivity of the effective bulk modulus of the pre-smectic lattice is a necessary condition for the stability of superheated FSSFs. Therefore, under an external force the thickness of a FSSF can vary continuously only within relatively narrow limits, but in any case within the corresponding allowed zone.

On account of the inequalities (4) the free energy (9) can be represented to within small terms in the simple form

$$\Delta F_{Sm}^{(N)}(L)=-S\frac{B_m^{(N)}}{q_0^2\xi_C}\cos[\phi(L)]_N , \tag{10}$$

where

$$B_m^{(N)}=2\alpha q_0^2\xi_{\|0}^2\varrho_{0s}^2\exp(-L_N/\xi_C) \tag{11}$$

is the bulk modulus at the center of the pre-smectic wave of mass density (see below). In the present approximation the allowed zones are centered at $L=L_N$ and are given by the inequality

$$|L-L_N|\leqslant d_0/4 . \tag{12}$$

We note that the ''melting through'' points discussed by de Gennes[6] in the middle of a pre-smectic lattice fall outside the allowed zones and therefore are unattainable in superheated FSSFs.

5. When an external force is applied to the free surfaces, an elastic restoring force

$$G_{Sm}^{(N)}(L)=-S\frac{B_m^{(N)}}{q_0\xi_C}\sin[\phi(L)]_N , \tag{13}$$

arises in the system. Here the values of L do not exceed the zones given by the inequality (12) (see Fig. 1). Near the minima of the energy (10) Eq. (13) simplifies and assumes

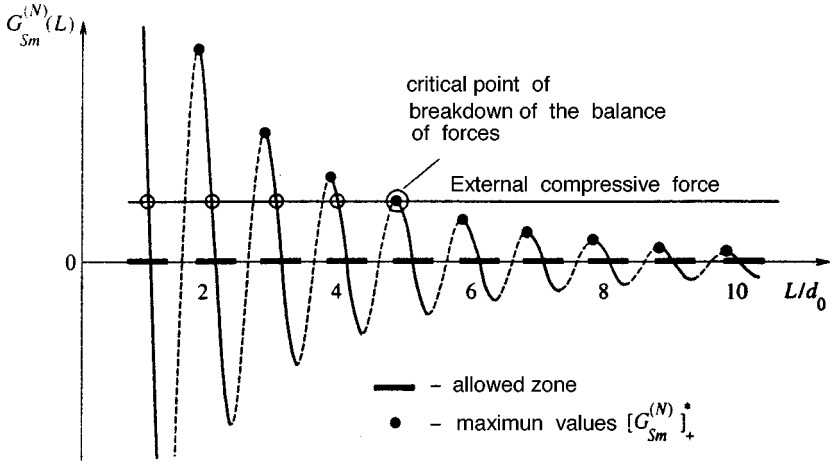


FIG. 1. Elastic force $G_{Sm}^{(N)}$ as a function of the reduced thickness L/d_0. The thick segments represent allowed zones centered at $L=Nd_0$, where N is the number of smectic layers (see Eq. (12)). The dashed curves show sections corresponding to thermodynamically unstable states. The horizontal line corresponds to an external compressive force $S\Delta p$. The points ($\bigcirc$) of intersection of the external compressive and elastic forces determine the spectrum of the equilibrium thicknesses. The existence of such a spectrum makes possible the restoration of the balance of the external compressive and elastic forces, which was destroyed at the critical point ($\odot$) by the spontaneous thinning of the superheated FSSF to an equilibrium thickness corresponding to a smaller number of smectic layers.

the form $G_{Sm}^{(N)} \approx -SB_m^{(N)}(\delta L/\xi_C)$, where $\delta L = L - L_N$. The latter agrees with the fact that when the inequalities (4) hold, the elastic deformation is entirely concentrated at the center of the FSSF with thickness of order ξ_C, where the bulk modulus $B_m^{(N)}$ of the pre-smectic wave of mass density is minimal (see Ref. 6).

It is obvious that because the allowed and forbidden zones alternate with one another, the elastic properties of a superheated FSSF are discontinuous functions of the thickness L. Specifically, the force (13) has upper and lower limits

$$[\,G_{Sm}^{(N)}\,]_{\pm}^{*} = \pm\; S\frac{B_m^{(N)}}{q_0\xi_C}\;, \tag{14}$$

which are reached, respectively, at the bottom ($+$) and top ($-$) boundaries of the corresponding allowed zones (see Fig. 1).

It is obvious, considering the relation (11), that the limiting values (14) decrease rapidly with increasing τ and increase with decreasing number N of smectic layers (see Fig. 1). Therefore the heating of a superheated FSSF in the presence of an external force inevitably terminates with the spontaneous breakdown of the balance of the external and elastic forces and, in consequence, mechanical instability of the FSSF. When the external force is compressive, the balance of forces which was disrupted can be restored and hence the stability of the FSSF can be restored as a result of spontaneous thinning of the film to a thickness corresponding to a smaller number of smectic layers (by shedding ''excess'' smectic layers into the surrounding reservoir). Moreover, when a superheated FSSF with adequate stability against disruption is heated in the presence of an external compressing force, in general a series of spontaneous thinning transitions at various temperatures should occur. It is obvious that the specific scenario of such successive discrete thinning, including disruption of the FSSF, is probabilistic and a dynamic theory is required to analyze it.

We note that the successive discrete thinning of superheated FSSFs is of a distinctly monotropic character. Indeed, on cooling, on the one hand the balance of the forces acting on a FSSF is not disrupted in any way, while on the other hand the transition to larger thicknesses is obviously energetically unfavorable (see Sec. 3). For heating in the presence of an external tensile force, the breakdown of the balance of the forces will inevitably be accompanied by disruption of the FSSF.

6. Let an external compressive force produce an additional pressure $\Delta p > 0$ on the free surfaces. Then the condition of mechanical equilibrium of a superheated FSSF (the condition for the balance of forces) is

$$-\;\frac{B_m^{(N)}}{q_0\xi_C}\sin[\,\phi(L)\,]_N = \Delta p\;, \tag{15}$$

where, as earlier, the values of L stay within the corresponding allowed zone. It is obvious that a superheated FSSF is stable only if $B_m^{(N)}/(q_0\xi_C) \geqslant \Delta p$ holds, which corresponds to an ''upper'' limiting value $[G_{Sm}^{(N)}]_+^*$ exceeding the external compressive force $S\Delta p$ (see Fig. 1).

It follows from Eq. (15) that when a superheated FSSF is heated in the presence of an external compressive force, the equilibrium value of L shifts continuously toward the lower limit of the corresponding allowed zone (the point where $L - L_N = -d_0/4$ holds); this corresponds to a negative thermal expansion coefficient of the superheated FSSF. This effect is a trivial consequence of the decrease in the effective bulk modulus of the pre-smectic lattice when heating occurs in the presence of a constant external compressive force. The lower limit of the allowed zone is itself a critical point where the balance of the external compressive and elastic forces breaks down, i.e., the point of mechanical instability of the FSSF (see Fig. 1). It is obvious that this point is reached for a critical superheating $\tau^*(L_N)$ given by the condition $B_m^{(N)}/(q_0\xi_C) = \Delta p$.

We note that the breakdown of the balance of the external compressive and elastic forces signifies a long-wavelength thickness instability of the superheated FSSF. Indeed, introducing a displacement of the free surfaces relative to the equilibrium positions $z = \pm L/2$ and adding to Eq. (3) terms which explicitly account for the increase in the free surface area (which are proportional to the surface tension γ), we can show that the critical points of breakdown of the balance of forces are simultaneously points of the long-wavelength instability of an antisymmetric capillary squeez-

ing mode, which is responsible for the thermal thickness fluctuations.[10,11] Formally, this is related to the fact that the second derivative of the interaction energy of the free surfaces with respect to thickness L (in this case, $[\Delta F_{Sm}]''_{LL}$) vanishes at the critical points of breakdown of the balance of forces and plays the role of a gap in the spectrum of the squeezing mode. Therefore anomalous enhancement of light scattering by the squeezing mode can occur as the critical points where the balance of forces breaks down are approached (see Ref. 12). It can also be shown that as these points are approached, the thermal expansion coefficient of a superheated FSSF should diverge (while remaining negative); this reflects the divergence of the isothermal compressibility that is natural for the thickness instability.

7. It is obvious that if the dislocation mechanism of thinning is neglected (see below), the critical points where the balance of forces breaks down are points with the maximum possible superheating of the FSSF. The envelope of these points is determined by the equation

$$\exp\left(\frac{L_N}{\xi_C^*}\right)=\frac{2B_s}{q_0\xi_C^*\Delta p}, \tag{16}$$

where $B_s=\alpha q_0^2\xi_{\parallel 0}^2\varrho_{0s}^2$ is the ''surface'' value of the bulk modulus and ξ_C^* is the bulk smectic correlation length on the envelope.

It follows from Eq. (16) that, to logarithmic accuracy, the envelope of the points of maximum possible superheating of a FSSF follows a simple ''scaling'' relation

$$L_N/\xi_C^*\approx\text{const}, \tag{17}$$

where the constant (the logarithm of the right-hand side of Eq. (16)) is large because of the left-hand inequality in Eq. (4). The latter signifies that near the points of the maximum possible superheating, which lie on the ''scaling'' section of the envelope, the thickness L of a superheated FSSF should be much greater (at least severalfold) than the bulk smectic correlation length ξ_C.

We note that the relation (17) is a trivial consequence of the fact that when the inequalities (4) holds, the elastic restoring force is short-range. Since the latter is unrelated to the simplified character of the de Gennes model, there is hope that the relation (17) is universal. Then for a power-law function $\xi_C(\tau)$ the envelope of the points of the maximum possible superheating $\tau^*(L_N)$ should be a power-law function of L_N with an exponent that is the reciprocal of the exponent of the bulk smectic correlation length.

8. Proceeding to numerical estimates, we note that the van der Waals force (which is always present in the real situation), the force due to the pseudo-Casimir effect (dimensional screening of thermal fluctuations of a nematic director),[13] and the force due to the pressure difference arising between the exterior and interior regions of a FSSF as a result of the curvature of the surrounding miniscus[7,14] could act as the ''external'' compressive force inducing successive discrete thinning of superheated FSSFs in conjunction with heating. The first two forces are long-range, and for thicknesses $L\sim 10^{-5}$ cm, which are typical of known experiments,[2,3] they are of order 1 dyne/cm^2, while the last force depends weakly on L and gives a characteristic value $\Delta p\sim 10^3$ dynes/cm^2.[7,14] Such a large ''external'' compressive force, together with the large value of $q_0\xi_C\sim 10^2$ (for $\tau\sim 10^{-3}$ see, for example, Ref. 15), makes it necessary to provide $B_m^{(N)}\sim 10^5$ dynes/cm^2 on the envelope of the critical points. The latter with $B_s\sim 10^7$ dynes/cm^2 (i.e., on the order of the typical value of the bulk modulus B (Ref. 4)) gives a ''reasonable'' value of the ratio $L_N/\xi_C^*\approx 4-5$. Unfortunately, the lack of measurements of the maximum possible superheating of FSSFs as a function of the number N of smectic layers and the temperature-dependence of the bulk smectic correlation length ξ_C for the same substance together with the uncertainty in the bulk modulus B_s make it impossible at the present time to discuss the experimental situation in greater detail, and specifically, to check the ''scaling'' relation (17).

9. We shall show that the dislocation mechanism of spontaneous thinning of superheated FSSFs, which is associated with the production of elementary edge dislocation loops in the plane of the central layer, where the effective uniaxial tension of such a dislocation is minimal, can become dangerous near the critical points where the balance of forces breaks down. The effective uniaxial tension evidently consists of ''volume'' ($E_m^{(N)}$) and ''surface'' ($E_s^{(N)}$) parts, which are related to the appearance of a distributed elastic deformation field in the interior of the FSSF and with the occurrence of transitional edge profiles at the free surfaces, causing the thickness of the FSSF tro vary continuously by an amount of order d_0 (see Refs. 16–18).

Since the amplitude $\varrho_0(z)$ varies slowly in the range $|z|\lesssim\xi_C$, the well-known ''parabolic'' formula of de Gennes can be used for the elastic deformation field produced by an edge dislocation,[4,19] and $E_m^{(N)}$ can be estimated qualitatively as

$$E_m^{(N)}\sim\frac{(B_m^{(N)})^{3/4}K^{1/4}d_0^2}{\xi_C^{1/2}}, \tag{18}$$

where K is the Frank constant, determining the energy of transverse bending of the director field in a nematic.[4] It can be shown in turn that the width of the transitional edge profile occurring at the free surface of the FSSF is given by the squeezing capillary length $\lambda_{sq}=(\gamma/[\Delta F_{Sm}]''_{LL})^{1/2}$. Using Eq. (10), $E_s^{(N)}$ can be estimated qualitatively as

$$E_s^{(N)}\simeq\gamma\frac{d_0^2}{\lambda_{sq}}\sim\frac{\gamma^{1/2}(B_m^{(N)})^{1/2}d_0^2}{\xi_C^{1/2}}. \tag{19}$$

Comparison of (18) and (19) shows that for $\gamma\gg(KB_m^{(N)})^{1/2}$ the ''surface'' contribution to the effective uniaxial tension of an edge dislocation dominates. This inequality obviously holds, since in the typical situation $\gamma\simeq 30$ ergs/cm^2,[4] $K\sim 10^{-6}$ dynes,[4] and $B_m^{(N)}\sim 10^5$ dynes/cm^2 (see Sec. 8).

It can also be shown that in a wide neighborhood of a critical point where the balance of forces breaks down, as determined by the inequality $B_m^{(N)}/(q_0\xi_C)^2\ll\Delta p$, the gain in the free-energy density of a superheated FSSF as a result of the thinning of the film by an amount of order d_0 is less than

the corresponding work $\Delta p d_0$ performed by the pressure difference. Then the activation energy of a critical dislocation loop can be estimated qualitatively as

$$W_c \simeq \pi \frac{(E_s^{(N)})^2}{\Delta p\, d_0} \sim \frac{B_m^{(N)}}{q_0 \xi_C \Delta p} \gamma d_0^2 \quad . \tag{20}$$

We note that $d_0 \sim 10^{-7}$ cm (Ref. 4) and therefore $\gamma d_0^2 \sim 10 k_B T$, i.e., it is on the order of the threshold activation energy satisfies $W_c^* (W_c^* \approx 60 k_B T$ (Ref. 4)). Therefore it follows from Eq. (20) that far from the critical points where the balance of forces breaks down $(B_m^{(N)}/(q_0 \xi_C) \gg \Delta p)$ the activation energy $W_c \gg W_c^*$ and so the probability of spontaneous thinning of a superheated FSSF as a result of nucleation of growing dislocation loops is vanishingly small. Conversely, near the critical points $(B_m^{(N)}/(q_0 \xi_C) \sim \Delta p)$ the activation energy satisfies $W_c \sim W_c^*$ and so dislocations can strongly influence the scenarios of successive discrete thinning of a superheated FSSF, determining, specifically, the stratified nature of the thinning. Unfortunately, the purely qualitative character of the estimates, together with the uncertainty in the character of the spontaneous thinning resulting from the squeezing instability, preclude us from making an unequivocal assertion about the real nature of the successive discrete thinning of superheated FSSFs. In our view the final word here belongs to experiments.

In conclusion, we note that discrete thinning of superheated smectic (pre-smectic) films should also be observed in Izraelishvili-type experiments,[9,20,21] if the temperature is raised under a fixed external load. The observation of this effect would be an additional confirmation of the theory presented here.

We thank V. K. Dolganov, E. I. Kats, V. V. Lebedev, and A. R. Muratov for their interest in this work and for a helpful discussion. This work was supported by the Russian Fund for Fundamental Research (Project 96-02-18235).

*)E-mail: vita@fluid.msk.ru

[1] T. Stoebe, P. Mach, and C. C. Huang, Phys. Rev. Lett. **73**, 1384 (1994).
[2] E. I. Demikhov, V. K. Dolganov, and K. P. Meletov, Phys. Rev. E **52**, R1285 (1995).
[3] V. K. Dolganov, E. I. Demikhov, R. Fouret, and C. Gors, Phys. Lett. A **220**, 242 (1996).
[4] P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, 1993).
[5] C. C. Huang, P. Mach, and P. M. Johnson, in *The 6th International Conference on Ferroelectric Liquid Crystals. Conf. Summaries* (Brest, France, 1997), p. 280.
[6] P. G. de Gennes, Langmuir **6**, 1448 (1990).
[7] P. Pieranski, L. Beliard, J.-Ph. Tournellec *et al.*, Physica A **194**, 364 (1993).
[8] P. G. de Gennes, Solid State Commun. **10**, 753 (1972).
[9] L. Moreau, P. Richetti, and P. Barois, Phys. Rev. Lett. **73**, 3556 (1994).
[10] A. Vrij, J. G. H. Joosten, and H. M. Fijnaut, Adv. Chem. Phys. **48**, 329 (1981).
[11] C. Y. Young and N. A. Clark, J. Chem. Phys. **74**, 4171 (1981).
[12] A. Böttger, D. Frenkel, J. G. H. Joosten, and G. Krooshof, Phys. Rev. A **38**, 6316 (1988).
[13] A. Ajdari, L. Peliti, and J. Prost, Phys. Rev. Lett. **66**, 1481 (1991).
[14] J.-C. Géminard, R. Hołyst, and P. Oswald, Phys. Rev. Lett. **78**, 1924 (1997).
[15] R. J. Birgeneau, C. W. Garland, G. B. Kasting, and B. M. Ocko, Phys. Rev. A **24**, 2624 (1981).
[16] P. S. Pershan, J. Appl. Phys. **45**, 1590 (1974).
[17] P. S. Pershan and J. Prost, J. Appl. Phys. **46**, 2343 (1975).
[18] R. Hołyst and P. Oswald, Int. J. Mod. Phys. B **9**, 1515 (1995).
[19] P. G. de Gennes, C. R. Acad. Sci. Paris B **275**, 939 (1972).
[20] R. G. Horn, J. N. Israelachvili, and E. Perez, J. de Phys. **42**, 39 (1981).
[21] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1992).

Translated by M. E. Alferieff

Critical velocity in ^{3}He-B vibrating wire experiments as analog of vacuum instability in a slowly oscillating electric field

A. Calogeracos

Low Temperature Laboratory, Helsinki University of Technology, 02150 Espoo, Finland; NCA Research Associates, Maroussi 15I 01 Athens, Greece

G. E. Volovik

Low Temperature Laboratory, Helsinki University of Technology, 02150 Espoo, Finland; L. D. Landau Institute for Theoretical Physics, 117940 Moscow, Russia

(Submitted 24 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 70–79 (January 1999)

The Lancaster experiments [C. A. M. Castelijns, K. F. Coates, A. M. Guenault, S. G. Mussett, and G. R. Pickett Phys. Rev. Lett. **56**, 69 (1986)] with a cylindrical wire moving in superfluid ^{3}He-B are discussed, where the measured critical velocity of pair creation was much below the Landau critical velocity. The phenomenon is shown to be analogous to the instability of the electron–positron vacuum in an adiabatically alternating strong electric potential of both signs, where the positive- and negative-root levels cross and thus the instability threshold is half the conventional value in a single static potential well. © *1999 American Institute of Physics.* [S1063-7761(99)00701-5]

1. INTRODUCTION

In superfluid Fermi systems pairs of quasiparticles are produced by a uniformly moving object if its velocity exceeds the Landau critical velocity, $v_L = \Delta_0 / p_F$. Here p_F is the Fermi momentum, and Δ_0 is the superfluid gap in bulk liquid. The critical velocity v_L is also called the pair-breaking velocity; it marks the threshold of instability of the superfluid vacuum: breaking of Cooper pairs which form the superfluid condensate. In the vacuum of high energy physics, a similar situation can occur (i) in a strong electric field;[1–3] (ii) in a strong gravitational field, for example near an event horizon;[4] (iii) if the hypothetical object, which is external to the physical vacuum, moves at superluminal speed. Here we consider pair creation in superfluid ^{3}He-B, which is analogous to the production of electron–positron pairs in a strong electric field.

Such experiments have been conducted in Lancaster,[5,6] where a cylindrical wire vibrating in superfluid ^{3}He-B has been used as a moving object. It turned out that the measured critical velocity, at which significant extra dissipation of the wire was observed due to particle creation, was considerably less than v_L (about $0.25 v_L$, independent of the material and radius of the wire).

It was originally suggested in Ref. 5 that such reduction has two origins: a geometrical factor 1/2 results from the local enhancement of velocity near the wire, while the other reduction is related to suppression of the gap in the vicinity of the surface of the wire, $\Delta < \Delta_0$. As a result, the Landau criterion for the filling of surface bound states is much less than v_L. To provide for momentum loss by the wire, however quasiparticles must escape to infinity, which is why the production of scattering states at subcritical velocity must be explained. This scenario was developed by Lambert, who shows[7] that adiabatic oscillation can do this job if the speed of the wire exceeds some value, which was estimated to be $v^* = (1/5) v_L$.

Here, we further develop these arguments, noting firstly that in ^{3}He-B the surface leads to splitting of the gap into $\Delta_\parallel$ and $\Delta_\perp$, and secondly that the classical description of the bound state in the surface layer must be replaced by a quantum mechanical description. We obtain a modified value for v^*, which depends on the gap suppression. Since Bogolyubov–Nambu fermions in ^{3}He-B are in many respects similar to Dirac electrons, we associate the critical radiation of quasiparticles by a slowly vibrating wire with instability of the electron–positron vacuum in the presence of a strong electric field. The present case corresponds to a slowly alternating electric potential, which allows for electron–positron production in much weaker fields than the conventional mechanism discussed by Gershtein and Zel'dovich.[2] In this scenario the classical positive- and negative-root solutions cross, which leads to particle–antiparticle production (see also the discussion in Ref. 8). We construct a simple time-dependent potential for Dirac electrons that enables us to model the proposed scenario.

2. FERMIONS IN A VIBRATING WIRE

2.1. Fermionic spectrum in ^{3}He-B

In bulk superfluid ^{3}He-B the fermionic spectrum is defined by the following 4×4 matrix Hamiltonian (Bogolyubov–Nambu Hamiltonian)[9,10]:

$$H(\mathbf{p}) = \beta M(p) + c\mathbf{p}\cdot\boldsymbol{\alpha}, \quad M(p) = v_F(p - p_F),$$

$$c = \Delta_0 / p_F. \tag{1}$$

Here β and $\boldsymbol{\alpha}$ are Dirac matrices, composed from the 2×2 Pauli matrices $\boldsymbol{\tau}$ describing the Bogolyubov–Nambu spin in particle-hole space and 2×2 Pauli matrices $\boldsymbol{\sigma}$ for conventional spin:

$$\beta=\tau_3, \quad \boldsymbol{\alpha}=\tau_1\boldsymbol{\sigma}. \tag{2}$$

The energy spectrum is

$$E_{\pm}(p)=\pm\sqrt{M^2(p)+c^2p^2}. \tag{3}$$

The quantity c plays the part of the speed of light, but in contrast to the relativistic case, the mass M depends on the momentum p. Since $v_F\gg c$, the minimum of the positive energy occurs not at $p=0$ but at $p=p_F$, with $\min E_+(p)=\Delta_0$.

According to the Landau criterion, if the external body moves at a velocity greater than $v_L=\min(E_+(p)/p)=c$, it will produce quasiparticles. In contrast to the relativistic case, where the minimum is realized at $p\to\infty$, in ^{3}He-B it occurs at $p=p_F$.

In the reference frame of the body, the energy spectrum is Doppler shifted:

$$H(\mathbf{p})=\mathbf{p}\cdot\mathbf{v}_s+\beta M(p)+c\mathbf{p}\cdot\boldsymbol{\alpha},$$

$$E_{\pm}(p)=\mathbf{p}\cdot\mathbf{v}_s\pm\sqrt{M^2(p)+c^2p^2}, \tag{4}$$

where $\mathbf{v}_s$ is the superfluid velocity in the body frame. If $v_s(\infty)>c$, the positive square-root continuum merges with the negative square-root continuum, and thus the production from the vacuum of pairs of quasiparticles with momentum p_F becomes possible. Here we discuss the situation in which particle production is possible even well below the Landau criterion. This is a combined effect of enhancement of the local superfluid velocity in the vicinity of the surface of the object, the decrease in the ''speed of light'' near the surface, and adiabatic oscillation of the velocity of the body.

2.2. Fermions in the surface layer

Experimentally,[5,6] the external body moving in ^{3}He-B is a cylindrical wire of radius R from 2 to 50 μm, which is much larger than the coherence length $\xi\sim v_F/\Delta_0$. The velocity of the wire oscillates, $\mathbf{u}(t)=\hat{\mathbf{x}}u(t)$, $u(t)=u_0\cos(\omega t)$, at frequency $\omega\sim10^2-10^3$ Hz, which is much less than the characteristic quasiparticle energy of order Δ_0; the motion is thus extremely adiabatic.

The presence of a moving external object disturbs the vacuum state of the superfluid. First, the velocity field is modified by the moving wire. In the reference frame of the wire, the superfluid executes ideal dipole flow around the wire:

$$\mathbf{v}_s(\mathbf{r},t)=-\mathbf{u}(t)+\frac{R^2}{r^2}[2\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{u}(t))-\mathbf{u}(t)], \quad r>R, \tag{5}$$

where $\mathbf{r}=(x,y)$ is the 2D radius vector in the plane perpendicular to the wire, reckoned from the center of the wire; $\hat{\mathbf{r}}=\mathbf{r}/r$. At two lines at the surface of the wire the superfluid velocity is twice the value at infinity: $\mathbf{v}_s(\pm R\hat{\mathbf{y}})=-2\mathbf{u}(t)$.

The second effect is that the order parameter (gap) is suppressed near the surface of the wire in a layer whose thickness is of the order of the coherence length. In ^{3}He-B this suppression is anisotropic, which leads to the two ''speeds of light'' in the region $r-R\sim\xi\sim v_F/\Delta_0$ (see Fig. 1):

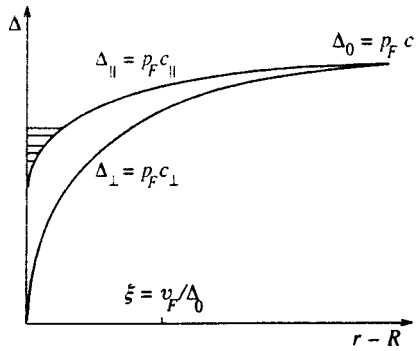


FIG. 1. Schematic illustration of gaps, ''speeds of light,'' and bound states near the surface of the wire.

$$H=\beta M(p)+[c_{\parallel}(\delta_{ij}-\hat{n}_i\hat{n}_j)+c_{\perp}\hat{n}_i\hat{n}_j]p_i\alpha_j,$$

$$E_{\pm}(p)=\mathbf{p}\cdot\mathbf{v}_s\pm\sqrt{M^2(p)+c_{\perp}^2(\hat{\mathbf{n}}\cdot\mathbf{p})^2+c_{\parallel}^2[\hat{\mathbf{n}}\cdot\mathbf{p}]^2}, \tag{6}$$

where $c_{\perp}=\Delta_{\perp}/p_F$ and $c_{\parallel}=\Delta_{\parallel}/p_F$ are the ''speeds of light'' along the normal $\hat{\mathbf{n}}=\hat{\mathbf{r}}$ to the surface of the wire and parallel to the surface, respectively. According to Ref. 11, where diffusive boundary conditions were considered, the transverse speed of light vanishes, $c_{\perp}(r=R)=0$, while $c_{\parallel}(r=R)\approx0.4c$ at $T=0$. Due to the suppression of the order parameter the surface layer serves as a potential well for quasiparticles, which contains bound states with energies below the gap[12] (see Fig. 1).

3. CRITICAL VELOCITIES AND NUCLEATION OF QUASIPARTICLES

3.1. Excitations of bound states

We first consider a uniformly moving wire with constant velocity u. The filling of the bound states can occur at a velocity smaller than the Landau velocity v_L for creation of fermions in the continuous spectrum. This velocity can be estimated from the Landau criterion for the classical spectrum in Eq. (6) for the surface fermions. Since the superfluid velocity is tangential near the wall, the Landau velocity for nucleation of quasiparticles in surface states is $v_L^{\rm surf}=\min(E_+(p)/p_{\parallel})=c_{\parallel}(r=R)$. The minimum first occurs at $p_{\parallel}=p_F$ and $E_+=p_Fc_{\parallel}(r=R)$; note that the transverse speed of light $c_{\perp}(r)$ does not enter into the criterion. Taking enhancement of the superfluid velocity near the wall into consideration, one obtains that negative energy levels appear in the surface layer if the velocity u exceeds

$$v_0^*=\frac{1}{2}c_{\parallel}(r=R)=v_L\frac{\Delta_{\parallel}(r=R)}{2\Delta_0}. \tag{7}$$

Here we used the Lambert notation for the various critical velocities[7] (in his paper, however, he does not take splitting of the gap into account, and he assumes that v_0^* is very small).

The situation does not change if instead of taking a classical approach to the energy spectrum in the surface layer, one takes into account quantization of quasiparticle motion along the normal to the wall. According to Ref. 12, the quasicontinuum of the subgap bound states starts above the energy $p_F c_\parallel(r=R)$ with $p_\parallel \approx p_F$, which again yields $(1/2)c_\parallel(r=R)$ as the Landau critical velocity for nucleation of surface fermions.

Can negative energy levels in the surface layer be filled by quasiparticles? For this it is necessary to connect to a reservoir of quasiparticles. It appears that this always occurs in the present situation. The negative square-root branch E_- of quasiparticles in Eq. (4) is always occupied. When the velocity u exceeds v_0^*, the energy of branch E_- can be positive, while the energy of branch E_+ can be negative, so the branches overlap and a quasiparticle from the filled branch E_- can jump to an empty level on E_+. Since momenta p_z of these states are opposite, this can happen only if the momentum p_x is not conserved, which is always the case because of surface roughness.

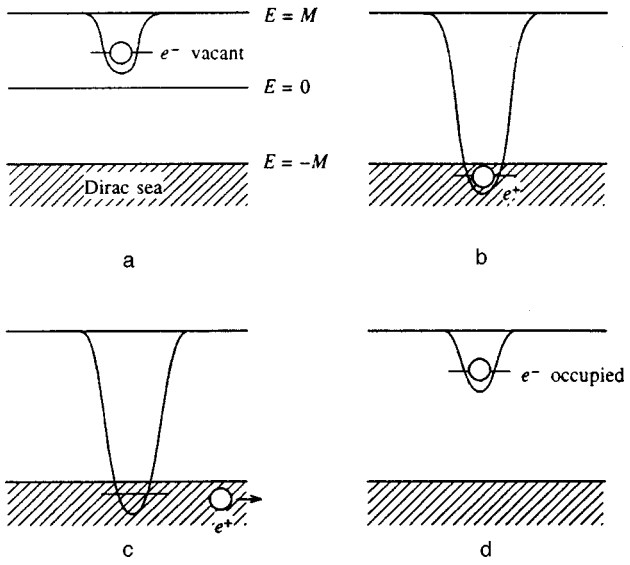


FIG. 2. Zel'dovich mechanism of positron creation.

3.2. Analog of Zel'dovich mechanism of positron nucleation

When the surface Landau velocity is reached, however, the created surface quasiparticles, which have zero energy in the wire reference frame, cannot escape to infinity, where the minimum energy of the scattering state is $\Delta_0 - p_F u = \Delta_0[1-(1/2)(c_\parallel(r=R)/c)] > 0$. For quasiparticles to escape to infinity, the velocity of wire must be considerably higher. This happens when the lowest energy of the bound state $p_F c_\parallel(r=R) - 2p_F u_0$ merges with the continuum of the negative root states, whose upper edge is at $-\Delta_0 + p_F u$. This yields a criterion for the emission of a quasihole, $u > v_1^*$, with

$$v_1^* = [c + c_\parallel(r=R)]/3. \tag{8}$$

This is equivalent to the production of a positron by the strong electrostatic potential well discussed by Zel'dovich, where the created electron fills the bound state, while the positron is emitted to infinity.

It may be helpful to remind the reader of the essential features of the Zel'dovich mechanism[2] (see also Ref. 13 for a detailed review). Consider an electron-attractive potential with a vacant discrete level (Fig. 2a). Suppose that the potential adiabatically increases in strength. The level will cross $E=0$ for some value V_1 of the potential ($V_1 = \pi/2$ for a δ-function potential). There is nothing critical happening during the crossing. At some greater value V_2 the level crosses $E=-M$, and thus merges with the negative energy continuum ($V_2 = \pi$ for a δ-function potential). The original electron vacancy is now interpreted as the presence of a positron, and since the positron occupies a scattering state, it can escape to infinity (Fig. 2c). If the potential now becomes weak again, we revert to the situation of a discrete energy level (Fig. 2d), which however is now electron-filled. The whole cycle clearly conserves charge; the positron escapes, however, when the potential is strong, and the electron is observed when the potential returns to its original weak value.

If the velocity of the object is kept constant, the emission of quasiparticles at $u > v_1^*$ will finally stop after all the negative levels become occupied. Then the object will move without dissipation, but its mass will be greater due to the quasiparticles that occupied the negative-energy bound states. In the case of moving vortices in superfluids and superconductors, a similar enhancement of the mass due to trapped quasiparticles is the origin of the so-called Kopnin mass of the vortex (see Ref. 14).

Thus, for a uniformly moving object, dissipation is absent even if its velocity exceeds v_1^*, and nothing happens until the Landau velocity $v_L = c$ is reached (if, however, hydrodynamic instability does not develop earlier[15]). The source of this instability can be the following: filling of the bound state leads to an increase in the normal component density, and thus to rearrangement of the whole superflow pattern due to mass conservation (see Ref. 14 for the effect of the backflow due to the normal component in the vortex core). At some velocity the superflow pattern becomes unstable, being unable to satisfy mass conservation. Such hydrodynamic instability usually leads to the production of vortices by the moving object.

Equation (8) is analogous to the criterion obtained by Lambert,[7] and transforms to his result if $c_\parallel(r=R)$ is neglected. However, in reality $c_\parallel(r=R)/c$ is not small: it is close to unity for specular boundary conditions, while for diffuse conditions it is about $c_\parallel(r=R)/c = 0.4$.[11] Thus, the most optimistic estimate yields $v_1^* = 0.47 v_L$, which is greater than the experimental value, demonstrating that supercritical dissipation starts at $\sim 0.25 v_L$. Thus, it turns out that the Zel'dovich mechanism in its simplest form is not responsible for supercritical behavior. Modification of this mechanism is required according to another scenario, also suggested by Lambert,[7] who exploited the adiabatic oscillations of the wire.

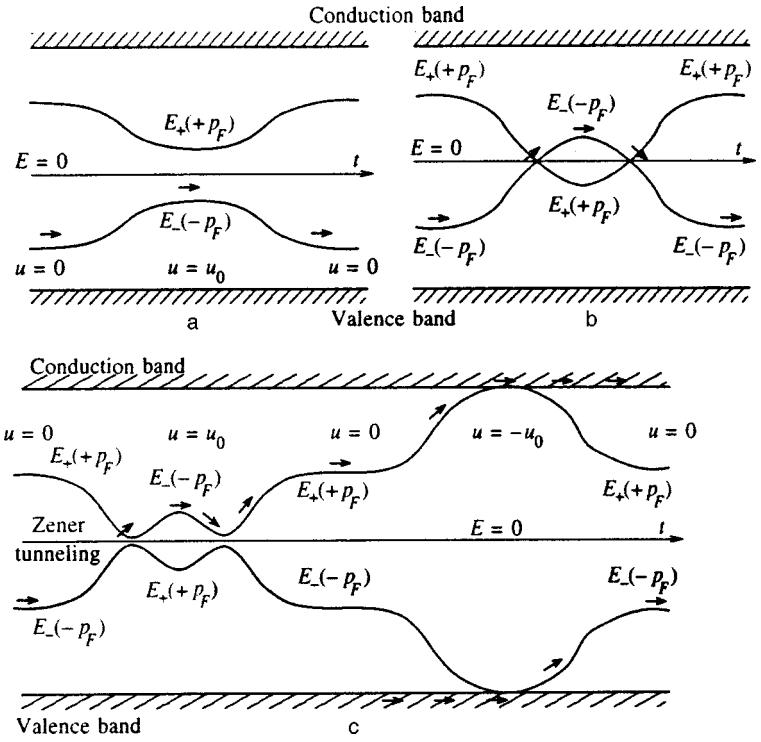


FIG. 3. Temporal evolution of two branches, $E_\pm(p_x=p_F)$ and $E_-(p_x=-p_F)$, of bound states. a) Subcritical regime. The half of the period is shown in which the velocity increases to u_0, and then decreases to zero. b) In the supercritical regime the two branches cross each other, but the evolution of the levels does not change if momentum p_x is conserved. c) Level flow in the presence of mixing of $+p_F$ and $-p_F$ states. The whole period of oscillations is shown, in which an ''electron–positron'' pair is created.

3.3. Radiation by adiabatically oscillating potential

This mechanism exploits the fact that in an oscillating wire $u=u_0\cos(\omega t)$, the velocity changes sign twice per period. Consider the case in which the amplitude of the velocity $u_0>v_0^*$ in Eq. (7). After the peak velocity $+u_0$ is reached, for example, the bound state with energy $E_+=\Delta_0 c_\parallel(r=R)/c-2p_F v_0^*=0$ will be filled by a quasiparticle. If the wire vibrates slowly, which is the case since $\omega \ll \Delta_0$, after half a period the energy of this quasiparticle will become $E_+=\Delta_0 c_\parallel(r=R)/c+2p_F v_0^*$. We must compare this energy to the minimum energy of the scattering states, which occurs for the opposite direction of the momentum: $E_+(\min\ \text{scattering})=\Delta_0-p_F v_0^*$. Thus, if

$$v_0^*>v_L/5, \quad \text{i.e.,} \quad c_\parallel(r=R)/c>2/5, \tag{9}$$

the continuum (conduction) energy band is achieved and quasiparticles will be emitted by the vibrating wire. If, however $c_\parallel(r=R)<(2/5)c$, then the same mechanism starts to work at higher velocity, with $u_0>(1/5)v_L$. The latter case corresponds to the Lambert result obtained under the assumption that v_0^* is very small. Thus, the criterion for the emission of quasiparticles by the vibrating wire is $u_0>v^*$, with

$$v^*=v_0^* \quad \text{if } v_0^*>v_L/5,$$
$$v^*=v_L/5 \quad \text{if } v_0^*<v_L/5. \tag{10}$$

The general scheme of particle production for $u_0>v^*$ is shown in Fig. 3. In the supercritical regime (b), as progresses, the two branches $E_+(p_x=p_F)$ and $E_-(p_x=-p_F)$ of bound states cross each other if the momentum p_x is conserved. In a real situation, surface roughness mixes $+p_F$ and $-p_F$ states, which leads to repulsion of levels. The temporal evolution of levels and one of the trajectories of a quasiparticle in the supercritical regime are shown in Fig. 3c for a full period of oscillation. The transition of a quasiparticle from branch $E_-(-)$ to branch $E_+(+)$ occurs either by scattering or Zener tunneling. In one cycle, the particle moves from the Dirac sea to the positive energy continuum via bound states. This corresponds to the production of an electron–positron pair via bound states.

This mechanism is different from the Zel'dovich mechanism, in which the bound-state energy touches the continuum spectrum of the Dirac sea, the electron occupies the bound state, and the positron is emitted. In the present case, criticality occurs when the bound-state energy of the branch E_+ reaches the zero energy and thus touches the occupied bound states of the branches E_-. In this process two particles in the scattering states are created (''electron'' and ''positron''), resulting in the production of momentum $2p_F$ from the vacuum. The level flow along two other branches, $E_-(p_x=p_F)$ and $E_+(p_x=-p_F)$, is similar, but is shifted by half a period. As a result, in this process the opposite momentum, $-2p_F$, can be produced during a cycle.

4. ANALOGY WITH FERMION PRODUCTION IN A STRONG ELECTRIC FIELD

Since close to the threshold velocity the relevant quasiparticle momentum p_x is greatest, $p_x=\pm p_F$, the term $\mathbf{p}\cdot\mathbf{v}_s$ in Eq. (4) serves as the timelike component of the 4-vector electromagnetic potential: $\mathbf{p}\cdot\mathbf{v}_s=\pm p_F v_{sx}(x,t)=eA_0(x,t)$. Here the sign of the momentum plays the part of the electric charge. Thus we have the problem of Dirac particles in a strong electric field. The above mechanism of particle creation requires five ingredients:

1) bound states;

2) for filling of the negative energy levels above v_0^* it is

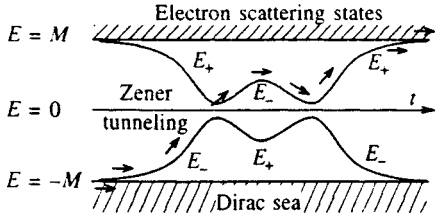


FIG. 4. Spectral flow and pair production in a system of potentials alternating in antiphase in the supercritical regime.

necessary to have the mirror image branch of quasiparticles with opposite momentum (i.e., with opposite e);

3) there must be an interaction that mixes the momenta p_F and $-p_F$, and thus allows the sign e to change;

4) the potential A_0 must be strong enough for the positive-root and negative-root branches to cross;

5) the potential A_0 must oscillate slowly in time. During one cycle, the positive-root and negative-root levels cross and then return to their respective (positive/negative) continua.

This is why, in mapping to the Dirac problem, we need particles with both negative and positive charges, which can transform into one another. One possibility is to use, instead of the timelike component of the four-vector electromagnetic potential, a time- and space-dependent mass term. In this case the spectrum is symmetric, so that positive and negative energy bound states can in principle approach one another,[3] in a manner similar to Fig. 3b.

The other possibility is to have the conventional electromagnetic field A_0, but in the form of two spatially separated potentials with opposite signs of A_0. In this case one has the required mirror image of states. This can be modeled by the conventional Dirac Hamiltonian with potential

$$A_0(x,t) = U\cos(\omega t)[\,\delta(x+a) - \delta(x-a)\,]. \tag{11}$$

Assuming that the Dirac mass $M=1$ and oscillations are adiabatic, $\omega \ll 1$, one obtains the time-dependent bound-state energy levels

$$E^2 = \cos^2\lambda + e^{-4ka}\sin^2\lambda, \quad \lambda = U\cos(\omega t), \quad k^2 = 1 - E^2. \tag{12}$$

If U exceeds the critical value $U_1 = \pi/2$, the first (positive energy) bound state crosses $E=0$. If the δ-potentials are well separated, $a \gg 1$, the time-dependent energy levels are as in Fig. 4. Here E_+ and E_- denote the bound state levels in the right and left δ-function potential, respectively. The probability of nucleation of electron–positron pairs is determined by the transition between the E_- and E_+ branches. For U slightly above but not very close to $U_1 = \pi/2$, one obtains a result similar to that for the Landau–Zener tunneling problem,[16] with pair-creation probability per cycle

$$2P(1-P), \quad P = \exp\left(-\frac{2\sqrt{\pi}\mathscr{T}^2}{\omega\sqrt{U-U_1}}\right), \tag{13}$$

$$\mathscr{T}^2 = e^{-4a} \ll U - U_1. \tag{14}$$

If ω is large enough, the transition between E_- and E_+ states is given by the matrix element $\mathscr{T}$, while for small ω the process is determined by Zener tunneling through the gap $2\mathscr{T}$ between repelling levels.

A similar effect in nuclear physics would correspond to a situation, different from that suggested by Gershtein and Zel'dovich. In their case positron production is possible during collision of two heavy bare nuclei with total charge Z greater that the supercritical Z_c at which the electron bound state with energy $E=-M$ appears. This would correspond to critical strength $U_2 = \pi$ of the δ-function potential. In our case the critical strength is $U_1 = \pi/2$. This means that we need considerably less total charge Z, at which the negative energy bound state for an electron appears, $E_+ < 0$. But in addition nearby one should have a similar hypothetical collision of the anti-nuclei, which produces the potential of the opposite sign. If the latter contains the bound state with $E_- = E_+$, an electron occupying this bound state can tunnel to the bound state of the positively charged nucleus. As a result, an electron–positron pair will appear after such a collision.

5. DISCUSSION

According to the present scenario, the observed critical velocity for pair nucleation by a vibrating wire, $v_0^* \approx 0.25 v_L$,[5] is determined by bound states near the surface of the wire, and thus by suppression of the parallel gap at the surface of the wire in Eq. (7). This yields an experimental estimate for the suppressed gap, $\Delta_\parallel(r=R) \approx 0.5\Delta_0$, which is comparable to the theoretical estimate $\Delta_\parallel(r=R) \approx 0.4\Delta_0$.[11] This consistency provides experimental evidence for a modified Zel'dovich pair creation mechanism in a strong field, in which particles can be created by a subcritical electric potential because of level crossing.

Other objects whose motion can be used to simulate particle production from the vacuum are topological objects, vortices and domain walls. For a discussion of the production of momentum from the vacuum by a moving vortex, due to the axial anomaly phenomenon, see Ref. 17. Quasiparticle production by a moving soliton in superfluid ^{3}He-A due to the combined effect of Schwinger pair production, the event horizon, and the ergoregion, is discussed in Ref. 18.

One of us (A. C.) wishes to thank the Low Temperature Laboratory of Helsinki University of Technology for hospitality, and the EU Training and Mobility of Research Program CHGECT94-0069 for its support.

[1] I. Pomeranchuk and Ya. Smorodinsky, Zh. Fiz. USSR **9**, 97 (1945).
[2] S. S. Gershtein and Ya. B. Zel'dovich, Zh. Eksp. Teor. Fiz. **57**, 674 (1969) [Sov. Phys. JETP **30**, 358 (1970)].
[3] W. Greiner, B. Müller, and J. Rafelski, *Quantum Electrodynamics of Strong Fields*, Springer-Verlag, Berlin (1985), Sec. 3.2.
[4] W. T. Zaumen, Nature (London) **247**, 530 (1974).
[5] C. A. M. Castelijns, K. F. Coates, A. M. Guenault, S. G. Mussett, and G. R. Pickett, Phys. Rev. Lett. **56**, 69 (1986).
[6] J. P. Carney, A. M. Guenault, G. R. Pickett, and G. F. Spencer, Phys. Rev. Lett. **62**, 3042 (1989).
[7] C. J. Lambert, Physica B **165-166**, 653 (1990).
[8] P. Jetzer, P. Liljenberg, and B.-S. Skagerstam, Astropart. Phys. **1**, 429 (1993).

[9] D. Vollhardt and P. Wölfle, *The superfluid phases of helium 3*, Francis & Taylor, London (1990).
[10] G. E. Volovik, *Exotic properties of superfluid* ^{3}He, World Scientific, Singapore, New Jersey, London, Hong Kong (1992).
[11] N. B. Kopnin, P. I. Soininen, and M. M. Salomaa, J. Low Temp. Phys. **85**, 267 (1991).
[12] I. A. Privorotskii, J. Low Temp. Phys. **26**, 379 (1977).
[13] A. Calogeracos, N. Dombey, and K. Imagawa, Yad. Fiz. **59**, 1331 (1996) [Phys. At. Nucl. **59**, 1275 (1996)].
[14] G. E. Volovik, JETP Lett. **67**, 528 (1998).
[15] Ü. Parts, V. M. H. Ruutu, J. H. Koivuniemi, Yu. N. Bunkov, V. V. Dmitriev, M. Fogelström, M. Huebner, Y. Kondo, N. B. Kopnin, J. S. Korhonen, M. Krusius, O. V. Lounasmaa, P. I. Soininen, and G. E. Volovik, Europhys. Lett. **31**, 449 (1995).
[16] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press (1997), §90.
[17] G. E. Volovik, Physica B **255**, 86 (1988)..
[18] T. A. Jacobson and G. E. Volovik, Phys. Rev. D **58**, 064021 (1998).

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

Analysis of the effect of oxygen doping and pressure on superconducting transition temperature in metal oxides

A. A. Kosov*)

Mari State University, 424001 Yoshkar-Ola, Mari El, Russia

R. I. Boughton†)

Bowling Green State University, Bowling Green, 43402 Ohio, USA

(Submitted 28 January 1998)
Zh. Éksp. Teor. Fiz. **115**, 80–88 (January 1999)

The two-orbital Hubbard model is used to obtain formulas for the fermion excitation spectrum in the energy bands hybridized by the Anderson interaction. An analysis of lower part of the energy spectrum leads to a formula for the superconducting transition temperature T_c associated with the pairing of quasiparticles in one of the correlated bands. The dependence of T_c on pressure is analyzed, and the individual influence of carrier density enhancement and interaction strength is obtained as a function of oxygen concentration. The experimental discrimination made by Honma *et al.* [Solid State Commun **98**, 395 (1996)] in $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_{7-\delta}$ by separating out the contributions due to carrier density and pairing strength can be reproduced quantitatively, and perhaps with further refinement, so can the carrier concentration. Although the prediction of the absolute value of the transition temperature using the present model is not accurate, it is clear that it furnishes a reasonably accurate description of the change in transition temperature with pressure. The component contributions due to the change in carrier concentration and due to the change in interaction strength as a function of oxygen concentration are also in reasonable agreement with the experimental results. © *1999 American Institute of Physics.* [S1063-7761(99)00801-X]

1. INTRODUCTION

The investigation of the mechanism of superconductivity in cuprate superconductors is related to the effects of oxygen doping and pressure on the superconducting transition temperature (T_c) of these substances. One of the advantages of using high-pressure techniques is the ability to change atomic distances without substitution of components, which often causes some side effects.[1,2] At the present time several review articles concerning high-pressure work in cuprate superconductors have been published. According to the experimental results and the conclusions of Shafer *et al.*,[3] and Kubo *et al.*,[4] the carrier concentration increases with increasing pressure in many high-T_c materials. This increase in carriers is considered to be due to charge transfer from a charge reservoir layer to the Cu–O plane. The evidence for the dependence of the carrier concentration on pressure comes from measurements of the Hall number $1/eR_H$ and the thermoelectric power under high pressure. Generally, T_c initially increases with increasing $1/eR_H$, but decreases when $1/eR_H$ exceeds a specific value. However, the relation between T_c and $1/eR_H$ in high-pressure experiments varies among different cuprate superconductors. For instance, in La–Sr–Cu–O ceramics, T_c increases with increasing pressure, but no change in $1/eR_H$ is observed. On the other hand, in the case of Y–Ba–Cu–O, a variation of T_c with $1/eR_H$ has been observed by several authors in high-pressure experiments. It was proposed[5] that the change in T_c due to pressure ΔT_c should be expressed as the sum of two terms, $(\Delta T_c)_c$ and $(\Delta T_c)_p$, where $(\Delta T_c)_c$ is the change in T_c due to pressure-induced changes in the carrier density, and $(\Delta T_c)_p$ is the change in T_c due to pressure-enhanced electron pairing (e.g., change in the electron-phonon coupling strength, or in the exchange coupling constant). Honma *et al.*,[6] investigated the dependence of T_c on the Hall number by changing the oxygen content and the pressure in $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$. They determined that the contribution of $(\Delta T_c)_c$ to ΔT_c increases with decreasing oxygen content.

In this work we use the idea that $\Delta T_c=(\Delta T_c)_c+(\Delta T_c)_p$ and apply the Anderson-Hubbard two orbital model[7] to describe the experimental results of Ref. 6. In Sec. 2 we introduce the Hamiltonian of the problem, the Green's functions of the quasiparticles in correlated bands, and the equation for T_c. Section 3 is devoted to the calculation of the pressure effect on the superconducting transition temperature. By proposing a simple relation between the value of pressure P and the width of the correlated band W, we can obtain the dependence of T_c on pressure. In addition, a comparison with experimental results on the dependence of T_c, ΔT_c, $(\Delta T_c)_c$ and $(\Delta T_c)_p$ on the concentration of carriers in YCaBaCuO is made. Good agreement between the theoretical calculation of the dependence of (ΔT_c), and $(\Delta T_c)_c$ on pressure and the experimental results is found. It is concluded that the model under consideration is quite promising for studying the effects of oxygen doping and pressure on the superconducting transition in cuprate superconductors.

2. FORMULATION

One of the popular models used for describing a strongly correlated system is the Hubbard model.[8] Recently Kosov and Shilov[7,9] studied the superconducting transition and pressure effects by using a unified Hamiltonian containing operators of the Hubbard two-orbital model and the Anderson interaction. The interaction considerably enhances the applicability of the Hubbard model and allowed the authors to describe the interaction of non-localized and localized electrons by proceeding from the mixing of their one-particle states.

The model in Refs. 7 and 8 is based on the following Hamiltonian:

$$H=H_0+H_{\rm int}=\sum_i H_{0i}+\sum_{ijs} t_{ij}c_{is}^+c_{is},$$

$$\begin{aligned}H_{0i}=&-\mu(n_{ia\uparrow}+n_{ia\downarrow}+n_{ic\uparrow}+n_{ic\downarrow})+E(n_{ia\uparrow}+n_{ia\downarrow})\\&-\mathscr{H}(n_{ia\uparrow}-n_{ia\downarrow}+n_{ic\uparrow}-n_{ic\downarrow})+In_{ia\uparrow}n_{ia\downarrow}U\\&\times(n_{ia\uparrow}+n_{ia\downarrow})(n_{ic\uparrow}+n_{ic\downarrow})+U_1n_{ic\uparrow}n_{ic\downarrow}\\&+V_0(a_{i\uparrow}^+c_{i\uparrow}+a_{i\downarrow}^+c_{i\downarrow}+\text{H.c.}),\end{aligned}\tag{1}$$

where c_{is}^+, c_{is} and a_{is}^+, a_{is} are field operators corresponding to free and localized electrons at the site i with spin projection s: $n_{ias}=a_{is}^+a_{is}$ and $n_{ics}=c_{is}^+c_{is}$ are the operators for the number of electrons: μ is the chemical potential; $\mathscr{H}$ is the applied magnetic field. E is the one-particle energy of the a-electrons; I, U, and U_1 are the energy parameters defining intra-atomic correlation; I is the Hubbard interaction between localized electrons; U is the interorbital Coulomb interaction of c- and a-electrons; U_1 is the repulsive interaction of c-electrons on one site; V_0 is the matrix element responsible for the hybridization of the c- and a-electronic states (Anderson's constant); and $H_{\rm int}$ describes the interstitial tunneling of c-electrons with transport integral t_{ij}.

We need to check our results according to separate condition: $E>0$ and $E<0$. For this reason, we use the intermediate symbols:

$$Q=\begin{cases}B, & E>0,\\ A, & E<0,\end{cases}\quad S=\begin{cases}D, & E>0,\\ C, & E<0.\end{cases}$$

This means, for example, that for value R_K^S we have

$$R_K^S=\begin{cases}R_K^D, & E>0,\\ R_K^C, & E<0.\end{cases}$$

We assume that the density of states in the dispersion region has a rectangular shape:[7–11]

$$\rho(\varepsilon)=(1/2W)\,\theta(W^2-\varepsilon^2),$$

where $2W$ is the width of the c-band.

Carrying out calculations similar to those in Refs. 7 and 11, we obtain the following expression for the chemical potential:

$$\mu=-\Delta/2-WB_+/2+2(n_K+n_S)W(R_K^S)^2-P_-/2,\tag{2}$$

$$P_\pm=(B_+^2W^2\pm2\Delta_1B_-W+\Delta_1^2)^{1/2},$$

$$B_\pm=\cos^2\alpha(n_0+n_Q)\pm(R_D^K)^2(n_K+n_S).$$

Here n_0, n_Q, n_K, and n_S are the population densities of the energy states under investigation, which satisfy the condition $n_0+n_Q+n_Kn_S=1$;

$$\begin{aligned}&R_C^K=R_A^K=\cos\alpha(C_1A_3-A_1C_3)/\sqrt{2}\\&\qquad+\sin\alpha(C_1A_2-C_2A_1),\\&A=(1+A_{21}^2+A_{31}^2)^{-1/2},\quad A_2=A_{21}A_1,\quad A_3=A_{31}A_1,\\&A_{21}=(E_L-2E-I)^2/2V_0^2,\quad A_{31}=(E_L-2E-I)^2/\\&\qquad(E_L-U_1)^2,\\&B_2=(1+B_{12}^2+B_{32}^2)^{-1/2},\quad B_1=B_{12}B_2,\quad B_3=B_{32}B_2,\\&B_{12}=\frac{(E_M+2E-U)(E_M-U_1)-2V_0^2}{\sqrt{2}V_0(E_M-U_1)},\\&B_{32}=\frac{\sqrt{2}V_0}{E_M-U_1},\\&C_3=(1+C_{12}^2+C_{23}^2)^{-1/2},\quad C_1=C_{13}C_3,\quad C_2=C_{23}C_3,\\&C_{13}=\frac{(E_K-E-U)(E_K-U_1)-2V_0^2}{2V_0^2},\quad C_{23}=\frac{E_K-U_1}{\sqrt{2}V_0},\\&E_K<E_M<E_L.\end{aligned}$$

The energy of the two-particle states E_K, E_L, and E_M can be obtained by using the cubic equation

$$\begin{aligned}&X^3+AX^2+BX+C=0,\\&A=-(I+U+U_1+3E),\quad B=(I+2E)(I+U)\\&\qquad+U_1(3E+U+I)-4(V_0)^2,\\&C=2(V_0)^2(I+2E+U_1)-U_1(E+U)(2E+I).\end{aligned}\tag{3}$$

The roots $X=\{X_m\}$ of Eq. (3) define the energy $E_{K,L,M}$: $X_m=2\mu+E_m$, $m=K,L,M$.

The chemical potential μ is determined by the concentration of electrons in the dispersed correlated band, which equals $n_c=(R_K^S)^2(n_K+n_S)$. The dependence of n_c and μ on V_0 shows different behaviors in the cases $E>0$ and $E<0$. An increase in the hybridization parameter V_0 leads to a decrease in n_c for $E>0$, and to an increase for $E<0$.

A transition to the Hubbard operators allows the use of the Green's temperature function technique to take the interstitial jump term into account in order to study the superconducting properties of the model. An analysis of the lower part of the energy spectrum leads to the following formula for the superconducting transition temperature associated with the pairing of quasiparticles in one of the correlated bands:

$$\frac{T_c}{2W}=0.57\left[-\frac{\xi_{10}(-W)\xi_{10}(W)}{W^2}\right]^{1/2}\exp\left[-\frac{1}{A(n,t_0)}\right],\tag{4}$$

$$\xi_{10}(-W)=-2(R_K^S)^2(n_K+n_S),$$

$$\xi_{10}(W)=-\mu+(B_+W-P_+-\Delta)/2,$$

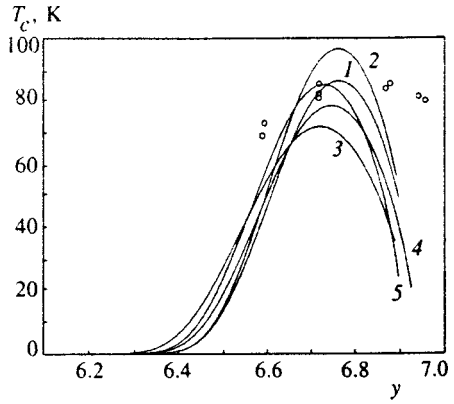


FIG. 1. Dependence of the function T_c on the oxygen content y for $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ obtained for the parameters $k_a=0.00024\,\mathrm{GPa}^{-1}$, $P=2$ GPa.

$$A(n,t_0)=\Gamma(n,t_0)/\Lambda(n,t_0),$$

$$\Gamma(n,t_0)=t_0[\cos^2\alpha(n_0+n_Q)(E_S-E_K)(E_S+B_-t_0)+(R_K^S)^2(n_K+n_S)E_Q(E_Q-E_K+B_-t_0)],$$

$$\Lambda(n,t_0)=W[B_+(2\mu+\Delta)+B_-\Delta_1](2\mu+\Delta-B_+t_0).$$

The quantity $A(n,t_0)$ plays the role of the quasiparticle scattering amplitude with different spin orientation. The attraction between quasiparticles in a correlated band takes place under the conditions:

$$A(n,t_0)>0, \quad -W<t_0<W, \quad \xi_{10}(W)\geqslant 0. \tag{5}$$

The conditions (5) can be used to determine the concentration $n_{S1}<n<n_{S2}$ for which $T_c\neq 0$ holds. Solving the equation $\xi_{10}(W)=0$ gives the following value for n_{S2}:

$$n_{S2}=\frac{2\cos^2\alpha}{\cos^2\alpha+(R_K^S)^2}. \tag{5a}$$

The condition $A(n,t_0)=0$ gives the following result for n_{S1}:

$$n_{S1}=\frac{2\cos^2\alpha}{\cos^2\alpha+2(R_K^S)^2}. \tag{5b}$$

In particular, if we put $V_0=0$ and $E>0$ in (5a) and (5b), we obtain $n_{S1}=2/3$ and $n_{S2}=1$. This result has been obtained by Zaitsev and Ivanov[12] in the framework of the one-orbital Hubbard model (the so-called «kinematic mechanism of superconductivity»).

3. PRESSURE EFFECTS

In applying the two-orbital Anderson-Hubbard model to describe the pressure dependence of T_c, we choose to examine the energy parameters, which more sensitively depend on the value of pressure. Since U, I, U_1, V_0, and E are on-site properties, their pressure dependence can be neglected. The transport integral t_p depends on the spatial distribution of atoms and is changed by applied pressure. Let us consider the region where W depends linearly on t_p ($W\propto t_p$). According to Marsiglo and Hirsh,[13] the transport integral in cuprate superconductors can be expressed trough the lattice parameters of the CuO_2 planes a, b, and c by the formulas

$$t_\parallel=\frac{\hbar^2}{2m_\parallel a}, \quad t_\perp=\frac{\hbar^2}{2m_\perp c},$$

where $m_\parallel$ and $m_\perp$ are the respective effective masses.

Neglecting any pressure dependence of the effective masses, we estimate the magnitude of dW/dP:

$$\frac{dW}{dP}\propto\frac{dt_p}{dP}\sim -2W\frac{d\ln a}{dP}=2Wk_a. \tag{6}$$

Here k_a, k_b, and k_c are the compressibility components along each crystallographic direction, defined by

$$k_a=-\frac{d\ln a}{dP}, \quad k_b=-\frac{d\ln b}{dP}, \quad k_c=-\frac{d\ln c}{dP}.$$

In order to simplify the numerical estimation of the results, we consider the case $k_a=k_b=k_c$. Using the relations in the equation (6), we obtain the following expression for the dependence of the bandwidth on pressure

$$W(P)=W(P=0)\exp(2k_aP). \tag{7}$$

Formulas (4) and (7) allow us to express T_c as a function of pressure. Using expression (6) we can obtain formulas for dT_c/dP and $d\ln T_c/dP$:

$$\frac{dT_c}{dP}=T_c\frac{d\ln T_c}{dP}\sim 2Wk_a\frac{dT_c}{dW}, \tag{8}$$

$$\frac{dT_c}{dW}=\frac{T_c}{2W}\left\{1+2W\times\left[B_-+\frac{B_+^2(P_+-P_-)W-\Delta_1B_-(P_++P_-)}{2P_+P_-}\right]+\frac{2W}{A^2(n,t_0)}\frac{dA(n,t_0)}{dW}\right\},$$

$$\frac{dA(n,t_0)}{dW}=-\frac{A(n,t_0)}{W}+\frac{1}{\Lambda(n,t_0)}\frac{d\mu}{dW}\{[2(2\mu+\Delta)-B_+t_0][\cos^2\alpha(n_0+n_Q)(E_s-E_K)(E_S+2B_-t_0)+(R_K^S)^2(n_K+n_S)E_Q(E_Q-E_K+2B_-t_0)+A(n,t_0)B_+W(B_+(2\mu+\Delta)+\Delta_1B_-)]-2WA(n,t_0)[2B_+(2\mu+\Delta)+\Delta_1B_--B_+^2t_0]\},$$

$$\frac{d\mu}{dW}=2(R_K^S)^2(n_K+n_S)-\frac{B_+}{2}-\frac{WB_+^2-\Delta_1B_-}{2P_-}.$$

By means of formula (8) we obtain a theoretical expression for $(\Delta T_c)_p$:

$$(\Delta T_c)_p=\frac{dT_c}{dP}\Delta P=\frac{dT_c}{dW}\Delta W. \tag{9}$$

To compare our results with the experimental data of Ref. 6 we obtain the value $(\Delta T_c)_c$ from formula (4):

TABLE I. Parameter values used.

Curve	$-E/W$	I/W	U_1/W	U/W	V_0/W
1	−1.9	6.0	3.5	2.4	1.8
2	−1.9	6.0	3.5	2.4	2.0
3	−1.7	6.0	3.5	2.4	1.8
4	−1.8	5.0	2.5	2.8	1.8
5	−1.9	6.0	2.4	3.5	2.0

$$(\Delta T_c)_c = \frac{dT_c}{dn}\Delta n = \frac{dT_c}{dn}\frac{dn}{dP}\Delta P = \frac{dT_c}{dn}\frac{dn}{dW}\Delta W. \qquad (10)$$

4. DISCUSSION

We have chosen to compare the theoretical results presented in the previous section with the recent experimental results of Honma *et al.*,[6] who studied the pressure dependence and the effect of oxygen doping on the transition temperature of the $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ (YCBCO) system. These authors systematically investigated the critical temperature dependence and the dependence of the Hall number $1/eR_H$, on pressure, in order to be able to distinguish between those effects which result from changes in carrier density $(\Delta T_c)_c$, and changes in the coupling strength (electron pairing) $(\Delta T_c)_p$.

The first comparison we shall examine is the dependence of the critical temperature T_c, on the oxygen content *y*. The theoretical results for a pressure of 2 GPa and a compressibility k_a of 0.0024 GPa^{-1} are displayed in Fig. 1, where five curves are plotted, each with a variation in one or more of the following bandwidth-normalized parameters: the single-particle energy E/W, the interatomic correlation energies I/W, U_1/W, and U/W; and the hybridization energy V_0/W. The values of the parameters used in the various curves used to fit the data are listed in Table I below.

The data of Fig. 1 in the paper by Honma *et al.*, are represented by open circles. As can be seen from the figure, the general concave-downward form of the experimental results can be qualitatively reproduced by the theoretical curves with the same maximum value, and with little variation exhibited for different parameter combinations. The requisite width of the curve exhibited by the experimental data, however, is not readily attained with any reasonable variation in the parameters. The difficulty in fitting the theory to an absolute quantity like the critical temperature is not unexpected, since the absolute values of the potentials chosen are involved in making such a comparison.

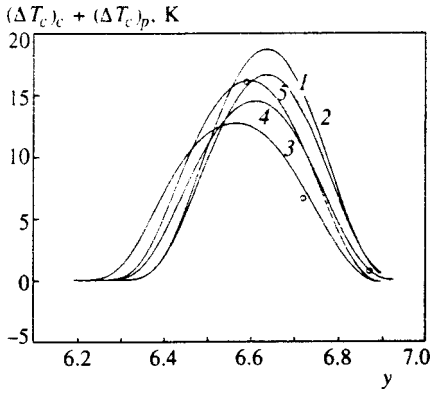


FIG. 3. Dependence of $(\Delta T_c)_p + (\Delta T_c)_c$ on the oxygen content *y* for $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ obtained for the parameters $k_a = 0.00024\ GPa^{-1}$, $P = 2$ GPa.

In Figs. 2a and 2b are plotted the critical temperature change at a pressure of 2 GPa and a compressibility k_a of 0.0024 GPa^{-1}, due to changes in carrier density and pairing strength $(\Delta T_c)_c$ and $(\Delta T_c)_p$, respectively, as a function of oxygen content *y*. Here, the agreement with the experimental results of Honma *et al.* (open circles) is fairly good. Once again, the theoretical curves exhibit a concave-downward behavior over the range of *y* that was measured. Of the parameter sets chosen for illustration, it appears that the parameters corresponding to curve *5* ($-E/W = -1.9$, $I/W = 6.0$, $U_1/W = 2.4$, $U/W = 3.5$, and $V_0/W = 2.0$) give the closest fit to the experimental data in both cases. Figure 3 illustrates the sum of the two effects to give the total change in temperature, $(\Delta T_c)_c + (\Delta T_c)_p = \Delta T_c$ plotted vs. oxygen content *y*, along with the experimental data. Once again the theoretical treatment yields a set of concave-downward curves which qualitatively approximate the data with the best fit to the experimental data given by curve *5*. Overall, the theoretical model appears to quite accurately account for the observed experimental variation with oxygen content. In contrast with the first plot, these comparisons involve the change in critical temperature with pressure and so can be more realistically accounted for by the theoretical model.

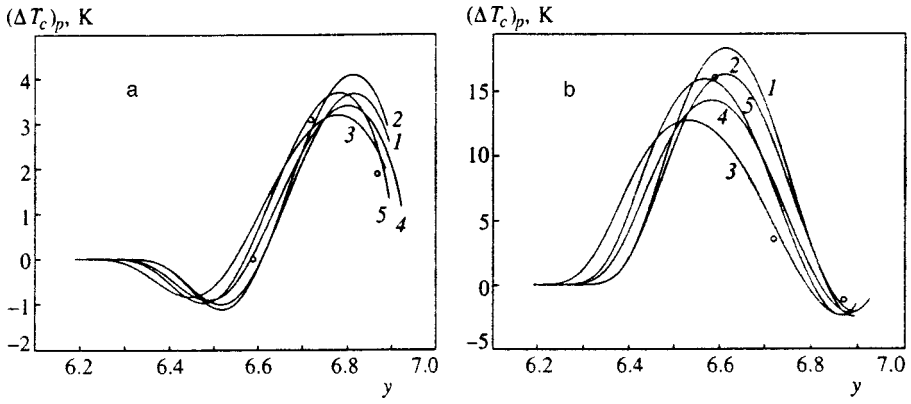


FIG. 2. Dependences of $(\Delta T_c)_p$ (a) and $(\Delta T_c)_c$ (b) on the oxygen content *y* for $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ obtained for the parameters $k_a = 0.00024\ GPa^{-1}$, $P = 2$ GPa.

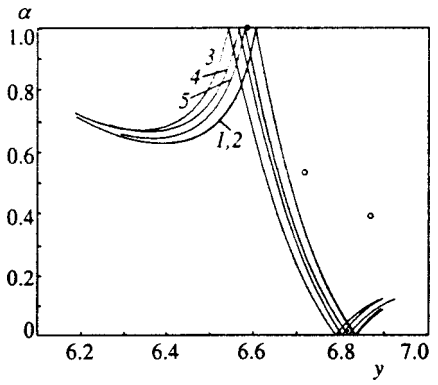


FIG. 4. Dependence of the ratio $\alpha=|(\Delta T_c)_c|/(|(\Delta T_c)_p|+|(\Delta T_c)_c|)$ on the oxygen content y for $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ obtained for the parameters $k_a=0.00024\ \mathrm{GPa}^{-1}$, $P=2$ GPa.

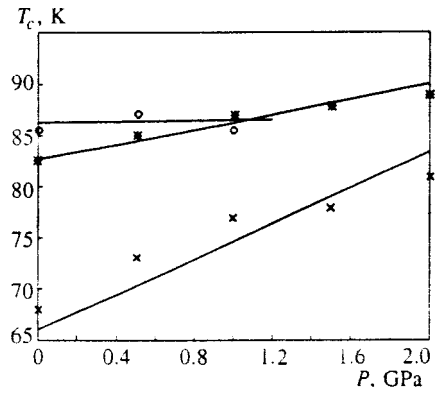


FIG. 5. T_c as a function of pressure in $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ ($\times$—$y=6.59$, $*$—$y=6.72$, $\circ$—$y=6.87$—experimental data[6]) obtained for the parameter $k_a=0.0024\ \mathrm{GPa}^{-1}$ (used for curve *5*).

For the Hall number $1/eR_H$, which should scale with the ratio of the carrier density enhancement of the critical temperature $(\Delta T_c)_c$ to the sum of the absolute values of the individual critical temperature enhancements $|(\Delta T_c)_c|+|(\Delta T_c)_p|$, the plot vs. oxygen concentration y shown in Fig. 4 illustrates the comparison between theory and experimental data of Honma *et al.*, who labelled this ratio α. Here the fit is not good with any of the parameter variations that were tried. Although the experimental data fall on curve *4* near $\Delta T_c/(|(\Delta T_c)_c|+|(\Delta T_c)_p|)=1$, where the temperature change is almost entirely due to change in carrier concentration, the slopes are clearly not in agreement. We believe this discrepancy arises from the reduced accuracy in evaluating this fraction when the Hall coefficient changes sign, which it does in the case where $(y,\alpha)=(6.87,0.39)$. The theoretical curve nevertheless gives a reasonable qualitative description of the variation in this parameter, as a decreasing function of oxygen content. In order to determine how dT_c/dn varies with oxygen concentration y, we have used a third-order polynomial fit of the form: $dn(y)=a_3y^3+a_2y^2+a_1y+a_0$. The fitting parameters are obtained by using the data of Honma *et al.*, at $y=6.59$, 6.72 and 6.87, respectively, and by setting $dn(7.0)=0$. The values of the fitting parameters are $a_0=-461.1850$, $a_1=141.5920$, $a_2=-11.6100$, $a_3=0.1135$. This relation provides a reasonable idea about the strength of the effect of charge carrier density upon ΔT_c.

Finally, in Fig. 5, we plot the total change in the transition temperature vs. pressure as obtained from the theoretical model. The experimental results are shown as data points according to the legend. It is apparent that the quality of the agreement is good, with no more than 10% discrepancy. The rate of increase in T_c with pressure is faithfully reproduced for all three fractional oxygen contents studied by Honma *et al.* It is important to note the nearly zero slope at the highest concentration, $y=6.87$.

5. CONCLUSIONS

In this paper we have demonstrated that the application of the Hubbard model provides a good basis for describing the observed variation of the pressure dependence of the critical temperature on oxygen concentration in the $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_y$ (YCBCO) system. It appears that the experimental discrimination made by Honma *et al.*, in separating out the contributions due to carrier density and pairing strength can be reproduced quantitatively, and perhaps with further refinement, as can the carrier concentration. Although the prediction of the absolute value of the transition temperature is not accurate using the present model, it is clear that it furnishes a reasonably accurate description of the change in transition temperature with pressure. The component contributions due to change in carrier concentration and due to change in interaction strength as a function of oxygen concentration are also in reasonable agreement with the experimental results.

*)E-mail: kosov@margu.mari.ru
†)E-mail: boughton@bgnet.bgsu.edu

[1] J. S. Schilling and S. Klotz, *Physical Properties of High Temperature Superconductors*, Vol. III, ed. by D. M. Ginsberg, World Scientific, Singapore (1992).
[2] H. Takahashi and N. Mori, *Studies of High Temperature Superconductors*, Vol. 16, ed. by Anant Narlikar, Nova Science Publishers (1996).
[3] M. W. Shafer, T. Penny, B. L. Olson, R. L. Greene, and R. H. Koch, Phys. Rev. B **39**, 2914 (1988).
[4] Y. Kubo, Y. Shimakawa, T. Monako, and H. Igarashi, Phys. Rev. B **43**, 7875 (1991).
[5] Y. Iye, J. Phys. Chem. Solids **53**, 1561 (1992).
[6] T. Honma, K. Yamaya, N. Mori, and M. Tanimoto, Solid State Commun. **98**, 395 (1996).
[7] A. A. Kosov and V. E. Shilov, Fiz. Nizk. Temp. **22**, 1032 (1996) [Low Temp. Phys. **22**, 787 (1996)].
[8] J. Zielinski, M. Matlak, and P. Entel, Phys. Lett. A **136**, 441 (1989).
[9] A. A. Kosov, Fiz. Nizk. Temp. **24**(3), (1998) [Low Temp. Phys. **24**, (3) (1998)].
[10] Yu. A. Izyumov, M. I. Katsnel'son, and Yu. N. Skryabin, *Magnetism of Collectivized Electrons*, (in Russian), Nauka, Moscow (1994).
[11] P. B. Zyubin, V. A. Ivanov, and E. A. Ugolkova, Teor. Mat. Fiz. **101**, 304 (1994).
[12] R. O. Zaitsev and V. A. Ivanov, Fiz. Tverd. Tela **29**, 2554 (1987) [Sov. Phys. Solid State **29**, 1475 (1987)].
[13] F. Marsiglo and J. E. Hirsh, Phys. Rev. B **41**, 6435 (1990).

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

Resonant acceptor states and terahertz stimulated emission of uniaxially strained germanium

I. V. Altukhov, M. S. Kagan,*) K. A. Korolev, V. P. Sinis, and E. G. Chirkova

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

M. A. Odnoblyudov and I. N. Yassievich

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 9 February 1998)
Zh. Éksp. Teor. Fiz. **115**, 89–100 (January 1999)

The stimulated emission spectrum of uniaxially strained *p*-Ge is presented. The energy spectrum of the states of a shallow acceptor in Ge under uniaxial compression is calculated. The threshold pressure at which the acceptor state split off from the ground state becomes resonant is found. The pressure dependence of the width of this resonant level is calculated. The stimulated emission lines are identified. In particular, it is shown that the principal emission peak corresponds to the transition of holes from the resonant $1s$ $(1s_r)$ state to the local $p_{\pm 1}$ state. The probabilities of optical transitions are calculated. A mechanism of population inversion due to the intense resonant scattering of hot holes with an energy corresponding to the position of the $1s_r$ level is proposed.
[S1063-7761(99)00901-4]

1. INTRODUCTION

Solid-state sources of electromagnetic radiation in the terahertz range, which corresponds to wavelengths of 10–1000 μm, have been undergoing rapid development. The first pulsed semiconductor lasers with wavelengths in the range 100–300 μm were developed in the mid-1980s using emission from *p*-Ge under the simultaneous action of strong electric and magnetic fields at liquid-helium temperatures (see, for example, Ref. 1 and the work cited there). It was shown that the stimulated emission is caused by an inverted hole population in momentum space. A cascade laser based on intraband transitions in narrow quantum wells was recently developed and can, in principle, operate at wavelengths from the mid-IR range to 100 μm.[2,3]

The stimulated emission of *p*-Ge subjected to uniaxial compression was observed in a strong electric field in Ref. 4. It was suggested that such stimulated emission is due to the appearance of resonant states as a result of the strain-induced splitting of the fourfold degenerate acceptor level.[5] This system is of unquestionable interest for developing a new type of lasers for the terahertz range. Uniaxial strain appears, for example, in heterostructures based on semiconductors with lattice parameter mismatch, particularly in Ge–Si structures.

This paper compares the stimulated emission spectra of uniaxially strained *p*-Ge with the calculated energy spectrum of the strain-split levels of a shallow acceptor. Analysis of the emission spectrum reveals that the stimulated emission is associated with the appearance of resonant acceptor states. The possibility of adjusting the emission energy in the range from 10 to 42 meV by varying the pressure is demonstrated. A mechanism of population inversion as a consequence of the accumulation of holes near the lower resonant state due to the intense resonant scattering of free holes at that energy is proposed. The probabilities of intra-impurity optical transitions and transitions from the continuous spectrum to local states are calculated. It is shown that the principal peak in the stimulated emission spectrum corresponds to a radiative transition from the lowest resonant $1s$ $(1s_r)$ state to the first excited local $2p_{\pm 1}$ state. The probability of this transition is calculated; in particular, when the pressure along the [111] axis is $P=6.85$ kbar, for which the transition energy should be 24 meV, the radiative transition time is $\tau=2.2\times10^{-6}$ s.

2. EXPERIMENT

Gallium-doped *p*-Ge crystals with a Ga concentration from 3×10^{13} to $10^{14}\,\mathrm{cm}^{-3}$ were investigated at liquid-helium temperatures. Match-shaped samples having a length of 6–10 mm and a cross-sectional area of 0.5–1 mm^2 were cut in the [111] or [100] crystallographic direction. A pressure *P* was applied along the sample in either of these directions. Voltage pulses of duration 0.2–1 μs, which created an electric field *E* parallel to the pressure, were applied to contacts deposited on a lateral (long) face of the sample. The distance between the contacts was 4–9 mm.

The terahertz emission of the samples was detected by a cooled Ga-doped Ge photodetector with a sensitivity band $h\nu>10$ meV. Figure 1a shows the pressure dependence of the photodetector signal, which is proportional to the integrated luminescence intensity in the sensitivity band of the detector, for $\mathbf{E}\|\mathbf{P}\|[100]$ at various voltages. For samples such that the long faces deviated from parallel by less than $4'$ the intensity increased abruptly at a certain threshold pressure P_c. The intensity jump was also accompanied by a current jump (by up to ten fold).

The high-intensity emission observed is stimulated, as is shown by the following facts: 1) there is a threshold pres-

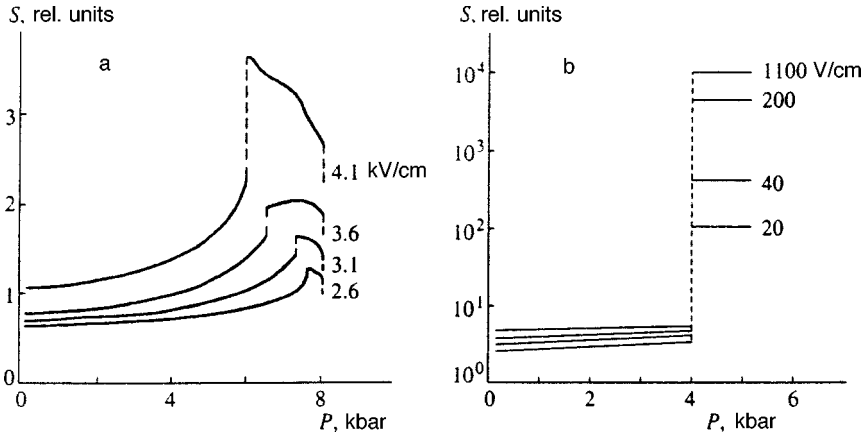


FIG. 1. Intensity S of far-IR luminescence as a function of the pressure P for various values of the mean field U/L. Here U is the applied voltage and L is the sample length; $\mathbf{E}\|\mathbf{P}\|[100]$. The nonparallelism of the lateral (long) faces of the sample equals about $4'$ (a) and less than $20''$ (b).

sure; 2) a resonator, which is formed in our case by parallel faces of the sample owing to total internal reflection (see, for example, Ref. 6), is needed for the appearance of the intensity jump. A simple experiment showed that this is, in fact, the case. Rough grinding of one of the lateral faces of the sample removed both the intensity jump and the current jump. Repeated polishing and etching of that face permitted restoration of the resonator, and the stimulated emission reappeared.[4]

The current jump appearing simultaneously with the intensity jump, as well as the voltage dependence of P_c, can be explained in the following manner. In uniaxially strained Ge the redistribution of hot holes between different branches of the Ge valence band with different effective masses produces a negative differential conductivity, which leads to the formation of electric domains.[7–9] In this case the distribution of the electric field along the sample is very nonuniform and consists of strong- and weak-field regions.

As the applied voltage is increased the length of the strong-field domain increases, but the electric field intensities inside and outside the domain scarcely depend on the applied voltage;[10] therefore, the current–voltage characteristic of the sample has a current-saturation segment. Stimulated emission appears at a certain critical domain length. As was shown in Ref. 10, the domain length increases with both increasing voltage and increasing pressure at a fixed voltage. Therefore, the smaller is the applied voltage, the greater is the pressure which must be applied so that the domain length would reach the critical value (see Fig. 1). When the stimulated emission intensity is sufficiently high, the domain disappears, the field distribution in the sample becomes homogeneous, and the current abruptly increases to the value corresponding to the homogeneous field.[10]

Stimulated emission could be obtained at a lower pressure by improving the resonator. The best result, which was obtained for a sample with the faces about $20''$ out of parallel, is shown in Fig. 1b. The intensity jump was observed at $P \approx 4$ kbar and at a significantly smaller voltage (below the domain formation threshold), beginning at the impurity breakdown voltage. We note that in this case a domain could not form at any applied voltage due to suppression of the negative differential conductivity by the high-intensity stimulated emission. At low voltages [due to the small amount of power dissipated in the sample, which was equal to ~ 100 mW (Ref. 10)] stimulated emission could be excited in a continuous regime.

The stimulated emission spectrum measured using a grating monochromator for $P = 6.85$ kbar is shown in Fig. 2. In these measurements a cryostat with the sample was placed at the entrance, and a cryostat with the photodetector was placed at the exit from the monochromator. The spectrum consists of several peaks. The energy at which the principal maximum occurs varies from 21.2 meV at $P = 6.85$ kbar to 40.2 meV at $P = 11.5$ kbar (see the points in Fig. 8). The width of the maxima is fairly large and amounts to 0.2–0.5 meV for different peaks.

The maxima in the spectra measured in greater detail exhibit a mode structure which results from the resonator modes. Figure 3 shows the principal stimulated emission maximum at $P = 7.1$ kbar for a sample with a 1×1 mm^2 cross section. The inset shows the optical path in the sample at resonance due to total internal reflection. It can be seen that the distance between the lines in the spectrum (≈ 0.11 meV) coincides with the value found from the condition $K\lambda = nL$, where λ is the emission wavelength, n is the refractive index ($n = 4$ for Ge), L is the optical path length, and K is an integer. Thus, in our case, in analogy to Ref. 6, an optical resonator is formed as a result of total internal reflection from parallel longitudinal faces of the crystal.

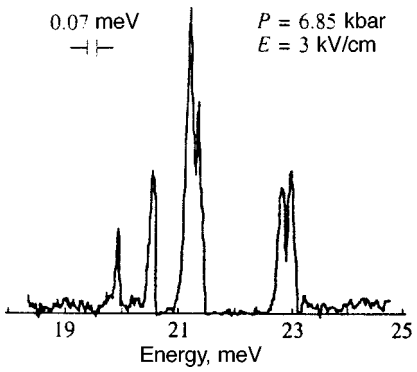


FIG. 2. Stimulated emission spectrum. $\mathbf{E}\|\mathbf{P}\|[111]$. The peak at 21.2 meV corresponds to the optical transition between the resonant $1s$ state and the $2p_{\pm 1}$ acceptor state.

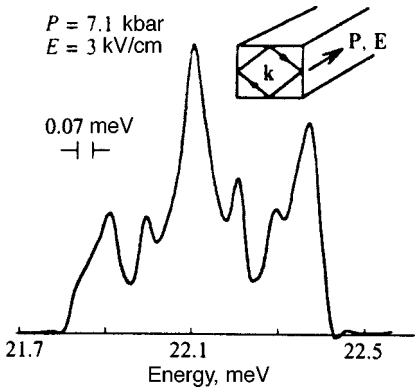


FIG. 3. Mode structure of the principal peak in the stimulated emission spectrum. The cross section of the sample measures 1×1 mm^2. The optical path in the sample is shown in the inset.

3. ANALYSIS OF THE EMISSION SPECTRUM

To analyze the emission spectrum we calculated the positions of the levels of a shallow acceptor in uniaxially compressed Ge. As is well known, uniaxial strain removes the degeneracy of the Ge valence band at $k=0$ and splits it into two subbands with the momentum projections $M=\pm3/2$ and $M=\pm1/2$ on the axis parallel to P (the z axis), which are separated by an energy gap proportional to the applied pressure. The degenerate acceptor ground state is similarly split into two states, whose energy difference also increases with pressure. Figure 4 schematically shows the structure of the Ge valence band and the positions of the ground state of a shallow acceptor and the state split off from it by uniaxial compression at various pressures. Above a certain pressure ($P\approx4$ kbar for $\mathbf{P}\|[111]$ and $P\approx3$ kbar for $\mathbf{P}\|[100]$) the split-off acceptor state is in the continuous spectrum and forms a resonant level (see the band diagrams for $P\geqslant4$ kbar in Fig. 4), while the ground state remains in the band gap. There should be two series of excited states belonging to split valence subbands in the band gap and in the continuum.

For the calculations we used the Luttinger Hamiltonian in the spherical approximation.[11,12] Diagonalization of this Hamiltonian for uniaxial strain gives a valence-band spectrum consisting of two subbands, which can be called the heavy-hole (ε_h) and light-hole (ε_l) subbands. Their extrema are separated by the energy gap

$$\varepsilon_{\rm def}=\frac{\hbar^2\zeta}{m_0}=bP, \tag{1}$$

where b is the deformation potential, ζ is the deformation parameter, and m_0 is the free electron mass. In germanium we have $b\approx4$ and 6 meV/kbar for compression along the [111] and [100] axes, respectively. The spectrum of the valence subbands in a strained crystal has the form

$$\varepsilon_{l,h}(k)=-\frac{\hbar^2}{2m_0}[-\gamma_1(k_x^2+k_y^2+k_z^2) \pm\sqrt{\zeta^2-2\gamma\zeta(2k_z^2-k_x^2-k_y^2)+4\gamma^2(k_x^2+k_y^2+k_z^2)^2}], \tag{2}$$

where $\gamma=(3\gamma_3+2\gamma_2)/5$, and γ_1, γ_2, and γ_3 are the Luttinger parameters: $\gamma_1=13.38$, $\gamma_2=4.24$, and $\gamma_3=5.69$.[13]

The positions of the levels of a shallow acceptor in uniaxially compressed Ge were calculated both within the zero-radius potential model and by the variational method for Coulomb centers in the large-strain limit, where only one subband can be taken into account in the treatment of each series of levels. The details of the calculation of the splitting of the Γ_8 acceptor state in a strained semiconductor within the zero-radius potential model and of the lifetime of the resonant states appearing in such a system were given in Ref. 14. Here we present only the results and some brief explanations.

The acceptor ground state in the unstrained semiconductor is fourfold degenerate with respect to the projection of the total momentum onto the z axis. According to the zero-radius potential method, the wave function of the impurity state is constructed as the Green's function of the Luttinger Hamiltonian. The energies of the impurity levels appear in the expressions for the wave functions as parameters for an assigned value of the binding energy in the unstrained semiconductor. Above a certain pressure the energy of the impurity level with $M=\pm3/2$ has a complex value $\varepsilon^{\pm3/2}-i\Gamma/2$, which corresponds to the passage of this level into the con-

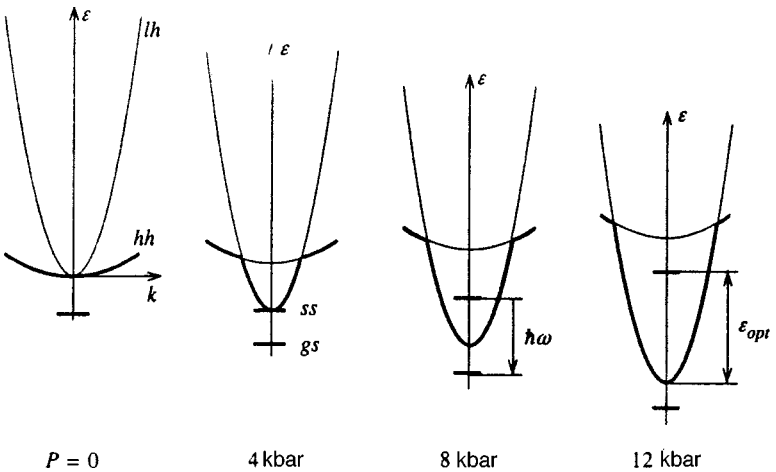


FIG. 4. Structure of the valence band and position of the acceptor levels for various values of P. The heavy-hole and light-hole bands are denoted by hh and lh, the acceptor ground and split-off states are denoted by gs and ss, and the energy of the optical phonon is denoted by $\varepsilon_{\rm opt}$. For convenience the hole bands are shown as in the electron case.

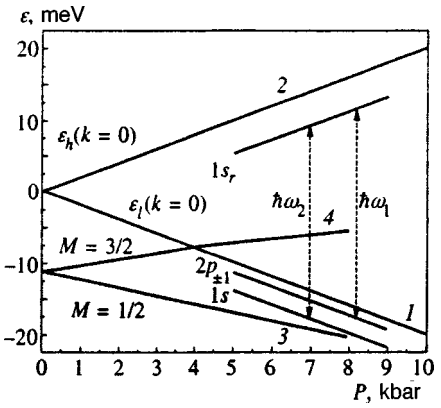


FIG. 5. Positions of the tops of the light-hole and heavy-hole subbands (*1, 2*) and binding energies of the $1s$ and $1s_r$ impurity levels calculated by the zero-radius potential method (*3, 4*), together with the $1s$, $2p_1$, and $1s_r$ levels obtained from a variational calculation, as functions of pressure. $\mathbf{P}\|[111]$.

tinuous spectrum of the light-hole subband and the appearance of resonant states with lifetime $\tau=\hbar/\Gamma$. The calculation for Ga impurities in Ge (the binding energy is 11.3 meV) in the zero-radius potential model gives the threshold value $\varepsilon_{\rm def}=15.7$ meV, at which the split-off $1s$ state becomes resonant. When there is compression along the [111] direction, this occurs at $P=3.9$ kbar.

Figure 5 presents plots of the dependence of the position of the tops of the light-hole (l) and heavy-hole (h) subbands (lines *1* and *2*) and the energies of the strain-split $M=\pm1/2$ and $M=\pm3/2$ levels (curves *3* and *4*) of an impurity center on the pressure applied to the sample, which were calculated by the zero-radius potential method, as well as the results of a variational calculation of the energies of the resonant $1s$ ($1s_r$) state and the local $2p_{\pm1}$ and $1s$ states. (We note that a variational calculation of the potential of the $1s$ and $1s_r$ states of a shallow acceptor was previously performed in Ref. 11.) It is seen that the zero-radius potential model poorly describes the position of the resonant level at large pressures. However, it makes it possible to estimate the magnitude of the decay of the resonant state. Figure 6 presents the dependence of the width Γ of the resonant state on applied pressure.

A variational calculation of the level splitting, including the positions of the excited states of a Coulomb impurity center in a uniaxially compressed crystal, has been performed only for fairly high pressures. In that limit, retaining only the terms quadraticin k in the expression under the radical sign in Eq. (2) and expanding it in a series in the small parameter k^2/ζ, we obtain a valence-band spectrum consisting of two noninteracting ellipsoidal subbands:

$$\varepsilon_l(k)=\frac{\hbar^2}{2m_0}[(\gamma_1+2\gamma)k_z^2+(\gamma_1-\gamma)(k_x^2+k_y^2)-\zeta],$$
$$\varepsilon_h(k)=\frac{\hbar^2}{2m_0}[(\gamma_1-2\gamma)k_z^2+(\gamma_1+\gamma)(k_x^2+k_y^2)+\zeta]. \quad (3)$$

The large-strain limit corresponds to the transition from a four-component basis set to two two-component basis sets $u_{\pm3/2}$ and $u_{\pm1/2}$ of the Bloch functions, i.e., to the elimination of the off-diagonal terms corresponding to the interaction of states with different values of the projection of the hole spin onto the z axis, which is parallel to P, from the Luttinger Hamiltonian. In this approximation the Coulomb potential of the shallow acceptor impurity creates two series of acceptor levels below the bottom of each subband. We calculated the energies of the four lowest local states ($1s$, $2p_{\pm1}$, $2p_0$, and $2s$) below the bottom of the ellipsoidal l band, following Ref. 15, in which the energy spectra of a shallow donor in Si and Ge were calculated. A similar approximation was employed to calculate the ground-state splitting in Ref. 11. We used the variational functions listed in Table I in the calculation.

TABLE I.

States	Variational functions
$1s$	$C\exp(-\sqrt{\rho^2/a^2+z^2/b^2})$
$2p_0$	$Cz\exp(-\sqrt{\rho^2/a^2+z^2/b^2})$
$2s$	$(C_1+C_2\rho^2+C_3z^2)\exp(-\sqrt{\rho^2/a^2+z^2/b^2})$
$2p_{\pm1}$	$C(x\pm iy)\exp(-\sqrt{\rho^2/a^2+z^2/b^2})$

The energies of the levels of the shallow acceptor in Ge obtained are presented in Table II. The energies of the states are calculated relative to the edges of the respective subbands (see Fig. 5). We recall that the $1s$ state below the bottom of the h subband is a resonant state. The last column contains the energies corresponding to possible optical transitions from the resonant state ($1s_r$) to the local states (i) indicated in the first column for splitting of the valence subbands by 27.4 meV, which corresponds to $P=6.85$ kbar. The scheme of acceptor levels is presented in Fig. 7. The figure also points out optical transitions that are split in the dipole approximation from the resonant $1s_r$ state to local states. Comparing these data with the spectrum in Fig. 2, we can assign the principal peak in the stimulated emission spectrum to the optical transition from the resonant $1s_r$ state to the local $2p_{\pm1}$ excited state and the peak at 20.5 meV to the transition between the resonant $1s_r$ state and the local $2p_0$ state. We attribute the peak at 19.9 meV to the transition from the $1s_r$ state to shallow states which are located near the edge of the l band and are not resolved in the present experiment. The peak at 23 meV is close to the expected value for the optical transition between the resonant $1s_r$ state and the local $1s$ state. However, transitions between the $1s_r$

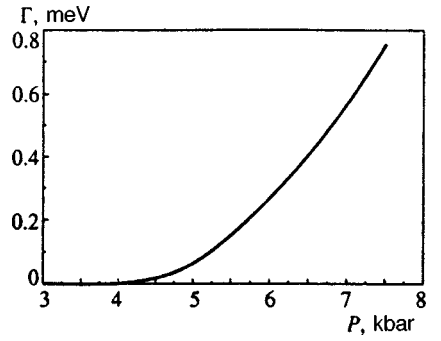


FIG. 6. Pressure dependence of the width Γ of the resonant level. $\mathbf{P}\|[111]$.

TABLE II.

	l subband $M=\pm 1/2$			*h* subband $M=\pm 3/2$			
States	ε_i, meV	$a\cdot 10^6$, cm	$b\cdot 10^6$ cm	ε, meV	$a\cdot 10^6$ cm	$b\cdot 10^6$ cm	$\varepsilon_{\text{def}}-\varepsilon_{1s_r}+\varepsilon_i$, meV ($P=6.85$ kbar)
$1s$	3.8	1.137	1.137	4.76	1.145	0.51	26.5
$2p_{\pm 1}$	1.3	1.57	2.3	0.9	2.71	1.338	24
$2s$	1.2	1.227	1.8	1.53	1.51	0.72	23.8
$2p_0$	0.8	2.195	3.185	2	1.56	0.744	23.4

state and local *s* states are forbidden in the dipole approximation. Therefore, we assume that this peak is caused by optical transitions of carriers from the continuous spectrum to the $1s$ level. This is possible only if the energy distribution of the free carriers has a local maximum near the energy of the resonant state.

Figure 8 shows the calculated energies of the $1s_r \rightarrow 2p_{\pm 1}$ optical transition for holes (the solid line) and the energy of the principal stimulated emission peak (points) as functions of pressure. It should be noted that the distances between the peaks in the stimulated emission spectrum agree well with the calculated values, but the entire spectrum is shifted relative to the calculated spectrum by about 3 meV. This may be because the interaction of the light-hole and heavy-hole bands, which gives rise to the decay Γ and to displacement of the levels to smaller energies, was not taken into account in the calculation performed.

The participation of the resonant $1s_r$ state in the transition is confirmed by several additional facts. The minimum pressure at which stimulated emission could be excited (Fig. 1b) corresponds exactly to the pressure at which the acceptor $1s_r$ state split off from the ground state passes into the continuous spectrum (see the diagrams in Fig. 4). The energy splitting of the ground acceptor state at that pressure amounts to about 10 meV (Fig. 4). On the other hand, as can be seen from Fig. 1a, the intensity of the stimulated emission decreases sharply at a pressure of about 8 kbar for $\mathbf{P}\|[100]$. Depopulation of the $1s_r$ state begins at that pressure because holes pass to the valence-band edge with emission of an optical phonon. The corresponding hole transition is shown in Fig. 4 for $\mathbf{P}\|[111]$ (for this crystallographic direction the energy of the split-off state measured from the valence-band edge becomes equal to the energy of the optical phonon at $P=12$ kbar). The splitting energy of the ground acceptor state is then $\simeq 42$ meV. Thus, the energy of the stimulated emission can vary with pressure in the range from 10 to 42 meV.

4. POPULATION INVERSION MECHANISM

Stimulated emission can appear only when there is an inverted energy distribution of the carriers. In our case there must be inverted population of the resonant $1s_r$ state with respect to the local states in the band gap, which are depopulated by impact ionization. We associate the appearance of inversion with strong resonant scattering by acceptors of free holes with an energy ε close to ε_0, which corresponds to the position of the resonant $1s_r$ level (see Fig. 7). For the probability of resonant scattering we obtained the following expression within the zero-radius potential model:

$$w_{\varepsilon,\theta_0,\theta}=N\frac{\hbar^2\gamma_1^{3/2}}{m_0^{3/2}}\frac{\varepsilon_0}{\varepsilon^{3/2}}\frac{(\varepsilon/\varepsilon_0)^2}{(1-\varepsilon/\varepsilon_0)^2+(\Gamma/\varepsilon_0)^2}\times G\left(\frac{\varepsilon_{\text{def}}}{\varepsilon},\theta_0,\theta\right). \tag{4}$$

Here N is the impurity concentration, Γ is the broadening of the resonant state ($\Gamma \ll \varepsilon_0$), G is a function which specifies the angular dependence of the scattering, and θ_0 and θ are the angles of incidence and scattering relative to the $\mathbf{z}\|\mathbf{P}\|\mathbf{E}$ direction. Figure 9 shows the angular dependence of G for three values of θ_0 at $\varepsilon_{\text{def}}=27.4$ meV, and $\varepsilon=\varepsilon_0=22.6$ meV (which corresponds to ε_{1s_r}). It is seen that the

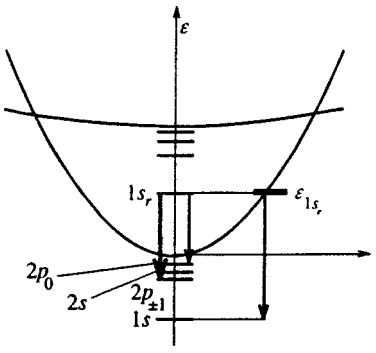


FIG. 7. Impurity level diagram and intracenter optical transitions in uniaxially strained Ge.

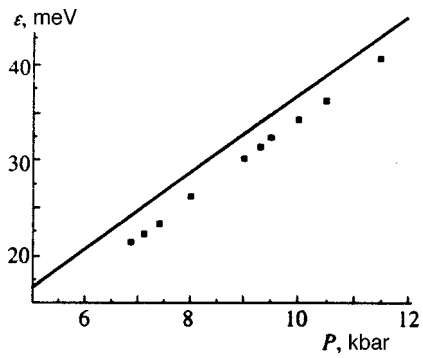


FIG. 8. Energy of the principal stimulated emission peak (points) and calculated energies of the optical transitions from the resonant $1s_r$ acceptor state to the local $2p_{\pm 1}$ state (straight line).

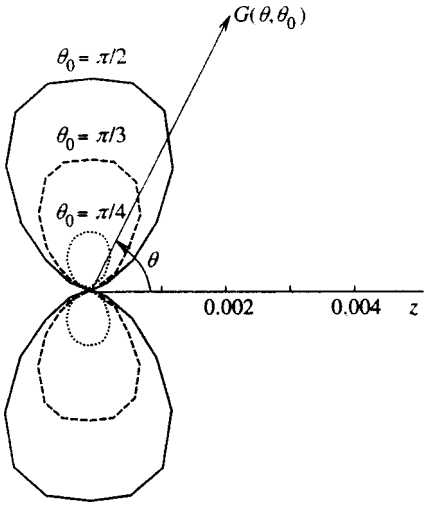


FIG. 9. Angular dependence of the probability of the resonant scattering of holes with an energy $\varepsilon\approx\varepsilon_0$ on the scattering angle θ for three values of the angle of incidence θ_0: solid line — $\theta_0=\pi/2$; dashed line — $\theta_0=\pi/3$; dotted line — $\theta_0=\pi/4$. $\mathbf{P}\|[111]$.

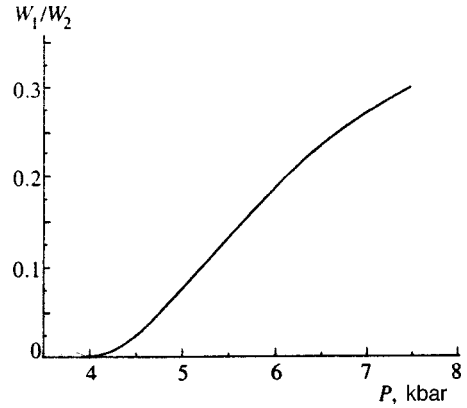


FIG. 10. Pressure dependence of the transition probability ratio: continuum $(\varepsilon=\varepsilon_{1s_r})\rightarrow 1s$ level (W_1) and quasilocal $1s_r$ level$\rightarrow 2p_{\pm1}$ level (W_2).

momentum of the holes after scattering is directed, for the most part, perpendicularly to the applied field and thus also promotes the accumulation of holes with an energy close to ε_0.

5. PROBABILITIES OF OPTICAL TRANSITIONS

Let us now consider the relationship between the probabilities of the $1s_r\rightarrow 2p_{\pm1}$ intracenter transitions and the transitions of free holes with an energy $\varepsilon\approx\varepsilon_0$ to the local $1s$ state. In the continuous spectrum there are two types of states at the resonance energy ε_0, viz., quasilocal (resonant) states and continuum states, which are specified by solutions of the unperturbed Luttinger Hamiltonian, and transitions to local states in the band gap are possible for them. According to the selection rules, the $1s_r\rightarrow 1s$ transition is forbidden in the dipole approximation; therefore, transitions to the ground state are possible only from continuum states near the resonance energy. The transition from $1s_r$ to the local $2p_{\pm1}$ state is allowed in the dipole approximation. For transitions from states in the continuous spectrum to the $1s$ state we assume that the initial energy of the carriers lies within an interval of width Γ near ε_0. We suppose that the intense exchange between the resonant level and the band establishes a quasiequilibrium between these types of states, which allows us to introduce a single distribution function f_ε. Then the concentration p of holes in the interval of width Γ near ε_0 is specified by the expression

$$p=[\eta(\varepsilon_0)\Gamma+N]f_\varepsilon, \tag{5}$$

where N is the concentration of centers and $\eta(\varepsilon_0)=(2\pi)^{-3}\int\delta(\varepsilon(k)-\varepsilon_0)\,d^3k$ is the density of states in the continuous spectrum.

Since the radiation emerges from the sample perpendicularly to the z axis in the experiment, we present expressions for the probability W_1 of the spontaneous $1s_r\rightarrow 2p_{\pm1}$ optical dipole transition, as well as for the probability W_2 of transitions between continuum states with an energy ε in the interval $(\varepsilon_0-\Gamma/2,\varepsilon_0+\Gamma/2)$ and the local $1s$ state (these transitions are shown in Fig. 7):

$$W_1=Nf_\varepsilon\frac{e^2\gamma_1^2}{c^3m_0^2}\hbar\omega_1 n\frac{1}{a_1^2}\frac{b_{2p}}{b_1}I_1\,d\Omega, \tag{6}$$

$$W_2=Nf_\varepsilon\frac{e^2\gamma_1^2}{c^3m_0^2}\kappa^2\hbar\omega_2 n\frac{\Gamma}{\varepsilon_0}(a_{1s}\kappa)^2(b_{1s}\kappa)I_2\,d\Omega, \tag{7}$$

where $\hbar\omega_1$ and $\hbar\omega_2$ are the photon energies for these transitions, a_1, b_1 and a_{1s}, b_{1s} are the characteristic sizes of the resonant $1s_r$ state and the $1s$ state, respectively, b_{2p} is the characteristic size of the local $2p_{\pm1}$ state (see the tables), n is the refractive index, and I_1 and I_2 are dimensionless quantities. A numerical calculation gives $I_1=0.056$ and $I_2=0.0215$ for $P=6.85$ kbar. The parameter κ is related to ε_0 by the expression $\varepsilon_0=\hbar^2\kappa^2\gamma_1/2m_0$.

Figure 10 presents the pressure dependence of the ratio of transition probabilities W_2/W_1. The level width Γ found within the zero-radius potential model was used in the calculations. For $P=6.85$ kbar we have $W_2/W_1\approx 0.3$, which is close to the intensity ratio of the peaks at 23 and 20.5 meV in the stimulated emission spectrum (Fig. 2).

To conclude this paper, we present the expression for the radiative lifetime of the spontaneous intracenter $1s_r\rightarrow 2p_{\pm1}$ transition:

$$\tau[\mathrm{s}]=5\times10^{-5}(4P[\mathrm{kbar}]-3.5)^{-1}. \tag{8}$$

For $P=6.85$ kbar we obtain $\tau=2.2\times10^{-6}$ s.

6. CONCLUSION

The experimental data presented and a comparison with the results of calculations show that the terahertz stimulated emission of uniaxially strained p-Ge is caused by population inversion of pressure-split acceptor levels, which, in turn, is caused by the resonant scattering of holes heated by the electric field. One necessary condition for such inversion is that the acceptors have resonant states, i.e., states located in the continuous energy spectrum of the valence band. The lines in the spectrum have been identified. It has been shown, in particular, that the principal line in the stimulated emission

spectrum is caused by optical transitions of holes from the resonant $1s$ state to the local $2p_{\pm 1}$ excited state of the acceptor. The strong alteration of the frequency by external pressure has been demonstrated.

This work was supported by grants from the Russian Fund for Fundamental Research (Grants Nos. 97-02-16820 and 96-02-17532) and the Physics of Solid-State Nanostructures Program (Grant No. 97-1055).

*)E-mail: kagan@landau.ac.ru, kagan@mail.splire.ru

[1] A. A. Andronov, in *Spectroscopy of Nonequilibrium Electrons and Phonons*, C. V. Shank and B. P. Zakharchenya (eds.), Elsevier, Amsterdam–New York (1992), p. 1.
[2] J. Faist, F. Capasso, D. L. Sivco *et al.*, Science **264**, 553 (1994).
[3] O. Gauthier-Lafaye, S. Sauvage, P. Boucaud *et al.*, Appl. Phys. Lett. **70**, 3197 (1997).
[4] I. V. Altukhoy, M. S. Kagan, K. A. Korolev, V. P. Sinis, and F. A. Smirnov, Zh. Éksp. Teor. Fiz. **101**, 756 (1992) [Sov. Phys. JETP **74**, 404 (1992)].
[5] I. V. Altukhov, M. S. Kagan, K. A. Korolev, and V. P. Sinis, JETP Lett. **59**, 476 (1994).
[6] A. A. Andronov, I. V. Zverev, V. A. Kozlov, Yu. N. Nozdrin, S. A. Pavlov, and V. N. Shastin, JETP Lett. **40**, 804 (1984).
[7] A. A. Kastal'skiĭ and S. M. Ryvkin, Fiz. Tekh. Poluprovodn. **1**, 622 (1967) [Sov. Phys. Semicond. **1**, 523 (1967)].
[8] J. E. Smith, Jr., J. C. McGroddy, and M. I. Nathan, Phys. Rev. **186**, 727 (1969).
[9] N. O. Gram and N. I. Meyer, Phys. Status Solidi **1**, 237 (1970).
[10] I. V. Altukhov, M. S. Kagan, K. A. Korolev, and V. P. Sinis, Zh. Éksp. Teor. Fiz. **103**, 1829 (1993) [JETP **76**, 903 (1993)].
[11] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, Wiley, New York; Israel Program for Scientific Translations, Jerusalem (1974).
[12] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).
[13] *Landolt-Börnstein, New Series, Group III, Vol. 22b: Semiconductors, Impurities and Defects in Group IV Elements and III-V Compounds*, O. Madelung (ed.), Springer Verlag, Berlin–Heidelberg (1984).
[14] M. A. Odnoblyudov, A. A. Pakhomov, V. M. Chistyakov, and I. N. Yassievich, Fiz. Tekh. Poluprovodn. **31**, 1180 (1997) [Semiconductors **31**, 1014 (1997)].
[15] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

Translated by P. Shelnitz

Longitudinal dynamic susceptibility of superparamagnetic particles with cubic anisotropy

Yu. P. Kalmykov*) and S. V. Titov†)

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 141120 Fryazino, Moscow Region, Russia

(Submitted 23 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 101–114 (January 1999)

We study the linear response of a system of single-domain ferromagnetic particles with cubic magnetic anisotropy to a weak external a.c. magnetic field. By averaging the Gilbert equation with a fluctuating field for the magnetization of an individual particle we derive a system of recurrence equations for the spectra of equilibrium correlation functions describing the longitudinal relaxation of the system. We find the solution of this system by using matrix continued fractions. We also evaluate the longitudinal relaxation time and the spectrum of the complex-valued magnetic susceptibility. Finally, we show that the nature of susceptibility dispersion is determined by the anisotropy and dissipation parameters. © *1999 American Institute of Physics.* [S1063-7761(99)01001-X]

1. Single-domain ferromagnetic particles are characterized by an internal anisotropy potential, which may have several positions of local equilibrium with potential barriers separating them. If the particles are small (~ 100 Å) and as a result the potential barriers are low, thermal fluctuations may cause the magnetization vector to reorient itself over the barriers from one equilibrium position to another.[1] The thermal instability of magnetization leads to what is known as superparamagnetism.[2] Studies of thermal fluctuations and the relaxation of magnetization of single-domain particles have attracted much attention in connection with the problem of improving the characteristics of magnetic storage elements.[3]

When relaxation processes in superparamagnets are studied theoretically, to simplify the mathematics one usually examines the case of uniformly magnetized uniaxial particles.[2,4–12] Although using a uniaxial anisotropy potential simplifies the analysis significantly, the results obtained in this approximation are of limited value.[13] For other types of anisotropy, such as cubic, either the discrete orientation approximation has been employed or solutions for the continuous diffusion model that are only asymptotic have been studied (see, e.g., Refs. 2 and 13–19). However, neither approach can be used in the most interesting case where the anisotropy energy is comparable to the thermal energy kT.

In the diffusion model the dynamics of the magnetization vector $\mathbf{M}(t)$ of a single-domain particle is similar to the Brownian rotation of a macromolecule in a liquid and is described by the Fokker–Planck equation for the probability distribution density $W(\{\mathbf{M}\},t)$ of magnetization.[2,20,21] The Fokker–Planck equation is derived from the Gilbert equation[2,20] with a fluctuating field that allows for the thermal fluctuations of the magnetization of an individual particle. For the case of cubic anisotropy, the Fokker–Planck equation can by formally solved by, say, expanding the distribution function W in spherical harmonics.[19] In this approach, we need only solve an infinite system of recurrence equations for averaged spherical harmonics (moments).[19] The system of equations for the moments can be derived by averaging the Gilbert equation without using the Fokker–Planck equation.[12] It is difficult to employ the well-known methods of solving such a system of equations in the case of weak dissipation (characteristic of single-domain particles),[1)] since such calculations require using 10^4–10^5 and more equations to achieve convergence. For this reason, calculations and analysis of the spectrum of the complex-valued magnetic susceptibility in the case of cubic anisotropy for the diffusion model have yet to be done. The problem, however, can be simplified significantly if we use the method of matrix continued fractions developed in Refs. 22 and 23 to solve infinite systems of recurrence equations for the moments. In the present paper we use this method to calculate the relaxation time $\tau_\parallel$ of the longitudinal component of magnetization and the dynamic magnetic susceptibility $\chi_\parallel(\omega)$ of a system of noninteracting single-domain particles for arbitrary values of the anisotropy and dissipation energy parameters, σ and α. We determine and study the behavior of $\tau_\parallel$ and $\chi_\parallel(\omega)$ over the entire range of values of σ and α.

2. If we allow for thermal fluctuations, the Gilbert equation for the magnetization $\mathbf{M}$ of a single-domain particle has the form[2,24]

$$\frac{d}{dt}\mathbf{M}(t)=\gamma\mathbf{M}(t)\times[\mathbf{H}(t)+\mathbf{h}(t)-\eta\mathbf{M}(t)], \tag{1}$$

where γ is the gyromagnetic ratio, η is the friction coefficient, $\mathbf{H}$ is the total magnetic field, which an consist of external fields applied to the system and the effective magnetic field of the anisotropy, and $\mathbf{h}(t)$ is a random field having the properties of white noise:

$$\overline{h_i(t)}=0, \quad \overline{h_i(t_1)h_j(t_2)}=\frac{2kT\eta}{v}\,\delta_{ij}\delta(t_1-t_2). \tag{2}$$

Here v is the volume of the particle, and the horizontal bar stands for statistical averaging over an ensemble of particles

having at time t the same magnetization $\mathbf{M}(t)$. The order of magnitude of the amplitude of $\mathbf{h}(t)$ can be estimated at kT/vM_s (M_s is the magnetization of the material of the particle), which at room temperature yields a value $\geqslant 100$ Oe and, the strength of the random field is thus comparable to that of the magnetic anisotropy field.[8]

If V is the free energy per unit volume expressed in terms of the components of $\mathbf{M}$, the field $\mathbf{H}$ is determined by the equation

$$\mathbf{H}=-\frac{\partial V}{\partial \mathbf{M}}. \tag{3}$$

In the present case of weak cubic anisotropy we use the representation[2,23]

$$V=K(u_x^2u_y^2+u_x^2u_z^2+u_y^2u_z^2)$$
$$=\frac{K}{4}(\sin^4\vartheta\sin^2 2\varphi+\sin^2 2\vartheta), \tag{4}$$

where K is the anisotropy constant, which is either positive or negative (below we use the dimensionless anisotropy parameter $\sigma=vK/4kT$). When K is positive, the potential (4) has 6 minima, 8 maxima, and 12 saddle points (e.g., in the directions [100], [111] and [110], respectively).[2] When K is negative, the minima and maxima change places. For $K>0$ both the heights of all the potential barriers and the energy at the saddle points are equal to σ, while for $K<0$ the barrier heights are $|\sigma|/3$ and the energy at the saddle points is $|\sigma|$ (see Ref. 2). Below we limit ourselves to the case of positive anisotropy, $K>0$. The case of $K<0$ can be examined along the same lines.

By transforming the Gilbert equation (1) to the Landau–Lifshitz equation[2] and writing the components of this equation in the laboratory system of coordinates, we arrive at[12]

$$\frac{1}{\alpha g' M_s}\frac{du_x(t)}{dt}=[1-u_x^2(t)]h_x(t)$$
$$-[\alpha^{-1}u_z(t)+u_x(t)u_y(t)]h_y(t)$$
$$+[\alpha^{-1}u_y(t)-u_z(t)u_x(t)]h_z(t)$$
$$+[1-u_x^2(t)]H_x(t)$$
$$-[\alpha^{-1}u_z(t)+u_x(t)u_y(t)]H_y(t)$$
$$+[\alpha^{-1}u_y(t)-u_z(t)u_x(t)]H_z(t), \tag{5}$$

$$\frac{1}{\alpha g' M_s}\frac{du_y(t)}{dt}=[\alpha^{-1}u_z(t)-u_x(t)u_y(t)]h_x(t)$$
$$+[1-u_y^2(t)]h_y(t)$$
$$-[\alpha^{-1}u_x(t)+u_y(t)u_x(t)]h_z(t)$$
$$+[\alpha^{-1}u_z(t)-u_x(t)u_y(t)]H_x(t)$$
$$+[1-u_y^2(t)]H_y(t)$$
$$-[\alpha^{-1}u_x(t)+u_y(t)u_x(t)]H_z(t), \tag{6}$$

$$\frac{1}{\alpha g' M_s}\frac{du_z(t)}{dt}=-[\alpha^{-1}u_y(t)+u_x(t)u_z(t)]h_x(t)$$
$$+[\alpha^{-1}u_x(t)-u_y(t)u_z(t)]h_y(t)$$
$$+[1-u_z^2(t)]h_z(t)$$
$$-[\alpha^{-1}u_y(t)+u_x(t)u_z(t)]H_x(t)$$
$$+[\alpha^{-1}u_x(t)-u_y(t)u_z(t)]H_y(t)$$
$$+[1-u_z^2(t)]H_z(t), \tag{7}$$

where $\alpha=\gamma\eta M_s$ is the dimensionless dissipation coefficient,

$$g'=\frac{\gamma}{(1+\alpha^2)M_s}, \tag{8}$$

$$u_x=\sin\vartheta\cos\varphi,\quad u_y=\sin\vartheta\sin\varphi,\quad u_z=\cos\vartheta, \tag{9}$$

with φ and ϑ the azimuthal and polar angles, respectively. Here we have ignored surface effects and assumed that inside the particle the magnetization is uniform.

Below we use the spherical harmonics $Y_{n,m}$ (Ref. 26), which in terms of the variables u_x, u_y, and u_z are

$$Y_{n,m}=(-1)^m\sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}}(u_x+iu_y)^m$$
$$\times\frac{d^mP_n(u_z)}{du_z^m},\quad m\geqslant 0, \tag{10}$$

$$Y_{n,-m}=(-1)^mY_{n,m}^*, \tag{11}$$

where the $P_n(x)$ are Legendre polynomials.[26] Moreover, when averaging and transforming the stochastic differential equations (5)–(7) with multiplicative noise, it is convenient to use the Stratonovich approach.[25] In particular, in this case there is no need to first transform Eqs. (5)–(7) into the equivalent form of Itô equations.[22] Thus, bearing in mind that in transformations of stochastic differential equations via the Stratonovich approach[25] we can use the rules of ordinary analysis,[22,23] we easily arrive at a stochastic differential equation for the spherical harmonics:

$$\frac{dY_{n,m}(t)}{dt}=\frac{1}{u_x(t)+iu_y(t)}$$
$$\times\left[mY_{n,m}(t)\left(\frac{du_x(t)}{dt}+i\frac{du_y(t)}{dt}\right)\right.$$
$$\left.-Y_{n,m+1}(t)\sqrt{\frac{n+m+1}{n-m-1}}\frac{du_z(t)}{dt}\right], \tag{12}$$

where $\dot{u}_x$, $\dot{u}_y$, and $\dot{u}_z$ can be found from (5)–(7), respectively. Next, using the method developed in Refs. 12, 23, and 27 for solving nonlinear Langevin equations with multiplicative noise, we can derive from (12), after doing a series of algebraic transformations, a system of coupled equations for the equilibrium correlation functions characterizing the linear response of the system:

$$\tau_N \frac{dc_{n,m}(t)}{dt} = \sum_{s=-1}^{1} \sum_{r=-4}^{4} d_{n,m,r,s} c_{n+r,m+4s}(t), \tag{13}$$

where

$$c_{n,m}(t) = \langle \cos\vartheta(0) Y_{n,m}(t) \rangle_0, \tag{14}$$

with angle brackets $\langle\ \rangle_0$ standing for average over the equilibrium state at time $t=0$, and

$$\tau_N = \frac{v}{2kT\alpha g'} \tag{15}$$

is the characteristic time of thermal fluctuations of magnetization. The expressions for the coefficients $d_{n,m,r,s}$ are given in Appendix A. The system of equations (13) can also be derived from the corresponding Fokker–Planck equation[19]

$$2\tau_N \frac{\partial W}{\partial t} = \frac{1}{\sin\vartheta} \frac{\partial}{\partial\vartheta} \left\{ \sin\vartheta \left[\frac{v}{kT} \left(\frac{\partial V}{\partial\vartheta} - \frac{1}{\alpha \sin\vartheta} \frac{\partial V}{\partial\varphi} \right) W + \frac{\partial W}{\partial\vartheta} \right] \right\} + \frac{1}{\sin\vartheta} \frac{\partial}{\partial\varphi} \left[\frac{v}{kT} \left(\frac{1}{\alpha} \frac{\partial V}{\partial\vartheta} + \frac{1}{\sin\vartheta} \frac{\partial V}{\partial\varphi} \right) W + \frac{1}{\sin\vartheta} \frac{\partial W}{\partial\varphi} \right]. \tag{16}$$

Using (13) to find $c_{1,0}(t)$, we can calculate the longitudinal dynamic susceptibility and relaxation time, since according to linear-reaction theory[23] the decrease in the magnetization $\langle \mathbf{M} \rangle(t)$ due to a sudden switch-on at time $t=0$ of a weak external d.c. magnetic field $\mathbf{H}_1$ parallel to the z axis of the laboratory system of coordinates is of the form

$$\langle M_\parallel \rangle(t) = \chi_\parallel H_1 C_\parallel(t), \tag{17}$$

where

$$C_\parallel(t) = \frac{c_{1,0}(t)}{c_{1,0}(0)} \tag{18}$$

is the normalized relaxation function of the longitudinal magnetization component, and

$$\chi_\parallel = \frac{v^2 M_s^2 N_0}{3kT} \tag{19}$$

is the static magnetic susceptibility (here we have allowed for the fact that in the case of cubic anisotropy, $\langle u_x^2 \rangle_0 = \langle u_y^2 \rangle_0 = \langle u_z^2 \rangle_0 = 1/3$), with N_0 the number of particles per unit volume. The longitudinal dynamic magnetic susceptibility $\chi_\parallel(\omega)$ can be expressed in terms of the spectrum $C_\parallel(t)$ as follows:

$$\chi_\parallel(\omega) = \chi_\parallel'(\omega) - i\chi_\parallel''(\omega) = \chi_\parallel \{ 1 - i\omega \tilde{C}_\parallel(i\omega) \}, \tag{20}$$

where

$$\tilde{C}_\parallel(i\omega) = \int_0^\infty C_\parallel(t) e^{-i\omega t}\, dt. \tag{21}$$

Moreover, the relaxation time $\tau_\parallel$ of the longitudinal magnetization component, defined as the area under the curve $C_\parallel(t)$, or

$$\tau_\parallel = \int_0^\infty C_\parallel(t)\, dt = \tilde{C}_\parallel(0), \tag{22}$$

can be measured in experiments or calculated by (21). We assumed all along that the particles are identical. To allow for the polydispersity of the particles we must also average the susceptibility and relaxation time over the respective distribution functions.[8]

3. A formal approach that uses matrix continued fractions in the solution of recurrence equations of type (13), where two indices vary, was proposed in Refs. 22 and 23. However, it proved to be extremely difficult to use this approach to solve practical problems, since matrices of indefinite dimension have to be introduced into the transformations.[28] Below we use a modified method, which makes it it possible to reduce the solution procedure to operations involving finite-dimensional matrices. Let us introduce a vector $\mathbf{C}_n(t)$ by the formula

$$\mathbf{C}_n(t) = \begin{pmatrix} \mathbf{c}_{4n}(t) \\ \mathbf{c}_{4n-1}(t) \\ \mathbf{c}_{4n-2}(t) \\ \mathbf{c}_{4n-3}(t) \end{pmatrix}, \qquad \mathbf{c}_{4n-i}(t) = \begin{pmatrix} c_{4n-i,-4(n-1+\delta_{i0})}(t) \\ c_{4n-i,-4(n-2+\delta_{i0})}(t) \\ \vdots \\ c_{4n-i,-4(n-1+\delta_{i0})}(t) \end{pmatrix}, \quad i=0,1,2,3. \tag{23}$$

The vector $\mathbf{C}_n(t)$ has $8n-2$ elements. Thus, Eq. (13) reduces to the matrix equation

$$\tau_N \frac{d\mathbf{C}_n(t)}{dt} = \mathbf{Q}_n^- \mathbf{C}_{n-1}(t) + \mathbf{Q}_n \mathbf{C}_n(t) + \mathbf{Q}_n^+ \mathbf{C}_{n+1}(t), \quad n=1,2,3,\ldots, \tag{24}$$

where

$$\mathbf{C}_0(t) = \mathbf{0}, \quad \mathbf{C}_1(t) = \begin{pmatrix} c_{4,-4}(t) \\ c_{4,0}(t) \\ c_{4,4}(t) \\ c_{3,0}(t) \\ c_{2,0}(t) \\ c_{1,0}(t) \end{pmatrix}. \tag{25}$$

The explicit form of the matrices $\mathbf{Q}_n^-$, $\mathbf{Q}_n$, and $\mathbf{Q}_n^+$ is given in Appendix A.

Using the general method of solving matrix recurrence equation of Ref. 23 (see Appendix B), we obtain the exact solution for the Laplace transform of $\mathbf{C}_1(t)$ in the form

$$\tilde{\mathbf{C}}_1(s) = \tau_N [\tau_N s\mathbf{I} - \mathbf{Q}_1 - \mathbf{Q}_1^+ \mathbf{S}_2(s)]^{-1} \left\{ \mathbf{C}_1(0) + \sum_{n=2}^{\infty} \left[\prod_{k=2}^{n} \mathbf{Q}_{k-1}^+ \mathbf{S}_k(s) (\mathbf{Q}_k^-)^{-1} \right] \mathbf{C}_n(0) \right\}, \tag{26}$$

where $\mathbf{I}$ is the identity matrix, and the matrix continued fraction $\mathbf{S}_n(s)$ is given by the formula

$$\mathbf{S}_n(s)=\cfrac{\mathbf{I}}{\tau_N s\mathbf{I}-\mathbf{Q}_n-\mathbf{Q}_n^+\cfrac{\mathbf{I}}{\tau_N s\mathbf{I}-\mathbf{Q}_{n+1}-\mathbf{Q}_{n+1}^+\cfrac{\mathbf{I}}{\tau_N s\mathbf{I}-\mathbf{Q}_{n+2}-\cdots}\mathbf{Q}_{n+2}^-}\mathbf{Q}_{n+1}^-}\mathbf{Q}_n^-. \tag{27}$$

The method of calculating the initial-value vectors $\mathbf{C}_n(0)$ via matrix continued fractions is given in Appendix C.

Formula (26) is the exact solution of Eq. (24) expressed in terms of matrix continued fractions. In this form the solution was obtained in Ref. 23. By its very nature, formula (26) is the analytical representation of the numerical algorithm of matrix continued fractions used in Risken's monograph.[22] An essential development of the results of Ref. 22 is, first, that the solution is obtained in analytic form and is expressed by formula (26) and, second, that we have generalized the method to the case where the dimension of $\mathbf{Q}_n$, $\mathbf{Q}_n^-$, or $\mathbf{Q}_n^+$ depends on n. As shown in Ref. 23 by many examples, formula (26) is convenient for calculations. In our problem the maximum dimension of all matrices required by the calculations is of order 10^2, which makes it possible to use a personal computer for all calculations.

4. First we examine the dependence of the relaxation time $\tau_\parallel$ on the anisotropy parameter $\sigma=vK/4kT$ for different values of the dissipation parameter α. The curves representing this dependence calculated by (22) and (26) are depicted in Fig. 1. By its very physical meaning, $\tau_\parallel$ is determined primarily by the lowest-frequency longitudinal relaxation mode related to the transition of the magnetization vector over the barrier separating one potential well from another. The characteristic relaxation time τ of this low-frequency mode is determined by the reciprocal value of the smallest eigenvalue λ_1 of the Fokker–Planck operator in (16). In the low-temperature limit ($\sigma\gg1$) and for strong and/or moderate dissipation ($\alpha\geqslant1$), the estimate of τ is given by a relationship derived in Refs. 2 and 15, which in the present notation can be written

$$\tau\sim\frac{\tau_N\pi e^\sigma}{2\sqrt{2}\,\sigma(\sqrt{9+8/\alpha^2}+1)},\quad \sigma>0. \tag{28}$$

The corresponding formula[13,18] for weak dissipation ($\alpha\ll1$) is

$$\tau\sim\frac{\pi kTe^\sigma}{2\omega_A\Delta E}\approx\frac{\tau_N\pi e^\sigma}{8\sigma^2},\quad \sigma>0, \tag{29}$$

where $\omega_A=8\sigma\gamma kT/vM_s$ is the frequency of oscillations in a potential well, and $\Delta E\approx\alpha vK/4$ is the energy loss per period of the almost periodic motion of $\mathbf{M}(t)$ (see Ref. 13). Figure 1 shows that in contrast to uniaxial particles, where $\tau_\parallel/\tau_N$ is independent of α (see Ref. 6 and 9), in the case of cubic anisotropy the ratio $\tau_\parallel/\tau_N$ strongly depends on α. This dependence is due to the interaction of transverse and longitudinal relaxation intrawell modes. This interaction leads, in particular, to a nonmonotonic dependence of $\tau_\parallel/\tau_N$ on σ for small values of α (for $\sigma<1$, the contribution of longitudinal modes leads to a decrease in $\tau_\parallel/\tau_N$ with increasing σ; a further increase in σ leads to a situation in which the contribution of the low-frequency longitudinal relaxation mode becomes dominant and $\tau_\parallel/\tau_N$ begins to grow exponentially). Note that at intermediate values of α (say, $\alpha\approx0.1$) neither (28) nor (29) yields correct values of $\tau_\parallel/\tau_N$. In this case a more exact analysis is needed, and there are no formulas that would make it possible to estimate the values of the ratio.[13]

The dependence of longitudinal relaxation on the dissipation parameter α also manifests itself in the spectra $\chi_\parallel''(\omega)$ depicted in Figs. 2 and 3. Two peaks are clearly visible in the loss spectrum (the necessary calculations were done at $v^2M_s^2N_0/kT=1$). The first (low-frequency) peak appears at frequencies on the order of the average frequency of the reorientation of the magnetization vector. The characteristic

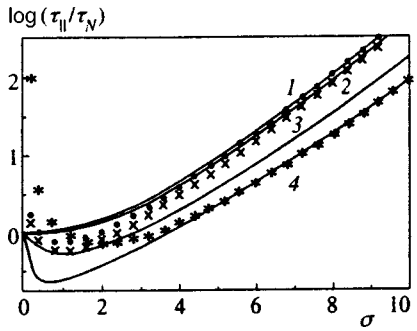


FIG. 1. Dependence of $\log(\tau/\tau_N)$ on the anisotropy parameter σ for different values of the dissipation parameter α. Solid curves represent the results of calculations by (22) and (26) for $\alpha\to\infty$ (curve *1*), $\alpha=1$ (curve *2*), $\alpha=0.1$ (curve *3*), and $\alpha=0.01$ (curve *4*); the ● and × represent the results of calculations by the asymptotic formula(28) for $\alpha\to\infty$ and $\alpha=1$, respectively; and the * represent the results of calculations by (29) at $\alpha=0.01$.

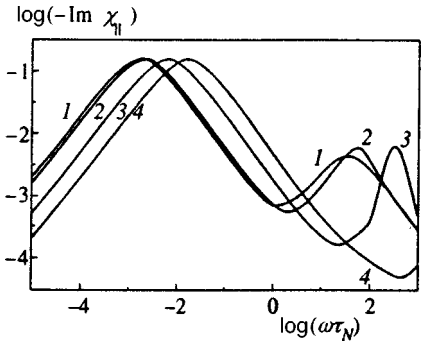


FIG. 2. $\log(-\mathrm{Im}\chi_\parallel'')$ vs. $\log(\omega\tau_N)$ at $\sigma=10$ for different values of the dissipation parameter: $\alpha=\infty$ (curve *1*), $\alpha=1$ (curve *2*), $\alpha=0.1$ (curve *3*), and $\alpha=0.01$ (curve *4*).

frequency and halfwidth of this band are determined by τ. The susceptibility dispersion in this frequency range is of a purely relaxational nature. The second, lower, peak appears because of the contribution of high-frequency transverse and longitudinal modes. As σ decreases, this high-frequency band narrows and shifts toward higher frequencies, and the nature of the dispersion changes from relaxation to resonant. On the other hand, as σ increases, the band also shifts toward higher frequencies, but it does not narrow significantly in the process. Such behavior can be explained by the strongest effect of transverse modes on this band, and these modes determine the transverse susceptibility spectrum and the ferromagnetic resonance at frequencies coinciding with those at which the magnetization vector precesses, $\omega_0 \sim \sigma(\alpha\tau_N)^{-1}$, with damping $\sim \alpha^{-1}$ (see Ref. 6).

The model we have developed can be used to explain the results of measurements of the dynamic susceptibility of systems of single-domain particles with cubic anisotropy. Until now the interpretation of experiments involving such systems has been done within the uniaxial-particle model (see, e.g., Refs. 29 and 30, where the frequency and temperature curves of the linear and nonlinear dynamic susceptibilities of systems of single-domain particles are studied). However, in the present paper we have shown that the behavior of the response of particles with cubic anisotropy differs from that of uniaxial particles. In particular, one must take into account the dependence of the response on the dissipation parameter α. Bitoh *et al.*[29] probably were unable to achieve quantitative agreement with the experimental data because they ignored this.

We believe that our approach will enable us to quantitatively describe the experiments of Bitoh *et al.*[29,30] and similar experiments. In a future paper we propose comparing in every detail the theory and experiment in relation to the frequency and temperature curves of the linear and nonlinear dynamic susceptibilities of systems of single-domain particles with cubic anisotropy, since this goal requires calculating the transverse component $\chi_\perp$ of the linear susceptibility, [2)] and comparison with the experimental data on the nonlinear response requires calculating the nonlinear dynamic susceptibility. One must also take into account the volume distribution of the particles. All these problems can be solved by our method, but such analysis lies outside the scope of the present paper.[3)]

5. Thus, the longitudinal dynamic susceptibility $\chi_\parallel(\omega)$ and the relaxation time $\tau_\parallel$ in the case of cubic anisotropy can be calculated by formula (26) by using matrix continued fractions in all ranges of the anisotropy and dissipation parameters. Here, in contrast to uniaxial particles, the spectrum $\chi_\parallel(\omega)$ and the relaxation time $\tau_\parallel$ of particles with cubic anisotropy are strongly dependent on α, owing to the interaction of the longitudinal and transverse modes.

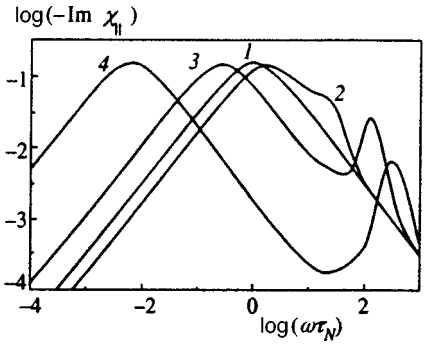


FIG. 3. $\log(-\mathrm{Im}\chi_\parallel'')$ vs. $\log(\omega\tau_N)$ at $\alpha=0.1$ for different values of the anisotropy parameter: $\sigma=0$ (curve *1*), $\sigma=1$ (curve *2*), $\sigma=5$ (curve *3*), and $\sigma=10$ (curve *4*).

The authors are grateful to W. T. Coffey and Yu. L. Raĭkher for fruitful discussions. The work was supported by a grant from the Russian Fund for Fundamental Research (Grant No. 96-02-16762-a).

APPENDIX A: EXPLICIT FORM OF THE MATRICES $\mathbf{Q}_n^-$, $\mathbf{Q}_n$, AND $\mathbf{Q}_n^+$ AND THEIR ELEMENTS

The matrices $\mathbf{Q}_n^-$, $\mathbf{Q}_n$, and $\mathbf{Q}_n^+$ of (24) are given by the formulas

$$\mathbf{Q}_n^- = \begin{pmatrix} \mathbf{J}_{4n} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{4n-1} & \mathbf{J}_{4n-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{P}_{4n-2} & \mathbf{D}_{4n-2} & \mathbf{J}_{4n-2} & \mathbf{0} \\ \mathbf{B}_{4n-3} & \mathbf{P}_{4n-3} & \mathbf{D}_{4n-3} & \mathbf{J}_{4n-3} \end{pmatrix}, \tag{A1}$$

$$\mathbf{Q}_n = \begin{pmatrix} \mathbf{A}_{4n} & \mathbf{B}_{4n} & \mathbf{P}_{4n} & \mathbf{D}_{4n} \\ \mathbf{B}_{4n}^{\mathrm{T}} & \mathbf{A}_{4n-1} & \mathbf{B}_{4n-1} & \mathbf{P}_{4n-1} \\ f_{4n}\mathbf{P}_{4n}^{\mathrm{T}} & \mathbf{B}_{4n-1}^{\mathrm{T}} & \mathbf{A}_{4n-2} & \mathbf{B}_{4n-2} \\ \mathbf{D}_{4n}^{\mathrm{T}} & f_{4n-1}\mathbf{P}_{4n-1}^{\mathrm{T}} & \mathbf{B}_{4n-2}^{\mathrm{T}} & \mathbf{A}_{4n-3} \end{pmatrix}, \tag{A2}$$

$$\mathbf{Q}_n^+ = \begin{pmatrix} g_{4n+4}\mathbf{J}_{4n+4}^{\mathrm{T}} & \mathbf{D}_{4n+3}^{\mathrm{T}} & f_{4n+2}\mathbf{P}_{4n+2}^{\mathrm{T}} & \mathbf{B}_{4n+3}^{\mathrm{T}} \\ \mathbf{0} & g_{4n+3}\mathbf{J}_{4n+3}^{\mathrm{T}} & \mathbf{D}_{4n+2}^{\mathrm{T}} & f_{4n+1}\mathbf{P}_{4n+1}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{0} & g_{4n+2}\mathbf{J}_{4n+2}^{\mathrm{T}} & \mathbf{D}_{4n+1}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & g_{4n+1}\mathbf{J}_{4n+1}^{\mathrm{T}} \end{pmatrix}, \tag{A3}$$

where the superscript T stands for ''transposed,'' and

$$f_n=-\frac{2n-11}{2n+9},\quad g_n=-\frac{n-4}{n+1}. \tag{A4}$$

The dimensions of the matrices $\mathbf{Q}_n$, $\mathbf{Q}_n^+$, and $\mathbf{Q}_n^-$ are, respectively, $(8n-2)\times(8n-2)$, $(8n-2)\times(8n+6)$, and $(8n-2)\times(8n-10)$. An exception is the matrix $\mathbf{Q}_1^-$, which degenerates into a vector of dimension 6.

In Eqs. (A1)–(A3), the submatrices $\mathbf{A}_{4n}$, $\mathbf{A}_{4n-1}$, $\mathbf{A}_{4n-2}$, $\mathbf{A}_{4n-3}$, $\mathbf{B}_{4n-1}$, $\mathbf{B}_{4n-2}$, $\mathbf{B}_{4n-3}$, $\mathbf{D}_{4n-1}$, $\mathbf{P}_{4n-1}$, and $\mathbf{P}_{4n-2}$ can be represented as follows:

$$\mathbf{X}_{4n-i}=\begin{pmatrix} x_{4n-i,-4(n-1+\delta_{i0})} & x^+_{4n-i,-4(n-1+\delta_{i0})} & 0 & \cdots & 0 \\ x^-_{4n-i,-4(n-2+\delta_{i0})} & x_{4n-i,-4(n-2+\delta_{i0})} & x^+_{4n-i,-4(n-2+\delta_{i0})} & \cdots & 0 \\ 0 & x^-_{4n-i,-4(n-3+\delta_{i0})} & x_{4n-i,-4(n-3+\delta_{i0})} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x_{4n-i,4(n-1+\delta_{i0})} \end{pmatrix} \tag{A5}$$

($i=0,1,2,3$). They are of dimension $[2(n+\delta_{i0})-1]\times[2(n+\delta_{i0})-1]$. The submatrices $\mathbf{B}_{4n}$, $\mathbf{D}_{4n}$, $\mathbf{J}_{4n}$, $\mathbf{P}_{4n}$, $\mathbf{D}_{4n-2}$, $\mathbf{D}_{4n-3}$, $\mathbf{J}_{4n-1}$, $\mathbf{J}_{4n-2}$, $\mathbf{J}_{4n-3}$, and $\mathbf{P}_{4n-3}$ have the form

$$\mathbf{X}_{4n-i}=\begin{pmatrix} x^+_{4n-i,-4(n-1+\delta_{i0})} & 0 & 0 & \cdots & 0 \\ x_{4n-i,-4(n-2+\delta_{i0})} & x^+_{4n-i,-4(n-2+\delta_{i0})} & 0 & \cdots & 0 \\ x^-_{4n-i,-4(n-3+\delta_{i0})} & x_{4n-i,-4(n-3+\delta_{i0})} & x^+_{4n-i,-4(n-3+\delta_{i0})} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x^-_{4n-i,-4(n-1+\delta_{i0})} \end{pmatrix} \tag{A6}$$

($i=0,1,2,3$) and are of dimension $[2(n+\delta_{i0})-1]\times[2(n+\delta_{i0})-3]$. The submatrix elements in (A5) and (A6) are given by the formulas

$$a_{n,m}=d_{n,m,0,0}=\sigma\frac{9(n-1)n(n+1)(n+2)-15m^2[6n(n+1)-5-7m^2]}{(2n-3)(2n-1)(2n+3)(2n+5)}-\frac{n(n+1)}{2},$$

$$a^-_{n,m}=a^+_{n,-m}=d_{n,m,0,-1}=\frac{15\sigma\sqrt{(n+m)(n-m+4)[n^2-(m-3)^2][n^2-(m-2)^2][n^2-(m-1)^2]}}{2(2n-3)(2n-1)(2n+3)(2n+5)},$$

$$b_{n,m}=d_{n,m,-1,0}=-\frac{3i\sigma m(3n^2-5-7m^2)}{\alpha(4n^2-9)}\sqrt{\frac{n^2-m^2}{4n^2-1}},$$

$$b^-_{n,m}=-b^+_{n,-m}=d_{n,m,-1,-1}=-\frac{3i\sigma}{2\alpha(4n^2-9)}\sqrt{\frac{(n+m-4)(n+m)[n^2-(m-3)^2][n^2-(m-2)^2][n^2-(m-1)^2]}{4n^2-1}},$$

$$p_{n,m}=d_{n,m,-2,0}=\frac{\sigma(2n+9)(n^2-n-2-7m^2)}{(2n-5)(2n-1)(2n+3)}\sqrt{\frac{(n^2-m^2)[(n-1)^2-m^2]}{(2n+1)(2n-3)}},$$

$$p^-_{n,m}=p^+_{n,-m}=d_{n,m,-2,-1}$$
$$=-\frac{\sigma(2n+9)}{2(2n-5)(2n-1)(2n+3)}\sqrt{\frac{(n+m-5)(n+m-4)(n+m-3)(n+m)[n^2-(m-2)^2][n^2-(m-1)^2]}{(2n+1)(2n-3)}},$$

$$d_{n,m}=d_{n,m,-3,0}=-\frac{7i\sigma m}{\alpha(2n-3)(2n-1)}\sqrt{\frac{(n^2-m^2)[(n-1)^2-m^2][(n-2)^2-m^2]}{(2n-5)(2n+1)}},$$

$$d^-_{n,m}=-d^+_{n,-m}=d_{n,m,-3,-1}=\frac{i\sigma}{2\alpha(2n-3)(2n-1)}\sqrt{\frac{(n+m-6)(n+m-5)\cdots(n+m-1)(n+m)(n-m+1)}{(2n-5)(2n+1)}},$$

$$j_{n,m}=d_{n,m,-4,0}=\frac{7\sigma(n+1)}{(2n-5)(2n-3)(2n-1)}\sqrt{\frac{[(n-3)^2-m^2][(n-2)^2-m^2][(n-1)^2-m^2](n^2-m^2)}{(2n-7)(2n+1)}},$$

$$j^{-}_{n,m}=j^{+}_{n,-m}=d_{n,m,-4,-1}=\frac{\sigma(n+1)}{2(2n-5)(2n-3)(2n-1)}\sqrt{\frac{(n+m-7)(n+m-6)\cdots(n+m-1)(n+m)}{(2n-7)(2n+1)}}$$

we have allowed for the fact that the coefficients $d_{n,m,r,s}$ obey the following relationships:

$$d_{n,m,r,1}=d^{*}_{n,-m,r,-1},\quad d_{n,m,1,s}=d^{*}_{n+1,-m-4s,-1,s},$$

$$d_{n,m,3,s}=d^{*}_{n+3,-m-4s,-3,s},$$

$$d_{n,m,2,s}=f_{n+2}d_{n+2,-m-4s,-2,s},$$

$$d_{n,m,4,s}=g_{n+4}d_{n+4,-m-4s,-4,s},$$

where $s=0,-1$, and f_n and g_n are defined in (A4).

APPENDIX B: SOLUTION OF EQUATION (24)

By using the Laplace transformation we can reduce Eq. (24) to

$$\mathbf{Q}^{-}_{n}\tilde{\mathbf{C}}_{n-1}(s)+|\mathbf{Q}_n-s\tau_N\mathbf{I}|\tilde{\mathbf{C}}_n(s)+\mathbf{Q}^{+}_{n}\tilde{\mathbf{C}}_{n+1}(s)=-\tau_N\mathbf{C}_n(0), \tag{B1}$$

where $\mathbf{I}$ is the identity matrix, and

$$\tilde{\mathbf{C}}_n(s)=\int_0^\infty \mathbf{C}_n(t)e^{-st}\,dt. \tag{B2}$$

Following Refs. 22 and 23, we seek the solution $\tilde{\mathbf{C}}_n(s)$ in the form

$$\tilde{\mathbf{C}}_n(s)=\mathbf{S}_n(s)\tilde{\mathbf{C}}_{n-1}(s)+\mathbf{U}_n(s), \tag{B3}$$

where $\mathbf{S}_n(s)$ is a matrix continued fraction given by (27). Substituting (B3) into (B1) and allowing for the fact that, according to the definition of the continued fraction (27)

$$\mathbf{S}_n(s)=[s\tau_N\mathbf{I}-\mathbf{Q}_n-\mathbf{Q}^{+}_{n}\mathbf{S}_{n+1}(s)]^{-1}\mathbf{Q}^{-}_{n},$$

we arrive at the recurrence equation

$$\mathbf{U}_n(s)=\mathbf{S}_n(s)(\mathbf{Q}^{-}_{n})^{-1}[\tau_N\mathbf{C}_n(0)+\mathbf{Q}^{+}_{n}\mathbf{U}_{n+1}(s)]. \tag{B4}$$

This equation can be solved by successive substitutions and has the form

$$\mathbf{U}_n(s)=\tau_N\mathbf{S}_n(s)(\mathbf{Q}^{-}_{n})^{-1}\Bigg\{\mathbf{C}_n(0)+\sum_{k=1}^{\infty}\Bigg[\prod_{m=1}^{k}\mathbf{Q}^{+}_{n+m-1}\mathbf{S}_{n+m}(s)\times(\mathbf{Q}^{-}_{n+m})^{-1}\Bigg]\mathbf{C}_{n+k}(0)\Bigg\}. \tag{B5}$$

Thus, allowing for (B3) and (B5), we obtain the desired solution (26) if we put $n=1$ and $\tilde{\mathbf{C}}_0(0)=0$.

APPENDIX C: CALCULATING THE INITIAL- VALUE VECTORS

It is convenient to calculate the initial-value vectors $\mathbf{C}_n(0)$ in (26) by using matrix continued fractions.[22,23] According to (14), the initial values $c_{n,m}(0)$ have the form

$$c_{n,m}(0)=\langle\cos\vartheta(0)Y_{n,m}(0)\rangle_0=\sqrt{\frac{(n+1)^2-m^2}{(2n+1)(2n+3)}}\,\langle Y_{n+1,m}\rangle_0+\sqrt{\frac{n^2-m^2}{(2n+1)(2n-1)}}\,\langle Y_{n-1,m}\rangle_0. \tag{C1}$$

In accordance with (13), the equilibrium averages $\langle Y_{n,m}\rangle_0$ satisfy the recurrence relation

$$\sum_{s=-1}^{1}\sum_{r=-4}^{4}d_{n,m,r,s}\langle Y_{n+r,m+4s}\rangle_0=0, \tag{C2}$$

which can be written in the form of a matrix recurrence relation:

$$\mathbf{Q}^{-}_{n}\mathbf{R}_{n-1}+\mathbf{Q}_n\mathbf{R}_n+\mathbf{Q}^{+}_{n}\mathbf{R}_{n+1}=\mathbf{0},\quad n=1,2,3,\ldots, \tag{C3}$$

where the matrices $\mathbf{Q}_n$, $\mathbf{Q}^{+}_{n}$, and $\mathbf{Q}^{-}_{n}$ are given in (A1)–(A3), and

$$\mathbf{R}_n=\begin{pmatrix}\mathbf{r}_{4n}\\ \mathbf{r}_{4n-1}\\ \mathbf{r}_{4n-2}\\ \mathbf{r}_{4n-3}\end{pmatrix},\quad \mathbf{r}_{4n-i}=\begin{pmatrix}\langle Y_{4n-i,-4(n-1+\delta_{i0})}\rangle_0\\ \langle Y_{4n-i,-4(n-2+\delta_{i0})}\rangle_0\\ \vdots\\ \langle Y_{4n-i,4(n-1+\delta_{i0})}\rangle_0\end{pmatrix},$$

$i=0,1,2,3,\ldots$.

The solution of Eq. (C3) has the form

$$\mathbf{R}_n=\mathbf{S}_n(0)\mathbf{R}_{n-1}=\frac{\mathbf{S}_n(0)\mathbf{S}_{n-1}(0)\cdots\mathbf{S}_2(0)\mathbf{S}_1(0)}{\sqrt{4\pi}}, \tag{C4}$$

where $\mathbf{S}_n(0)$ is the matrix continued fraction defined by (27) with $s=0$ and where we have allowed for the fact that $\mathbf{R}_0=1/\sqrt{4\pi}$.

Using (C4), we can write the initial-value vectors $\mathbf{C}_n(0)$ as

$$\mathbf{C}_n(0)=\frac{1}{\sqrt{4\pi}}[\hat{\mathbf{K}}_n+[\mathbf{K}_n+\hat{\mathbf{K}}^{\mathrm{T}}_{n+1}\mathbf{S}_{n+1}(0)]\times\mathbf{S}_n(0)]\mathbf{S}_{n-1}(0)]\cdots\mathbf{S}_1(0), \tag{C5}$$

where the matrices $\mathbf{K}_n$ and $\hat{\mathbf{K}}_n$ have the form

$$\mathbf{K}_n=\begin{pmatrix} \mathbf{0} & \mathbf{U}_{4n} & \mathbf{0} & \mathbf{0} \\ \mathbf{U}_{4n}^{\mathrm{T}} & \mathbf{0} & \mathbf{U}_{4n-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{4n-1}^{\mathrm{T}} & \mathbf{0} & \mathbf{U}_{4n-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_{4n-2}^{\mathrm{T}} & \mathbf{0} \end{pmatrix}, \quad \hat{\mathbf{K}}_n=\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{U}_{4n-3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{C6}$$

with the matrix $\hat{\mathbf{K}}_1$ degenerating into a vector of dimension 6. The submatrices $\mathbf{U}_{4n-1}$, $\mathbf{U}_{4n-2}$, and $\mathbf{U}_{4n-3}$ in (C6) are

$$\mathbf{U}_{4n-i}=\begin{pmatrix} u_{4n-i,-4(n-1)} & 0 & 0 & \cdots & 0 \\ 0 & u_{4n-i,-4(n-2)} & 0 & \cdots & 0 \\ 0 & 0 & u_{4n-i,-4(n-3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{4n-i,4(n-1)} \end{pmatrix} \tag{C7}$$

$(i=1,2,3)$ and are of dimension $(2n-1)\times(2n-1)$. The submatrix $\mathbf{U}_{4n}$ is

$$\mathbf{U}_{4n}=\begin{pmatrix} 0 & 0 & \cdots & 0 \\ u_{4n,-4n+4} & 0 & \cdots & 0 \\ 0 & u_{4n,-4n+8} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{4n,4n-4} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{C8}$$

and is of dimension $(2n+1)\times(2n-1)$. The elements of the submatrices (C7) and (C8) are

$$u_{n,m}=\sqrt{\frac{n^2-m^2}{4n^2-1}}.$$

*)E-mail: ypk169@ire216.msk.su
†)E-mail: svt245@ire216.msk.su

[1)]Discussions of the theoretical and experimental methods used in estimating the dissipation parameter α can be found, say, in Refs. 6 and 13. These methods yield values of α of order 0.01–0.1.

[2)]In experiments involving dynamic susceptibility measurements, the common system of particles is the one in which the anisotropy axes of the particles are oriented at random. In this case the magnetic susceptibility has the form $\chi=(\chi_\parallel+\chi_\perp)/3$.

[3)]The method of calculating the dynamic susceptibility used in this paper is of a very general nature and can be employed in studies of relaxation processes and ferromagnetic resonance of single-domain particles with magnetic anisotropy of various types in systems in strong external d.c. and a.c. magnetic fields.

[1] L. Neel, Ann. Geophys. (C.N.R.S.) **5**, 99 (1949).
[2] W. F. Brown Jr., IEEE Trans. Magn. **15**, 1196 (1979).
[3] H. B. Braun and H. N. Bertram, J. Appl. Phys. **75**, 4609 (1994).
[4] W. F. Brown Jr., J. Appl. Phys. **30**, 130S (1959).
[5] A. Aharoni, Phys. Rev. **177**, 763 (1969).
[6] Yu. L. Raĭkher and M. I. Shliomis, Zh. Éksp. Teor. Fiz. **67**, 1060 (1974) [Sov. Phys. JETP **40**, 526 (1975)].
[7] D. A. Garanin, V. V. Ishchenko, and L. V. Panina, Teor. Mat. Fiz. **82**, 242 (1990).
[8] Yu. L. Raĭkher and V. I. Stepanov, Zh. Éksp. Teor. Fiz. **102**, 1409 (1992) [Sov. Phys. JETP **75**, 764 (1992)].
[9] W. T. Coffey, D. S. F. Crothers, Yu. P. Kalmykov, and J. T. Waldron, Phys. Rev. B **51**, 15 947 (1995).
[10] D. A. Garanin, Phys. Rev. E **54**, 3250 (1996).
[11] É. K. Sadykov and A. G. Isavnin, Fiz. Tverd. Tela (St. Petersburg) **38**, 2104 (1997) [Phys. Solid State **38**, 1160 (1996)].
[12] Yu. P. Kalmykov and W. T. Coffey, Phys. Rev. B **56**, 3325 (1997).
[13] I. Klik and L. Gunther, J. Stat. Phys. **60**, 473 (1990).
[14] A. Aharoni, Phys. Rev. B **7**, 1103 (1973).
[15] D. A. Smith and F. A. de Rosario, J. Magn. Magn. Mater. **3**, 219 (1976).
[16] I. Eisenstein and A. Aharoni, Phys. Rev. B **16**, 1278 (1977).
[17] I. Eisenstein and A. Aharoni, Phys. Rev. B **16**, 1285 (1977).
[18] I. Klik and L. Gunther, J. Appl. Phys. **67**, 4509 (1990).
[19] L. J. Geoghegan, W. T. Coffey, and B. Mulligan, in *Advances in Chemical Physics*, Vol. 100, I. Prigogine and S. A. Rice (Eds.), Wiley, New York (1997), p. 475.
[20] W. F. Brown Jr., Phys. Rev. **130**, 1677 (1964).
[21] Yu. L. Raikher and M. I. Shliomis, in *Advances in Chemical Physics*, Vol. 87, W. T. Coffey (Ed.), Wiley, New York (1994), p. 595.
[22] H. Risken, *The Fokker–Planck Equation. Method of Solution and Applications*, 2nd ed., Springer, Berlin (1989).
[23] W. T. Coffey, Yu. P. Kalmykov, and J. T. Waldron, *The Langevin Equation*, World Scientific, Singapore (1996).
[24] T. L. Gilbert, Phys. Rev. **100**, 1243 (1956).
[25] R. L. Stratonovich of, *Selected Problems on Fluctuation Theory in Radio Engineering* [in Russian], Sovet-skoe Radio, Moscow (1961).
[26] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[27] Yu. P. Kalmykov, Khim. Fiz. **16**, 130 (1997).
[28] H. J. Breymayer, H. Risken, H. D. Vollmer, and W. Wonneberger, Appl. Phys. B **28**, 335 (1982).
[29] T. Bitoh, K. Ohba, M. Takamatsu, T. Shirane, and S. Chikazawa, J. Phys. Soc. Jpn. **64**, 1311 (1995).
[30] T. Bitoh, K. Ohba, M. Takamatsu, T. Shirane, and S. Chikazawa, J. Magn. Magn. Mater. **154**, 59 (1996).

Translated by Eugene Yankovsky

Spontaneous optical-coherence transfer and the nonlinear spectroscopy of gases

S. G. Rautian*)

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 9 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 12–29 (January 1999)

Nonlinear-spectroscopy population effects due to spontaneous transfer of optical coherence are examined. The existence in the velocity distribution of new resonance elements of a specific form is established. The occurrence of these effects in the nonlinear resonance of a counterpropagating wave is analyzed. © *1999 American Institute of Physics.* [S1063-7761(99)00201-2]

1. INTRODUCTION

The main physical ''signature'' of the nonlinear spectroscopy of rarefied gases is the Bennett structure, which is a system of fairly sharp ''peaks'' and ''dips'' in the distribution of the atoms in the velocity projected on the wave vector of a monochromatic field resonantly interacting with the atoms.[1] The spectral manifestations of the Bennett structure are closely linked to two other fundamental phenomena, the field-induced splitting of levels and the nonlinear interference effects, and critically depend on many factors, such as the type of radiative process (absorption, scattering etc.), the spectral and spatial properties of the field, the field intensity and polarization, the nature of the collisions, the angular momenta of the levels, the oscillator strengths, and the external fields (see, e.g., Refs. 2–4).

The shape of the components of the Bennett structure follows the spectral lineshape of an immobile atom and in many cases is Lorentzian. Generally speaking, variations in the atomic velocities due to collisions affect the contours of the Bennett peaks and dips, sometimes significantly, but do not alter the dome-like shape of the contours. This ''stability'' of the shape of Bennett structures results from allowing only for radiative and collisional transitions in the number of particles. If magnetic- and optical-coherence transfers are taken into account, the situation changes dramatically. This is a fact well known in the theory of collisional collapse of spectral lines and in the spectroscopy of magnetooptic resonances. For instance, a spontaneous magnetic-coherence cascade generates a resonance whose contour is described by the product of two complex-valued Lorentzians and alternates in sign (see, e.g., Refs. 5 and 6). The spectral line reflecting radiative optical-coherence transfer undergoes similar changes. More precisely, in Refs. 7–9 it was established that radiative optical-coherence transfer from a high-lying transition m_1–n_1 (Fig. 1) to a low-lying transition m–n constitutes a special type of radiative process, which differs from the Bohr–Einstein process (particle transfer) and the Barrat–Cohen-Tannoudji process (magnetic-coherence transfer). While the latter processes determine the intensity and polarization, optical-coherence transfer generates a characteristic interference, alternating-in-sign, structure, as was shown in Refs. 7–9 for linear emission, absorption and refraction spectra.

In view of all this, we should expect that in the event of optical-coherence transfer the fine nonlinear structure in the velocity distribution will differ from the ordinary Bennett structure due to the effect of the strong field on the level populations. In this paper we will focus on this problem. In Secs. 2 and 3 we will use the simple model of nondegenerate states to derive a velocity distribution that allows for coherence transfer. In Sec. 4 we analyze, under the same conditions, the nonlinear resonance of a counterpropagating probe wave. In Sec. 5 we generalize the results to degenerate levels and discuss polarization effects.

2. THE MAIN RELATIONSHIPS

Radiative optical-coherence transfer has much in common with collisional transfer[10] but differs from the latter in the origin or mechanism of transfer. In contrast to the collisional process, radiative spontaneous optical-coherence transfer has a unidirectional cascade nature. The spontaneous transfer of optical coherence is universal and is virtually independent of the external conditions (one exception is radiating systems in a high-Q cavity that interact with a small number of modes). Below we examine the process of spontaneous optical-coherence transfer for a system with a level diagram sketched in Fig. 1.

To make things clear and simple, we begin with the model of nondegenerate states. The fact that degeneracy actually exists strongly complicates analysis due to the interaction of transitions with $\Delta M=0$ and $\Delta M=\pm 1$ (M is the magnetic quantum number), but at the same time has little effect on the qualitative picture of the phenomena. It is also known that some simple systems with small angular momenta are described directly by the model of nondegenerate states. In Sec. 5 we will show that a similar rule holds for our problem. Note that in the level diagram depicted in Fig. 1 four transitions are allowed: m_1–n_1, m–n, m_1–m, and n_1–n. Moreover, for optical-coherence transfer to manifest itself the difference $\Delta=\omega_{m_1n_1}-\omega_{mn}$ must be small.

Suppose that the electromagnetic field is a plane monochromatic wave (frequency ω, wave vector $\mathbf{k}$, and amplitude

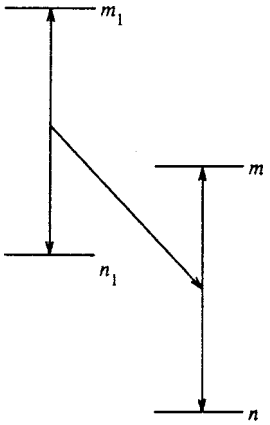


FIG. 1. The transition diagram for a four-level system. The vertical arrows stand for polarizations induced by a resonant field, and the connecting slanted arrow stands for the optical-coherence transfer.

E). We use the standard resonance approximation and the model of relaxation constants (see, e.g., Refs. 2 and 4). Under the above restrictions and steady-state conditions, the off-diagonal elements ρ and ρ_1 of the density matrix for the transitions m–n and m_1–n_1, respectively, obey the equations[8]

$$\begin{aligned}&(\Gamma - i\Omega')\rho = iG(\rho_n - \rho_m) + A\rho_1,\\ &(\Gamma_1 - i\Omega_1')\rho_1 = iG_1(\rho_{n_1} - \rho_{m_1}).\end{aligned} \tag{2.1}$$

Here the ρ_j $(j = m,n,m_1,n_1)$ are the diagonal elements of the density matrix, and Γ and Γ_1 are the relaxation constants. The term $A\rho_1$ describes optical-coherence transfer process, with A the rate of spontaneous optical-coherence transfer:

$$\begin{aligned}&A = K\sqrt{A_{m_1 m}A_{n_1 n}}, \qquad (2.2)\\ &G = d_{mn}E/2h, \quad G_1 = d_{m_1 n_1}E/2h, \quad \Omega' = \Omega - \mathbf{k}\cdot\mathbf{v},\\ &\Omega = \omega - \omega_{mn}, \quad \Omega_1' = \Omega' - \Delta = \Omega_1 - \mathbf{k}\cdot\mathbf{v},\\ &\Omega_1 = \omega - \omega_{m_1 n_1}, \quad \Delta = \omega_{m_1 n_1} - \omega_{mn} = \omega_{m_1 m} - \omega_{n_1 n},\end{aligned}$$

where d_{ij}, ω_{ij}, and A_{ij} are the dipole-moment matrix elements, the Bohr frequencies, and Einstein's first coefficients for the i–j transition, and K is a proportionality coefficient of order unity, which actually depends on the degeneracy of the level and will be specified in Sec. 5. According to Eqs. (2.1),

$$\begin{aligned}&\rho_1 = \frac{iG_1}{\Gamma_1 - i\Omega_1'}(\rho_{n_1} - \rho_{m_1}),\\ &\rho = \frac{i}{\Gamma - i\Omega'}\left[G(\rho_n - \rho_m) + \frac{AG_1}{\Gamma_1 - i\Omega_1'}(\rho_{n_1} - \rho_{m_1})\right].\end{aligned} \tag{2.3}$$

In weak fields the level populations ρ_j are assumed fixed. Equation (2.3) has been used in Refs. 7 and 8 to describe the role of optical-coherence transfer in this approximation. To analyze the nonlinear phenomena, which are of interest to us and are related to the variations in the populations caused by the field, we examine the equations for ρ_j:

$$\begin{aligned}&\Gamma_{m_1}\rho_{m_1} = -2\mathrm{Re}(iG_1^*\rho_1) + Q_{m_1},\\ &\Gamma_{n_1}\rho_{n_1} = 2\mathrm{Re}(iG_1^*\rho_1) + Q_{n_1} + A_{m_1 n_1}\rho_{m_1},\end{aligned} \tag{2.4}$$

$$\begin{aligned}&\Gamma_m\rho_m = -2\mathrm{Re}(iG^*\rho) + Q_m + A_{m_1 m}\rho_{m_1},\\ &\Gamma_n\rho_n = 2\mathrm{Re}(iG^*\rho) + Q_n + A_{mn}\rho_m + A_{n_1 n}\rho_{n_1},\end{aligned} \tag{2.5}$$

where the Q_j are the excitation rates of the levels j. The terms $A_{ij}\rho_i$ describe spontaneous population cascades, which, obviously, should be taken into account in examining optical-coherence transfer. Substituting (2.3) in Eqs. (2.4) and (2.5) yields the equations

$$\begin{aligned}&\Gamma_{m_1}\rho_{m_1} + w_1(\rho_{m_1} - \rho_{n_1}) = Q_{m_1},\\ &(\Gamma_{m_1} - A_{m_1 n_1})\rho_{m_1} + \Gamma_{n_1}\rho_{n_1} = Q_{m_1} + Q_{n_1},\end{aligned} \tag{2.6}$$

$$\begin{aligned}&\Gamma_m\rho_m + w(\rho_m - \rho_n) = Q_m + A_{m_1 m}\rho_{m_1} + W(\rho_{n_1} - \rho_{m_1}),\\ &(\Gamma_m - A_{mn})\rho_m + \Gamma_n\rho_n = Q_m + Q_n + A_{m_1 m}\rho_{m_1} + A_{n_1 n}\rho_{n_1},\end{aligned} \tag{2.7}$$

where we have introduced the following notation for the stimulated transition rates:

$$\begin{aligned}&w = \frac{2\Gamma|G|^2}{\Gamma^2 + \Omega'^2}, \quad w_1 = \frac{2\Gamma_1|G_1|^2}{\Gamma_1^2 + \Omega_1'^2},\\ &W = \mathrm{Re}\,\frac{2G^*G_1A}{(\Gamma - i\Omega')(\Gamma_1 - i\Omega_1')}.\end{aligned} \tag{2.8}$$

The system of equations (2.6) contains only the populations ρ_{m_1} and ρ_{n_1} of the levels of the ''upper'' transition. Its solutions enter into the right-hand sides of the equations of system (2.7), in accordance with the cascade nature of the spontaneous optical-coherence transfer. The physical meaning of Eqs. (2.6) and (2.7) is well known. An essentially new term corresponding to optical-coherence transfer is $W(\rho_{n_1} - \rho_{m_1})$. We see that the coherent process of optical-coherence transfer leads to a cascade population flux. Similar to Einstein stimulated processes (which are transitions due to absorption and stimulated emission), the ''coherent'' transfer has no effect on the total excitation flux for the levels m and n [the second equation in (2.7)]. As in the case of Einstein processes, the coherence transfer flux is proportional to the difference $\rho_{n_1} - \rho_{m_1}$ of the populations at the ''upper'' transition, which emphasizes its cascade origin. From (2.8) we see that the rate W is proportional to the field intensity $|E|^2$, and in this respect optical-coherence transfer is no different from other nonlinear effects. Also, $W \propto d_{mn}d_{m_1 n_1}A \propto \sqrt{A_{mn}A_{m_1 n_1}A_{m_1 m}A_{n_1 n}}$, which reflects the interference properties of optical-coherence transfer as a coherent process. In contrast to w and w_1, which are proportional to ordinary Lorentz factors, the frequency dependence of W is determined by a product of two complex-valued Lorentzians and is found to alternate in sign. One can easily show that the integral of W with respect to the frequency ω vanishes, as it should for an interference effect. However, the integral of the

flux $W(\rho_{n_1}-\rho_{m_1})$ with respect to ω is finite, since in the presence of nonlinear phenomena the populations are frequency-dependent.

Absorption (amplification) of the field is determined by the work done by the field:

$$P=-2h\omega\mathrm{Re}\langle i(G^*\rho+G_1^*\rho_1)\rangle$$
$$=h\omega\langle w(\rho_n-\rho_m)+(w_1+W)(\rho_{n_1}-\rho_{m_1})\rangle, \qquad (2.9)$$

where the angle brackets stand for averaging over the velocities. The dependence of P on the frequency Ω stems from two factors: the lineshape of a single atom, which is specified by w, w_1, and W, and the variation of the difference of populations as functions of Ω. Like the classical statement of the Karplus–Schwinger problem,[11] the absorption lineshape for a strong monochromatic field changes, according to (2.9), because of the second factor, the frequency dependence of the field-induced variation $\rho_i-\rho_j$. Thus, optical-coherence transfer does not alter this fundamental conclusion of the theory.

3. VELOCITY DISTRIBUTION OF THE ATOMS

When the atoms interact with a plane monochromatic wave and the Doppler broadening is large, the velocity distribution of the atoms acquires a characteristic Bennett structure, which has been studied in detail in the two-level model (see, e.g., Refs. 2–4). Below we analyze such a structure in the four-level system with optical-coherence transfer.

The solution of the system of equations (2.7) for $\rho_n-\rho_m$ can be represented as

$$\rho_n-\rho_m=N_n-N_m$$
$$-\frac{1}{1+Tw}\left\{Tw(N_n-N_m)+TW(\rho_{n_1}-\rho_{m_1})\right.$$
$$-\frac{A_{n_1n}(\rho_{n_1}-N_{n_1})}{\Gamma_n}$$
$$\left.+\frac{A_{m_1m}(\Gamma_n-A_{mn})(\rho_{m_1}-N_{m_1})}{\Gamma_m\Gamma_n}\right\}, \qquad (3.1)$$

where the N_j have the meaning of populations of the levels $j=m,n,m_1,n_1$ in the absence of a field, and T is the effective time of interaction of the field and the two-level system m,n:

$$N_{m_1}=\frac{Q_{m_1}}{\Gamma_{m_1}}, \quad N_{n_1}=\frac{Q_{n_1}+A_{m_1n_1}N_{m_1}}{\Gamma_{n_1}},$$
$$N_m=\frac{Q_m+A_{m_1m}N_{m_1}}{\Gamma_m}, \quad N_n=\frac{Q_n+A_{mn}N_m+A_{n_1n}N_{n_1}}{\Gamma_n}, \qquad (3.2)$$
$$T=\frac{\Gamma_m+\Gamma_n-A_{mn}}{\Gamma_m\Gamma_n}.$$

We will assume that the N_j have a Maxwellian velocity distribution:

$$N_j=\frac{N_{j0}}{(\sqrt{\pi}\,\overline{v})^3}\exp\left(-\frac{v^2}{\overline{v^2}}\right), \quad \overline{v^2}=\frac{2T_a}{m}, \qquad (3.3)$$

where T_a and m are the atomic temperature and mass.

The expression inside the braces in Eq. (3.1) consists of terms describing the field-induced variation of the population difference, where the last two terms are related to cascade population transfer (effects of the type discussed in Ref. 12) and of interest to us only in second order. We focus on the term containing the rate W; this term describes the variation of the populations due to coherence transfer from the m_1–n_1 transition to the m–n transition. By its very appearance this term differs little from its analog related to ''ordinary'' transitions: instead of the rate w of stimulated Einstein transitions we have W, and instead of the difference N_n-N_m we have the saturated population difference of the ''upper'' transition, $\rho_{n_1}-\rho_{m_1}$. Hence the population variation due to optical-coherence transfer experiences saturation both on the ''upper'' transition and on the ''lower'' transition [the factor $1/(1+Tw)$]. The components of $\rho_n-\rho_m$ due to a population cascade have the same property [the last two terms in Eq. (3.1)].

We now write the well-known expressions for the populations of the levels belonging to the ''upper'' transition, which enter into (3.1) and are simpler because they are not ''burdened'' by cascade processes:

$$\rho_{m_1}-N_{m_1}=\frac{\Gamma_1^2\kappa_1}{\Gamma_{1s}^2+\Omega_1'^2}\frac{1}{T_1\Gamma_{m_1}}(N_{n_1}-N_{m_1}),$$
$$T_1=\frac{\Gamma_{m_1}+\Gamma_{n_1}-A_{m_1n_1}}{\Gamma_{m_1}\Gamma_{n_1}},$$
$$\rho_{n_1}-N_{n_1}=-\frac{\Gamma_1^2\kappa_1}{\Gamma_{1s}^2+\Omega_1'^2}\left(1-\frac{1}{T_1\Gamma_{m_1}}\right)(N_{n_1}-N_{m_1}),$$
$$\kappa_1=\frac{2|G_1|^2T_1}{\Gamma_1}, \qquad (3.4)$$
$$\rho_{n_1}-\rho_{m_1}=\left(1-\frac{\Gamma_1^2\kappa_1}{\Gamma_{1s}^2+\Omega_1'^2}\right)(N_{n_1}-N_{m_1}),$$
$$\kappa=\frac{2|G|^2T}{\Gamma},$$
$$\Gamma_{1s}^2=\Gamma_1^2(1+\kappa_1), \quad \Gamma_s^2=\Gamma^2(1+\kappa).$$

Equations (2.9), (3.1), and (3.4) imply that as the field intensity tends to infinity, the population differences $\rho_n-\rho_m$ and $\rho_{n_1}-\rho_{m_1}$ tend to zero as $1/|E|^2$, while the work done by the field, P, tends to a finite value. Hence these properties of the classical saturation effect are not violated by optical-coherence transfer.

It should be recalled that Ω' and Ω_1' are linear functions of the projection of the atomic velocity $\mathbf{v}$ on the wave vector $\mathbf{k}$ [see Eq. (2.2)], so that the dependence on Ω' and Ω_1' represents the velocity distribution. From Eqs. (2.8), (3.1), and (3.4) it follows that the distinctive pattern in the four-level system considered here consists of Bennett ''peaks''

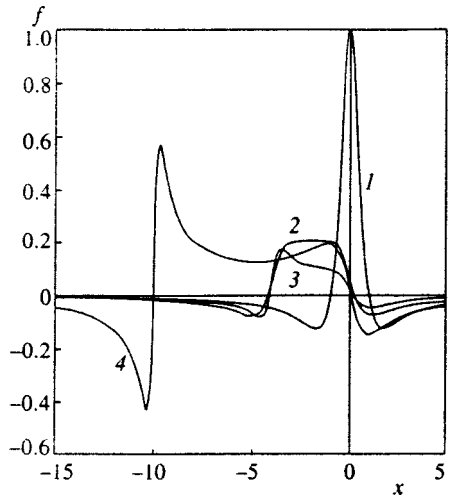


FIG. 2. The graphs of $f(v)$ as functions of $x=(kv-\Omega)/\Gamma$: curve *1*, $\Delta=0$ and $\Gamma=\Gamma_1$; curve *2*, $\Delta=4\Gamma$ and $\Gamma=\Gamma_1$; curve *3*, $\Delta=4\Gamma$ and $\Gamma=2\Gamma_1$; and curve *4*, $\Delta=10\Gamma$ and $\Gamma=3\Gamma_1$ (the function $f(v)$ has been multiplied by 10).

and ''dips'' and specific components reflecting the optical-coherence transfer process. A Bennett structure has the same properties as a structure without optical-coherence transfer (Lorentzian contours with characteristic halfwidths Γ_s and Γ_{1s}). The structure with optical-coherence transfer is described by the term on the right-hand side of Eq. (3.1) containing W. The transition rate W as a function of $v=\mathbf{k}\cdot\mathbf{v}/k$ has a nonstandard form. We write

$$W=\frac{AG^*G_1}{\Gamma\Gamma_1}f(v),$$
$$f(v)=\frac{\Gamma\Gamma_1[\Gamma\Gamma_1+(\Delta/2)^2-(kv-\Omega+\Delta/2)^2]}{[\Gamma^2+(kv-\Omega)^2][\Gamma_1^2+(kv-\Omega+\Delta)^2]}. \tag{3.5}$$

In Fig. 2 the function $f(v)$ is plotted for some values of $|\Delta|/\Gamma$ and Γ/Γ_1. The integral of $f(v)$ with respect to v vanishes, and the function changes sign at

$$kv=\Omega-\Delta/2\pm\sqrt{(\Delta/2)^2+\Gamma/\Gamma_1}. \tag{3.6}$$

If $\Delta=0$ holds, the function $f(v)$ is symmetric about the point $kv=\Omega$,

$$f(v)=\frac{\Gamma\Gamma_1[\Gamma\Gamma_1-(kv-\Omega)^2]}{[\Gamma^2+(kv-\Omega)^2][\Gamma_1^2+(kv-\Omega)^2]}, \tag{3.7}$$

and (if we discard the negative tails) resembles a Lorentzian. But in the limit $|\Delta|\gg\Gamma,\Gamma_1$, near the points $kv=\Omega$ and $kv=\Omega-\Delta$ the function $f(v)$ is described fairly well not by Lorentzians but by dispersion curves:

$$f(v)\approx-\frac{\Gamma_1}{\Delta}\frac{\Gamma(kv-\Omega)}{\Gamma^2+(kv-\Omega)^2},\quad |kv-\Omega|\approx\Gamma,$$
$$f(v)\approx\frac{\Gamma}{\Delta}\frac{\Gamma_1(kv-\Omega+\Delta)}{\Gamma_1^2+(kv-\Omega+\Delta)^2},\quad |kv-\Omega+\Delta|\approx\Gamma_1 \tag{3.8}$$

(curve *4* in Fig. 2). The function $f(v)$ depends on three parameters, e.g., Γ/Γ_1, Δ/Γ, and Ω/Γ, but not on the characteristics of the Maxwell distribution. The center of this distribution corresponds to $kv=0$ and can be at any point on the horizontal axis in Fig. 2. Thus, optical-coherence transfer generates elements of a specific shape in the velocity distribution, elements that differ from the Bennett structure proper resulting from the interaction with the field and from the structure reflecting a cascade transfer of particles (populations).

4. NONLINEAR RESONANCE OF A COUNTERPROPAGATING PROBE WAVE

As is well known, when Doppler broadening is large, the work P done by a traveling wave, regarded as a function of frequency, has a Gaussian shape with a Doppler width,[2,4] i.e., a Bennett structure does not lead to narrow nonlinear resonances. Using Eqs. (2.8), (2.9), and (3.4), we can easily show that the above conclusion is also true in the case of optical-coherence transfer. The fine structure of the velocity distribution of the atoms can be detected by different versions of the probe-field method.[2–4] Here we use the method of a counterpropagating probe wave of the same frequency,[3] which is convenient from the experimental viewpoint since it requires the use of only one laser.

As is known, three types of nonlinear effects manifest themselves in the probe-field method: variation of populations by a strong field, field-induced splitting of the levels into quasienergy sublevels, and nonlinear interference effects.[2,4] When there is a counterpropagating probe wave and large Doppler broadening, field-induced splitting of levels and nonlinear interference effects do not manifest themselves in the first nonvanishing nonlinear corrections, i.e., only population effects are important in this approximation. Below we examine this simple case.

For the polarizations ρ_μ and $\rho_{1\mu}$ induced by the probe wave at the ''lower'' and ''upper'' transitions, respectively, we have equations of the form (2.1):

$$(\Gamma_1-i\Omega'_{1\mu})\rho_{1\mu}=iG_{1\mu}(\rho_{n_1}-\rho_{m_1}),$$
$$(\Gamma-i\Omega'_\mu)\rho=iG_\mu(\rho_n-\rho_m)+A\rho_{1\mu},$$
$$G_\mu=d_{mn}E_\mu/2h,\quad G_{1\mu}=d_{m_1n_1}E_\mu/2h, \tag{4.1}$$
$$\Omega'_\mu=\Omega+kv,\quad \Omega'_{1\mu}=\Omega_1+kv,$$

where E_μ is the amplitude of the probe wave, and the populations are independent of the characteristics of the probe wave and are given by formulas (3.1) and (3.4). The work done by the probe field is described by the expression

$$P_\mu=-2h\omega\mathrm{Re}\langle iG^*_\mu\rho_\mu+iG^*_{1\mu}\rho_{1\mu}\rangle$$
$$=2h\omega\mathrm{Re}\left\langle\frac{|G_\mu|^2}{\Gamma-i(\Omega+kv)}(\rho_n-\rho_m)+\left\{\frac{|G_{1\mu}|^2}{\Gamma_1-i(\Omega_1+kv)}\right.\right.$$
$$\left.\left.+\frac{AG^*_\mu G_{1\mu}}{[\Gamma-i(\Omega+kv)][\Gamma_1-i(\Omega_1+kv)]}\right\}(\rho_{n_1}-\rho_{m_1})\right\rangle. \tag{4.2}$$

Combining this with (3.1) and (3.4) are keeping the first nonlinear correction terms, we obtain

$$P_\mu = 2h\omega\left\langle \frac{\Gamma|G_\mu|^2}{\Gamma^2+(\Omega+kv)^2}\left[\left(1-\frac{\Gamma^2\kappa}{\Gamma^2+(\Omega-kv)^2}\right)\times(N_n-N_m)-\frac{2AT}{\Gamma\Gamma_1}G^*G_1 f(v)(N_{n_1}-N_{m_1})-\frac{T_2}{T_1}\frac{\Gamma_1^2\kappa_1}{\Gamma_1^2+(\Omega_1-kv)^2}(N_{n_1}-N_{m_1})\right]+\left[\frac{|G_{1\mu}|^2\Gamma_1}{\Gamma_1^2+(\Omega_1+kv)^2}+\frac{G_\mu^*G_{1\mu}A}{\Gamma\Gamma_1}f(-v)\right]\times\left[1-\frac{\Gamma_1^2\kappa_1}{\Gamma_1^2+(\Omega_1-kv)^2}\right](N_{n_1}-N_{m_1})\right\rangle, \tag{4.3}$$

where

$$T_2=\left(1-\frac{A_{mn}}{\Gamma_n}\right)\frac{A_{m_1m}}{\Gamma_{m_1}\Gamma_m}+\left(1-\frac{A_{m_1n_1}}{\Gamma_{m_1}}\right)\frac{A_{n_1n}}{\Gamma_n\Gamma_{n_1}}. \tag{4.4}$$

Equation (4.3) has eight different terms. Three terms correspond to linear absorption of the probe wave: two Voigt contours I_1 and I_2 for the doublet $\omega_{mn},\omega_{m_1n_1}$, and the interference contour I_3 described in Ref. 8:

$$I_1=\left\langle\frac{\Gamma^2}{\Gamma^2+(\Omega+kv)^2}\right\rangle,$$
$$I_2=\left\langle\frac{\Gamma_1^2}{\Gamma_1^2+(\Omega_1+kv)^2}\right\rangle,\quad I_3=\langle f(-v)\rangle. \tag{4.5}$$

The counterpropagating-wave resonances due to the ordinary Bennett structure, I_4 and I_5, and the cross-resonance due to a Bennett-structure population cascade, I_6, are proportional to the factors

$$I_4=\left\langle\frac{\Gamma^2}{\Gamma^2+(\Omega+kv)^2}\frac{\Gamma^2}{\Gamma^2+(\Omega-kv)^2}\right\rangle,$$
$$I_5=\left\langle\frac{\Gamma_1^2}{\Gamma_1^2+(\Omega_1+kv)^2}\frac{\Gamma_1^2}{\Gamma_1^2+(\Omega_1-kv)^2}\right\rangle, \tag{4.6}$$

$$I_6=\left\langle\frac{\Gamma^2}{\Gamma^2+(\Omega+kv)^2}\frac{\Gamma_1^2}{\Gamma_1^2+(\Omega_1-kv)^2}\right\rangle. \tag{4.7}$$

Finally, there are two nonlinear terms related to optical-coherence transfer: one is the nonlinear resonance of the counterpropagating wave in the population variation due to optical-coherence transfer,

$$I_7=\left\langle\frac{\Gamma^2}{\Gamma^2+(\Omega+kv)^2}\times\mathrm{Re}\frac{\Gamma\Gamma_1}{[\Gamma-i(\Omega-kv)][\Gamma_1-i(\Omega_1-kv)]}\right\rangle, \tag{4.8}$$

and the other is the nonlinear resonance in the Bennett structure of the ''upper transition'' m_1–n_1 of the part of the polarization that is induced by the probe wave and has experienced optical-coherence transfer,

$$I_8=\left\langle\mathrm{Re}\frac{\Gamma\Gamma_1}{[\Gamma-i(\Omega+kv)][\Gamma_1-i(\Omega_1+kv)]}\frac{\Gamma_1^2}{\Gamma_1^2+(\Omega_1-kv)^2}\right\rangle, \tag{4.9}$$

The integrals I_1,I_4 and I_2,I_5 characterize the spectrum of absorption of the probe wave by the two-level systems m,n and m_1,n_1, respectively, provided that these systems are not coupled; the quantities I_3 and I_6–I_8 reflect population transfer (I_6) and optical-coherence transfer.

The quantities I_1–I_8 can be expressed in terms of the error function (probability integral) of a complex-valued argument.[13] For instance,

$$I_1=\mathrm{Re}\left\langle\frac{\Gamma}{\Gamma-i(\Omega-kv)}\right\rangle=\frac{\sqrt{\pi}\,\Gamma}{k\bar{v}}\,\mathrm{Re}[w(p)],\quad p=\frac{\Gamma-i\Omega}{k\bar{v}}, \tag{4.10}$$

$$w(p)=\exp(p^2)\,[1-\Phi(p)],\quad \Phi(p)=\frac{2}{\sqrt{\pi}}\int_0^p\exp(-t^2)\,dt, \tag{4.11}$$

$$I_2=\frac{\sqrt{\pi}\,\Gamma_1}{k\bar{v}}\,\mathrm{Re}[w(p_1)],\quad p_1=\frac{\Gamma_1-i\Omega_1}{k\bar{v}},$$
$$I_3=\frac{\sqrt{\pi}\,\Gamma\Gamma_1}{(k\bar{v})^2}\,\mathrm{Re}\,\frac{w(p)-w(p_1)}{p_1-p}. \tag{4.12}$$

All averaged expressions in (4.5)–(4.9) can be represented by partial fraction expansions (as functions of kv) and hence can be written in the form of linear combinations of the functions $w(p)$ with different arguments. The quantities I_1–I_5 have been thoroughly studied, so that we need to consider only I_6–I_8, which reflect the interaction of the monochromatic waves with both transitions and describe the contribution of the stimulated population transfer and optical-coherence transfer to the nonlinear resonance of the counterpropagating probe wave. Here we will not write the lengthy expressions for I_6–I_8 in terms of the functions $w(p)$; we confine ourselves to an analysis of simple and physically clear limits and, primarily, the limit of large Doppler broadening ($\Gamma,\Gamma_1\ll k\bar{v}$).

The value of the parameter $|\Delta|/k\bar{v}$ is important here. For $|\Delta|<k\bar{v}$ and $|\Omega|<k\bar{v}$ a Bennett structure involving the levels of both transitions and an optical-coherence transfer structure develop within the Maxwell distribution. In this case the approximate expressions are

$$I_6=\frac{\sqrt{\pi}\,\Gamma}{2k\bar{v}}\,\frac{\Gamma_1(\Gamma+\Gamma_1)/2}{(\Gamma+\Gamma_1)^2/4+(\Omega-\Delta/2)^2}\times\exp\left\{-\left[\frac{\Omega-\Delta/2}{k\bar{v}}\right]^2\right\}, \tag{4.13}$$

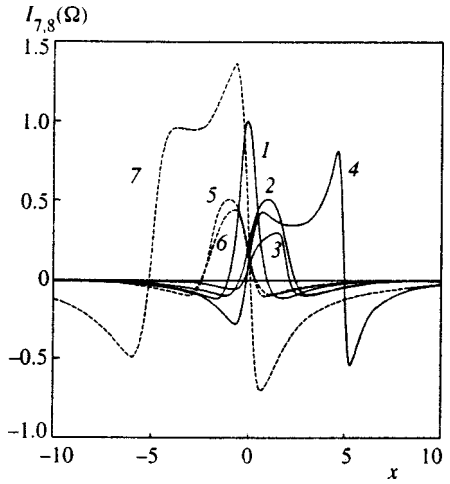


FIG. 3. The graphs of $I_7(\Omega)$ (solid curves) and $I_8(\Omega)$ (dashed curves) as functions of $x=(\Omega-\Delta/2)/\Gamma$: curve *1*, $\Delta=0$ and $\Gamma=\Gamma_1$; curves *2* and *5*, $\Delta=4\Gamma$ and $\Gamma=\Gamma_1$; curves *3* and *6*, $\Delta=4\Gamma$ and $\Gamma=2\Gamma_1$; and curves *4* and *7*, $\Delta=10\Gamma$ and $\Gamma=3\Gamma_1$ (the functions I_7 and I_8 have been multiplied by 10).

$$I_7=\frac{\sqrt{\pi}\,\Gamma}{2k\bar{v}}\,\mathrm{Re}\,\frac{\Gamma\Gamma_1}{(\Gamma-i\Omega)[(\Gamma+\Gamma_1)/2-i(\Omega-\Delta/2)]}\times\exp\left[-\left(\frac{\Omega}{k\bar{v}}\right)^2\right], \tag{4.14}$$

$$I_8=\frac{\sqrt{\pi}\Gamma_1}{2k\bar{v}}\times\mathrm{Re}\,\frac{\Gamma\Gamma_1\exp\{-[(\Omega-\Delta)/k\bar{v}]^2\}}{[(\Gamma+\Gamma_1)/2-i(\Omega-\Delta/2)][\Gamma_1-i(\Omega-\Delta)]}. \tag{4.15}$$

All three formulas have characteristic resonance singularities at the frequency $\Omega=\Delta/2$, right in the middle between the components of the doublet $\Omega=0$ and $\Omega=\Delta$, right out of a clear sky, so to say. The reason is that the resonant rates of the Bennett structure for the ''upper'' transition are $kv=\Omega-\Delta$, and after the spontaneous transfer to the ''lower'' transition they are in resonance with the counterpropagating wave at $kv=-\Omega$. The two conditions match at $\Omega=\Delta/2$, a fact revealed in (4.13)–(4.15). The same arguments explain the splitting of the Lamb dip in gas lasers with a magnetic field applied to their active medium.[14,2]

Formula (4.13) describes a nonlinear resonance of the ordinary Lorentzian shape. The contours of the nonlinear resonances (4.14) and (4.15), related optical-coherence transfer, are characteristic of an interference effect (Fig. 3) and qualitatively resemble $f(v)$, differing only in width and position of singularities (cf. Figs. 2 and 3). The resonances I_7 and I_8 are shifted in relation to each other and overlap only in the region $\Omega\approx\Delta/2$, where they are partially balanced when $|\Delta|\gg\Gamma,\Gamma_1$.

Formulas (4.13)–(4.15) are valid for $|\Delta|<k\bar{v}$, while the halfwidths Γ and Γ_1 can be either larger or smaller than $|\Delta|$. In other words, the resonances near the frequencies $\Omega=0$, $\Delta/2$, and Δ are spectrally resolvable.

In the limit $|\Delta|\gg k\bar{v}$, the value of $|\Omega|/k\bar{v}$ plays an important role. If $|\Omega|,|\Omega_1|>k\bar{v}$, all four levels m, n, m_1, and n_1 have no Bennett structure, and the Lorentz factors in (4.7)–(4.9) overlap by distant tails. Then

$$I_6=\frac{\Gamma^2\Gamma_1^2}{\Omega^2\Omega_1^2},\quad I_7=-\frac{\Gamma^2\Gamma_1^2}{\Omega^3\Omega_1},\quad I_8=-\frac{\Gamma^2\Gamma_1^2}{\Omega\Omega_1^3},\quad |\Omega|,|\Omega_1|>k\bar{v}. \tag{4.16}$$

If $|\Delta|$ is still much larger than $k\bar{v}$ but we have either $|\Omega|<k\bar{v}$ or $|\Omega_1|<k\bar{v}$, there is a Bennett structure involving the levels of one transition, and because of the resonance conditions the values of the I_j increase substantially:

$$I_6=\frac{\sqrt{\pi}\,\Gamma}{k\bar{v}}\frac{\Gamma_1^2}{\Delta^2}\exp\left[-\left(\frac{\Omega}{k\bar{v}}\right)^2\right],\quad I_8=-\frac{\sqrt{\pi}\,\Gamma}{k\bar{v}}\frac{\Gamma_1^3}{\Delta^3}\mathrm{Im}\,w(p), \tag{4.17}$$

$$I_7=\frac{\sqrt{\pi}\,\Gamma}{2k\bar{v}}\frac{\Gamma_1}{\Delta}\frac{\Gamma\Omega}{\Gamma^2+\Omega^2}\exp\left[-\left(\frac{\Omega}{k\bar{v}}\right)^2\right],\quad |\Omega|<k\bar{v},$$

$$I_6=\frac{\sqrt{\pi}\,\Gamma_1}{k\bar{v}}\frac{\Gamma^2}{\Delta^2}\exp\left[-\left(\frac{\Omega}{k\bar{v}}\right)^2\right],$$

$$I_7=-\frac{\sqrt{\pi}\,\Gamma_1}{k\bar{v}}\frac{\Gamma^3}{\Delta^3}\mathrm{Im}\;w(p_1), \tag{4.18}$$

$$I_8=-\frac{\sqrt{\pi}\,\Gamma_1}{2k\bar{v}}\frac{\Gamma}{\Delta}\frac{\Gamma_1\Omega_1}{\Gamma_1^2+\Omega_1^2}\exp\left[-\left(\frac{\Omega_1}{k\bar{v}}\right)^2\right],\quad |\Omega_1|<k\bar{v}.$$

Hence, when the splitting of the doublet is so large that it appreciably exceeds the Doppler width, the largest amplitude of the resonance due to optical-coherence transfer is attained when the field frequency is scanned within the limits of the Doppler width of the doublet components [I_7 in (4.17) and I_8 in (4.18)].

Let us discuss the amplitude properties of the resonances. We begin by noting their dependence on the Einstein coefficients A_{ij} of the four allowed transitions in the system of levels depicted in Fig. 1 (note that $(A_{ij})^{1/2}$ is proportional to d_{ij}). The resonances of the two-level systems, which are described by the integrals I_1,I_2 and I_4,I_5, are proportional to $A_{mn},A_{m_1n_1}$ and $A_{mn}^2,A_{m_1n_1}^2$, respectively. The other four resonances are closely related to the radiative transfer of particles or polarizations and contain products of all four coefficients A_{ij}:

$$|G_\mu^*G_{1\mu}|A\propto\sqrt{A_{mn}A_{m_1n_1}A_{m_1m}A_{n_1n}}\quad (I_3),$$

$$T_2|G_\mu|^2\kappa_1\propto A_{m_1m}A_{mn}A_{m_1n_1},\ A_{m_1n_1}A_{n_1n}A_{mn}\quad (I_6),$$

$$|G_\mu|^2G_1G^*A\propto\sqrt{A_{mn}^3A_{m_1n_1}A_{m_1m}A_{n_1n}}\quad (I_7),$$

$$|G_1|^2G_\mu^*G_{1\mu}A\propto\sqrt{A_{mn}A_{m_1n_1}^3A_{m_1m}A_{n_1n}}\quad (I_8). \tag{4.19}$$

The coefficients A_{ij} may differ substantially, by a factor of ten or more, and so the ratios of amplitudes of different resonances may vary within broad limits.

Now let us examine the effect of Doppler broadening on the amplitude properties of linear and nonlinear resonances. We take the amplitude of a Doppler broadened spectral line in the absence of saturation (i.e., $\sqrt{\pi}\Gamma/k\bar{v}$) and use as the unit in estimating the amplitudes of the other resonances, ignoring all other factors. Since $f(v)$ alternates in sign, the contribution of optical-coherence transfer to the linear absorption coefficient contains an additional factor $\Gamma_1/k\bar{v}$ (if the conditions for resonance are met: $|\Delta|,|\Omega|,|\Omega_1|<k\bar{v}$) or $\Gamma/|\Delta|$ (if $|\Delta|>k\bar{v}$ and $|\Omega|\approx\Gamma$) or $\Gamma k\bar{v}/\Delta^2$ (if $|\Delta|\gg k\bar{v}$ and $|\Omega|,|\Omega_1|>k\bar{v}$). On the other hand, the nonlinear resonances due to optical-coherence transfer have relative amplitudes $\sqrt{\kappa\kappa_1}$ (I_7) and κ_1 (I_8), which may exceed the amplitude of the linear resonances. Thus, the selectivity of the Bennett structure compensates for the negative effect of the fact that $f(v)$ alternates in sign by replacing the factor $\Gamma_1/k\bar{v}$ with $\sqrt{\kappa\kappa_1}$ or κ_1 and, to within these factors, the amplitude of the nonlinear resonances due to optical-coherence transfer remains essentially the same as for immobile atoms.

5. DEGENERATE LEVELS

The analysis of Secs. 2–4 can also be applied to degenerate states. To this end, each magnetic sublevel should be considered a separate state. We then deal with a set of four-level systems each of which resembles the one depicted in Fig. 1. We introduce the density matrix elements $\rho(J_iM_i,J_jM_j)$ (M_i and M_j are the magnetic quantum numbers of the states i and j) and replace the matrix elements G and G_1 of the interaction with the field and the optical-coherence transfer rate via the well-known formulas (see, e.g., Refs. 2, 4, and 8)

$$G\to G_{MM'}=\frac{d_{mn}}{2\sqrt{3}\,h}\times\sum_{\sigma}(-1)^{J_n-M'}\langle J_mM\,J_n-M'|1\sigma\rangle E_\sigma,$$
$$G_1\to G_{1MM'}=\frac{d_{m_1n_1}}{2\sqrt{3}\,h}\times\sum_{\sigma}(-1)^{J_{n_1}-M'}\langle J_{m_1}M\,J_{n_1}-M'|1\sigma\rangle E_\sigma,$$
$$A\to A(J_mMJ_nM'|J_{m_1}M_1J_{n_1}M_1')=\sqrt{A_{m_1m}A_{n_1n}}\sum_{\sigma}\langle J_mM1\sigma|J_{m_1}M_1\rangle\times\langle J_nM'1\sigma|J_{n_1}M_1'\rangle. \tag{5.1, 5.2}$$

Here E_σ is the circular component of the field, J_i is the total angular momentum of state i, d_{ij} is the reduced dipole-moment matrix element,[15] and $\langle\cdots|\cdots\rangle$ stands for a vector addition coefficient. Also, G_μ and $G_{1\mu}$ of Sec. 4 must be interpreted as the quantities (5.1) after E_σ is replaced by $E_{\mu\sigma}$. Spontaneous magnetic-coherence transfer, which is absent in the model of nondegenerate states, proceeds with a rate[2,4,5]

$$A(J_mMJ_mM'|J_{m_1}M_1J_{m_1}M_1')=A_{m_1m}\sum_{\sigma}\langle J_mM1\sigma|J_{m_1}M_1\rangle\langle J_mM'1\sigma|J_{m_1}M_1'\rangle. \tag{5.3}$$

Thus, the problem reduces to a set of four-level systems jJ_jM_j ($j=m,n,m_1,n_1$) that to a certain extent are coupled by spontaneous and stimulated processes. The problem becomes extremely cumbersome for arbitrary values of J_j and in strong fields, especially because of the inclusion of all types of spontaneous cascade processes. Hence, as in other areas of nonlinear polarization spectroscopy, we limit ourselves to the analysis of specific cases.

First we must mention the states of field polarization and systems of levels with small values of J_j, which can be directly reduced to the model of nondegenerate states and to which the first three rows in Table I correspond. For the combinations represented by rows 4–8 of Table I, the problem of the interaction with a strong field also reduces to the model of degenerate states and can be solved exactly for arbitrary field intensities. The distinctive nonlinear structure of the velocity distribution contains the same elements as in the model of nondegenerate states (see Eq. (3.4) and the discussion that follows). However, in the cases corresponding to rows 4–8 of Table I there are additional population cascade channels, and because of this the relative amplitudes of the various structure components change. The difference between the cases represented by rows 1–3 and 4–8 is illustrated in Figs. 4a and 4b: the dashed arrows in Fig. 4b indicate additional (relative to the model of nondegenerate states) spontaneous cascades, which contribute to the Bennett structure in the levels involved in optical-coherence transfer. Hence the systems listed Table I exhibit nothing new in comparison to the model of nondegenerate states, so that we will not analyze then here in greater detail.

In diagrams of levels with large values of angular momenta J_j, the problem of the interaction with a strong field becomes more complicated, which naturally leads to a more complicated Bennett structure: each of the subsystems m,n and m_1,n_1 acquire Bennett dips and peaks, and the squares of the corresponding widths are nonlinear functions of the

TABLE I.

No.	J_{m_1}	J_{n_1}	J_m	J_n	ΔM
1	1	0	0	1	$0,\pm1$
2	1	0	1	1	$0,\pm1$
3	1	1	0	1	$0,\pm1$
4	0	1	1	0	$0,\pm1$
5	0	1	1	1	$0,\pm1$
6	1	1	1	0	$0,\pm1$
7	1	1	1	1	$0,\pm1$
8	1/2	1/2	1/2	1/2	±1

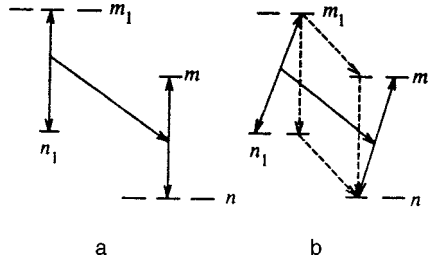


FIG. 4. Diagrams of optical coherence transfer in the case of row 1 in Table I (a) and in the case of row 8 in Table I and an additional spontaneous population cascade (dashed arrows) (b).

power of the field. An example is the interaction of a high-power linearly polarized field and the system with $J_m=J_n=2$ discussed in Ref. 16. However, the frequency dependence of the optical-coherence transfer rate similar to W has the form $1/(\Gamma-i\Omega)(\Gamma_1-i\Omega_1)$, as in the model of nondegenerate states, i.e., we can say that the Karplus–Schwinger idea,[11] according to which in the strong-field problem the broadening is due to light-induced population variation, remains valid. At moderate intensities, when we can limit ourselves to the first nonlinear corrections, the theory leads to a universal result for arbitrary values of J_j and arbitrary polarization states of the field. We will not write the expressions for the velocity distribution and the work done by the strong field, and we limit ourselves to an analysis of the work P_μ done by the probe field in the scheme with a counterpropagating prove wave of the same frequency:

$$P_\mu=2h\omega\left|\frac{d_{mn}}{2h}\right|^2\frac{E_\mu^2}{3}\left\{\frac{1}{\Gamma}I_1(\Omega)N_{nm}+\left[\frac{d^2}{\Gamma_1}I_2(\Omega)+\frac{dA}{\Gamma\Gamma_1}I_3(\Omega)\right]N_{n_1m_1}-\left|\frac{d_{mn}}{2h}\right|^2 2E^2\times\sum_{\kappa q}I(\kappa q)I_\mu^*(\kappa q)\left[\frac{T_\kappa}{\Gamma^2}I_4(\Omega)N_{nm}+\frac{d^2T_{1\kappa}}{\Gamma_1^2}I_5(\Omega)N_{n_1m_1}+\frac{d^2T_{2\kappa}}{\Gamma\Gamma_1}I_6(\Omega)N_{n_1m_1}+\frac{dA}{\Gamma\Gamma_1}\mathrm{Re}\left(\frac{T_\kappa}{\Gamma}I_7(\Omega)+\frac{d^2T_{1\kappa}}{\Gamma_1}I_8(\Omega)\right)N_{n_1m_1}\right]\right\}. \tag{5.4}$$

Here $I(\kappa q)$ and $I_\mu(\kappa q)$ are the normalized field polarization tensors,

$$I(\kappa q)=\sqrt{3}\sum_{\sigma\sigma_1}(-1)^{1-\sigma}\langle 1\sigma_1\,1-\sigma|\kappa q\rangle\frac{E_{\sigma_1}E_\sigma^*}{E^2},$$

$$E^2=\sum_\sigma|E_\sigma|^2,$$

$$I_\mu(\kappa q)=\sqrt{3}\sum_{\sigma\sigma_1}(-1)^{1-\sigma}\langle 1\sigma_1\,1-\sigma|\kappa q\rangle\frac{E_{\mu\sigma_1}E_{\mu\sigma}^*}{E_\mu^2},$$

$$E_\mu^2=\sum_\sigma|E_{\mu\sigma}|^2, \tag{5.5}$$

and the effective times T_κ, $T_{1\kappa}$, and $T_{2\kappa}$ are similar to T, T_1, and T_2 of the model of nondegenerate states (see Eqs. (3.2), (3.4), and (4.4)):

$$T_\kappa=\frac{a_{mn\kappa}^2}{\Gamma_m}+\frac{a_{nm\kappa}^2}{\Gamma_n}-\frac{a_{mn\kappa}a_{nm\kappa}A_{mn\kappa}}{\Gamma_m\Gamma_n},$$

$$T_{1\kappa}=\frac{a_{m_1n_1}^2}{\Gamma_{m_1}}+\frac{a_{n_1m_1\kappa}^2}{\Gamma_{n_1}}-\frac{a_{m_1n_1\kappa}a_{n_1m_1\kappa}A_{m_1n_1\kappa}}{\Gamma_{m_1}\Gamma_{n_1}},$$

$$T_{2\kappa}=a_{m_1n_1\kappa}\left(a_{mn\kappa}-\frac{a_{nm\kappa}A_{mn\kappa}}{\Gamma_n}\right)\frac{A_{m_1m\kappa}}{\Gamma_{m_1}\Gamma_m}+a_{nm\kappa}\left(a_{n_1m_1\kappa}-\frac{a_{m_1n_1\kappa}A_{m_1n_1\kappa}}{\Gamma_{m_1}}\right)\frac{A_{n_1n\kappa}}{\Gamma_{n_1}\Gamma_n}, \tag{5.6}$$

$$a_{mn\kappa}=3(-1)^{1+\kappa+J_m+J_n}\begin{Bmatrix}J_n & J_m & 1\\ \kappa & 1 & J_m\end{Bmatrix},$$

$$a_{nm\kappa}=3(-1)^{1+J_n+J_m}\begin{Bmatrix}J_m & J_n & 1\\ \kappa & 1 & J_n\end{Bmatrix},$$

$$A=\sqrt{A_{m_1m}A_{n_1n}}(-1)^{J_{m_1}+J_n}\sqrt{2J_{m_1}+1}\times\sqrt{2J_{n_1}+1}\begin{Bmatrix}J_{m_1} & J_{n_1} & 1\\ J_n & J_m & 1\end{Bmatrix},$$

$$A_{ij\kappa}=A_{ij}(-1)^{1+\kappa+J_i+J_j}(2J_i+1)\begin{Bmatrix}J_i & J_i & \kappa\\ J_j & J_j & 1\end{Bmatrix}, \tag{5.7}$$

$$d=\left|\frac{d_{m_1n_1}}{d_{mn}}\right|,\quad N_{ij}=N_i-N_j.$$

The dependence of P_μ on the frequency Ω is given by the same integrals $I_j(\Omega)$ as in the model of nondegenerate states defined by Eqs. (4.5)–(4.9). The $I_j(\Omega)$ acquire other coefficients (in comparison to the model of nondegenerate states), which depend on field polarization and degeneracy of the levels. The quantities T_κ and $T_{1\kappa}$ are the effective times of the interaction of the field and the polarizations moments in the transitions m–n and m_1–n_1, and $T_{2\kappa}$ is the effective time of interaction for the spontaneous cascades through the channels $m_1\to m$, $m_1\to m\to n$, $n_1\to n$, and $m_1\to n_1\to n$, a parameter specific to four-level systems (in two-level systems there are only T_κ and $T_{1\kappa}$; see Refs. 2 and 4). Formulas (5.4) and (5.6) remain valid of the decay times of the polarization moments of the levels depend on κ. In this case Γ_j must be replaced by $\Gamma_{j\kappa}$ in (5.6) (see Refs. 2 and 4).

Note that the times T_κ and $T_{1\kappa}$ specify the amplitudes of ''ordinary'' nonlinear resonances due to saturation ($I_4(\Omega)$ and $I_5(\Omega)$) and nonlinear resonances due to optical-coherence transfer ($I_7(\Omega)$ and $I_8(\Omega)$).In the event this seems quite obvious, but predicting it would be difficult. We conclude that the relationship between the resonance

TABLE II. Values of $I(\kappa q)I_\mu(\kappa q)$.

κq	$\uparrow\uparrow$	$\uparrow\rightarrow$	$++$	$+-$
00	1	1	1	1
10	0	0	3/2	−3/2
20	2	−1	1/2	1/2

[a]*Note*: The ''plus'' and ''minus'' stand for circular polarizations, and the arrows stand for linear polarizations. The combinations $\uparrow\uparrow$ and $\uparrow\rightarrow$ correspond to parallel and orthogonal linear polarizations of the counterpropagating waves. The values for linear polarizations are given in a system whose z axis is directed along the field strength in one of the waves, and the values for circular polarizations are given in a system whose z axisis directed along **k**.

TABLE III. Values of K.

J_{m_1}	J_{n_1}	J_m	J_n	K
J	J	J	J	$1-1/J(J+1)$
J	J	J	$J+1$	$1/(J+1)$
J	J	$J+1$	J	$-\sqrt{1+2/(2J+1)}/(J+1)$
J	J	$J+1$	$J+1$	$\sqrt{1-1/(J+1)^2}$
J	$J+1$	$J+1$	J	$1/(J+1)(2J+1)$
$J+1$	J	J	$J+1$	$1/(J+1)(2J+3)$
$J+1$	$J+1$	$J+1$	J	$-1/(J+1)$
$J+1$	$J+1$	J	$J+1$	$\sqrt{1-2/(2J+3)}/(J+1)$
$J+1$	J	$J+2$	$J+1$	$\sqrt{1-4/(2J+3)^2}$
J	$J+1$	$J+1$	$J+2$	1

$I_4(\Omega),I_5(\Omega)$ and $I_7(\Omega),I_8(\Omega)$ is weakly dependent on the angular momenta of the levels and is fixed by the values d and Γ/Γ_1.

In the method of a counterpropagating probe wave within the adopted approximation in the intensity of the strong field, the nonlinear resonances of all three types depend in the same manner on the polarization of the fields, with the dependence expressed by the product of polarization tensors, $I(\kappa q)I^*_\mu(\kappa q)$. The values of $I(\kappa q)I_\mu(\kappa q)$ for the ordinary (simplest) combinations of linear and circular polarizations are listed in Table II. This table readily shows that the $a^2_{ij\kappa}$ entering into T_κ and $T_{1\kappa}$ may be present in the following combinations acting as factors of $1/\Gamma_i$:

$$\begin{aligned}
&\uparrow\uparrow: \quad a^2_{ij0}+2a^2_{ij2},\\
&\uparrow\rightarrow: \quad a^2_{ij0}-a^2_{ij2},\\
&++: \quad a^2_{ij0}+\frac{3}{2}a^2_{ij1}+\frac{1}{2}a^2_{ij2},\\
&+-: \quad a^2_{ij0}-\frac{3}{2}a^2_{ij1}+\frac{1}{2}a^2_{ij2}.
\end{aligned} \tag{5.8}$$

The validity of the following relationships can easily be proved:

$$\begin{aligned}
&a^2_{mn0}+2a^2_{mn2}=a^2_{nm0}+2a^2_{nm2},\\
&a^2_{mn1}-a^2_{mn2}=a^2_{nm1}-a^2_{nm2},\\
&a^2_{mn0}+\frac{3}{2}a^2_{mn1}+\frac{1}{2}a^2_{mn2}=a^2_{nm0}+\frac{3}{2}a^2_{nm1}+\frac{1}{2}a^2_{nm2}.
\end{aligned} \tag{5.9}$$

As a result, for counterpropagating waves of like polarization ($\uparrow\uparrow$ and $++$), the level lifetimes Γ_i^{-1} enter into T_κ and $T_{1\kappa}$ in the combinations $\Gamma_m^{-1}+\Gamma_n^{-1}$ and $\Gamma_{m_1}^{-1}+\Gamma_{n_1}^{-1}$. This is also the case for arbitrary polarizations when $J_m=J_n$. The quantities in (5.8) have been thoroughly studied in Refs. 2 and 4, where the cases of strong dependence of the relative amplitudes on the field polarizations are also investigated.

In view of the inequalities

$$\begin{aligned}
&\left|\sqrt{2J_m+1}\sqrt{2J_{n_1}+1}\begin{Bmatrix}\kappa & J_{m_1} & J_{n_1}\\ 1 & J_n & J_m\end{Bmatrix}\right|\leqslant 1,\\
&\left|\sqrt{2J_{m_1}+1}\sqrt{2J_n+1}\begin{Bmatrix}\kappa & J_{m_1} & J_{n_1}\\ 1 & J_n & J_m\end{Bmatrix}\right|\leqslant 1,
\end{aligned} \tag{5.10}$$

which follow from the orthonormality of $6j$-symbols, we have

$$\begin{aligned}
&|A_{ij\kappa}|\leqslant\sqrt{\frac{2J_i+1}{2J_j+1}}A_{ij},\\
&|A|\leqslant\sqrt{\frac{2J_{m_1}+1}{2J_m+1}}\sqrt{A_{m_1m}A_{n_1n}},
\end{aligned} \tag{5.11}$$

and, as can easily be shown, the effect of the cascade terms $A_{ij\kappa}/\Gamma_i\Gamma_j$ on T_κ and $T_{1\kappa}$ diminishes as κ increases. Note also that in some cases $A_{ij\kappa}$ may be negative (in the adopted approximation this happens when $\kappa=2$ and $J_m=J_n=1$), and then the population cascade extends the time of interaction with the field for the corresponding polarization moment.

It is convenient to write the factor A (the optical-coherence transfer rate) in the form

$$A=\sqrt{\frac{2J_{m_1}+1}{2J_m+1}}\sqrt{A_{m_1m}A_{n_1n}}K, \tag{5.12}$$

$$K=(-1)^{J_{m_1}+J_n}\sqrt{2J_m+1}\sqrt{2J_{n_1}+1}\begin{Bmatrix}J_{m_1} & J_{n_1} & 1\\ J_n & J_m & 1\end{Bmatrix},$$

since the coefficient K obeys the inequality $|K|\leqslant 1$ and is symmetric with respect to the permutations $J_{m_1}\leftrightarrow J_n$ and $J_{n_1}\leftrightarrow J_m$. Table III lists the values of K for possible combinations of angular momenta J_j. Other combinations reduce to those listed if we use the symmetry properties. We see that there is a remarkable case, $K=1$, in which the polarization of the m_1-n_1 transition is shifted to the $m-n$ transition with a probability equal to unity for all values of J. For other combinations of angular momenta ($|K|<1$), the values of K may be positive or negative, and they can increase or decrease with increasing J.

6. DISCUSSION

One of the main conclusions that can be drawn from the fact that there is spontaneous optical-coherence transfer is that the common two-level approximation must be discarded even under exact resonance conditions. Joint analysis of at least a pair of two-level systems coupled by optical-coherence transfer is required. This becomes especially evident when an external field (e.g., a magnetic field) is applied

to the systems, with each quadruplet of the Zeeman sublevels of the states m_1, n_1, m, and n playing an independent role.

Spontaneous optical-coherence transfer introduces new features into the physical picture of the processes that play an important role in nonlinear saturation spectroscopy. The velocity distribution of the atoms acquires, in addition to a system of Bennett peaks and dips, elements of the form (3.5) with specific shapes, alternating in sign and having a zero total area (Fig. 2). The amplitudes of this optical-coherence transfer structure depend on many factors; in particular, they decrease with increasing relative splitting $|\Delta|/\Gamma$ of the doublet.

Nonlinear resonances of a counterpropagating probe wave contain two terms due to optical-coherence transfer. These terms are in the frequency range $0 \leqslant |\Omega| \leqslant \Delta$ and adjoin each other. The ratio of their amplitudes is strongly dependent on the polarizations of the fields and many parameters of the $m-n$ and m_1-n_1 transitions: population differences, Einstein coefficients, angular momenta of the levels, etc.

In linear spectra (absorption and the like), due to the alternation in sign of the velocity distribution of the optical-coherence transfer structure, the spectral manifestation of this structure is masked by Doppler broadening, and because of this the amplitudes of the spectral resonances decrease (in comparison to those of immobile atoms) in the Γ-to-$k\bar{v}$ ratio, which in typical conditions amounts to something of the order 10^{-2}–10^{-3}. Nonlinear Bennett peaks and dips eliminate the Doppler flattening of spectral optical-coherence transfer structures and lead to a paradoxical excess of nonlinear resonances over linear.

The resonance of a counterpropagating prove wave is the simplest type of nonlinear resonances in gas saturation spectroscopy. In the many variants of the probe-field method, including those that use spontaneous and stimulated transitions to ''third'' levels, one can expect other manifestations of optical-coherence transfer, manifestations in which, in addition to the population effects discussed in this paper, nonlinear interference effects and field-induced level splitting play a significant role. But if the population factor is predominant, the phenomena resemble those discussed in Secs.4 and 5. In particular, field-induced population variations are important for fluorescence resonance,[3] which contains components due to optical-coherence transfer that are proportional to I_7 and I_8.

As is well known, atomic collisions affect nonlinear resonances in various ways (see, e.g., Refs. 2 and 4). If we ignore variations in rate, collisions may be taken into account by redefining the relaxation constants of the levels and transitions, i.e., by including radiative and collisional terms in the relaxation constants (the model of relaxation constants). But if rate variations are significant, the optical-coherence transfer structure occupies a special position due to alternation of its sign. In particular, in the model of intense collisions, optical-coherence transfer provides almost nothing to the homogeneous saturation band.

In conclusion we note that in this paper we focused on the doublet $\omega_{m_1 n_1}, \omega_{mn}$. Clearly, similar phenomena occur near the doublet $\omega_{m_1 m}, \omega_{n_1 n}$: all the relationships of this paper remain valid when we go over to this spectral region if we interchange the indices as follows: $m \leftrightarrow n_1$, $m_1 \rightarrow m_1$, and $n \rightarrow n$. These two doublets carry common information, since the properties if each doublet depend on the characteristics of all four transitions between the levels m, n, m_1, and n_1.

The author is grateful to M. P. Chaĭka, S. N. Seleznev, A. M. Shalagin, D. A. Shapiro, and V. A. Sorokin for fruitful discussions of the topics touched on in this paper. The work was supported by grants from the Fundamental Spectroscopy Section of the State Science and Technology Program (Grant 08.02.31) and the Federal Incentive Program *Integration* (Grant No. 274).

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

[1] W. R. Bennett, Jr., Appl. Opt. Suppl. No. 1, 24 (1962).
[2] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).
[3] V. S. Letokhov and V. P. Chebotaev, *Nonlinear Laser Spectroscopy* (Springer-Verlag, Berlin, 1977) [Russ. original, Nauka, Moscow, 1975; 2nd ed. 1990].
[4] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Non-linear Spectroscopy*, North-Holland, Amsterdam (1991).
[5] M. P. Chaĭka, *Interference of Degenerate Atomic States* [in Russian], Leningrad Univ. Press, Leningrad (1975).
[6] E. B. Aleksandrov, G. I. Khvostenko, and M. P. Chaĭka, *Interference of Atomic States*, Springer, New York (1993).
[7] S. G. Rautian, JETP Lett. **61**, 473 (1995).
[8] S. G. Rautian, Zh. Éksp. Teor. Fiz. **108**, 1186 (1995) [JETP **81**, 651 (1995)].
[9] S. G. Rautian, Zh. Éksp. Teor. Fiz. **110**, 462 (1996) [JETP **83**, 246 (1996)].
[10] I. I. Sobelman, L. A. Vainshtein, and E. A. Yukov, *Excitation of Atoms and Broadening of Spectral Lines*, Springer, Berlin (1981).
[11] R. Karplus and J. Schwinger, Phys. Rev. **73**, 1020 (1948).
[12] R. I. Sokolovskiĭ, Opt. Spektrosk. **27**, 1017 (1969) [Opt. Spectrosc. **27**, 553 (1969)].
[13] V. N. Faddeeva and N. M. Terent'ev, *Tables of Values of the Function* $w(z)=e^{-z^2}*(1+(2i/\sqrt{\pi})\int_0^z e^{t^2}\,dt)$, Pergamon Press, Elmsford, N.Y. (1961).
[14] M. I. D'yakonov and V. I. Perel', Opt. Spektrosk. **20**, 472 (1966) [Opt. Spectrosc. **20**, 257 (1966)].
[15] I. I. Sobelman, *Atomic Spectra and Radiative Transitions*, Springer, Berlin (1979).
[16] S. G. Rautian and A. V. Shishaev, Zh. Éksp. Teor. Fiz. **108**, 807 (1995) [JETP **81**, 440 (1995)].

Translated by Eugene Yankovsky

Electrical conductivity of germanium with dislocation grids

S. A. Shevchenko*)

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Moscow, Russia
(Submitted 13 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 115–125 (January 1999)

Samples of *n*-type germanium with a donor concentration $N_d = 2.4 \times 10^{16}\,\mathrm{cm}^{-3}$ are plastically deformed to a degree of strain equal to 18–40% to detect static conduction by electrons trapped on dislocations in a system of dislocation grids. In samples with $20\% < \delta < 31\%$, which retain an electronic type of conductivity, the conductivity for $T < 8$ K, which is weakly temperature-dependent, is associated with conduction by electrons trapped on dislocations. The nonmonotonic dependence of the conductivity at 4.2 K on the degree of strain as the latter increases from 18% to 40% attests to the existence of an energy gap between the donor and acceptor dislocation states in strongly plastically deformed germanium. © *1999 American Institute of Physics.* [S1063-7761(99)01101-4]

1. INTRODUCTION

The interest in the study of the conductivity of plastically deformed germanium stems from the possibility of quasi-one-dimensional conduction along dislocations and its relationship to the structure of the energy spectrum of dislocations in semiconductors.

In Ref. 1 plastically deformed *p*-type germanium was found to exhibit low-temperature static conductivity, which was characterized by a weak temperature dependence, the essential absence of a Hall emf, and a conductivity at 4.2 K exceeding that of the control sample by several orders of magnitude. It was assumed in Ref. 1 that the specific dislocation conductivity due to the motion of charge carriers trapped on dislocations in strongly deformed crystals is superposed on the conductivity due to free charge carriers, which decreases as a function of temperature. It was subsequently established[2–4] that dislocation conductivity appears at a threshold (at a degree of strain $\delta > 15\%$ in Ref. 3 and $\delta > 30\%$ in Ref. 4) and becomes a measurable Hall emf in the dislocation conduction region as δ increases further and that the signs of the Hall emf and the thermopower correspond to hole-type conductivity. Structural investigations[3,5,6] showed that the high-temperature deformation of germanium and silicon to $\delta = 15 - 40\%$ promotes the formation of a block (cellular) structure and that 60° and screw dislocations (with a density $\sim 10^{10}\,\mathrm{cm}^{-2}$) are arranged, for the most part, in the form of grids in block walls. This permitted associating dislocation conduction with the motion of holes trapped on dislocations along a branched system of dislocation segments in block walls.

The dislocation conductivity of strongly plastically deformed germanium was investigated in Refs. 4 and 7–9 over broad ranges of temperatures and electric and magnetic fields, as well as in the frequency range 0.1–10 MHz. The appearance of an activationless temperature dependence of the dislocation conductivity when a certain value of δ^* is attained, which depends on the deformation conditions, was attributed in Ref. 4 to the delocalization of carriers in a ''dislocation'' band as a consequence of the semiconductor-metal transition caused by plastic deformation.

The idea that conduction is possible along dislocations was also confirmed by observing the features of the microwave conductivity of germanium with anisotropically arranged 60° dislocations, whose density was less than $2 \times 10^7\,\mathrm{cm}^{-2}$ (Refs. 10,11). Investigations of the dislocation microwave conductivity in silicon[12] and germanium aided in the development of theories regarding the structure of the energy spectrum of dislocation states in covalent semiconductors. In Shockley's picture[13] linear dislocation segments are represented in the form of rows of atoms with dangling chemical bonds. These atoms can trap electrons from the conduction band or donate unpaired electrons to other centers and consequently exhibit acceptor or donor activity, respectively. The states in a dislocation core were represented at first by one level or a half-filled one-dimensional band and later on in the form of two bands separated by a gap. According to Refs. 14–17, the donor (lower) E_1 and acceptor (upper) E_2 dislocation bands are separated by a gap $\Delta_{12} \sim 0.18$ eV and are located in the germanium band gap (unlike the model used to explain hole conduction on germanium bicrystal boundaries[18,19]). There are localized E_1' acceptor states (a narrow band) at a distance $\Delta_1' < 0.03$ eV above the top of the donor band.[14–17] In Refs. 15 and 16 they were hypothetically assigned to states of holes bound by Coulomb attraction near negatively charged dislocation defects, such as steps, kinks, and sites of intersection with other dislocations.

Within this scheme it was possible to account for the decrease in the microwave conductivity in *n*-type germanium with increasing dislocation density,[10,11] as well as the dependence of the microwave conductivity on the concentration and type of the dopant present.[20] The dislocation conductivity in strongly plastically deformed germanium was associated in Ref. 3 with the same dislocation states as the microwave conductivity. In this case there is probably static conductivity due to the motion of electrons trapped on dislocation grids in *n*-type germanium. In the present work an

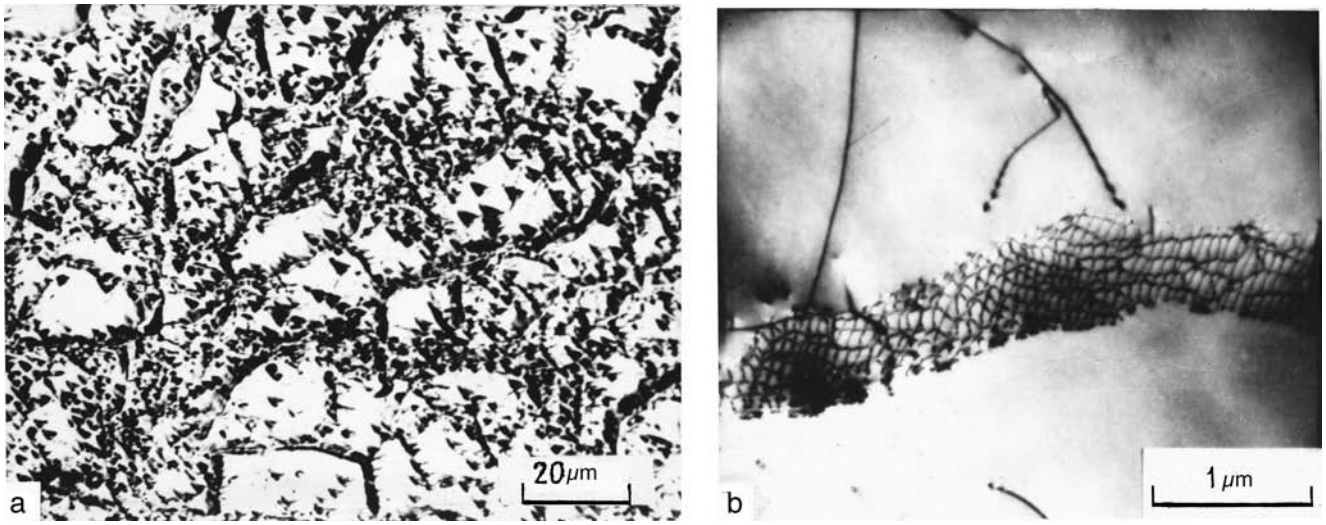


FIG. 1. Optical-microscopic image of the {111} plane in a sample from group II ($\delta=18\%$) after chemical etching (a) and image of a fragment of a dislocation boundary in a {112} plane obtained by transmission electron microscopy on a germanium sample with $\delta=29\%$ from group Ib (b).

attempt was made to detect such conductivity.

2. EXPERIMENT

The investigations were performed on single crystals of *n*-type germanium cut from a GÉS-0.1 ingot with a difference concentration of chemical donors (antimony) $N_d=2.4\times10^{16}\,\mathrm{cm}^{-3}$ and a density of growth dislocations less than $10\,\mathrm{cm}^{-2}$. According to the data from spark mass spectrometry, the oxygen and carbon concentrations in this ingot do not exceed 1.2×10^{17} and $4\times10^{16}\,\mathrm{cm}^{-3}$, respectively, and the concentration of other elements is less than $3\times10^{16}\,\mathrm{cm}^{-3}$.

Two groups of crystals in the form of parallelepipeds measuring $10\times6\times2.5\,\mathrm{mm}^3$ were prepared for deformation. They differed with respect to the direction of the compression axis, which coincided with the [100] direction in the crystals from group I and deviated by 10° in the crystals from group II. Consequently, the crystals from groups I and II contained two primary {111} slip planes or one such plane, respectively. All other deformation conditions being equal, this difference can influence the connectivity of the dislocation system. The other edges of the crystals were directed along the [011] and $[01\bar{1}]$ axes. These crystals were polished mechanically and chemically, were coated by a thin layer of gold (to prevent contamination by impurities during deformation) and were deformed at $T_d=800\,^\circ\mathrm{C}$ in a dynamic regime to $\delta=18-40\%$. The strained crystals were cooled for 30 min at T_d and then cooled at a rate of 1–2 deg/min to room temperature (group Ia) or were cooled together with the furnace without additional annealing (groups Ib and II).

To measure the static conductivity and the Hall effect in the temperature range 4.2–300 K, samples of length ~9 mm and cross section $\sim2\times1.2\,\mathrm{mm}^2$ were cut from the central part of the strained crystals. Six molten indium contacts were deposited on chemically polished surfaces of these samples. After high-voltage ($\sim25\,\mathrm{kV}$) pulses were supplied to each pair of current and potential contacts, their ohmic properties improved. The measurements were performed on the linear portions of the current–voltage characteristics. The magnetic field strength could be varied from 0.05 to 0.7 T. The temperature in the helium cryostat was held to within ±0.05 K. The maximum error in the determination of the conductivity and the Hall coefficient was $\pm20\%$ and $\pm10\%$, respectively.

We previously investigated the dislocation structure of strongly plastically deformed germanium[3] and silicon[5] by x-ray structural analysis, as well as optical and transmission electron microscopy. It was established that the samples with $\delta=5-40\%$ are composed of numerous blocks (cells) measuring 1–20 μm (see Fig. 1a, as well as Fig. 2 in Refs. 3 and 5). The dislocations are concentrated mainly in the low-angle boundaries separating blocks, i.e., in the block walls, whose structure is well resolved by transmission electron microscopy.[3,5,21,22] These boundaries are formed from partially ordered rows of 60° and screw dislocations, which move during deformation, and dislocation segments formed as products of their interaction, i.e., they have the form of two-dimensional dislocation grids. As δ rises, the number of disordered rows of dislocations and incomplete fragments of dislocation grids decreases, and the fraction of regular (com-

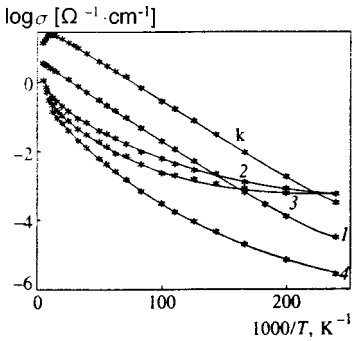


FIG. 2. Temperature dependence of the conductivity in control (k) and strained samples of *n*-type germanium ($N_d=2.4\times10^{16}\,\mathrm{cm}^{-3}$, group Ia) with various values of δ, %: *1* — 18.6, *2* — 20.9, *3* — 21.5, *4* — 28.

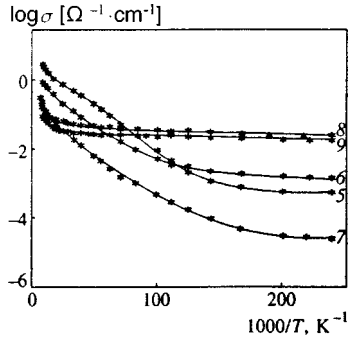


FIG. 3. Temperature dependence of the conductivity in strained samples of *n*-type germanium ($N_d=2.4\times10^{16}\,\mathrm{cm}^{-3}$, group Ib) with various values of δ, %: *5* — 27.6, *6* — 28.4, *7* — 31, *8* — 33.9 (p^*-type), *9* — 49.1 (p^*-type).

plete) fragments of dislocation grids like those shown in Fig. 1b increases. Therefore, the connectivity of the macroscopic system of dislocation segments covering distances of 300–1000 Å in block boundaries increases as a function of δ. The estimated values of the mean dislocation densities in samples with $\delta=18-40\%$ lie in the range $1\times10^9<N_D<5\times10^{10}\,\mathrm{cm}^{-2}$.

3. MEASUREMENT RESULTS

It follows from Fig. 2 that the conductivity in the control (k) sample decreases with decreasing temperature with an activation energy of 0.0048 ± 0.0004 eV, which is equal to half of the ionization energy of antimony atoms in germanium and attests to the weak degree of compensation of the original samples. In such crystals there is a transition to hopping conduction with a constant activation energy for $T<6$ K (Ref. 8).

The $\sigma(T)$ curves for strained samples from groups Ia and Ib are presented in Figs. 2 and 3. In strained samples *2–7*, which retained an electronic type of conduction, the conductivity for $T>8$ K decreases with increasing δ as a consequence of the decrease in the concentration and mobility of free electrons, as has been observed repeatedly in previous studies.[8,16] The number of electrons trapped on dislocations at 80 K equals $\sim0.25N_d$ for sample *1* and $(0.5-0.8)N_d$ for samples *2, 3, 5,* and *6*. In samples *4* and *7* with the lowest conductivity in groups Ia and Ib, respectively, the Hall coefficient has a negative sign, but does not provide information regarding the concentration of the charge carriers in them. For $T<8$ K the temperature dependence of the conductivity of the *n*-type strained samples weakens appreciably (except in the case of sample *1* in Fig. 2). In samples *2, 3, 5,* and *6* the conductivity for 4.2 K ($\sigma_{4.2}$) is 2–5 times greater than the value for the control sample ($\sigma_{4.2}=3\times10^{-4}\Omega^{-1}\cdot\mathrm{cm}^{-1}$), and in samples *4* and *7* $\sigma_{4.2}<2\times10^{-5}\Omega^{-1}\cdot\mathrm{cm}^{-1}$. In samples *2–7* the measurements of the Hall emf for $T<10$ K are unreliable because of the low values and instability of the signal at the $\sim5\,\mu$V level. However, the voltage on the potential contacts used to calculate the conductivity exceeds 0.5 mV.

We note that in the samples from group II (with one primary slip plane), which exhibit an electronic type of conduction for $20\%<\delta<29\%$, the value of $\sigma_{4.2}$ decreases with increasing δ and lies in the range $10^{-4}-10^{-6}\ \Omega^{-1}\cdot\mathrm{cm}^{-1}$.

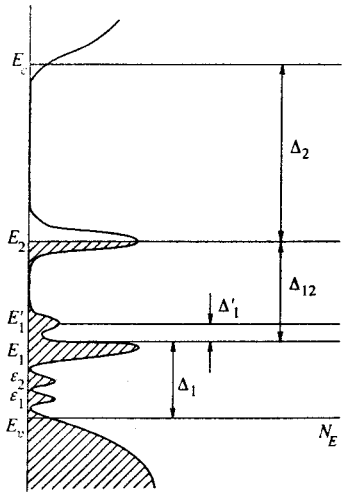


FIG. 4. Energy diagram of dislocation states in germanium according to Refs. 15 and 16. Here N_E is the density of states, $\Delta_1=0.07$ eV, $\Delta_1'<0.03$ eV, $\Delta_{12}=0.18$ eV, $\Delta_2=0.49$ eV.

The small increase in δ from 31% to 33.9% in samples *7* and *8* from group Ib, respectively, leads to inversion of the type of conduction in the sample over the entire temperature range and the appearance of dislocation conduction for $T<30$ K, which was previously observed in Refs. 3, 4, 7, and 8. In this case $\sigma_{4.2}$ increases by more than two orders of magnitude (compare curves *7* and *8* in Fig. 3). We shall call samples *8* and *9* p^*-type samples. Thus, in plastically deformed *n*-type germanium with $N_d=2.4\times10^{16}\,\mathrm{cm}^{-3}$ the value of $\sigma_{4.2}$ varies nonmonotonically as δ is gradually increased to 40%.

4. DISCUSSION

In the Shockley–Read models[13,23] some of the electrons from shallow chemical donors are trapped on states in dislocation cores in *n*-type germanium at $T=0$ K, and regions of positive space charge (Read cylinders) form around the dislocations. In accordance with the scheme of the energy spectrum of the dislocation states (Fig. 4) in strained samples of germanium with $N_d=2.4\times10^{16}\,\mathrm{cm}^{-3}$ and $\delta<30\%$, which retain an electronic type of conduction, the cutoff for filling of the dislocation states is found in the E_2 band. The free electrons within blocks are separated by dielectric regions from the electrons trapped on dislocations in the walls and within the blocks. Therefore, several conduction mechanisms are possible in such samples: conduction by free electrons and hopping conduction between shallow chemical donors by electrons outside the Read cylinders and conduction by electrons trapped on dislocations over the system of dislocation grids in block walls.

It follows from Figs. 2 and 3 that the conductivity by free electrons outside the Read cylinders (at $T>8$ K) decreases as the degree of strain δ increases. The exponential decrease in the concentration of free electrons as a function of temperature promotes a decrease in the contribution of

this conductivity at low temperatures. According to Ref. 8, the hopping conductivity among shallow donors with a constant activation energy (at $T<6$ K) in *n*-type germanium with $N_d=2.5\times10^{16}\,\mathrm{cm}^{-3}$ also decreases with increasing δ.

Since the dislocation conductivity in *p*- and *p**-type germanium is quite high in the range $22\%<\delta<30\%$ [$\sigma_{4.2}>5\times10^{-3}\Omega^{-1}\cdot\mathrm{cm}^{-1}$ (Ref. 3)], a connected system of conducting dislocation segments probably also forms in *n*-type germanium at these values of δ. Taking these arguments into account, we can assume that in the strained samples *2–4* in Fig. 2 and samples *5–7* in Fig. 3 the conductivity for $T<8$ K, which falls off slowly with temperature, is attributable to the motion of electrons trapped on dislocations in a continuous system of dislocation grids in block walls. It is distinguished from the dislocation conductivity in the *p*- and *p**-type samples (see Ref. 3) by the significantly smaller values of $\sigma_{4.2}$ at similar values of δ and by the fact that it arises at lower temperatures.

In a disordered system of dislocation grids generally the static conductivity is exponentially dependent on the connectivity of the dislocation segments.[24] The very low values of $\sigma_{4.2}$ ($<10^{-4}\Omega^{-1}\cdot\mathrm{cm}^{-1}$) in the samples from group II with $20<\delta<29\%$, which were deformed under the same conditions as the samples from groups Ia and Ib, can be regarded as a manifestation of the influence of the connectivity of the dislocation system on $\sigma_{4.2}$. In fact, the motion of dislocations in intersecting (111) and $(1\bar{1}\bar{1})$ planes during deformation of the samples from groups Ia and Ib promotes the joining of dislocation grids lying in parallel (111) and $(1\bar{1}\bar{1})$ planes into a single macroscopic dislocation grid penetrating the entire volume of the sample.

The connectivity of this system can probably be improved by annealing the strained samples at temperatures $T>T_d$, which promotes an increase in the area of the regular grids.[5,21,22] Preliminary investigations showed that after the annealing of sample *4* (*p**-type, $\delta=45\%$) from Ref. 2 at 890 °C for 5 h the value of $\sigma_{4.2}$ does, in fact, increase. Therefore, it can be assumed that the annealing of samples *2* ($\delta=20.9\%$) and *3* ($\delta=21.5\%$) from group Ia at 800 °C promotes improvement of the connectivity of the system of dislocation segments and an increase in $\sigma_{4.2}$ in them to values comparable to those for samples *5* and *6* from group Ib, which have higher values of δ (27.7 and 28.4%, respectively). The influence of the small difference in the number of electrons in the E_2 band in these samples is less significant. On the other hand, since the connectivity of the dislocation system increases as a function of δ, the sharp drop in the value of $\sigma_{4.2}$ in samples *4* ($\delta=28\%$) and *7* ($\delta=31\%$) in comparison to samples *3* and *6*, respectively, is due to the decrease in the concentration of electrons in the E_2 band with increasing δ.

According to Refs. 25 and 26, the point defects and polyatomic complexes appearing during the plastic deformation of germanium vanish after brief (for several minutes) annealing at a temperature $T\geqslant700$ °C. The bulk of the copper (one of the rapidly diffusing impurities in germanium) is found near dislocations in the form of isolated precipitates. The latter bind some of the oxygen and, possibly, carbon. Therefore, the variation of $\sigma_{4.2}$ in response to annealing or an increase in δ is caused, with a high probability, by changes in the system of dislocation segments itself.

The existence of a considerable dislocation conductivity in *n*-type germanium with $N_d\leqslant1\times10^{16}\,\mathrm{cm}^{-3}$ following inversion of the type of conduction was associated in Ref. 3 with the appearance of a new structural element in them, viz., dislocation intersection sites (grid sites). They belong to the category of dislocation line defects to which kinks, steps, and impurity atoms in dislocation cores are also assigned. It was theorized in Refs. 15 and 16 that the donor (ε_1) and acceptor (ε_2) states of such defects are located below the top of the E_1 donor band (Fig. 4). Some of the unpaired electrons in dislocation cores pass into these states, and holes form in the E_1 band. Each such hole is attracted to the nearest negatively charged dislocation defect and localized near it. The narrow E_1' acceptor band was associated in Refs. 15 and 16 with just such dislocation defects.

For $\delta<20\%$ the concentration of dislocation defects n_d is small ($n_d\ll N_d$), all the E_1' states are filled by electrons, the dislocation filling cutoff is located in the E_2 acceptor band (Fig. 4), and the low-temperature conductivity is mediated by electrons trapped on dislocations. If the value of N_d is fixed and n_d (i.e., the number of acceptor states in the E_1' band) is increased, the number of acceptors in the E_2 band decreases with increasing δ as a consequence of the filling of states in the E_1' band, and if $n_d\sim N_d$, the E_2 band is emptied. This case probably corresponds to the minimum value of $\sigma_{4.2}$ in Figs. 2 and 3. When $n_d>N_d$ holds, the cutoff for filling of the dislocation states shifts into the E_1' band with resultant inversion of the type of conduction and the appearance of hole-type dislocation conduction.[3] It follows from Fig. 3 that this occurs already for strained sample *8* ($\delta=33.9\%$). Therefore, at such values of δ the total concentration of dislocation defects exceeds $2.4\times10^{16}\,\mathrm{cm}^{-3}$. This conclusion is consistent with the estimate of the possible concentration of dislocation defects obtained by another method. For the maximum dislocation density in the samples with $\delta>30\%$ we take the value $N_D=5\times10^{10}\,\mathrm{cm}^{-2}$, which corresponds to a concentration of unpaired electrons $\sim10^{18}\,\mathrm{cm}^{-3}$. Taking into account that the filling factor of the dislocation states in *p*-type germanium equals 0.05,[14] we obtain a concentration of acceptor centers equal to $5\times10^{16}\,\mathrm{cm}^{-3}$.

The conducting cluster probably contains only some of the dislocation segments, i.e., the total number of dislocation holes is greater than the number of holes contributing to the dislocation conductivity. The concentration of the latter was identified in Ref. 3 with the Hall number $p_h=1/eR$ (R is the Hall coefficient, and e is the charge of an electron), which was determined from measurements of R in the dislocation conduction region. The values of p_h for sample *4* ($\delta=36\%$) in Fig. 3 and sample *7* ($\delta=39.3\%$) in Fig. 4 in Ref. 3 are equal to 4×10^{15} and $6\times10^{15}\,\mathrm{cm}^{-3}$, respectively, i.e., they are, in fact, lower than the estimated concentration of dislocation defects. However, the relation between p_h and R for a percolation system of quasi-one-dimensional dislocation segments is unknown.

It follows from Fig. 3 that the dislocation conductivity decreases as a function of temperature in the *p**-type

samples. It was established in Ref. 4 that the dislocation conductivity can be described by a dependence of the form $\sigma(T)\sim T^y$ over a very broad temperature range (0.01–40 K) and that the values of y decrease from 1.5 to 0.4 as δ increases from 30 to 43%. The values of y for the samples investigated in Ref. 3 decreased from 0.8 to 0.2 as δ increased from 22 to 39.3%. The values $y=0.35$ and 0.23 are obtained using such a description for samples *8* and *9* (Fig. 3). The disparity between the experimental values of y for samples with similar values of δ is due to the dependence of this parameter on the deformation conditions. It was noted in Ref. 3 that no difference between the empirical dependences $\sigma(T)\propto T^y$ and $\sigma(T)\propto \ln T$ can be traced for $y<0.2$ within the experimental error for $4.2<T<20$ K. If T^y is written in the form $\exp(y\ln T)$ and expanded in powers of $y\ln T$, it can easily be shown that the values of T^y and $1+y\ln T$ for $y<0.2$ are essentially indistinguishable in the temperature range 4.2–20 K. It is known[27] that a function of the form $1+y\ln T$ characterizes the conductivity of a two-dimensional degenerate electron gas (in the weak-disorder limit) with consideration of the quantum corrections in the weak-localization and electron-electron interaction theories and that the corrections themselves are considerably smaller than the value of the residual metallic conductivity σ_0.

According to Refs. 28 and 29, the variation of the resistivity of strongly plastically deformed *p*-type germanium in a magnetic field in the temperature range 0.1–4.2 K follows the laws characteristic of two-dimensional metallic systems. This fact is not surprising, since the localization radius of unpaired electrons on dislocations (~ 10 Å) is much smaller than the distance between dislocation segments in grids, which are extended two-dimensional formations (see the figures in References 6, 22, and 23 and Fig. 1b). Therefore, the dislocation system in strongly plastically deformed samples consists of macroscopic fragments of two-dimensional dislocation grids joined in a specific manner. The lack of a temperature dependence of the Hall coefficient in the dislocation conduction region at $T<10$ K (Refs. 2 and 3) can be evidence that the Fermi level is located within the dislocation band. If the carriers are delocalized at the Fermi level, the conductivity along dislocation grids has a metallic character, and the weak decrease in the dislocation conductivity as a function of temperature in *p*- and *p**-type germanium can be associated with the influence of the quantum corrections.

We estimate the relative addition to the residual conductivity $\alpha=[\sigma(T)-\sigma_0]/\sigma_0=y\ln T$ at 4.2 K for such a case using the experimental values of y presented above. The minimum value $\alpha_{4.2}=0.12$ was obtained for a *p*-type sample ($\delta=38\%$ and $y=0.08$), on which the magnetoresistance was measured in Refs. 28 and 29. The additions to σ_0 for $y\geqslant 0.2$ are found to be larger ($\alpha_{4.2}\geqslant 0.28$), and they increase with decreasing δ. Since macroscopic inhomogeneities in the samples influence the manifestations of the quantum effects,[30] the decrease in the connectivity of the dislocation cluster with decreasing δ probably has a significant influence on the character of $\sigma(T)$ for $T>4.2$. Then, according to Ref. 4, the plots of $\sigma(T)\propto T^y$ with $y<0.1$ at $T<30$ K describe the metallic state far from the metal–insulator transition, and the plots for $y>0.2$ characterize the conductivity of the dislocation cluster as the classical metal–insulator transition is approached. We note that in two-dimensional conducting channels based on silicon and gallium arsenide[31,32] the logarithmic dependence of σ on T in the metallic region (at a high concentration of free electrons) gives way to a power-law dependence with $y=0.8$ when the electron concentration decreases. This was regarded as a manifestation of the correction to the logarithmic dependence in the next approximation.

In this model the dislocation conductivity in the *p*- and *p**-type samples can be associated with the motion of delocalized holes, although the cutoff for filling of the dislocation states is located in the E_1' band. For $N_D<10^7$ cm^{-2} the values of n_d are small ($\sim 10^{13}$ cm^{-3}), and the hole states are positioned far from one another and are, in fact, localized. Consequently, microwave conductivity is not observed in overcompensated samples.[11] When n_d is increased to values of the order of 10^{16} cm^{-3}, the situation can change, because the dislocation defects are arranged in two-dimensional grids, rather than distributed uniformly throughout the volume. In heavily doped *p*-type semiconductors the impurity band merges with the valence band, if the mean distance between acceptors is equal to twice the radius r_a of the hole wave function (''complete'' overlap of the wave functions). For shallow chemical acceptors in germanium $r_a=85$ Å (Ref. 33). The estimated value of the transverse radius of the wave functions of unpaired electrons and holes localized on dislocation defects, whose states are ~ 0.1 eV above the top of the valence band, is ~ 10 Å. The binding energy of dislocation holes to negatively charged defects depends on the defect potential and probably does not exceed Δ_1' [$\Delta_1'<0.03$ eV (Refs. 3 and 5)]. If the distance occupied by hole wave functions along a dislocation is of the same order as r_a, then finding several kinks or steps on a dislocation segment with a length of ~ 500 Å (Refs. 3 and 5) is sufficient for ''complete'' overlap of the wave functions of the localized holes. Then the dislocation holes are delocalized as a consequence of the broadening of the E_1' band and its merger with the E_1 band.

Negatively charged dislocation defects are scattering centers for electrons in the E_2 band and can lead to the partial or complete localization of states in this band.[34] Therefore, the amplitude and character of the temperature dependence of the electron conductivity along dislocation grids can differ from those in the *p*- and *p**-type samples at equal values of δ. The narrowness of the temperature range precludes reliable determination of the character of the dependence of σ on T in the *n*-type samples for $T<8$ K. It follows from Fig. 3 for samples *5* ($\delta=27.7\%$) and *6* ($\delta=28.4\%$) that the value of $\sigma_{4.2}$ ($\sim 10^{-3}\Omega^{-1}\cdot$cm^{-1}) is, in fact, significantly smaller than the value $\sigma_{4.2}=4\times 10^{-2}\Omega^{-1}\cdot$cm^{-1} in the *p*-type ($\delta=27\%$, $N_a=2\times 10^{12}$ cm^{-3}) and *p**-type ($\delta=29\%$, $N_d=1\times 10^{16}$ cm^{-3}) samples from Ref. 3 prepared by the same method. Thus, at a fixed value of δ ($\sim 28\%$) the value of $\sigma_{4.2}$ depends on N_d, if N_d is comparable to n_d (for $1\times 10^{16}<N_d<2.4\times 10^{16}$ cm^{-3}), while no such dependence is observed when $N_d\ll n_d$.[3]

This finding, as well as the nonmonotonic variation of

$\sigma_{4.2}$ with δ (Fig. 3), reflects the presence of a gap between the donor and acceptor dislocation states. In this case dislocation conduction can be mediated by both electrons and holes. In this respect static conduction along dislocation grids differs fundamentally from hole conduction on germanium bicrystal boundaries.[18,19] The model of a boundary composed of edge dislocations with unpaired electrons, whose states are strongly localized and located in the valence band, was used in the papers just cited. In *p*-type crystals the negative charge of dislocations which have trapped electrons from the valence band, is screened by the cloud of mobile degenerate holes in the valence band (the radius of the cloud is about 30 Å). When the misorientation angle between the two parts of a bicrystal is large ($\theta = 20-30°$), the distance between neighboring dislocations amounts to 15–30 Å (Ref. 18), the hole clouds overlap, and a two-dimensional conducting plane forms. The concentration of degenerate holes determined from measurements of the Hall effect in bicrystals at low temperatures turned out to be fairly large ($10^{12}-10^{13}\,\mathrm{cm}^{-3}$) and not dependent on the dopant concentration when N_d, $N_a < 2\times10^{16}\,\mathrm{cm}^{-3}$ (Ref. 19).

An energy spectrum of quasi-one-dimensional states with a gap can also be characteristic of states in a dislocation deformation potential. In the absence of dopants the states near the bottom of the conduction band are then empty, and states near the top of the valence band are filled by electrons.[17] The attempt which we undertook in Ref. 5 to observe static dislocation conductivity along dislocation grids in low-doped *n*- and *p*-type silicon, in which there are no dangling bonds of linear dislocation segments, was unsuccessful.

We express our sincerest thanks to Yu A. Osip'yan, V. V. Kveder, A. I. Kolyubakin, I. A. Ryzhkin, V. D. Shikin, and V. M. Edel'shteĭn for discussing the results and offering some valuable comments, as well as I. I. Khodos for permission to publish the photograph shown in Fig. 1b.

*)E-mail: shevchen@issp.ac.ru

[1] Yu. A. Osip'yan and S. A. Shevchenko, JETP Lett. **20**, 328 (1974).
[2] Yu. A. Osip'yan and S. A. Shevchenko, JETP Lett. **33**, 207 (1981).
[3] V. A. Goncharov, Yu. A. Osip'yan, and S. A. Shevchenko, Fiz. Tverd. Tela (Leningrad) **29**, 1928 (1987) [Sov. Phys. Solid State **29**, 1110 (1987)].
[4] I. V. Klyatskina, M. L. Kozhukh, S. M. Ryvkin *et al.*, Fiz. Tekh. Poluprovodn. **13**, 1089 (1979) [Sov. Phys. Semicond. **13**, 638 (1979)].
[5] S. A. Shevchenko, Yu. A. Ossipyan, T. R. Mchedlidze *et al.*, Phys. Status Solidi A **146**, 745 (1994).
[6] H. G. Brion and P. Haasen, Philos. Mag. A **51**, 879 (1985).
[7] K. N. Zinov'eva, M. L. Kozhukh, V. A. Trunov *et al.*, JETP Lett. **30**, 281 (1979).
[8] I. V. Klyatskina, M. L. Kozhukh, S. M. Ryvkin *et al.*, Fiz. Tekh. Poluprovodn. **15**, 795 (1981) [Sov. Phys. Semicond. **15**, 454 (1981)].
[9] M. L. Kozhukh, S. M. Ryvkin, I. S. Shlimak *et al.*, Fiz. Tekh. Poluprovodn. **15**, 1423 (1981) [Sov. Phys. Semicond. **15**, 823 (1981)].
[10] Yu. A. Osip'yan, V. I. Tal'yanskiĭ, and S. A. Shevchenko, Zh. Éksp. Teor. Fiz. **72**, 1543 (1977) [Sov. Phys. JETP **45**, 810 (1977)].
[11] Yu. A. Osip'yan, V. I. Tal'yanskiĭ, A. A. Kharlamov, and S. A. Shevchenko, Zh. Éksp. Teor. Fiz. **76**, 1655 (1979) [Sov. Phys. JETP **49**, 840 (1979)].
[12] V. A. Grazulis, V. V. Kveder, and V. Yu. Mukhina, Phys. Status Solidi A **43**, 407 (1977); **44**, 107 (1977).
[13] W. Shockley, Phys. Rev. **91**, 228 (1953).
[14] A. I. Kolyubakin and S. A. Shevchenko, Phys. Status Solidi A **63**, 677 (1981).
[15] V. V. Kveder and Yu. A. Osip'yan, Zh. Éksp. Teor. Fiz. **80**, 1206 (1981) [Sov. Phys. JETP **53**, 618 (1981)].
[16] Yu. A. Ossip'yan, Sov. Sci. Rev., Sect. A **4**, 219 (1982).
[17] V. B. Shikin and Yu. I. Shikina, Usp. Fiz. Nauk **165**, 887 (1995) [Phys. Usp. **38**, 845 (1995)].
[18] B. M. Vul, É. I. Zavaritskaya, Yu. A. Bashkirov, and V. M. Vinogradova, JETP Lett. **25**, 187 (1977).
[19] B. M. Vul and É. I. Zavaritskaya, Zh. Éksp. Teor. Fiz. **76**, 1089 (1979) [Sov. Phys. JETP **49**, 551 (1979)].
[20] Yu. A. Osip'yan, V. M. Prokopenko, V. I. Tal'yanskiĭ *et al.*, JETP Lett. **30**, 111 (1979).
[21] R. Weber, Phys. Status Solidi A **24**, 575 (1974).
[22] D. Gwinner and G. Packeiser, Philos. Mag. A **42**, 645 (1980).
[23] W. T. Read, Philos. Mag. **45**, 775 (1954); **45**, 1119 (1954).
[24] I. A. Ryzhkin, Fiz. Tverd. Tela (Leningrad) **20**, 3612 (1978) [Sov. Phys. Solid State **20**, 2087 (1978)].
[25] A. I. Kolyubakin and S. A. Shevchenko, Fiz. Tekh. Poluprovodn. **13**, 1046 (1979) [Sov. Phys. Semicond. **13**, 613 (1979)].
[26] S. A. Shevchenko, Fiz. Tekh. Poluprovodn. **20**, 275 (1986) [Sov. Phys. Semicond. **20**, 172 (1986)].
[27] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, Elsevier, New York (1988) [Russ. original, Nauka, Moscow (1987), p. 182].
[28] O. V. Zharikov, Yu. A. Osip'yan, and S. A. Shevchenko, JETP Lett. **39**, 296 (1984).
[29] O. V. Zharikov, Yu. A. Osip'yan, and S. A. Shevchenko, in *Abstracts of Reports to the 23th All-Union Conference on Low-Temperature Physics* [in Russian], Tallin (1984), p. 72.
[30] A. G. Aronov, M. E. Gershenzon, and Yu. E. Zhuravlev, Zh. Éksp. Teor. Fiz. **87**, 971 (1984) [Sov. Phys. JETP **60**, 554 (1984)].
[31] R. A. Davis and M. Pepper, J. Phys. C: Solid State Phys. **15**, 371 (1982).
[32] D. J. Newson, Philos. Mag. **52**, 437 (1985).
[33] B. I. Shklovskiĭ, Fiz. Tekh. Poluprovodn. **6**, 1197 (1972) [Sov. Phys. Semicond. **6**, 1053 (1973)].
[34] I. A. Ryzhkin, Zh. Éksp. Teor. Fiz. **81**, 2192 (1981) [Sov. Phys. JETP **54**, 1165 (1981)].

Translated by P. Shelnitz

Cluster structure of the material surface as the cause of the selective enhancement of ultracold-neutron capture associated with subbarrier reflection

S. S. Arzumanov, S. T. Belyaev, L. N. Bondarenko,*) S. M. Ivanov, E. I. Korobkina, A. N. Lyubimov, V. I. Morozov, A. I. Ryazanov, Yu. N. Panin, A. I. Fomin, and S. M. Chernyavskiĭ

Kurchatov Institute Russian Scientific Center, 123182 Moscow, Russia

P. Geltenbort and J. Pendlebury

Institut Laue Langevin, 38042 Grenoble Cedex 9, France

K. Schreckenbach

Technische Universität München, D-85747 Garching, Germany

(Submitted 3 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 126–140 (January 1999)

Neutron-radiation analysis is used to investigate the subbarrier reflection of ultracold neutrons from the surface of a titanium-stabilized Fe–Ni–Cr stainless steel. A significant selective increase in the probability of neutron capture by nuclei of the medium in comparison to theory is discovered. An explanation is given for the effect, which is associated with the existence of titanium-containing clusters and structural defects that distort the form of the distribution of the effective interaction potential between ultracold neutrons and the material surface. © *1999 American Institute of Physics.* [S1063-7761(99)01201-9]

INTRODUCTION

It was reported in a preliminary publication[1] that theoretically unexpected enhancement of ultracold-neutron capture by the principal elements comprising 1Kh18N9T stainless steel was discovered during an investigation of the subbarrier reflection of such neutrons from this material. Further study of the structural features of this steel using electron microscopy provided an explanation for the physical mechanism of this enhancement as a manifestation of the cluster structure of the material, which causes the character of ultracold-neutron capture to differ from the simple model of this interaction for a homogeneous multielement medium.

For a medium containing several (q) elements uniformly distributed throughout the volume with the relative nuclear concentrations c_i, it is known that the total probability of the interaction of ultracold neutrons with the surface is equal to the sum of all the partial probabilities μ_c^i of capture by each specific element of the medium and the inelastic scattering probability μ_{ie}:

$$\mu=\mu_{ie}+\sum_{i=1}^{q}\mu_c^i, \tag{1}$$

where $\mu_{ie}=\eta_{ie}f(v)$ and $\mu_c^i=\eta_c^i f(v)$ are specified by the relations

$$f(v)=2y^{-2}[\arcsin y-y\sqrt{1-y^2}\,],$$

$$\eta_{ie}=k\sigma_{ie}/4\pi\overline{b}, \quad \eta_c^i=k\sigma_c^i c_i/4\pi\overline{b}.$$

Here v is the neutron velocity, $y=v/v_{\rm lim}$, $v_{\rm lim}=\sqrt{2E_{\rm lim}/m}$ is the cutoff velocity of the medium, m is the neutron mass, $E_{\rm lim}=h^2N\overline{b}/2\pi m$ is the cutoff energy of the medium, N is the number of nuclei per unit volume, $\overline{b}=\Sigma_{i=1}^{q}c_ib_i$ is the coherent scattering length averaged over the atomic composition of the medium, k is the wave number, σ_{ie} is the inelastic scattering cross section, and σ_c^i is the capture cross section of the ith element averaged over its isotopic composition.

Experimental testing of the theory of the interaction of ultracold neutrons with a medium in the presence of subbarrier reflection became possible following the appearance of neutron-radiation analysis for ultracold neutrons.[2] In the present work this method was used to study the interaction of ultracold neutrons with stainless steel both with capture and with inelastic scattering.

A diagram of the setup is shown in Fig. 1. The sample under investigation was irradiated by ultracold neutrons in a cylindrical stainless steel vessel of length 112 cm and diameter 8.8 cm. The sample, which was prepared from 1Kh18N9T stainless steel, was an electropolished foil of thickness 200 μm in the form of a spiral of width 10 cm and external diameter 8 cm with a total area $S=3120\,\text{cm}^2$.

Neutrons from an ultracold-neutron source entered the vessel along a vertical neutron guide through an entrance diaphragm with an aperture having an area $S_0=0.785\,\text{cm}^2$ or through the entire cross section of the vessel when the diaphragm was removed. The velocity spectrum of the ultracold neutrons in the vessel was concentrated in the range from 0 to 4.4 m/s with a mean velocity $\overline{v}=3.8(2)$ m/s. The flux density of ultracold neutrons in the vessel and at its entrance was measured by three gas proportional detectors D_1, D_2, and D_3, to which the vessel is exposed through apertures of area S_0. The γ rays appearing on the sample surface in con-

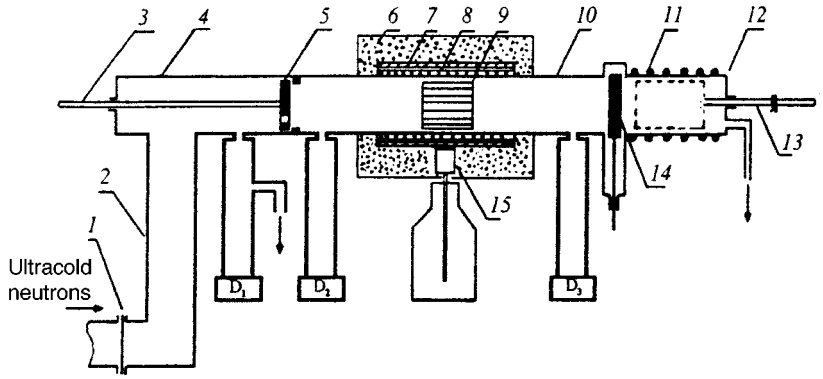


FIG. 1. Diagram of the setup: *1* — entrance Al foil, *2* — vertical neutron guide, *3* — rod for moving the entrance diaphragm, *4* — inlet chamber, *5* — movable entrance diaphragm, *6* — lead shielding, *7* — ^{10}B converter, *8* — jacket with a heating element and a heat shield, *9* – sample, *10* — vessel for ultracold neutrons, *11* — heating element, *12* — annealing chamber, *13* — rod for moving the sample, *14* — vacuum valve, *15* — Ge detector, D_1, D_2, D_3 — ultracold-neutron detectors.

junction with ultracold-neutron capture were detected by a Ge(Hp) detector made from ultrapure Ge with a resolution of 2 keV at a γ-ray energy of 1 MeV. Inelastically scattered ultracold neutrons were detected using a ^{10}B converter of thickness 1 cm positioned between the vessel and the entrance window of the detector. A ^{7}Li nucleus forms in an excited state with probability 0.96 in the reaction $n+{}^{10}\mathrm{B}=\alpha+{}^7\mathrm{Li}$ and emits 477 keV γ rays, which are picked up by the detector.

When the background was measured or the sample was degassed, the latter was withdrawn from the vessel and moved to a special chamber, which was separated from the vessel by a vacuum valve. The residual pressure in the vessel was maintained at $2\times10^{-5}-5\times10^{-6}$ Torr, and the residual pressure in the chamber was $\approx10^{-3}$ Torr.

When the total loss coefficient was measured, ultracold neutrons entered the vessel through the diaphragm. The counting rates j_1, j_2, and j_3 of detectors D_1, D_2, and D_3 were used to determine the quantity

$$\bar{\mu}S+\bar{\mu}_tS_t=\frac{2(j_1-2j_2-j_3)S_0}{j_2+j_3}, \tag{2}$$

where $\bar{\mu}$ and $\bar{\mu}_t$ are the total loss coefficients averaged over the ultracold-neutron flow for the sample surface and the vessel, respectively, and S_t is the surface area of the vessel. The value of $\bar{\mu}_tS_t$ was determined with the sample withdrawn from the vessel. The value of $\bar{\mu}$ for the sample was calculated from the results of two measurements by a difference method.

To measure the inelastic scattering probability and the partial capture coefficients of ultracold neutrons, the diaphragm was removed so that they would enter the vessel through the entire cross section. The inelastic scattering probability averaged over the neutron velocity spectrum was defined as

$$\bar{\mu}_{ie}=\frac{2j_{ie}S_0\varepsilon}{(j_2+j_3)S\varepsilon_{ie}}, \tag{3}$$

where j_{ie} is the pulse counting rate in the total absorption peak of 477 keV γ rays, ε_{ie} is the detection efficiency of ultracold neutrons that are inelastically scattered on the sample surface and manifested in j_{ie}, and ε is the detection efficiency of ultracold neutrons that have passed through the entrance apertures of the detectors. The ratio $\varepsilon_{ie}/\varepsilon$ was determined in an additional measurement with a polyethylene reference sample.

The probability of ultracold-neutron capture by the ith element averaged over the flow was defined as

$$\bar{\mu}_c^i=\frac{2j_i(E)\varepsilon S_0}{(j_2+j_3)\beta_i\varepsilon_\gamma(E)S}, \tag{4}$$

where $j_i(E)$ is the counting rate in the total absorption peak of γ rays with energy E, β_i is the yield of quanta with energy E per neutron capture act for a natural mixture of isotopes of the ith element, and $\varepsilon_\gamma(E)$ is the detection efficiency of γ rays with energy E emitted from the sample surface. The energy dependence of the ratio $\varepsilon_\gamma(E)/\varepsilon$ was determined by additional measurements using polyethylene, titanium, and aluminum reference samples.

The measurements were performed after chemical cleaning of the sample surface by etching in H_3PO_4 and vacuum annealing at about 1000 K for 2 h. Figure 2 presents fragments of the γ-ray spectrum. Total absorption peaks of 477 keV and 2.22 MeV γ rays, which are produced by the inelastic scattering and capture of ultracold neutrons on surface hydrogen, can be seen in the spectrum. The spectra obtained were treated with allowance for the external γ background and the γ background caused by the interaction of ultracold neutrons with the vessel surface.

The measurements yielded the following values: $\bar{\mu}=6.44(57)\times10^{-4}$, $\bar{\mu}_{ie}=1.57(26)\times10^{-4}$, and $\bar{\mu}_c^{\mathrm{H}}=2.2(3.8)\times10^{-6}$. There is a difference between the total ($\bar{\mu}$) and summed ($\bar{\mu}_{ie}+\bar{\mu}_c^{\mathrm{H}}$) probabilities, which equals $\Sigma_{i=1}^q\bar{\mu}_c^i=4.85(62)\times10^{-4}$ and specifies the total probability of ultracold-neutron capture by elements other than hydrogen.

Direct measurements of the partial capture probabilities were performed at the most intense γ transitions for ultracold-neutron capture by Fe, Ni, Cr, and Ti nuclei. For comparison with theory, the values of the parameters $\eta_c^i=\bar{\mu}_c^i/f(v)$ were determined from the values of $\bar{\mu}_c^i$ obtained in the approximation $\overline{f(v)}=f(\bar{v})=0.96$. The results are presented in Table I. It can be seen from the table that $\Sigma_{i=1}^4\bar{\mu}_c^i=4.99(18)\times10^{-4}$, which is consistent with the independent data obtained from measurements of $\bar{\mu}$, $\bar{\mu}_{ie}$, and

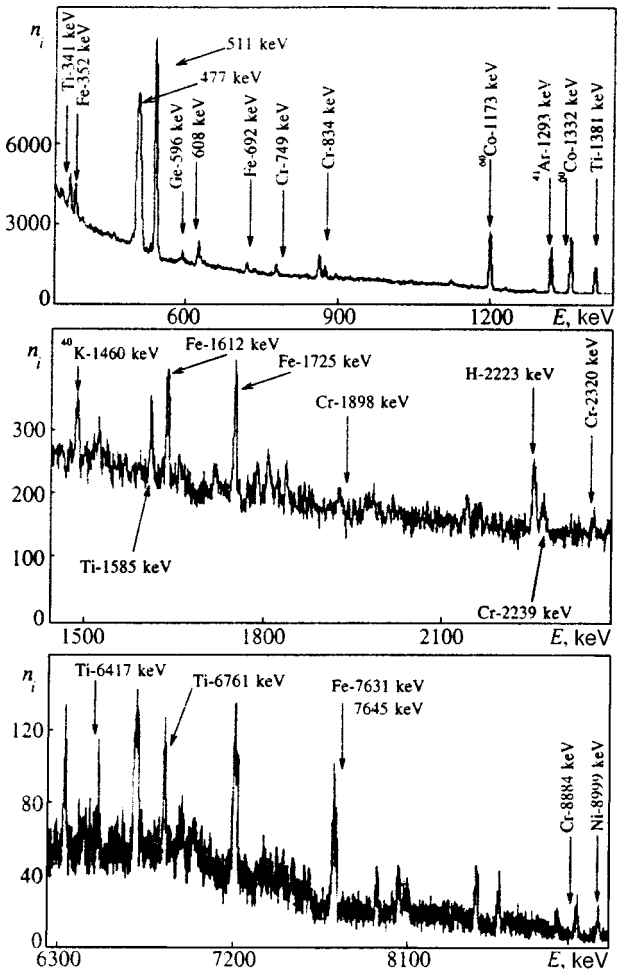


FIG. 2. Fragments of the γ-ray spectra for the irradiation of a sample by ultracold neutrons; n — number of pulses from the γ detector in relative units.

$\overline{\mu_c^{\mathrm{H}}}$. A comparison of the experimental and theoretical values of η_c^i reveals that the experimental capture probabilities are considerably higher than the theoretical values. This increase has a selective character and is most pronounced for titanium.

The observed enhancement cannot be attributed to a possible admixture in the spectrum of ultracold neutrons with $v \geqslant v_{\mathrm{lim}} = 6$ m/s, which could increase the partial capture probabilities as a result of the above-barrier penetration of ultracold neutrons into the bulk of the sample. Control measurements for a copper sample with $v_{\mathrm{lim}} = 5.65$ m/s showed that the experimental value of $\overline{\mu_c^{\mathrm{Cu}}}$ exceeds the theoretical value by no more than a factor of 2–2.5. While this upward deviation can be associated with the penetration of neutrons with $v > 5.65$ m/s into a copper sample, this effect will be weaker for stainless steel and can only partially account for the increase in capture on Fe, but has little effect on the observed enhancement for other elements.

TABLE I. Results of measurements of $\overline{\mu_c^i}$ and η_c^i.

Element	Ni	Ti	Fe	Cr
$\overline{\mu_c^i} \times 10^4$	0.83(10)	1.23(4)	1.74(10)	1.19(9)
$\eta_c^i \times 10^4$, experiment	0.86(11)	1.28(5)	1.81(11)	1.24(10)
$\eta_c^i \times 10^4$, theory	0.128	0.014	0.60	0.207
Exp./theor. ratio	6.7	91	3	6
$\Delta\eta_c^i \times 10^4$	0.73(11)	1.27(5)	1.21(11)	1.03(10)

TABLE II. Results of the atomic analysis of the samples.

Element	Ti	Fe	Ni	Cr	Si	C	O
Content, at. %, method (1)	0.7	70.6	8.7	20.0	–	–	–
Content, at. %, method (2)	0.6	70.7	8.3	20.4	–	–	–
Content, at. %, method (3)	0.7	69.5	8.5	19.7	1.6	–	–
Content, at. %, method (4)	0.3	32.4	4.2	11.3	–	28.8	23.0

Such enhancement might appear if there is a layer strongly enriched with titanium on the surface. Since $\overline{b} < 0$ holds for titanium, in such a model the potential at the surface has the form of a potential well in front of a positive potential jump of height E_{lim}. Considerable enhancement of ultracold-neutron capture on titanium nuclei would be possible in this case. To test this possibility, atomic analysis was performed within the sample and in its surface layer using (1) bulk neutron-radiation analysis, (2) x-ray fluorescence analysis at a depth of 50 μm, (3) an x-ray spectral electron probe microanalyzer at a depth of 1 μm, and (4) an x-ray photoelectron spectrometer at depths $\leqslant 10$ nm. The results are presented in Table II. According to the data obtained by methods (1)–(3), the content of the principle elements corresponds to 1Kh18N9T steel, for which η_c^i was calculated. No tendency for an increase in the Ti content is observed as the thickness of the layer being analyzed is decreased. Moreover, when the thickness of the layer is on the order of the wavelength of ultracold neutrons, the absolute content of Ti and the other elements of the stainless steel is less than the bulk value because of oxygen, carbon, and hydrogen atoms, to which method (4) is not sensitive. In addition, the ratio between the concentrations of Ti, Fe, Ni, and Cr corresponds to the bulk ratio.

The small value of the mean concentration of titanium on the surface did not rule out the possibile existence of regions where it is localized with an increased content. This raised the hypothesis that titanium forms clusters emerging on the surface, whose dimensions are greater than the wavelength of ultracold neutrons. Then neutrons freely pass into these clusters, move about in them, and are reflected from the boundaries until they are captured or escape back into the vacuum. If the clusters contain some Fe, Ni, and Cr nuclei, ultracold neutrons will also be efficiently captured by them.

To test this hypothesis, we investigated the structural features of 1Kh18N9T steel using electron microscopy. A JEOL-100CX transmission electron microscope and a DSM-960 digital scanning electron microscope equipped with an x-ray spectral analyzer were used. Thin samples with a ''translucent'' thickness (<2000 Å) were prepared electrolytically in an aqueous solution of sulfuric and orthophosphoric acid at room temperature.

Figure 3 presents photomicrographs of the same portions of a sample surface obtained using the DSM-960 digital scanning electron microscopy in various regimes: in second-

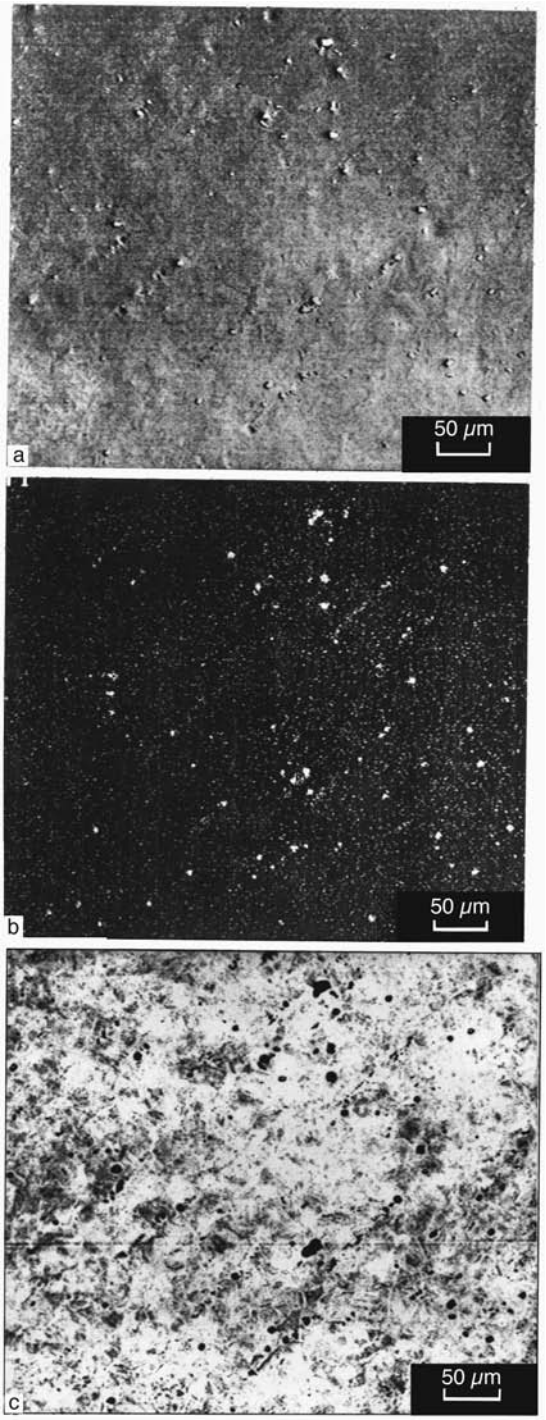


FIG. 3. Photomicrographs of a portion of the surface of a sample of 1Kh18N9T steel obtained using a DSM-960 digital scanning electron microscope in various regimes: a — in second electrons, b — in characteristic K_α Ti radiation, c — in back-reflected electrons.

ary electrons (a), in characteristic K_α Ti radiation (b), and in back-reflected electrons (c). Figure 3a reflects the surface relief and reveals the existence of second-phase inclusions (clusters) emerging on the surface with dimensions up to 9 μm. Figure 3b attests to the presence of an appreciable amount of titanium in the composition of the clusters. Figure 3c shows that the clusters contain an element with a small value of Z (probably carbon), which provides for good contrast between these clusters and the matrix of iron, nickel, and chromium. Estimation of the surface density of cluster outcrops with linear dimensions greater than 1 μm from a series of photographs showed that it is equal to $(2-6)\times 10^4\,\mathrm{cm}^{-2}$. The characteristic dimensions of the cluster outcrops vary from 1 to 9 μm with a mean value of 3 μm.

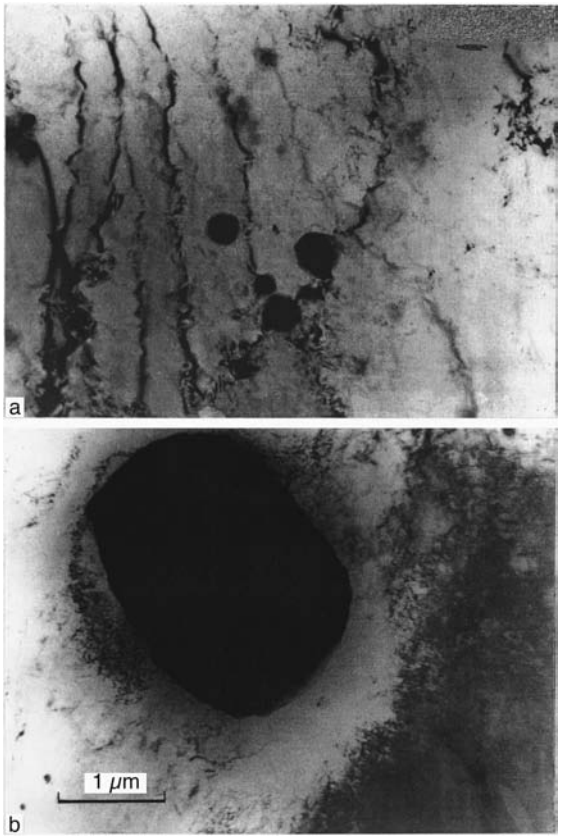


FIG. 4. Photomicrographs of the structure of a sample obtained using a JEOl-100CX transmission electron microscopy: a — group of small clusters, b — large cluster.

Figure 4 presents photomicrographs of the structure of a sample obtained using the JEOL-100CX transmission microscope on regions with a very small thickness (2000 Å). A group of small clusters measuring from 300 to 1000 Å can be clearly traced in Fig. 4a. An analysis of photomicrographs of different regions of the sample showed that the diameter of the small clusters varies in the range 150–1500 Å, that the mean diameter d equals 400 Å, and that the bulk density $n=1.7(5)\times 10^{14}\,\mathrm{cm}^{-3}$. When the diffraction patterns obtained from the small clusters were calculated, the values of the interplanar distances for the second-phase inclusions forming them were found to essentially coincide with the tabulated values for titanium carbide TiC (Table III).

A large precipitate with characteristic dimensions of 2–3 μm, which is assigned to a group of large titanium-containing clusters, was discovered on one of the translucent areas of the sample (Fig. 4b). Thus, two groups of titanium-containing clusters of different size, which are uniformly distributed throughout the volume of the sample, were discovered in it. It is known that titanium can be present in 1Kh18N9T steel in the form of small carbide precipitates, whose formation has a thermodynamic character. Their size,

TABLE III. Tabulated and experimental values of the interplanar distances d_{hkl} for titanium carbide.

hkl	d_{hkl}, tabulated	d_{hkl}, exptl.
111	2.49	2.30
200	2.15	2.10
220	1.52	1.49
311	1.30	–
222	1.245	1.22

number, and distribution in the structure of the metal are determined by the annealing temperature and time of the steel part. In addition, titanium can be present in the form of metallurgical inclusions of various composition, which form during the smelting and preparation of the steel. One characteristic feature of these inclusions is their dimensions, which usually exceed several microns. Apparently, just such inclusions are observed in Figs. 3 and 4b as clusters with characteristic dimensions greater than 1 μm.

The density of titanium carbide is 4.92 g/cm^3, and the scattering lengths of carbon and titanium are equal to 0.66×10^{-12} cm and -0.34×10^{-12} cm, respectively. The cutoff energy $E_{\rm lim}^{\rm TiC}=4.1\times10^{-8}$ eV, which is appreciably smaller than the cutoff energy of the stainless steel $E_{\rm lim}^{\rm ss}=1.8\times10^{-7}$ eV. Therefore, the clusters emerging on the surface deform the distribution of the effective potential for the interaction of ultracold neutrons with the surface. Neutrons with an energy $E>E_{\rm lim}^{\rm TiC}$ can penetrate a cluster, move about in it, and be reflected from its walls until they are captured or escape back into the vacuum. If the diameter of the cluster appreciably exceeds the wavelength of ultracold neutrons ($\lambda\simeq900$ Å), neutron capture will occur mainly on titanium nuclei.

In the classical approach the excess loss coefficient $\Delta\mu_c^{\rm Ti}$ due to ultracold-neutron capture by titanium nuclei in clusters having an outcrop on the surface can be represented in the form

$$\Delta\mu_c^{\rm Ti}=W_1W_pW_c,$$

where W_1 is the probability that ultracold neutrons strike the outcrop cross section of a cluster, W_p is the probability that they penetrate into the bulk of a cluster, and W_c is the probability of ultracold-neutron capture by titanium within a cluster. The bulk density and real diameter of a group of small clusters were determined from transmission photographs. Since the mean outcrop cross-sectional area of the clusters is $\pi d^2/6$ and the number of clusters emerging on a surface of unit area equals $2dn$, we have $W_1=\pi d^3n/3$. The density ρ of cluster outcrops on the sample surface and the mean diameter of these outcrops $\overline{d}$ were determined for the large clusters. In this case $W_1=\pi\overline{d^2}\rho/4$. To estimate W_p we assume that the ultracold-neutron flow has a Maxwellian spectrum, $F(E)\propto E$ from 0 to $E_{\max}=10^{-7}$ eV, which roughly corresponds to the conditions of the experiment. Only ultracold neutrons with a normal velocity component greater than $v_{\rm lim}^{\rm TiC}$ penetrate a cluster. If we neglect the quantum effects of the reflection of ultracold neutrons due to the potential jump at the entrance to a cluster, the value averaged over the spectrum and isotropic distribution of the ultracold-neutron flow is $W_p=(1-E_{\rm lim}^{\rm TiC}/E_{\max})^2=0.35$. Inside a cluster the mean velocity for the ultracold-neutron flow decreases to $v=2.7$ m/s, since the spectrum of the flow takes on the form $F(E)\propto E$ from 0 to $E_{\max}=0.59\times10^{-7}$ eV. The mean free path of the ultracold neutrons before escaping from a cluster can be estimated as being of order $2d$. Hence for the small clusters $W_c=2dN_{\rm Ti}\sigma v_{\rm th}/\overline{v}$, where $\sigma=5.9$ barn is the cross section for the capture of neutrons with a velocity $v_{\rm th}=2200$ m/s by titanium and $N_{\rm Ti}=5\times10^{22}$ cm^{-3} is the number of titanium nuclei in a unit volume of TiC. For the large clusters $W_c\simeq2.5\overline{d}N_{\rm Ti}\sigma v_{\rm th}/\overline{v}$. The value of $\Delta\mu_c^{\rm Ti}\simeq d^4nN_{\rm Ti}\sigma v_{\rm th}/\overline{v}$, which is due to the existence of the group of small clusters, equals $(1.05\pm0.32)\times10^{-5}$. Despite the roughness of the estimate, it can be seen that the existence of the small clusters makes only a 10% contribution to the selective enhancement of capture on titanium.

It should be noted that the closed small clusters are similar to pores in a material, in which bound states of ultracold neutrons are possible.[3,4] If the energy of the ultracold neutrons equals the resonance value, a neutron can penetrate a pore with a high probability, increasing the probability of losses. However, evaluation of the volume fraction occupied by small clusters in the sample material shows that this effect makes only an insignificant contribution to the observed enhancement.

Evaluation of the excess capture probability $\Delta\mu_c^{\rm Ti}$ associated with the penetration of ultracold neutrons into the large clusters gives $\Delta\mu_c^{\rm Ti}\simeq\overline{d^3}\rho N_{\rm Ti}\sigma v_{\rm th}/\overline{v}=(1.3-3.9)\times10^{-4}$. This value shows that the large titanium carbide clusters cause the required enhancement of ultracold-neutron capture by titanium nuclei.

The enhancement of capture by other elements in the steel (Fe, Cr, and Ni) could be associated, in principle, with their presence as impurities in the titanium carbide clusters. However, according to the literature data, the atomic concentration of these elements in second-phase precipitates of TiC is extremely small.[5] This is clearly seen on a qualitative level in Fig. 5, which shows photomicrographs of a sample surface with several large clusters obtained using the scanning electron microscope in secondary electrons (a), as well as in characteristic K_α Ti radiation (b) and characteristic K_α Cr radiation (c). The clusters appear as sharply contrasted dark spots superposed on the light field (Fig. 5c) produced by the characteristic radiation of Cr in the matrix, where its concentration is equal to 18%. This is evidence that the concentration of Cr is considerably smaller in the clusters than in the matrix. Since the cross sections for neutron capture by Fe, Cr, Ni, and Ti are comparable, the titanium-containing clusters can make only an insignificant contribution to the observed excess capture of ultracold neutrons by these nuclei.

A somewhat different mechanism for enhancement of ultracold-neutron capture by these nuclei, which is associated with the presence of phase boundaries between the individual grains of the stainless steel matrix, is possible in principle. Phase boundaries can be clearly seen in Fig. 6, which presents a typical photomicrograph of the surface relief of a sample obtained using the DSM-960 scanning elec-

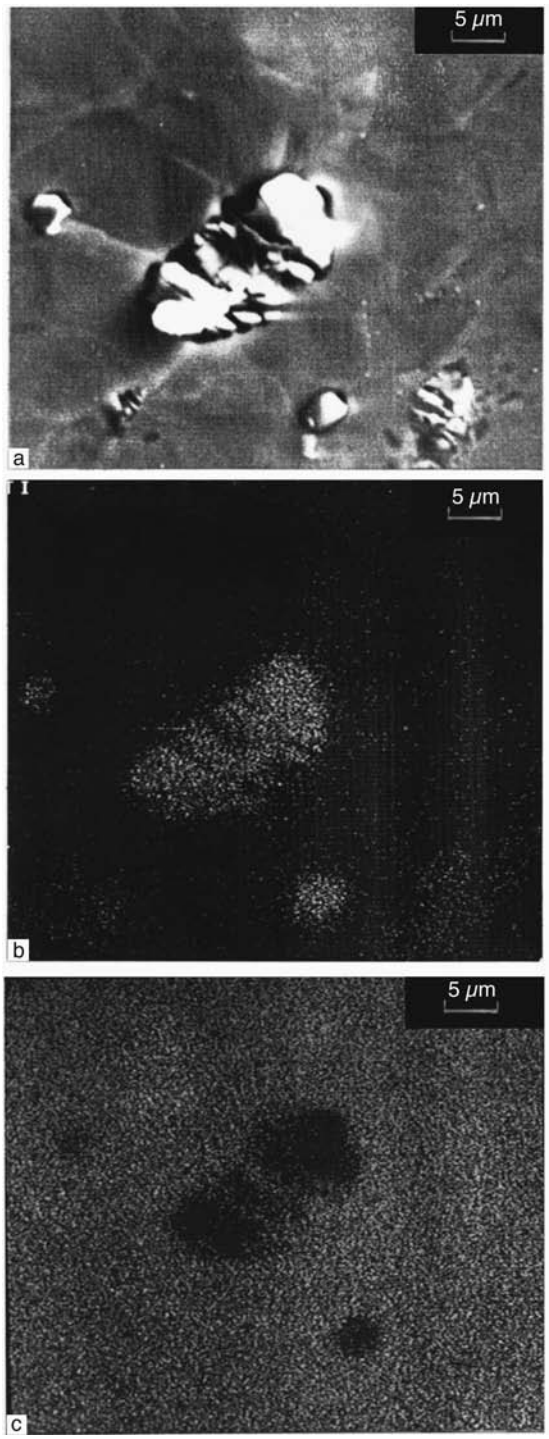


FIG. 5. Photomicrographs of a sample surface with several large clusters obtained using a digital scanning electron microscope: a — in secondary electrons, b — in characteristic K_α Ti radiation, c — in characteristic K_α Cr radiation.

tron microscope in secondary electrons. An analysis of several such photographs showed that the mean grain diameter is 10 μm. It was established in Ref. 6 for 12R72 Fe–Cr–Ni steel with a titanium content of 0.52 wt. % using electron-microscopic methods that the maximum titanium concentration near the grain boundaries reached 7.7 at. % after standard heat treatment. The width of the enrichment zone was 2000 Å, and the region occupied by grain-boundary precipitates measured 500 Å. Thus, the probability W_1 of ultracold neutrons striking an enrichment region is fairly high and can amount to 0.04. It is also known that holding austenitic stainless steel at 550–650 °C can lead to the precipitation of chromium carbides on the grain boundaries.[7] The increase in the concentration of titanium and chromium in grain-boundary regions adjacent to the surface can be promoted by the accelerated diffusion of carbon along grain boundaries away from the sample surface (Table II). Since chromium has a comparatively small scattering length ($b_{\mathrm{Cr}}=0.35\times10^{-12}$ cm, $b_{\mathrm{Ni}}=1.03\times10^{-12}$ cm, and $b_{\mathrm{Fe}}=0.95\times10^{-12}$ cm), a rise in the relative concentration of Ti and Cr in the intergrain regions can lower the value of the cutoff energy in comparison to the cutoff energy of the matrix. The same effect can be produced by an elevated hydrogen concentration in the intergrain regions, whose presence on the surface follows from the measurement results.

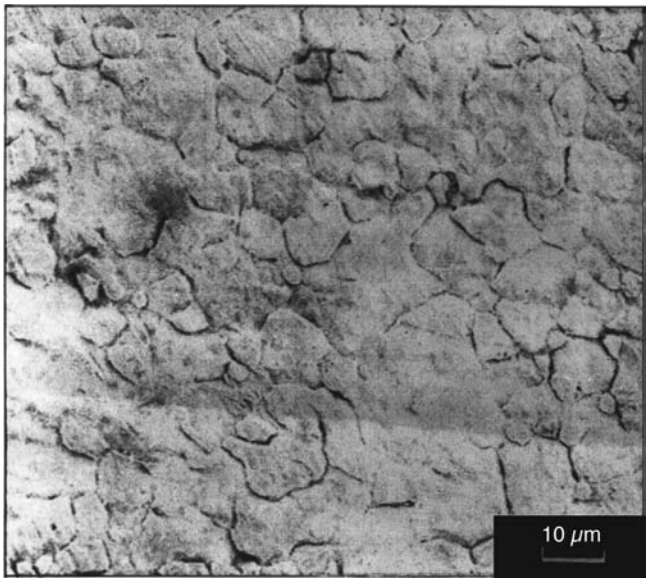


FIG. 6. Typical photomicrograph of the surface relief of a sample obtained using a DSM-960 electron microscope in secondary electrons. The mean grain diameter is $\sim 10\mu$m.

If some of the ultracold neutrons can pass into the intergrain regions and be captured, the excess values of the partial capture probabilities $\Delta\eta_c^i$ will be proportional to $Nc_i\sigma_c^i$. The experimentally observed ratio between the values of the excess capture of ultracold neutrons by Fe, Cr, and Ni nuclei requires that their atomic concentrations be in the ratio 2.9:2.0:1.0, which gives $\overline{b}=\Sigma_{i=1}^{q}c_ib_i=0.75\times10^{-12}$ cm. The corresponding effective interaction potential of ultracold neutrons is lower than the potential of the matrix; therefore, the observed enhancement of capture can take place. The cutoff energy of such a medium is determined by the number density of nuclei, which is no higher than the number density of nuclei in the grains, and therefore E_{lim} is at most 1.5×10^{-7} eV. Thus, the intergrain regions can also be regions of reduced effective potential, which extend into the sample to a depth on the order of 10 μm or more.

The estimate of E_{lim} obtained is above the edge of the ultracold-neutron spectrum (1×10^{-7} eV), but the spectrum contains an admixture of ultracold neutrons with a higher

energy, as is indicated by control measurements with a copper sample. If the relative fraction of ultracold neutrons with $E > 1.5 \times 10^{-7}$ eV in the spectrum is 10%, the probability of their reflection from a grain boundary equals 0.9, and the probability of their absorption in the intergrain space is close to 1, then the excess probability of losses as a result of capture by Fe, Ni, and Cr nuclei can be of order 4×10^{-4}, which is sufficient for explaining the observed enhancement.

Thus, the observed selective enhancement of capture does not contradict the existing theory of the interaction of ultracold neutrons with a multielement medium. Its occurrence indicates only that detailed information regarding the nuclear composition of the surface and its structure is needed to analyze experimental data and compare them with theory. The presence of localized regions having an increased concentration of elements with a negative or small positive scattering length and dimensions greater than λ in the medium can result in deformation of the effective interaction potential of ultracold neutrons.

We thank H. Just for assisting in the performance of the experiment, A. K. Churakov for providing the program for treating the gamma spectra, and S. A. Teterin for performing the atomic analysis of the samples.

This work was performed with the support of the International Association for the promotion of cooperation with scientists from the New Independent States of the former Soviet Union (INTAS) (Grant 93-298), the Russian Fund for Fundamental Research (Grant 93-02-927), and the Foundation for Supporting Research by Young People at the Kurchatov Institute Russian Scientific Center (Grant 20).

*)E-mail: bond@foton.polyn.kiae.su

[1] S. S. Arzumanov, L. N. Bondarenko, E. I. Korobkina *et al.*, JETP Lett. **65**, 1 (1997).
[2] S. V. Zhukov, V. L. Kuznetsov, V. I. Morozov *et al.*, JETP Lett. **57**, 464 (1993).
[3] Yu. M. Kagan, JETP Lett. **11**, 147 (1970).
[4] V. K. Ignatovich, *The Physics of Ultracold Neutrons*, Clarendon Press, Oxford; Oxford University Press, Oxford–New York (1990) [Russ. original, Nauka, Moscow (1986), p. 132].
[5] A. M. Parshin, *Structure, Strength, and Radiation Damage of Corrosion-Resistant Steels*, American Nuclear Society, La Grange Park, Illinois (1996) [Russ. original, Metallurgiya, Chelyabinsk (1988), p. 656].
[6] R. G. Faulkner, J. Mater. Sci. **16**, 373 (1981).
[7] Yu. M. Lakhtin and V. P. Leont'eva, *Materials Science* [in Russian], Mashinostroenie, Moscow (1990).

Translated by P. Shelnitz

Search for low-energy upscattering of ultracold neutrons from a beryllium surface

Al. Yu. Muzychka and Yu. N. Pokotilovski*)

Joint Institute for Nuclear Research, 141980 Dubno, Moscow Region, Russia

P. Geltenbort

Institut Laue-Langevin, Grenoble, France
(Submitted 13 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 141–148 (January 1999)

We present results of a search for anomalous low-energy upscattering of ultracold neutrons from a beryllium surface. This upscattering is considered one for the possible reasons for UCN ''disappearance'' from very cold beryllium bottles, as observed in experiments. The indium foil activation method was used to measure a very low intensity flux of upscattered UCN. The (15–300) m/s velocity range of upscattered UCN is ruled out by these measurements at a confidence level of 90%. © *1999 American Institute of Physics.* [S1063-7761(99)01301-3]

1. INTRODUCTION

There exists the well-known and long-standing puzzle of ultracold neutron (UCN) storage times in closed volumes, or equivalently, of anomalous losses of UCN upon reflection from the inner surfaces of UCN traps. The most surprisingly large discrepancy between experimental and predicted loss coefficients is observed in the most promising materials for long UCN storage times: cold beryllium[1,2] and solid oxygen.[3] The anomaly consists of an almost temperature-independent (in the temperature range 10–300 K) wall loss coefficient ($\sim 3\cdot 10^{-5}$), corresponding to an extrapolated inelastic thermal neutron cross section $\sigma^* \sim 0.9$b. This experimental figure for Be is two orders of magnitude higher than the theoretical one, the latter being completely determined at low temperatures by neutron capture in Be (0.008b). The experiment/theory ratio for a very cold oxygen surface reaches three orders of magnitude.[3] Approximate universality of the loss coefficient for beryllium and oxygen, and the temperature independence of the Be figures, forces one to suspect a universal reason for this anomaly.

A series of experiments to find the channel by which UCN leave the trap are described in Ref. 2. None of the suspected reasons has been confirmed: surface contamination by elements with large absorption cross-sections; penetration of UCN through possible micro-cracks in the surface layers of Be, hypothetical milliheating of UCN due to collisions with a low-frequency vibrating surface; upscattering of UCN due to thermal vibrations of the wall nuclei. The latter item deserves special and more careful consideration.

According to the description of the experiment in Ref. 4 (a subsequent conclusion[2] about the absence of UCN upscattering from the beryllium surface at liquid nitrogen temperature is based entirely on that experiment), upscattered neutrons passed through 1.5 mm of copper, 1.1 mm of stainless steel, and 2 mm of Al prior to entering the neutron detector. For an isotropic distribution of upscattered neutrons, this means that the detection efficiency of upscattered neutrons with energies of 0.5 meV was less than 0.2, and decreased at lower neutron energies. The reported[2] upscattering cross-section of UCN from a beryllium surface at a surface temperature of 80 K was 0.14b with an uncertainty of 30%, so it is quite possible that UCN upscattering takes place to the energy range below 1 meV. This hypothesis is consistent with the observed temperature independence[2] of anomalous losses of UCN if the vibrations causing this upscattering are not thermal in nature. The frequency of these vibrations (possibly surface waves) is in the range 10^8-10^{12} Hz. From a purely experimental point of view (without going into any hypotheses about the reasons for UCN anomalous losses), this low-energy upscattering channel is almost the only one that has not yet been investigated with conclusive results.

Additional qualitative considerations favoring possible high-frequency surface sound wave UCN upscattering comes from the rough coincidence of the typical surface roughness correlation length $T \sim 300-500$ Å, UCN wavelength λ, which is close to these values, and the possible surface sound wavelength. This coincidence may, in principle, increase the UCN upscattering probability due to some kind of «resonance». If the surface sound velocity $c \simeq 10^5$ cm/s then the upscattered neutron energy $E = hc/\lambda \simeq 10^{-4}$ eV, which is just outside of the investigated energy range.[2]

Recently, results have been published[5,6] describing searches for UCN upscattering from a beryllium foil surface in which gas counters were used for upscattered neutron detection. According to Ref. 5, the total (to the energy range $(10^{-7}-10^{-2})$eV) reduced upscattering cross-section was (0 ± 0.2)b at liquid nitrogen temperature, and (0 ± 0.3)b at room temperature. The first result is consistent with the early data,[2] but the second is in serious disagreement with the previous results. The authors of Ref. 5 point out that these figures are not final, and that ''these values were obtained after the subtraction of the large background from the trap walls and separating foil and it is necessary to increase the accuracy of measurements to establish these values.''

In contrast, Refs. 6 and 7 give quite different figures for the upscattering loss factor upon UCN reflection from a Be surface: $(1.47\pm 0.15)\cdot 10^{-4}$ and $(1.39\pm 0.18)\cdot 10^{-4}$ for two

different beryllium samples at room temperature, and $(2.6 \pm 0.3) \cdot 10^{-5}$ and $(1.7 \pm 0.2) \cdot 10^{-5}$ at liquid nitrogen temperature for the same two beryllium samples.

According to the usual formalism with which UCN interaction with a surface is considered, the UCN reflection probability as a function of the normal component of the neutron momentum $p_\perp$ is

$$w = 2\eta \cdot x/\sqrt{1-x^2}, \quad x = p_\perp / p_{\text{bound}}.$$

The loss factor is

$$\eta = \text{Im}\, b/\text{Re}\, b, \quad \text{Im}\, b = \sigma_{\text{inel}}/2\lambda,$$

where b is the coherent scattering length of the wall material, $p_{\text{bound}} = 2\hbar(\pi N b)^{1/2}$ is the boundary momentum of the UCN, characterizing the reflecting wall, and σ_{inel} is the total cross-section of all inelastic UCN interaction processes with the wall surface.

Using the above relations, it is easy to obtain from Refs. 6 and 7 the thermal neutron energy inelastic scattering cross sections for the two beryllium samples, $\sigma_{\text{inel}} = 2\lambda\eta\,\text{Re}\, b$: (4.1 ± 0.4)b and (3.9 ± 0.5)b for the room-temperature samples, and (0.73 ± 0.08)b and (0.48 ± 0.06)b for the liquid nitrogen temperature samples.

It is seen from the above that the results in Refs. 2, 5, 6, and 7 are mutually inconsistent.

For the sake of completeness, it is necessary also to mention measurements[8] of the UCN upscattering probability to the thermal energy range from the surface of a room-temperature beryllium foil after different high-temperature procedures: $18.5 \cdot 10^{-4}$ before any heating of the sample, $3.9 \cdot 10^{-4}$ after heating at the 450 °C, $4.1 \cdot 10^{-4}$ after heating at 700 °C with subsequent 5-min exposure to atmospheric air, and $2.2 \cdot 10^{-4}$ after heating at 700 °C in vacuum.

Taking into account that according to Ref. 8 the mean velocity of the stored UCN was measured to be $\simeq$3.2 m/s, it is possible to transform these figures into the room-temperature neutron upscattering cross-sections if we assume that the imaginary part of the wall potential can be attributed to this cross-section according Eq. (1). This procedure yields for the reduced upscattering cross-sections for the cases mentioned above 76, 16, 17, and 9b respectively, per atom of the wall. These and the previous[2,5,6,7] figures for the reduced upscattering cross-section lead to unrealistically high concentrations of hydrogen in the surface layer of wall if we take into account that the reduced UCN upscattering cross section per hydrogen atom at room temperature is $\simeq$(7–8)b. It means that the usual procedure that relates the upscattering cross-section to the imaginary part of the wall potential by means of Eq. (1) is incorrect, and can only serve to compare the results of different experiments.

The authors of Ref. 9, which used neutron-induced radiation analysis for the investigation of UCN interaction with beryllium samples, did not find any neutron-capture gamma-radiation from beryllium nuclei. From this fact they concluded that anomalous upscattering with reduced cross-section 0.9b (Gatchina anomaly) can not yield upscattered neutrons with velocity less than ~70 m/s in the final state.

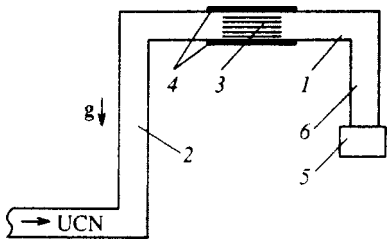


FIG. 1. The layout of an experiment to search for the low-energy upscattering of UCN from a Be surface. *1*—Vacuum stainless steel chamber $\phi 60 \times 0.5$ mm. *2*—Vertical UCN guide $\phi 60 \times 0.5$ mm, height 12 cm. *3*—Rouleau of aluminum foils with beryllium deposition. *4*—Cylindrical stack of In foils. *5*—Detector of UCN (^{3}He proportional counter). *6*—Vertical UCN guide $\phi 60 \times 0.5$ mm, height 60 cm. *7*—Shielding (borated polyethylene).

From this follows a lower bound on the velocity range of the upscattered neutrons: $v > 70$ m/s (with no indicated confidence level).

2. EXPERIMENTAL METHOD

The measurements were performed at the test channel of the UCN turbine source at the Institut Laue–Langevin.[10] The irradiation scheme is shown in Fig. 1.

Ultracold neutrons enter the stainless steel cylindrical chamber *1* ($\phi = 60$ mm, wall thickness 0.5 mm) through the vertical stainless steel neutron guide *2* (height 120 cm), and rebound from the surface of the specimen *3*, made of aluminum foil covered with a beryllium layer (film thickness $(2-3) \cdot 10^3$ Å). The specimens had the form of a corrugated ribbon rolled into a spiral with an overall area (two sides) of ~0.5 m^2 or ~0.25 m^2. The upscattered neutrons leaving the trap penetrate the cylindrical stack of indium foils surrounding the tube and activate them with an activation cross-section that conforms an inverse velocity law. The indium foils were 5–50 μm thick and were manufactured by means of electrolytic deposition on the surface of 10 μm copper foil. The homogeneity of the In thickness was thoroughly verified by cutting the test foils into numerous small specimens and weighing them, and was found to be better than 5%. The density of the UCN flux in the trap was calibrated by means of the activation measurement of the flux of upscattered UCN from small polyethylene samples located at the center of the irradiation chamber, and monitored with a ^{3}He proportional counter *5* located after the UCN trap and connected to the trap by a vertical neutron guide *6* through a small (0.5 cm^2) hole. The UCN flux at the beryllium sample measured in this way was $\simeq$40 cm^{-2} s^{-1}.

The efficiency of the UCN detector in the geometry of the experiment was simulated by the Monte Carlo method under various assumptions about the probability of diffuse UCN reflection from the neutron guide walls between the small hole and the membrane of the detector. The results of this simulation show almost constant (>90%) efficiency in the UCN energy interval of interest, 0–150 neV.

With the known efficiency of the UCN detector it is possible to determine from these measurements the effective areas of the polyethylene samples and to compare them with the actual ones. Figures 2 and 3 show the results of the

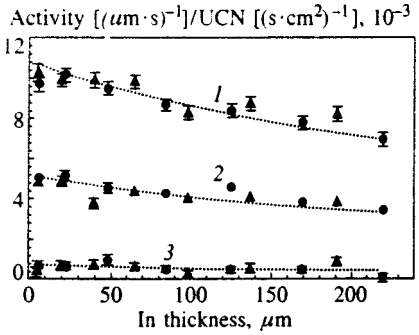


FIG. 2. Activation of the stack of indium foils as a result of UCN upscattering from polyethylene scatterers with surface area: line *1*—8.8 cm^2; line *2*—5.3 cm^2; line *3*—the empty stainless steel chamber.

activation measurements of indium stacks for two polyethylene samples with surface areas of 5.3 and 8.8 cm^2.

The response function of the activation of the stack of indium foils was calculated by the Monte Carlo method. The upscattered neutrons were assumed to emerge from the Be scatterer isotropically, having their starting points on the surface of the Be spiral. Reflection and absorption of upscattered neutrons along their trajectories were rigorously taken into account.

Detailed results of these simulations will be published elsewhere.[11]

Activity of the In foils was measured with a high efficiency (~70%) 4π scintillation β-counter with active (4π plastic anticoincidence counter) and passive (lead) shielding. The area of the In foils whose activity was measured simultaneously was ~200 cm^2. The counter background was about 1.05 s^{-1} in these measurements. The counter efficiency was carefully measured for different thicknesses of irradiated In and Cu foils. A description of the counter and results of the calibration will be published in Ref. 11.

This method of measuring slow neutron spectra via activation of a stack of In foils was calibrated by irradiating the stack with a beam of monochromatic thermal neutrons or a precisely measured (time-of-flight method) quasi-Maxwellian spectrum of cold and thermal neutrons. The measured and calculated distributions of foil activity along the stacks were in good agreement.

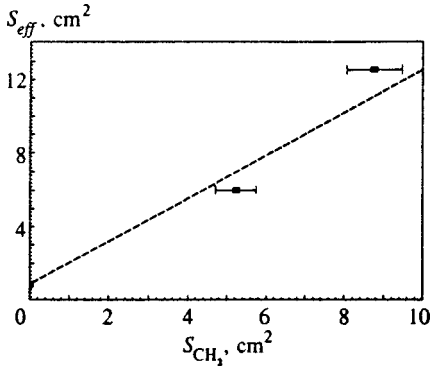


FIG. 3. Measured effective polyethylene sample area obtained from indium foil activation measurements as a function of the polyethylene scatterer surface area.

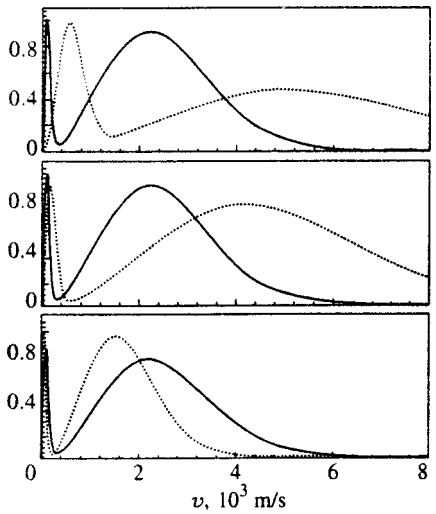


FIG. 4. Results of computer indium foil stack ''activation experiments'' (solid curves) and subsequent restoration of the incoming upscattered neutrons spectra (in relative units) under the assumption that the spectrum consists of two Maxwellian flux components: one with $v_{th}=2.2\cdot10^5$ cm/s and the other with $v_0=10^4$ cm/s, the latter had a weight 1/20 of the former (dotted curves). The results of computer ''activation'' of indium foil stack before the restoration procedure were statistically Gaussian distributed with a standard deviation of 5% for each foil.

In the absence of low-energy anomalous upscattering, UCN acquire energy from the thermal vibrations of the beryllium lattice, and with higher probability, from vibrations of surface contaminants (mostly hydrogenous). The spectrum of the upscattered neutrons has a «thermal» character, but it is not known. Information about the spectrum of possible ''anomalous'' upscattering is even more obscure. Therefore, In foil activity measured as a function of position in the stack (thickness coordinate) was approximated under the very general assumption that the spectrum of upscattered neutrons consists of two Maxwellian flux components, one with $v_{th}=2.2\cdot10^5$ cm/s (''normal'' upscattering from room temperature Be) and the other with low v_0, the latter being chosen in the range 10–300 m/s (anomalous upscattering).

The overall thickness of the indium stacks did not exceed 250 μm in our measurements, ($n\sigma\sim0.3$ for the isotropic thermal neutron flux), so the accuracy of the upscattered thermal neutron spectrum is not high. But it was demonstrated by rigorous Monte Carlo simulation[11] that it is possible not only to distinguish the low energy component of the upscattering from the high thermal background, but also to carry out rough spectrometry of this low-energy part of the spectrum.

Figure 4 demonstrates some results of the computer indium stack «activation experiment» and restoration of the incoming spectrum of upscattered UCN under the assumption that the spectrum consists of two Maxwellian flux components, one with $v_0=2.2\cdot10^5$ cm/s and the second one with $v_0=10^4$ cm/s, the latter had a weight 1/20 of the former.

The results of computer ''activation'' of the indium foil stack before the restoration procedure were statistically Gaussian-distributed with a standard deviation of 5% for each foil. It can be seen that the method is able to reconstruct with high confidence the small low energy-admixture to the

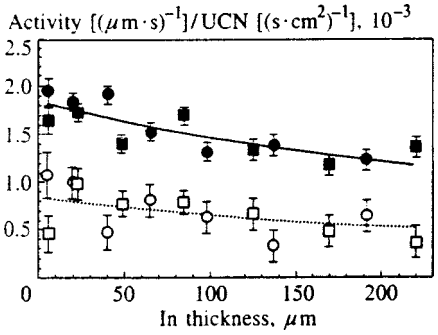


FIG. 5. Measured In foil activation points as a function of indium thickness coordinate for two different beryllium samples areas: full poins points 0.5 m^2, empty points 0.2 m^2.

intensive thermal background, but it is not dependable in extracting the thermal component of the spectrum from the indium activation data.

3. RESULTS AND DISCUSSION

Figure 5 shows the measured activity of irradiated In foils as a function of the thickness coordinate for two beryllium samples with different areas.

In our measurements, we used the compact surface sample with the enhanced area, so that the mean gap between the adjacent turns of our helical ribbon sample with area ~ 0.5 m^2 was about 1 mm. In UCN upscattering how effectively is the full surface area of the sample used with such narrow channels for UCN diffusion between the adjacent turns? Additional activation measurements were carried out with a sample of area ~ 0.2 m^2 with ~ 2.5 times larger gaps between the adjacent turns. Figure 6, representing the measured indium stack activity as a function of sample surface area, shows good proportionality between area and activity, attesting to evidence of the uniform and effective UCN upscattering over the full sample area.

The total measured flux of upscattered UCN from the beryllium sample with area 0.5 m^2 was $\simeq 50$ s^{-1}.

As mentioned above, this method has low reproducibility in extracting the spectrum of the thermal component of upscattered UCN. Therefore, the experimental data were processed under a different reasonable assumption about the temperature of the thermal component. Figure 7 represents the 90% exclusion contours for the cross-sections σ^*_{anom} and

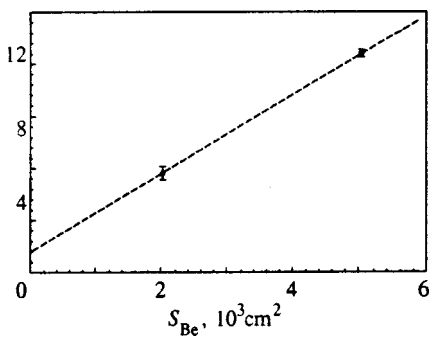


FIG. 6. Flux intensity of upscattered UCN normalized per primary UCN flux (relative units) from the beryllium sample as a function of sample area.

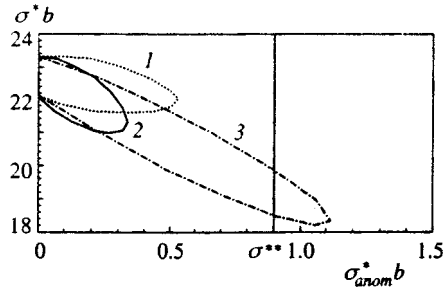


FIG. 7. The 90% exclusion contours for the cross sections σ^*_{anom} and σ^*_{th} deduced from In foil activation measurement of upscattered UCN flux from normal-temperature Be surface, under the assumption that the thermal component of upscattering represents a Maxwellian neutron flux with room temperature. The contours are presented for three different characteristic velocities v_0 of anomalously (low-energy) upscattered UCN, assumed to have a Maxwellian flux form: line *1*—$v_0 = 15$ m/s; line *2*—$v_0 = 50$ m/s; line *3*—$v_0 = 200$ m/s.

σ^*_{th} deduced from the In foil activation measurement of upscattered UCN fluxes from a normal temperature Be surface, assuming that the thermal component of upscattering represents a Maxwellian neutron flux with room temperature.

The contours are presented for three different characteristic velocities v_0 of anomalously (low-energy) upscattered UCN, assumed to have a Maxwellian flux form: $v_0 = 15$ m/s; $v_0 = 50$ m/s; $v_0 = 200$ m/s.

As is seen from Fig. 7 the room-temperature-adjusted UCN upscattering cross-section to the final thermal energy range is very high for non-outgassed beryllium ($\sim$22b), which is consistent with the result of Ref. 5 and some of the results of Ref. 8. We attribute so large an upscattering cross-section partly to the presence in the incoming UCN spectrum of neutrons with energies higher than the boundary energy of beryllium, but mostly to upscattering from the aluminum cuts of sample ribbons not covered with a beryllium layer. In both cases, this upscattering takes place not at the sample surface but in the bulk of the aluminum. This enhanced thermal upscattering was not very significant in our search for the low-energy anomalous component in the upscattered neutron spectrum, but it increased the thermal background.

Figure 8 shows the 90% confidence restriction curves for

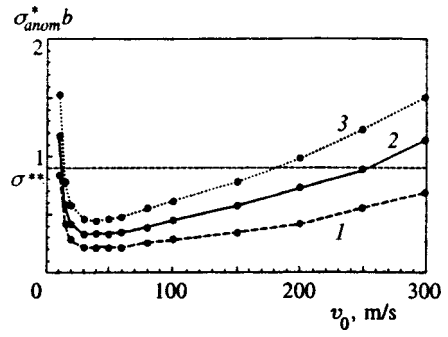


FIG. 8. The 90% confidence restriction curves for the reduced cross-section as a function of the characteristic velocity v_0 of the Maxwellian flux of anomalous low-energy UCN upscattering from room-temperature beryllium sample under different assumptions about the characteristic velocity v_{th} of the thermal flux: line *1*—$v_{\text{th}} = 1600$ m/s, line *2*—$v_{\text{th}} = 2200$ m/s; line *3*—$v_{\text{th}} = 2800$ m/s.

the reduced cross-section as a function of the characteristic velocity v_0 of the Maxwellian flux of anomalous low energy UCN upscattering from a room-temperature beryllium sample under various assumptions about the characteristic velocity v_{th} of the thermal flux.

In addition to (1) the following formulas were used in the data processing. The loss probability of UCN with the velocity v, averaged over an isotropic angular distribution upon reflection from the surface with boundary velocity v_b: $\overline{\mu}_{loss}=2\eta(\arcsin(y)-y\sqrt{1-y^2})/y^2$, where $y=v/v_b$, $\overline{\mu}=\int_0^{v_{lim}}\overline{\mu}f(v)dv$ is the UCN loss coefficient averaged over the normalized UCN flux spectrum $f(v)=4v^3/v_{lim}^4$, which is the low-energy tail of the Maxwellian spectrum. In our case, $v_{lim}=3.9$ m/s.

As may be seen from Fig. 8 we were not able to completely rule out in this experiment low-energy UCN upscattering over the entire energy range of interest, $0.1-10^3$ μeV but a significant part of this energy range, $\simeq(1-200)$ μeV is ruled out by our measurements.

Preliminary In foil activation measurements with upscattered UCN and partial calibrations of the method in cold and thermal neutron beams were performed at the reactor of the St. Petersburg Nuclear Physics Institute (SPNPI) at Gatchina. We are grateful to Drs. A. P. Serebrov, A. G. Kharitonov, V. V. Nesvizhevsky, and R. R. Taldaev for their kind permission to use the UCN channel of SPNPI and for their very valuable help. We thank them also for placing the Be sample at our disposal. Efficiency measurements of the beta-counter were performed via irradiation of indium and copper foils in the thermal neutron beam of the IBR-2 reactor of FLNP JINR and the microtron neutron source of FLNR JINR. The authors are grateful to Drs. V. V. Nitz and A. G. Belov for their permission to use these installations for our calibrations. We also express our appreciation to the ILL reactor staff.

*)E-mail: POKOT@nf.jinr.ru

[1] P. Ageron, W. Mampe, and A. I. Kilvington, Z. Phys. B **59**, 261 (1985).
[2] V. P. Alfimenkov, V. V. Nesvizhevski, A. P. Serebrov *et al.*, LNPI Preprint No. 1729, Gatchina, Russia (1991); JETP Lett. **55**, 84 (1992).
[3] V. P. Alfimenkov, V. V. Nesvizhevski, A. P. Serebrov *et al.*, LNPI Preprint No. 1629 (1991), PNPI Preprint No. 1756 (1992), Gatchina, Russia.
[4] V. K. Ignatovich, Kim Zun Bok, V. I. Lushchikov *et al.*, JINR Comm. P3-82-811, Dubna (1982).
[5] A. V. Strelkov, G. N. Nekhaev, V. N. Shvetsov *et al.*, in *Neutron Spectroscopy, Nuclear Structure, and Related Topics*, (Proc. IV Int. Seminar on Interaction of Neutrons with Nuclei, Dubna, April 27–30, 1996), p. 299.
[6] A. P. Serebrov, in Neutron Spectroscopy, Nuclear Structure, and Related Topics (Proc. V Int. Seminar on Interaction of Neutrons with Nuclei, Dubna, May 14–17, 1997), p. 67.
[7] V. E. Varlamov, P. Geltenbort, V. V. Nesvizhevsky *et al.*, Preprint PNPI 2216 Gatchina (1998).
[8] L. Bondarenko, S. Chernyavsky, A. Fomin *et al.*, Physica B **234–236**, 1189 (1997).
[9] S. S. Arzumanov, L. N. Bondarenko, E. I. Korobkina *et al.*, Preprint of Kurcharov Institute IAE-6010/2, Moscow, 1996; S. S. Arzumanov, L. N. Bondarenko, S. M. Chernyavsky *et al.*, Proc. of the Int. Seminar on Interaction of Neutrons with Nuclei ISINN-5: Neutron Spectroscopy, Nuclear Structure, and Related Topics, Dubna, 14–17 May, 1997), p. 91.
[10] A. Steyerl, H. Nagel, F.-X. Schreiber *et al.*, Phys. Lett. A **116**, 347 (1986).
[11] Al. Yu. Muzychka and Yu. N. Pokotilovski, submitted to Nucl. Instr. Meth.

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

Electron–lattice kinetics of metals heated by ultrashort laser pulses

L. A. Falkovsky and E. G. Mishchenko

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 14 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 149–157 (January 1999)

We propose a kinetic model of transient nonequilibrium phenomena in metals exposed to ultrashort laser pulses when heated electrons affect the lattice through direct electron-phonon interaction. This model describes the destruction of a metal under intense laser pumping. We derive the system of equations for the metal, which consists of hot electrons and a cold lattice. Hot electrons are described with the help of the Boltzmann equation and equation of thermoconductivity. We use the equations of motion for lattice displacements with the electron force included. The lattice deformation is estimated immediately after the laser pulse up to the time of electron temperature relaxation. An estimate shows that the ablation regime can be achieved. © *1999 American Institute of Physics.* [S1063-7761(99)01401-8]

1. INTRODUCTION

The first theoretical prediction of transient laser-induced nonequilibrium electron temperature phenomena in metals was made more than twenty years ago.[1] It was shown that an ultrashort laser pulse ($\sim 10^{-13}$–10^{-12} s) produces a nonequilibrium state of the electron gas near a metal surface. However, experimental picosecond ($\sim 10^{-12}$ s) laser studies of thermally assisted multiphonon photoemission were unable to measure, and even failed to observe this nonequilibrium electron state.[2] This failure had a simple explanation in terms of the theory of electron–lattice thermal relaxation,[3] which yields a relaxation time $\tau_{e-l} \sim 10^{-12}$ s. It was necessary to use power pulses shorter than τ_{e-l}. Such measurements with subpicosecond ($\sim 10^{-13}$ s) pulses revealed a transient nonequilibrium regime in transmittivity and IR reflection,[4–8] giant electron emission[9–11] and the emission of light.[12–14]

We briefly summarize the physical process. The ultrashort laser pulse ($\Delta t \sim 10^{-14}$–10^{-13} s) absorbed in a metal raises the electron temperature T_e considerably higher than the lattice temperature because of the difference in their specific heats ($c_e \ll c_l$). Subsequent electron cooling results mainly from two processes, namely electron-lattice thermal relaxation and electron thermoconductivity. These are usually modeled with a set of coupled thermoconductivity equations for the electron and lattice components. These equations are nonlinear and can generally be solved numerically, yielding the electron temperature relaxation. The solution also shows that the subsequent ablation regime can be achieved, which involves the «cold» destruction of a metal into the parts consisting of different phases. ''Hot'' destruction, namely melting, can also be studied with the help of this solution.[15] However, such an approach has several shortcomings. First, the question remains as to whether the equations of thermoconductivity are still hold at such high frequencies ($\sim 1/\Delta t$). Second and more importantly, these equations can only describe the temperature dynamics of a metal but not electron transport, lattice deformation, thermionic emission, etc. It is evident that a strict kinetic approach is needed to describe the various transport phenomena properly and derive thoroughly the equation of thermoconductivity.[16]

In this paper we present a theory of transient nonequilibrium phenomena in metals subject to ultrashort laser pulses. Our theory is based on the Boltzmann equation for the nonequilibrium electronic partition function. We focus mainly on times shorter than the electron-lattice relaxation time τ_{e-l}. Electrons therefore affect the lattice via direct electron-phonon interactions. To consider lattice deformations, we use the equations of the so-called dynamical theory of elasticity. Lattice deformation is due to the nonequilibrium electron state and results from the effective ''driving'' force (proportional to ∇T_e^2) on the lattice. This force also governs the renormalization (depending on T_e) of the lattice constants (sound velocity and optical phonon gap). We show that the driving force leads to large lattice deformations, and can destroy the crystal. These results are in agreement with measurements of time-resolved *X*-ray diffraction synchronized with laser pumping.[17] A nonstationary increase in lattice parameters of Au(111) and Pt(111) single crystals was detected. Measurements of the shift and intensity variation of Bragg peaks, as well as the Debye-Waller factor, enables one to separate the effects of lattice deformation and heating. The transformation of elastic into plastic deformation was also observed.

The plan of the paper is as follows. In Sec. 2 we present the kinetic theory of the process under study: the Boltzmann equation for an electron gas and the elastic equation for the lattice are derived, along with the equation for thermoconductivity. In Sec. 3, the solutions of the proposed equations are found for the times of interest. The lattice deformation is calculated. In Sec. 4 the solutions are analyzed. The lattice deformation is estimated analytically in various limiting cases. The possibility of crystal destruction under laser pumping in discussed.

2. THEORETICAL FRAMEWORK

Let us briefly recapitulate the main equations of our problem. For the lattice deformation we use the so-called equation of dynamical theory of elasticity[18,19]

$$\rho\frac{\partial^2 u_i}{\partial t^2}-\lambda_{ijlm}\frac{\partial^2 u_l}{\partial x_j\partial x_m}=G_i, \tag{1}$$

where ρ is the lattice density, λ_{ijlm} is the tensor of elastic constants, and the driving force describes the effect of free carriers on the lattice,

$$G_i=\frac{\partial}{\partial x_j}\int\frac{2d^3p}{(2\pi)^3}\lambda_{ij}(\mathbf{p})f_p(\mathbf{r},t). \tag{2}$$

The deformation potential $\lambda_{ij}(\mathbf{p})$ yields the change in the local electron spectrum,

$$\delta\varepsilon(\mathbf{p},\mathbf{r},t)=\lambda_{ij}(\mathbf{p})u_{ij}(\mathbf{r},t).$$

To find the electron distribution function $f_p(\mathbf{r},t)$, we use the Boltzmann equation with the electron–phonon collision integral

$$\begin{aligned}\mathrm{St}\,f_p=&\sum_n\int\frac{d^3k}{(2\pi)^3}w_{pk}^{(n)}\delta(\varepsilon_{p+k}-\varepsilon_p+\omega_k^{(n)})\\&\times[(1-f_p)f_{p'}N_k^{(n)}-f_p(1-f_{p'})(1+N_k^{(n)})]\\&+\sum_n\int\frac{d^3k}{(2\pi)^3}w_{pk}^{(n)}\delta(\varepsilon_{p+k}-\varepsilon_p-\omega_k^{(n)})\\&\times[(1-f_p)f_{p'}(1+N_k^{(n)})-f_p(1-f_{p'})N_k^{(n)}],\end{aligned} \tag{3}$$

with the probability of a scattering process involving a phonon of the nth branch,

$$w_{pk}^{(n)}=\frac{\pi}{\rho\omega_k^{(n)}}|e_i^{(n)}\lambda_{ij}(\mathbf{p})k_j|^2,$$

where $e_i^{(n)}$ and $\omega_k^{(n)}$ are the polarization and spectrum of phonons of the nth branch, respectively.

Since the phonon-phonon relaxation time is large ($\sim10^{-11}$ s) compared with the times of interest, the phonon distribution function $N_k^{(n)}$ takes its equilibrium value at the lattice temperature T_l,

$$N_k^{(n)}(T_l)=\frac{1}{\exp(\omega^{(n)}/T_l)-1}.$$

The electron–electron relaxation time due to scattering on phonons $\tau\sim T_l^{-1}\sim10^{-14}$ s (see below) is much less than the characteristic time of laser pumping. Therefore the electron gas is nearly in thermal equilibrium at the local temperature $T_e(\mathbf{r},t)$. We seek a solution of the Boltzmann equation in the form

$$f_p=f_0\left(\frac{\varepsilon_p-\mu}{T_e}\right)+\chi_p\frac{\partial f_0}{\partial\varepsilon}, \tag{4}$$

where f_0 is the local equilibrium Fermi–Dirac partition function and χ_p is the nonequilibrium part. We obtain for the collision integral (3)

$$\mathrm{St}\,f_p=\mathrm{St}\,f_0-\tau^{-1}\left(\chi_p-\frac{\langle\chi_p\rangle}{\langle1\rangle}\right)\frac{\partial f_0}{\partial\varepsilon}, \tag{5}$$

where the scattering rate

$$\tau^{-1}=T_l\sum_n\left\langle\frac{w_{pk}^{(n)}}{\omega_k^{(n)}}\right\rangle\sim\pi g^2T_l.$$

The latter estimate is valid when the ion temperature T_l is considerably higher than the Debye temperature; the dimensionless electron–phonon coupling constant $g\sim\lambda/\varepsilon_F\sim1$. The brackets denote integration over the Fermi surface

$$\langle\ldots\rangle=\int\frac{2dS_F}{v(2\pi)^3}(\ldots).$$

The first term in (5) comes from the contribution of the local equilibrium partition function:

$$\begin{aligned}\mathrm{St}\,f_0=&\sum_n\int\frac{d^3p'}{(2\pi)^3}w_{pk}^{(n)}[f_0(\varepsilon_p)-f_0(\varepsilon_{p'})]\\&\times[N_k^{(n)}(T_e)-N_k^{(n)}(T_l)][\delta(\varepsilon_p-\varepsilon_{p'}-\omega_k^{(n)})\\&+\delta(\varepsilon_p-\varepsilon_{p'}+\omega_k^{(n)})].\end{aligned} \tag{6}$$

This term describes the energy flow from electrons to phonons when they are at different temperatures. This term is absent if the temperatures of the electron and lattice subsystems coincide.

The nonequilibrium part of the electron distribution function has to satisfy two conditions. The first is indeed the conservation law of the number of carriers:

$$\int\frac{d^3p}{(2\pi)^3}\chi_p\frac{\partial f_0}{\partial\varepsilon}=0.$$

This expression determines the chemical potential μ and results in the renormalization of the deformation potential: $\lambda(\mathbf{p})\to\lambda(\mathbf{p})-\langle\lambda(\mathbf{p})\rangle/\langle1\rangle$.

The second condition

$$\int\frac{d^3p}{(2\pi)^3}(\varepsilon_p-\mu)\chi_p\frac{\partial f_0}{\partial\varepsilon}=0 \tag{7}$$

enables us to define the local temperature T_e (see Ref. 20), i.e., to write the equation of thermoconductivity.

Substituting Eq. (4) into the Boltzmann equation, we get

$$\begin{aligned}\frac{\partial\chi_p}{\partial t}+\mathbf{v}\frac{\partial\chi_p}{\partial\mathbf{r}}+\frac{\chi_p-\langle\chi_p\rangle/\langle1\rangle}{\tau}=&-e\mathbf{v}\mathbf{E}-\lambda_{ij}(\mathbf{p})\frac{\partial u_{ij}}{\partial t}\\&+\frac{\varepsilon_p-\mu}{T_e}\left(\frac{\partial T_e}{\partial t}+\mathbf{v}\frac{\partial T_e}{\partial\mathbf{r}}\right)+\mathrm{St}\,f_0\Big/\frac{\partial f_0}{\partial\varepsilon}.\end{aligned} \tag{8}$$

To obtain the equation for the local temperature $T_e(\mathbf{r},t)$, we multiply the Boltzmann equation (8) by $(\varepsilon_p-\mu)\partial f_0/\partial\varepsilon$ and integrate over $\mathbf{p}$. With the help of Eq. (7) we find the equation of thermoconductivity

$$c_e(T_e)\frac{\partial T_e}{\partial t}+\mathrm{div}\,\mathbf{q}=Q-\alpha(T_e-T_l), \tag{9}$$

where $c_e(T_e)=\pi^2\langle1\rangle T_e/3\equiv\beta T_e$ is the electron heat capacity and $\mathbf{q}$ is the heat flow:

$$\mathbf{q}=\int \frac{2d^3p}{(2\pi)^3}\mathbf{v}(\varepsilon_p-\mu)\chi_p\frac{\partial f_0}{\partial\varepsilon}. \tag{10}$$

Using Eq. (6), we can find the last (relaxation) term on the right-hand side of Eq. (9). For high temperatures $T_l, T_e \gg \omega_D$, the electron–lattice relaxation constant α is

$$\alpha=\sum_n \int \frac{2dS_F dS_F'}{vv'(2\pi)^6} w_{pk}^{(n)}\omega_k^{(n)}.$$

The density Q of laser energy absorbed by electrons can be taken in the form

$$Q(z,t)=I(t)(1-R)\kappa e^{-\kappa z}, \tag{11}$$

where R is the reflectivity and κ is the inverse penetration depth. The function $I(t)$ describes the pulse shape.

Equations (1) and (9) must be supplemented by the proper boundary conditions. We assume the simplest geometry: the metal occupies the half-space $z<0$. Hence, the boundary conditions for the above equations are

$$\left.\frac{\partial T_e}{\partial z}\right|_{z=0}=0, \quad \left.\frac{\partial u_z}{\partial z}\right|_{z=0}=0, \tag{12}$$

signifying that the heat flow through the surface and the normal stress tensor component vanish on the surface.

We also need the boundary conditions for the kinetic equation (8) at the metal surface. These boundary conditions depend on the type of electron reflection from the surface. We assume the specular reflection for simplicity.

3. DYNAMICS OF ELECTRON TEMPERATURE AND LATTICE DEFORMATION

The above equations are nonlinear and very complicated. However, it is possible to solve them in an important limiting case. Below we are interested in times shorter than the electron-lattice relaxation time $\tau_{e-l}\sim c_e(T_e)/\alpha$. In this case the lattice temperature can be set to the initial temperature T_0, and the last terms in Eqs. (8) and (9) can be omitted.

To solve the system (1), (8) for the half-space with the boundary condition (12), we use the even continuation of the temperature $T_e(\mathbf{r},t)$ and the partition function χ_p, and the odd continuation of $u_z(\mathbf{r},t)$, into the half-space $z<0$:

$$T_e(z<0)=T_e(-z,0), \quad u_z(z<0)=-u_z(-z,0). \tag{13}$$

For the parallel components u_x and u_y one must use the even continuation, but owing to the fact that the external heat (11) depends only on z, these components vanish. In Eq. (8) we discard $\langle\chi_p\rangle$, which represents the ''in-term'' in the collision integral. This term accounts for carrier conservation, i.e., for the isotropic channel of collision processes. Therefore it does not affect the heat flow and lattice driving force.

The solution of Eq. (8) has the form

$$\chi_p(\mathbf{r},t)=\int_{-\infty}^{t} dt' X_p(\mathbf{r}-\mathbf{v}(t-t'),t')\exp\left(-\frac{t-t'}{\tau}\right). \tag{14}$$

where X_p is the right-hand side of Eq. (8),

$$X_p=-\lambda_{ij}(\mathbf{p})\frac{\partial u_{ij}}{\partial t}+\frac{\varepsilon_p-\mu}{T_e}\left(\frac{\partial T_e}{\partial t}+\mathbf{v}\frac{\partial T_e}{\partial \mathbf{r}}\right). \tag{15}$$

Substituting the solution (14), (15) into the heat flow (10), and integrating over the energy variable according to $d^3p=d(\varepsilon_p-\mu)dS_F/v$, we obtain

$$\mathbf{q}(\mathbf{r},t)=-\frac{\pi^2}{6}\left\langle \int_{-\infty}^{t} dt' \exp\left(-\frac{t-t'}{\tau}\right)\mathbf{v}\left(\frac{\partial}{\partial t'}+\mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right) \times T_e^2(\mathbf{r}-\mathbf{v}(t-t'),t')\right\rangle. \tag{16}$$

The expression (16) is linear in T_e^2. It is convenient to introduce the new function $\Theta(\mathbf{r},t)=T_e^2(\mathbf{r},t)$ and take the Fourier transform with respect to space and time variables. Then Eq. (16) yields the Fourier component of the heat flow:

$$\mathbf{q}(\mathbf{k},\omega)=\frac{i\pi^2}{6}\left\langle\frac{(\omega-\mathbf{vk})\mathbf{v}}{\omega-\mathbf{vk}+i\tau^{-1}}\right\rangle\Theta(\mathbf{k},\omega). \tag{17}$$

Substituting this result into the equation of thermoconductivity (9), we obtain its Fourier component

$$-i\frac{\pi^2}{3}\left\langle\omega+\frac{(\omega-\mathbf{vk})\mathbf{vk}}{\omega-\mathbf{vk}+i\tau^{-1}}\right\rangle\Theta(\mathbf{k},\omega) =2\kappa(1-R)I(\omega)U(\mathbf{k}), \tag{18}$$

where $I(\omega)$ is the Fourier transform of the pulse shape $I(t)$. The factor $Q(\mathbf{k})$ describes the spatial distribution of the laser field (11), and

$$U(\mathbf{k})=\frac{2\kappa}{\kappa^2+k_z^2},$$

which depends only on k_z. Equation (18) yields the temperature dynamics of metals under laser heating with the time and space dispersion.

We now turn to the equation for lattice displacements (1). The driving force $G_i(\mathbf{r},t)$ can be evaluated as in the derivation of Eqs. (16) and (17). Both the local equilibrium partition function and nonequilibrium part (4) contribute to the integral (2). Expanding the integrals over the energy variable in powers of T_e/ε_F up to the second order, we obtain

$$G_i(\mathbf{k},\omega)=-\frac{\pi^2}{6}\frac{\partial}{\partial\varepsilon_F}\left\langle\frac{\tau^{-1}\lambda_{ij}(\mathbf{p})k_j}{\omega-\mathbf{vk}+i\tau^{-1}}\right\rangle\Theta(\mathbf{k},\omega). \tag{19}$$

In addition to the electron force (19) we also obtain the temperature-dependent renormalization of the elastic constants λ_{ijlm} (sound velocities) due to the interaction with electrons (electron loop in the phonon self-energy function). The dominant contribution in the range of interest comes from the local equilibrium partition function:

$$\lambda_{ijlm}\to\lambda_{ijlm}-\langle\lambda_{ij}(\mathbf{p})\lambda_{lm}(\mathbf{p})\rangle -\frac{\pi T_e^2}{6}\frac{\partial^2}{\partial\varepsilon_F^2}\langle\lambda_{ij}(\mathbf{p})\lambda_{lm}(\mathbf{p})\rangle.$$

The electron contribution to the sound velocity is second order in the electron temperature, $\Delta s/s\sim T_e^2/\varepsilon_F^2$.

Taking the Fourier transform of the left-hand side of Eq. (1), one needs to keep in mind the singularity at $z=0$ after continuation (13) of the function u_z. This singularity contributes the term $d\delta(z)/dz$ in the second derivative d^2u_z/dz^2; such a term accounts for surface effects. The Fourier transform with respect to the coordinate z over the entire space gives

$$-\rho(\omega^2-s^2k^2)u_z(k,\omega)=G_z(k,\omega)+kC(\omega), \tag{20}$$

where $s=\lambda_{zzzz}/\rho$ is the longitudinal sound velocity in the z-direction. In the last term, $C(\omega)$ must be determined from the boundary condition (12), and takes the form

$$C(\omega)=-2i\omega s\int\frac{dk}{2\pi}\frac{G_z(k,\omega)}{\omega^2-s^2k^2}. \tag{21}$$

We next proceed to the electron temperature and lattice deformations represented by Eqs. (18) and (20) in various limiting cases.

4. SPATIAL VARIATION OF ELECTRON TEMPERATURE AND LATTICE DEFORMATION

Equation (18) describes the electron temperature evolution under ultrashort laser heating of metals. This equation generalizes the usual thermoconductivity equation.[1] We are interested in the wave vector k, which is the greater of the inverse skin depth κ ($\sim10^5\,\mathrm{cm}^{-1}$) and the electron diffusion length $v\sqrt{\tau t_0}$ during the laser pulse t_0. In the usual experimental situation $\tau^{-1}\sim10^{14}\,\mathrm{s}^{-1}$, $\kappa v\sim10^{13}\,\mathrm{s}^{-1}$, and the hydrodynamic regime $\kappa v\ll\tau^{-1}$ is obtained. Thus, one can omit the term $\mathbf{kv}$ in the denominator of the left-hand side of Eq. (18). The dominant contribution comes from the diffusion pole $\omega\sim\tau v^2k^2\ll\tau^{-1}$. Therefore, we can also omit ω everywhere in comparison with τ^{-1} or kv. The solution of the thermoconductivity equation reads

$$\Theta(z,t)=\Theta_0+\frac{i}{\beta}\int_{-\infty}^{\infty}\frac{dk\,d\omega\,dt'\,dz'}{(2\pi)^2(\omega+iDk^2)}I(t')\times\exp[-i\omega(t-t')+ik(z-z')-\kappa|z'|], \tag{22}$$

where the diffusion coefficient $D=\tau\langle v_z^2\rangle/\langle 1\rangle$ is introduced. The constant $\Theta_0=T_0^2$ comes from the solution of the corresponding homogeneous equation, and represents the initial temperature. Evaluating the integral (22) with respect to ω and k, we obtain

$$\Theta(z,t)=\Theta_0+\int_{-\infty}^{t}dt'\int_{-\infty}^{\infty}dz'\frac{Q(|z'|,t')}{\beta\sqrt{\pi(t-t')D}}\times\exp\left(-\frac{(z-z')^2}{4(t-t')D}\right). \tag{23}$$

We see immediately that the function (23) satisfies the boundary condition (12). For the temperature at the surface $z=0$, Eq. (23) gives

$$T_e^2(0,t)=T_0^2+\frac{4}{\pi\beta}\int_0^t dt'\,Q(0,t-t')\times\exp(\kappa^2Dt')\,\mathrm{erfc}(\sqrt{\kappa^2Dt'}).$$

The electron temperature (23) just after the pulse peaks at the surface:

$$T_{\max}^2\sim\frac{It_0(1-R)}{\beta}\min(\kappa,(Dt_0)^{-1/2}). \tag{24}$$

This result has a simple explanation. For short pulses $\kappa\sqrt{Dt_0}\ll1$, the time dependence of the temperature corresponds to the local laser intensity at the observation point. In the opposite case, $\kappa\sqrt{Dt_0}\gg1$, the temperature distribution is determined mainly by the diffusion process.

Consider now the equation for lattice displacements (20) with the force (19). Note that in the hydrodynamic regime, $\kappa v\ll\tau^{-1}$, the dominant contribution to the force G_i comes from the local equilibrium partition function, i.e., the first term in (4), if we consider times greater than the electron-electron relaxation time, $t\gg\tau$. In this case, the force has the simple expression

$$G_i(\mathbf{r},t)=\Lambda_{ij}\frac{\partial T_e^2(\mathbf{r},t)}{\partial x_j},$$

where the constants

$$\Lambda_{ij}=\frac{1}{32\pi}\frac{\partial}{\partial\varepsilon_F}\int\frac{dS}{v}\lambda_{ij}(\mathbf{p})\sim g\beta$$

are of the order of the electron density of states at the Fermi surface.

From Eq. (20) with the help of the expression (21) one can find the lattice deformation

$$\frac{du_z}{dz}=\frac{i\Lambda_{zz}\kappa(1-R)}{\rho\beta}\int\frac{d\omega dk}{(2\pi)^2}\frac{k^2U(k)I(\omega)}{(\omega+ik^2D)(\omega^2-s^2k^2)}\times[e^{ikz}-e^{i\omega|z|/s}]e^{-i\omega t}. \tag{25}$$

The first term in the brackets in Eq. (25) represents the particular solution of the inhomogeneous Eq. (1), while the second corresponds to the general solution of the homogeneous form of Eq. (1), and represents the effect of the surface. The integrand in (25) contains poles associated with the diffusion and sound-wave excitations. Sound singularities are bypassed using infinitesimal phonon damping, $\omega\to\omega+i0$.

5. EFFECT OF ACOUSTIC AND OPTICAL DISPLACEMENTS ON DESTRUCTION OF METALS

Equation (25) describes the effect of nonequilibrium electron heating on lattice deformations of acoustic type. This deformation vanishes at the surface $z=0$ according to the boundary condition (12). For $z\neq0$, the second term in brackets in (25) represents a deformation wave propagating from the surface into the bulk of the metal. It makes a nonzero contribution only at sufficiently small depths $z<st\sim10^{-7}\,\mathrm{cm}$. Thus, we see that the deformation (25) peaks at $z\sim10^{-7}\,\mathrm{cm}\ll\kappa^{-1}$. To obtain the order of the effect, we can drop the second term in parentheses. It is then convenient to integrate over ω, substituting the Fourier transform $I(\omega)$. We obtain

$$\frac{du_z}{dz}\sim\frac{\Lambda_{zz}\kappa(1-R)}{\rho\beta}\int_0^t dt' I(t')\int\frac{dk}{2\pi}U(k)\times\left(\frac{\exp[-isk(t-t')]}{s(s+ikD)}-\frac{\exp[-k^2D(t-t')]}{k^2D^2+s^2}\right)e^{ikz}. \tag{26}$$

Consider times greater than the duration of a pulse ($t>t_0$) but less than both the characteristic time of electron diffusion ($t<(\kappa^2 D)^{-1}\sim 10^{-12}$ s) and a sound-wave period ($t\ll(s\kappa)^{-1}\sim 10^{-11}$ s) with characteristic wave vector of the order of the inverse skin depth κ. In this range we can expand the exponentials in (26) in powers of t up to second order:

$$\frac{du_z}{dz}\sim\frac{\Lambda I t_0\kappa(1-R)}{2\rho\beta}t^2\int_{-\infty}^{\infty}\frac{dk}{2\pi}U(k)k^2e^{ikz}. \tag{27}$$

Using the estimate $\Lambda/\rho\sim gs^2/\varepsilon_F^2$ and Eq. (24) for the maximum electron temperature, we extrapolate our result up to electron diffusion times $t\sim(\kappa^2 D)^{-1}$:

$$\frac{du_z}{dz}\sim g(1-R)\frac{It_0}{\kappa\beta}\left(\frac{s}{\tau v^2\varepsilon_F}\right)^2\sim g\left(\frac{sT_{\max}}{\kappa\tau v^2\varepsilon_F}\right)^2. \tag{28}$$

Setting $s/v\sim 10^{-2}$, $\kappa\sim 10^5\ \mathrm{cm}^{-1}$, we arrive at the numerical estimate $du_z/dz\sim 10^{-2}g^5(T_e/\varepsilon_F)^2$.

Our result contains the natural factor T_e^2/ε_F^2, which means that laser heating is important as soon as the electron temperature is higher than the Fermi energy. Although the estimate was obtained for $T_e\ll\varepsilon_F$, it is still roughly correct up to $T_e\sim\varepsilon_F$. The additional small factor $s^2/(v\kappa l)^2$ is due to the fact that the characteristic period of the sound wave (10^{-11} s) is much greater than the characteristic times of electron diffusion (10^{-12} s) and laser heating (10^{-13} s). Therefore it would be of considerable interest to calculate the lattice deformation from high-frequency (nevertheless long-wavelength) excitations, i.e., optical phonons whose period is about 10^{-14} s. This case differs from the calculations above in the equations of lattice motion (1) and electron force (2) due to the different form of deformation potential (see Ref. 21). Estimates show that the relative optical displacement (with respect to the lattice constant) is of the order of T_e^2/ε_F^2.

6. CONCLUSIONS

Our result (28) for acoustic deformation agrees with the experiment reported by Rentzepis,[17] where a deformation $du_z/dz\sim 10^{-3}$ had been observed in the laser heating of noble metals. However, we see that the interaction of heated electrons with optical phonons can provide a more effective means of strong lattice deformation, but this case has yet to be studied experimentally. An ultrashort intense laser pulse can result in the destruction and ablation of metals, while only the electron component is heated, and the lattice stays cool at a considerably low temperature.

In conclusion, we emphasize two points. First, as follows from Eq. (9), the driving force for lattice expansion is proportional to $T_e\partial T_e/\partial z$. Because of the high absorption coefficient of metals in the UV ($\kappa\sim 10^5\ \mathrm{cm}^{-1}$), the temperature gradient reaches $\sim 10^9$ K/cm. Note that the extremely high values of this parameter (which is typical of metals) leads to nonequilibrium expansion of the lattice. Second, subpicosecond elastic deformation of the lattice of the order of 10^{-3}–10^{-2}, corresponding to an internal pressure 10–100 GPa, can provide an effective mechanism for subsequent laser fracture of metals.

We are grateful to S. I. Anisimov and V. A. Benderskii for many useful discussions and valuable comments. This work was supported in part by the Russian Fund for Fundamental Research (Grant No 97-02-16044). One of the authors (E. G. M.) also thanks KFA Forschungszentrum, Jülich, Germany, for a Landau Postdoctoral Fellowship.

[1] S. I. Anisimov, A. M. Bonch-Bruevich, M. A. El'yashevich *et al.*, Zh. Tekh. Fiz. **36**(7), 1273 (1967) [Sov. Phys. Tech. Phys. **11**, 945 (1967)]; S. I. Anisimov, B. L. Kapeliovich, and T. L. Perelman, Zh. Éksp. Teor. Fiz. **66**, 776 (1974) [Sov. Phys. JETP **39**, 375 (1974)].
[2] R. Yen, J. Liu, and N. Bloembergen, Opt. Commun. **35**, 277 (1980); R. Yen, J. Liu, N. Bloembergen *et al.*, Appl. Phys. Lett. **40**, 185 (1982).
[3] M. I. Kaganov, I. M. Lifshits, and L. V. Tanatarov, Zh. Éksp. Teor. Fiz. **31**, 232 (1957) [Sov. Phys. JETP **4**, 173 (1957)].
[4] H. Elsayed-Ali, M. Pessot, T. Norris, and G. Mourou, in *Ultrafast Phenomena V*, Proc. of the Fifth Optical Society of America Meeting, Snowmass, Colorado, 1986, G. R. Fleming and A. E. Slegman (Eds.), Springer Series in Chemical Physics, Vol. 46, Springer-Verlag, New York (1986), p. 264.
[5] I. G. Fugimoto, J. M. Liu, E. P. Ippen *et al.*, Phys. Rev. Lett. **53**, 1837 (1984).
[6] D. M. Riffe, X. Y. Wang, and M. C. Downer, J. Opt. Soc. Am. B **10**, 1424 (1993).
[7] J. P. Girardeau-Montaut and C. Girardeau-Montaut, Phys. Rev. B **51**, 13560 (1995).
[8] S. I. Anisimov, V. A. Benderskii, and G. Farkas, Usp. Fiz. Nauk **122**(2), 185 (1977) [Sov. Phys. Usp. **20**, 467 (1977)].
[9] A. G. Krivenko, J. Kruger, W. Hautex *et al.*, Benchte Bunsengeselsch. Phys. Chem. **99**, 1489 (1995).
[10] V. A. Benderskii, Russ. J. Electrochem. **33**, 417 (1997).
[11] A. G. Krivenko, J. Kruger, W. Hautex *et al.*, Russ. J. Electrochem. **33**, 426 (1997).
[12] M. B. Agranat, A. A. Benditskii, G. M. Gandel'man *et al.*, JETP Lett. **30**, 167 (1979).
[13] M. B. Agranat, S. I. Anisimov, and B. I. Makshantsev, Appl. Phys. B **47**, 209 (1988).
[14] M. B. Agranat, S. I. Anisimov, and B. I. Makshantsev, Appl. Phys. B **55**, 451 (1992).
[15] S. I. Anisimov and B. Rethfeld, Proc. SPIE **3093**, 192 (1997).
[16] L. A. Falkovsky and E. G. Mishchenko, JETP Lett. **66**, 208 (1997).
[17] P. Rentzepis, in *Abstracts of the Second Conference on Modern Trends in Chemical Kinetics*, Vol. 1, Novosibirsk (1995), p. 142.
[18] L. A. Falkovsky and E. G. Mishchenko, Phys. Rev. B **51**, 7239 (1995).
[19] V. M. Kontorovich, Usp. Fiz. Nauk **142**(2), 265 (1984) [Sov. Phys. Usp. **27**, 134 (1984)].
[20] E. M. Lifshitz and L. P. Pitaevski, *Physical Kinetics*, Nauka, Moscow (1979), p. 33.
[21] E. G. Mishchenko and L. A. Falkovsky, Zh. Éksp. Teor. Fiz. **107**, 936 (1995) [JETP **80**, 531 (1995)].

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

On accuracy of different cluster models used in describing ordering phase transitions in fcc alloys

V. G. Vaks*) and G. D. Samolyuk†)

Kurchatov Institute, 123182 Moscow, Russia
(Submitted 15 April 1998)
Zh. Éksp. Teor. Fiz. **115**, 158–179 (January 1999)

A general formulation of cluster methods applied to calculations of thermodynamic quantities of alloys in terms of renormalizing fields describing interaction between a cluster and its environment is given. We have shown that the well-known cluster variation method and the cluster field method, which was suggested earlier, are special cases of our approach. These methods have been used in calculations of phase diagrams of fcc alloys with $L1_2$ and $L1_0$ ordering transitions with several realistic interaction models. It turns out that, for all these models, the simple tetrahedron version of the cluster field method suggested in this paper describes the phase diagrams almost as accurately as more complicated cluster variation techniques. Possible applications of the tetrahedron version of the cluster field method to inhomogeneous states and kinetics of phase transitions in fcc alloys are discussed. © *1999 American Institute of Physics.* [S1063-7761(99)01501-2]

1. INTRODUCTION

Studies of thermodynamics and kinetics of phase transformations in alloys attract a lot of attention because, in particular, the related problems are important for technical applications. Since experimental research in phase transitions of this kind is usually complicated, development of adequate theoretical techniques for their description is deemed very important.[1–7] The simplest theoretical method in this area is the mean-field approximation. In real alloys, however, this approximation does not yield quantitatively accurate results, and in the analysis of ordering transitions in fcc alloys it yields phase diagrams with incorrect configurations.[1] This is caused by the presence of strong interatomic correlations among nearest neighbors, where the configurational energy is comparable to or larger than temperature T, whereas the mean-field approximation neglects such correlations. In order to describe such correlations, more accurate cluster techniques, primarily the well-known cluster variation method,[4,8–14] which are usually applied to calculations of alloy phase diagrams, have been developed. Some researchers have also used computer simulations based on the Monte Carlo method.[5] As concerns realistic models of alloys, however, such numerical calculations demand a lot of CPU time, therefore they are rarely used in practice.

Various formulations and versions of the cluster approximation have been discussed in numerous publications.[4,9–15] This method has been discussed mostly in the context of applications to ordering transitions in fcc alloys, in particular, the most common $L1_2$ or $L1_0$ ordering (like in Ni_3Al or TiAl), since the results obtained by the mean-field approximation proved unsatisfactory, as was stated above. The comparison with available Monte Carlo calculations has demonstrated that the results of commonly used methods of tetrahedron cluster variation (TCVM) and tetrahedron–octahedron cluster variation (TOCVM) taking into account interactions among nearest and second-nearest neighbors are fairly accurate,[10,11] especially when clusters of larger sizes are defined in the disordered phase.[4] At the same time, interactions among the third and farther neighbors are usually weak and can be described in the mean-field approximation.[2,3] Thus, the techniques for calculating phase diagrams of homogeneous alloys based on the cluster variation method can be considered self-sufficient and well-grounded.

Yet the physics of phase transformations is not limited to the study of homogeneous and equilibrium systems. For example, one topical problem in this field is kinetics of transformations, in particular, $L1_2$ and $L1_0$ ordering transitions mentioned above.[16] In reality, these phase transitions (for example, after quenching into the region of thermodynamic instability of the initial state) proceed via generation of microscopic nuclei of a new phase within the initial phase and further growth of these entities. Therefore, the process of phase transformation should be described in terms of evolution of essentially inhomogeneous and nonequilibrium states. As a matter of fact, fully homogeneous states are unattainable, and real alloys usually contain a lot of inhomogeneities, in particular, interphase and antiphase boundaries. The microstructure and macroscopic properties, such as plasticity and strength, strongly depend, as a rule, on the thermal and kinetic prehistory of a sample, in particular, on the kinetic trajectory of phase transformations. Therefore, topics concerned with structure and evolution of inhomogeneous and nonequilibrium alloys have been intensely investigated and discussed in literature.[6,7] In this connection, generalization of the cluster methods mentioned above to inhomogeneous and unsteady states, in particular, to ordering in fcc alloys, is very significant from the viewpoint of both fundamental research and applications.

As is stated in earlier publications[17,18] and below, such

generalizations of the conventional cluster variation method lead to very cumbersome equations, which can hardly be used in practical calculations. At the same time, these generalizations can be fairly simple and practicable if one uses a simplified version of the cluster approximation, namely the cluster field method, which was previously employed in studies of thermodynamics[19–22] and short-range order[23–26] in some strongly correlated alloys. The question arises, however, of the accuracy of this relatively simple method since, unlike the cluster variation technique, the accuracy of the cluster field method has yet been little investigated. One should also keep in mind that the calculation accuracy may depend on both the type of the adopted model and the specific version of the method (namely, tetrahedron, TCVM, or tetrahedron–octahedron, TOCVM), moreover, utilization of more complex clusters does not necessarily lead to a higher accuracy.[4] The accuracy of this technique based on some models and versions of the cluster field method in calculations of phase diagrams was tested by comparison to Monte Carlo or cluster variation calculations[15,19] and proved to be fairly high. Such investigations, however, have not been systematic. At the same time, the accuracy of the cluster field method in applications to simpler, equilibrium properties of materials should be investigated comprehensively in view of applying it to complex and little-known problem of phase transformation kinetics.

This publication is dedicated to these problems. We will discuss the accuracy of some simplest versions of the cluster field method in describing phase diagrams of fcc alloys with $L1_2$ or $L1_0$ ordering. We will consider several models of such alloys. The accuracy of the versions of the cluster method will be estimated by comparing calculations to results of one of the most accurate versions of the cluster method described earlier[4,10] and applied to the same models. It will be shown that, for models that seem most realistic, calculations based on a simple approximation corresponding to a generalization of Yang's ''quasi-chemical tetrahedron cluster method''[27] and dubbed QCTCM hereafter, are fairly close to results of the cluster variation method. At the same time, with less realistic models, such as those taking into account interaction only among nearest neighbors,[28] the difference between results of QCTCM and cluster variation method can be appreciable. Calculations of ordering spinodals, i.e., boundaries of disordered phase stability against ordering in the concentration–temperature plane $(c-T)$, will be also given. In most calculations of phase diagrams published in the literature these curves are not given,[2–4] whereas they are very important for studies in kinetics of phase transformations.[29]

Section 2 presents a general description of various cluster techniques in terms of renormalizing fields in clusters. This approach seems more simple and general than commonly used ones[8–14] and admits various generalizations, including those used in the reported work. In Sec. 3 general results of Sec. 2 will be applied to cluster techniques used in our calculations. Section 4 describes applications of these techniques to calculations of phase diagrams and ordering spinodals for several models of alloys with $L1_2$ or $L1_0$ ordering and analyzes the results. Basic results of this work will be summarized in the Conclusion.

2. DESCRIPTION OF CLUSTER TECHNIQUES IN TERMS OF RENORMALIZING FIELDS IN CLUSTERS

This section presents general formulations of various cluster techniques, including the cluster variation methods and cluster field methods, in which basic physical quantities are effective fields and interactions inside a cluster due to its interactions with the environment. This graphic approach, which was described in detail earlier,[15] allows one to see clearly both the assumptions on which various generalized versions are based and the opportunities offered by them. The feasibility of this approach in a special case of cluster variation method was mentioned previously.[12,14] Below we will generalize earlier results[15] to the case of multicomponent and, possibly, inhomogeneous alloys and rectify some inaccurate statements.

2.1. General formulation of cluster methods. Consider a substitutional alloy that contains atoms of m sorts, $p=p_1, p_2, \ldots, p_m$, including, possibly, vacancies. The distribution of atoms over lattice sites is characterized by various sets of occupancies $\{n_{pi}\}$, where $n_{pi}=1$ if the ith site is occupied by an atom of sort p and $n_{pi}=0$ if otherwise. At all i operators n_{pi} satisfy the condition $\Sigma_p n_{pi}=1$, so that only $m-1$ of them are independent. Hereafter we assume that for one sort of atoms, specifically, $p=p_m$, this operator is expressed in terms of the rest, $n_{p_m i}=1-\Sigma_{p\neq p_m} n_{pi}$, so that below the sums over p include only $m-1$ terms with $p=p_1, p_2, \ldots, p_{m-1}$. In the case of a binary alloy, subscript p can have only one value and is omitted.

A general expression for the configurational Hamiltonian H' (i.e., it depends only on the distribution of atoms among lattice sites) in terms of occupancies n_{pi} is

$$H'=\sum_{pi} \varphi_i^p n_{pi}+\sum_{pq,i<j} v_{ij}^{pq} n_{pi} n_{qj} + \sum_{pqr,i<j<k} v_{ijk}^{pqr} n_{pi} n_{qj} n_{rk}+\cdots \qquad (1)$$

Here ''external fields'' φ_i^p are due to possible nonequivalence of sites, for example, owing to lattice defects. In calculations of equilibrium parameters, one should calculate thermodynamic potential Ω as a function of temperature T and relative chemical potentials $\mu_p=\mu_p^0-\mu_{p_m}^0$, where μ^0 are ''absolute'' chemical potentials (for example, measured with respect to the vacuum):

$$\Omega=-T\ln \mathrm{Tr}\exp(-\beta H), \quad H=H'-\sum_{pi} \mu_p n_{pi}. \qquad (2)$$

Here $\beta=1/T$, and Tr denotes summation over all possible configurations $\{n_{pi}\}$. For brevity, the effective Hamiltonian H of the grand canonical distribution in Eq. (2) will be dubbed simply Hamiltonian, and its mean value $\langle H\rangle$ will be called energy E.

Cluster techniques deal with clusters, i.e., sets of specific lattice sites $i,j,\ldots,l$, which will be denoted for brevity by first letters of the Greek alphabet, e.g., $\{i,j,\ldots,l\}=\alpha$. The

distribution ρ_α of various configurations in cluster α is characterized by its effective Hamiltonian H_α and the corresponding potential Ω_α, which is obtained, strictly speaking, by averaging the full Gibbs distribution over variables n_{pi} for all sites not included in the cluster:

$$\rho_\alpha = \exp[\beta(\Omega_\alpha - H_\alpha)] = \operatorname*{Tr}_{i \notin \alpha} \exp[\beta(\Omega - H)], \tag{3}$$

$$\Omega_\alpha = -T \ln Z_\alpha = -T \ln \operatorname*{Tr}_{i \in \alpha} \exp(-\beta H_\alpha). \tag{4}$$

If we denote for brevity the operator characterizing the occupancy of state $\xi = \{pi, qj, \ldots, rk\}$ in a certain subcluster $\gamma = \{i, j, \ldots, k\}$ of cluster α by $n_\xi = n_{pi} n_{qj} \ldots n_{rk}$, Hamiltonian H_α can be written in a compact form as a sum over all possible configurations ξ of the cluster:

$$H_\alpha = \sum_{p,i} (\varphi_{pi} + \psi^\alpha_{pi} - \mu_p) n_{pi} + \sum_{pq, i<j} (v^{pq}_{ij} + \psi^\alpha_{pi,qj}) \times n_{pi} n_{qj} + \ldots \equiv \sum_{\xi \subseteq \alpha} (v_\xi + \psi^\alpha_\xi) n_\xi . \tag{5}$$

Here the sums are performed over sites and configurations of cluster α, $v_\xi = \varphi_{pi} - \mu_p$ for ''one-atom'' occupancies $\xi = pi$, and ψ^α_ξ describes renormalization of variables v_ξ, i.e., fields φ_{pi} and potentials $v^{p \ldots q}_{i \ldots j}$ owing to interaction between the cluster α and its environment. It follows from Eqs. (4) and (5) that the mean occupancy $g_\xi = \langle n_\xi \rangle$ for all clusters $\alpha \supseteq \xi$ is related to field ψ^α_ξ in this cluster by the formula

$$g_\xi = (\partial \Omega_\alpha / \partial \psi^\alpha_\xi)_{T, v_\xi}. \tag{6}$$

Hamiltonian H_α accurately describes interactions v_ξ within the cluster. Its interaction with the environment, however, is described by variables ψ^α_ξ, which can be calculated in practice using some approximation. The contribution of interactions within the cluster, as compared with fields ψ^α_ξ, grows, generally speaking, with the cluster size, and the accuracy improves, but calculations become more complicated. In calculating thermodynamic quantities, such as thermodynamic potential $\Omega = -T \ln \operatorname{Tr} \exp(-\beta H)$, one should perform the following operations: (a) express the full Hamiltonian H in terms of a combination of cluster Hamiltonians H_α as a series fast converging with the cluster size; (b) select a decomposition of the lattice into clusters and an approximation for calculating fields ψ^α_ξ.

A consistent approach that should be used on step (a) is the method of cumulant expansions,[15] whose underlying ideas were discussed in several papers.[4,9,12] The physical indicator of convergence in this technique is the reduction in the contributions from irreducible correlations to all physical quantities with the number of sites in the cluster. These contributions are obtained by subtracting from full n-particle correlators the contributions of all correlations involving smaller numbers of particles, $m = n-1, n-2, \ldots, 1$, which compose subclusters of this cluster. This means that the expression for the energy should include, alongside H_α, ''cluster cumulant'' Hamiltonians $\tilde{H}_\alpha$ (for brevity called simply cumulants), which are related to H_α by the formulas

$$\tilde{H}_\alpha = \sum_{\gamma \subseteq \alpha} (-1)^{|\alpha| - |\gamma|} H_\gamma, \tag{7}$$

$$H_\alpha = \sum_{\gamma \subseteq \alpha} \tilde{H}_\gamma, \tag{8}$$

where $|\alpha|$ is the number of lattice sites in the cluster α, and the sum is performed over all subclusters γ of cluster α obtained by excluding $0,1,2,\ldots,|\alpha|-1$ different sites. In particular, for double and triple cumulants and one-particle cluster fields $\tilde{\psi}^\alpha_{pi}$, we have the following expressions in terms of cumulants defined by Eq. (7):

$$\tilde{H}_{ij} = H_{ij} - H_i - H_j,$$

$$\tilde{H}_{ijk} = H_{ijk} - H_{ij} - H_{jk} - H_{ki} + H_i + H_j + H_k; \tag{9}$$

$$\tilde{\psi}^{ij}_{pi} = \psi^{ij}_{pi} - \psi^i_{pi}, \quad \tilde{\psi}^{ijk}_{pi} = \psi^{ijk}_{pi} - \psi^{ij}_{pi} - \psi^{ik}_{pi} + \psi^i_{pi}. \tag{10}$$

Equations (5) and (7)–(10) clearly show that cumulants $\tilde{H}_\alpha$ and fields $\tilde{\psi}^\alpha_\xi$ do not contain interactions and fields included in the subclusters of the cluster. Therefore, it is natural to expect that the contributions of these cumulants to all physical quantities should decrease fairly rapidly with the cumulant size $|\alpha|$.

If the entire lattice is treated as the largest cluster, Eq. (8) reads as

$$H = \sum_\alpha \tilde{H}_\alpha . \tag{11}$$

Selection of the approximation on step (a) means retention of a limited set of cumulants $\{\alpha\}$ on the right-hand side of Eq. (11) with small norms $|\alpha| \leqslant |\alpha|_{\max}$ and rejection of the rest $\tilde{H}_\alpha$. After this operation, exact expressions for cluster Hamiltonians H_α, i.e., fields ψ^α_ξ in Eq. (5), are replaced by approximate ones. By expressing the retained cumulants $\tilde{H}_\alpha$ in Eq. (11) in terms of cluster Hamiltonians H_α, we obtain the basic equations of the cluster methods:

$$H = \sum_\alpha \nu_\alpha H_\alpha . \tag{12}$$

Here the sum is performed over the retained clusters of the maximal size and their subclusters, and factors ν_α, as was shown by Sanchez *et al.*,[12] for each subcluster γ of the basic clusters satisfy the condition

$$\sum_{\alpha \supseteq \gamma} \nu_\alpha = 1, \tag{13}$$

which ensures equality of the factors in front of all interactions $v_\xi n_\xi$ on both sides of Eq. (12). The left-hand side of Eq. (12), on the other hand, does not contain renormalizing fields ψ^α_ξ, so for all occupancies ξ these fields should be related by the equation

$$\sum_{\alpha \supseteq \xi} \nu_\alpha \psi^\alpha_\xi = 0. \tag{14}$$

The largest, or ''basic,'' clusters α_m with $|\alpha_m| = |\alpha|_{\max}$, which are not considered as parts of other clusters, contribute to Eqs. (13) and (14) only one term each with $\alpha = \alpha_m$.

Hence follows that $\nu_{\alpha_m}=1$ and $\psi_{\xi_m}^{\alpha_m}=0$, where the ξ_m denote different $|\alpha_m|$-particle occupancies of basis clusters. Thus, interactions among $|\alpha_m|$ particles, v_{ξ_m}, in Eqs. (3)–(5) for the maximal clusters are not renormalized. This approximation, which limits the sum in Eq. (11) to cumulants with $\alpha\subseteq\alpha_m$ is the basic approximation of all cluster techniques.

The geometrical factors ν_α in Eqs. (12) and (14) are determined by Eq. (13), i.e., they depend on the lattice configuration (bcc, fcc, etc.) and selection of basic clusters, but they are independent of the atomic distribution in the lattice, in particular, on its order. These factors for the fcc lattice in the most common versions of the cluster variation technique and cluster field method will be given in Sec. 3. The general properties of factors ν_α were discussed in Refs. 4 and 12, where they were denoted by a_α. Those papers mentioned, in particular, an important property of such factors for ''inner'' subclusters of basic clusters, i.e., such subclusters γ that are parts of only one basic cluster α_m. In order to discover this property, let us apply Eq. (13) to an inner subcluster γ. We obtain $\nu_{\alpha_m}+\nu_\gamma=1$ and $\nu_\gamma=0$, since in the basic cluster $\nu_{\alpha_m}=1$. Equation (14) indicates that in this case the fields $\psi_\xi^{\alpha_m}$, corresponding to occupancies ξ in the inner subcluster of the basis cluster α_m, also go to zero. Thus, the factors ν_γ for the inner subclusters in Eqs. (12)–(16) and the corresponding fields $\psi_\xi^{\alpha_m}$ in the basic clusters equal zero, so these subclusters do not contribute to observable quantities and can be discarded. This result, which we dub for brevity the inner-subcluster theorem, simplifies considerably all cluster calculations and will be used in Sec. 3.

By integrating over β the thermodynamic relation

$$\langle H\rangle=\frac{\partial}{\partial\beta}(\beta\Omega)=-\frac{\partial}{\partial\beta}\ln\mathrm{Tr}\exp(-\beta H), \tag{15}$$

where $\langle H\rangle$ is calculated using approximation (12)–(14), we obtain

$$\Omega=\sum_\alpha \nu_\alpha\Omega_\alpha, \tag{16}$$

where Ω_α is defined by Eq. (4). Here we have taken into account that [as follows from Eqs. (6) and (14)] the full derivative of expression (16) with respect to each independent variable ψ_ξ^α is zero.[20] Really, if we express, using Eq. (14), some ψ_ξ^β in terms of the rest of independent variables ψ_ξ^α, where $\alpha\neq\beta$, the derivative $d\psi_\xi^\beta/d\psi_\xi^\alpha$ equals $(-\nu_\alpha/\nu_\beta)$ and, given Eqs. (6) and (14), we obtain

$$\frac{d\Omega}{d\psi_\xi^\alpha}=g_\xi\left(\nu_\alpha+\nu_\beta\frac{d\psi_\xi^\beta}{d\psi_\xi^\alpha}\right)=0. \tag{17}$$

Note in this connection that the self-consistency conditions (6) are equivalent to the condition of minimal thermodynamic potential (16) when independent parameters ψ_ξ^α are varied at specified average occupancies g_ξ. Therefore, expression (16) can also be used in variational versions of the cluster technique, where fields ψ_ξ^α are varied parameters and the averages g_ξ are Lagrange multipliers, the same for all clusters α.

2.2. Cluster variation method. The cluster variation technique relies only on those approximations that have led to Eqs. (12)–(14). Therefore, in calculations of thermodynamic parameters, i.e., average occupancies g_ξ and potential Ω in Eq. (16) as functions of T and v_ξ, one should solve equation system (6) and (14).

Equation (6), which determines fields ψ_ξ^α and potentials Ω_α in Eq. (4) as functions of average occupancies g_ξ, can be solved directly. With this end in view, note that, if the contribution from a state with occupancy n_ξ to the partition function Z_α in Eq. (4) is denoted by S_ξ^α, i.e., $Z_\alpha=1+\Sigma_\xi S_\xi^\alpha$, the quantity S_ξ^α is related to $y_\xi^\alpha=\exp[-\beta(v_\xi+\psi_\xi^\alpha)]$ by the equation

$$\ln S_\xi^\alpha=\sum_{\eta\subseteq\xi\subseteq\alpha}\ln y_\eta^\alpha, \tag{18}$$

and Eq. (6), which determines S_ξ^α in terms of mean occupancies g_η, takes the form

$$Z_\alpha g_\xi=\sum_{\xi\subseteq\eta\subseteq\alpha} S_\eta^\alpha.$$

This equation can be solved with respect to S_ξ^α by induction, starting with the maximal $|\xi|=|\alpha|$ and reducing $|\xi|$, with the help of the following identity for the polynomial coefficients:

$$\sum_{k=1}^{m} C_m^k(-1)^{k-1}=1-(1-x)^m|_{x=1}=1.$$

As a result, we have

$$S_\xi^\alpha=Z_\alpha\sum_{\eta\subseteq\xi\subseteq\alpha}(-1)^{|\eta|-|\xi|}g_\eta,$$

where

$$Z_\alpha\equiv\exp(-\beta\Omega_\alpha)=\left[1+\sum_{\eta\subseteq\alpha}(-1)^{|\eta|}g_\eta\right]^{-1}. \tag{19}$$

Further, by solving Eq. (18) with respect to $\ln y_\xi^\alpha$ by induction on $|\xi|$ starting with the minimal $|\xi|=1$, and increasing $|\xi|$ similarly to the procedure described above, and expressing $\ln S_\xi^\alpha$ in terms of g_η using the above expression, we obtain expression for fields ψ_ξ^α in terms of mean occupancies g_η:

$$\beta(v_\xi+\psi_\xi^\alpha)=(-1)^{|\xi|}\ln Z_\alpha-\sum_{\eta\subseteq\xi\subseteq\alpha}(-1)^{|\xi|-|\eta|}\times\ln\left[\sum_{\eta\subseteq\zeta\subseteq\alpha}(-1)^{|\zeta|-|\eta|}g_\zeta\right], \tag{20}$$

where $|\xi$, $|\eta|$, or $|\zeta|$ is the number of sites in subcluster ξ, η, or ζ. Note that Eqs. (14b) and (15) in Ref. 15, which are similar to Eqs. (20) and (23) of this paper, contain errors.

After multiplying Eq. (20) by ν_α, performing the sum over all $\alpha\supseteq\xi$, and using Eqs. (13) and (14), we obtain an equation system for all mean occupancies g_ξ with $|\xi|\leqslant|\alpha_m|$:

$$\beta v_\xi = \sum_{\alpha \supseteq \xi} \nu_\alpha l_\xi^\alpha \{g\}, \tag{21}$$

where $l_\xi^\alpha\{g\}$ denotes the right-hand side of Eq. (20).

Substitution of g_ξ derived from Eq. (21) in Eqs. (19) and (16) allows one to calculate the thermodynamic potential $\Omega = \Omega(T, \mu_p)$. In solving physical problems, however, it is more convenient to use instead of Ω the free energy F, which is related to Ω by the formula

$$F(T, c_{pi}) = \Omega + \sum_{pi} \mu_p c_{pi}, \tag{22}$$

where $c_{pi} = \langle n_{pi} \rangle$ is the mean occupancy of site i by an atom of type p in the alloy, which can be either homogeneous or inhomogeneous.[17,18,30] Then the one-particle means $g_{pi} = c_{pi}$ in Eq. (21) should be treated not as sought-for functions of chemical potentials μ_p, but as arguments of sought-for functions Ω, μ_p, and g_ξ for $|\xi| \geqslant 2$. In this approach, Eq. (21) for the one-particle occupancy $\xi = pi$ determines the chemical potential $\mu_p(c_{qi}, T)$, and Eqs. (19) and (16) determine function $\Omega(c_{pi}, T)$.

The main computational problem in the cluster variation technique is solving nonlinear equation system (21) with respect to multiparticle means g_ξ for $|\xi| \geqslant 2$. We have solved this equation system by Newton's method.[31] This procedure is simplified by the opportunity to calculate analytically the matrix of derivatives $l_{\xi\eta}^\alpha = \partial l_\xi^\alpha / \partial g_\eta$ of functions l_ξ^α on the right of Eqs. (20) and (21):

$$l_{\xi\eta}^\alpha = (-1)^{|\xi| - |\eta| + 1} \times \left\{ Z_\alpha^{-1} + \sum_{\theta \subseteq (\xi \cap \eta)} \left[\sum_{\theta \subseteq \zeta \subseteq \alpha} g_\zeta (-1)^{|\zeta| - |\theta|} \right]^{-1} \right\}, \tag{23}$$

where $\theta \subseteq (\xi \cap \eta)$ means that the sum is performed over clusters θ, which are subclusters of both ξ and η clusters. The convergence of Newton's method in this case is fairly fast, but this requires inversion of the matrices $l_{\xi\eta} = \Sigma_\alpha \nu_\alpha l_{\xi\eta}^\alpha$, which are often ill-conditioned and sensitive to initial conditions. Therefore, it seems that, given modern high-performance computers, one should use more simple and stable methods, such as the method of conjugate gradients.[32]

Now let us discuss how our approach based on Eqs. (3)–(5) is related to the conventional cluster variation techniques[4,12–14] based on distribution functions ρ_α in clusters and their entropies S_α. In accordance with Eqs. (3)–(5), expressions for the cluster mean energy E_α and its entropy S_α can be written as

$$E_\alpha = \langle H_\alpha \rangle = \sum_{\xi \subseteq \alpha} (v_\xi + \psi_\xi^\alpha) g_\xi, \tag{24}$$

$$S_\alpha = \beta(E_\alpha - \Omega_\alpha) = -\operatorname{Tr} \rho_\alpha \ln \rho_\alpha. \tag{25}$$

Then, using Eqs. (12)–(14) and (16), we express the full system entropy $S = \beta(E - \Omega)$:

$$S = \beta \left(\sum_\xi v_\xi g_\xi - \sum_\alpha \nu_\alpha \Omega_\alpha \right) = -\sum_\alpha \nu_\alpha \operatorname{Tr} \rho_\alpha \ln \rho_\alpha, \tag{26}$$

which is the basic equations in the approaches developed earlier.[4,10–14]

Note also that, in conventional versions of the cluster variation technique,[4,8–14] ''pseudospin'' variables σ_{pi}, which are related to n_{pi} by formulas like $\sigma_{pi} = 2n_{pi} - 1$, are substituted for occupancy operators n_{pi}. In our opinion, utilization of natural variables n_{pi} instead of pseudospin variables σ_{pi} simplifies both the calculations[15] and physical interpretation of results, especially in multicomponent alloys.[18,30]

2.3. Cluster field method. The cluster field method is a simplified version of the cluster variation method, when cluster Hamiltonians (5) include only on-site fields ψ_{pi}^α due to environment, whereas renormalizations of interactions ψ_ξ^α with $(|\alpha_m| - 1) \geqslant |\xi| \geqslant 2$ are ignored. In this case the equality among multiparticle means g_ξ with $|\xi| \geqslant 2$ in clusters of different sizes $|\alpha|$ (which is achieved in the full cluster variation method by introducing renormalizations of ψ_ξ^α), generally speaking, no longer holds. Therefore the degree of self-consistency, hence the accuracy of the cluster field method, should be, in a general case, lower than in similar versions of the cluster variation method. Nonetheless, when cluster sizes are sufficiently large, the contribution of renormalizations to interactions within one cluster is usually reduced, so errors due to omission of such renormalizations can be limited. This circumstance was illustrated by comparing cluster field calculations to cluster variation and Monte Carlo calculations.[15,19] The more important is the fact that, in some specially selected basic clusters discussed below, interaction renormalizations and their thermodynamic contributions vanish, i.e., field ψ_ξ^α with $|\xi| \geqslant 2$ or factors ν_α in Eqs. (12) and (16) for clusters containing these fields go to zero. In this case, the cluster field method is equivalent to the cluster variation method, so both the degree of self-consistency and calculation accuracy can be fairly high. The QCTCM approximation mentioned above and discussed in detail in Secs. 3 and 4 is one example. Therefore, the good accuracy of the QCTCM calculations of simple models, which will be discussed in the following sections, is quite natural.

Equation system (19)–(21) is simplified considerably in the cluster field method. It is convenient to operate, instead of ψ_{pi}^α, with variables defined as ''activities'' y_{pi}^α of lattice sites and introduce them to Eq. (4), defining Ω_α:

$$y_{pi}^\alpha = \exp[\beta(\mu_p - \varphi_{pi} - \psi_{pi}^\alpha)]. \tag{27}$$

The cluster partition function Z_α in Eq. (4) takes the form of a polynomial of power $|\alpha|$ with y^α:

$$Z_\alpha = \sum_{\xi \subseteq \alpha} \exp(-\beta v_\xi) \prod_{pi \in \xi} y_{pi}^\alpha, \tag{28}$$

which contains the product of all activities y^α corresponding to the occupancy $\xi = \{p_1 i_1, p_2 i_2, \ldots, p_m i_m\}$, i.e., $y_{p_1 i_1}^\alpha y_{p_2 i_2}^\alpha \ldots y_{p_m i_m}^\alpha$. Equation (6) transforms to

$$c_{pi}^\alpha = y_{pi}^\alpha \partial \ln Z_\alpha / \partial y_{pi}^\alpha. \tag{29}$$

According to Eq. (28), the right-hand side of Eq. (29) is the ratio of two polynomials with y^α, whose numerator includes only those terms of the denominator which contain

factor y_{pi}^{α}. For all clusters α equation system (29) can be easily solved with respect to functions $y_{pi}^{\alpha}\{c_{qi}\}$ using Newton's method. Then thermodynamic potential $\Omega\{c_{pi}\}$ is derived from Eqs. (4), (16), and (28), and Eqs. (14) and (27) for one site, $\xi=pi$, determine the chemical potential $\mu_p\{c_{qi}\}$:

$$\mu_p=\varphi_{pi}+T\sum_{\alpha}\ \nu_{\alpha}\ln y_{pi}^{\alpha}. \tag{30}$$

Now let us comment on applications of basic equations of the cluster variation technique (21) and cluster field method (29) to inhomogeneous states. The mean occupancies g_{ξ} for subclusters ξ corresponding to nonequivalent sites are different. Therefore, the number of sought-for functions g_{ξ} in Eq. (21) is large even for simple, ordered phases. So, in the tetrahedron–octahedron cluster variation method, analysis of pure $L1_0$ or $L1_2$ phases requires calculations of 27 or 22 functions g_{ξ}, respectively, and a description of a simple inhomogeneous system, namely a plane antiphase boundary in the $L1_2$ phase, based on the tetrahedron cluster variation method demands calculation of several thousands of g_{ξ} and a lot of computer time.[4,33] Analysis of kinetic properties of alloys requires solution of the kinetic master equation, which is a system of differential equations containing derivatives with respect to time for all occupancies g_{ξ} in a lattice where all N sites are nonequivalent.[17,18] Solution of this problem by the cluster variation technique would require solutions of equation systems like Eq. (21) for all occupancies g_{ξ} on each time step (labeled by number s), and for interesting systems[29] with $N\gtrsim 10^5$ and $s\gtrsim 10^3$ such numerical calculations are hardly feasible, at least, in the foreseeable future. At the same time, equations (29) of the cluster field method are fairly simple, as was noted above, and algorithms for their solutions are fast convergent and stable. Therefore, application of these equations to kinetic problems mentioned above is quite feasible.

3. SELECTION OF BASIS CLUSTERS IN DIFFERENT VERSIONS OF CLUSTER TECHNIQUES

In this section we will discuss several commonly used versions of the cluster variation and cluster field methods, which will be used in Sec. 4. These versions are characterized by selection of maximal or basic clusters, which was discussed in Sec. 2.1.

3.1. Approximation of pair clusters. First let us consider the simplest approximation of pair clusters (2-clusters) in an alloy with pair interactions v_{ij}^{pq}. The basic clusters in this case are all pairs of sites $\{i,j\}$ whose interactions v_{ij}^{pq} are nonvanishing. As was proven above, interactions v_{ξ_m} for $|\xi_m|=|\alpha_m|$ in clusters with the maximal sizes $|\alpha_m|$ are not renormalized. Since in the approximation of 2-clusters we have $|\alpha_m|=2$, the renormalizing fields ψ_{ξ}^{α} in Eq. (5) are nonzero only for one-site subclusters, $\xi=pi$, and the cluster variation method is equivalent to the cluster field method.

As follows from Eq. (13), one-site coefficients ν_i are related to pair factors ν_{ij} by the formula $\nu_i=1-\Sigma_j\nu_{ij}$, where $\nu_{ij}=1$ if $v_{ij}^{pq}\neq 0$ and $\nu_{ij}=0$ for $v_{ij}^{pq}=0$. Equation (14) relating fields ψ_{pi}^{ij} and ψ_{pi}^{i} in 2- and 1-clusters can be rewritten as

$$\sum_{j}\ \nu_{ij}(\psi_{pi}^{ij}-\psi_{pi}^{i})+\psi_{pi}^{i}=0. \tag{31}$$

This equation illustrates additivity of fields δ_{pi}^{ij} acting on an atom of type p at site i and generated by all ''external'' sites j. Then the difference between field ψ_{pi}^{ij} in a 2-cluster $\{ij\}$ and total field ψ_{pi}^{i} generated by all sites is only the contribution of interaction (ij): $\delta_{pi}^{ij}=\psi_{pi}^{i}-\psi_{pi}^{ij}$, and summation of all such contributions over all interactions (ij) yields the total field ψ_{pi}^{i}, i.e., Eq. (31).

Equations (16) and (30) for Ω and μ_p in the pair cluster approximation have the form

$$\Omega=\sum_{i}\ \Omega_i+\sum_{i<j}\ \tilde{\Omega}_{ij}, \tag{32}$$

$$\mu_p-\varphi_{pi}=\mu_{pi}+\sum_{j}\ \tilde{\mu}_{pi}^{ij}. \tag{33}$$

Here Ω_i and μ_{pi} correspond to one-site contributions, i.e., they describe an ideal solution, whereas $\tilde{\Omega}_{ij}$ and $\tilde{\mu}_{pi}^{ij}$ are due to interactions v_{ij}^{pq}:

$$\Omega_i=T\ln\left(1-\sum_{p}\ c_{pi}\right),$$

$$\mu_{pi}=T\ln y_{pi}^{i}=T\ln\left[c_{pi}/\left(1-\sum_{q}\ c_{qi}\right)\right]; \tag{34}$$

$$\tilde{\Omega}_{ij}=\Omega_{ij}-\Omega_i-\Omega_j,\quad \tilde{\mu}_{pi}^{ij}=\ln(y_{pi}^{ij}/y_{pi}^{i}). \tag{35}$$

In the important case of a binary alloy AB, Eqs. (27)–(29) can be solved analytically for an arbitrary distribution of local concentrations $c_i=\langle n_{\mathrm{A}i}\rangle$.[20,23] In Eqs. (33)–(35) subscript p of the chemical potential $\mu_p=\mu_{\mathrm{A}}-\mu_{\mathrm{B}}$ can be omitted: $\Omega_i=T\ln(1-c_i)$, $\mu_i=T\ln[c_i/(1-c_i)]$, and quantities $\tilde{\Omega}_{ij}$ and $\tilde{\mu}_{i}^{ij}$ in Eqs. (32)–(35) are expressed by

$$\tilde{\Omega}_{ij}=-T\ln(1-c_ic_jg_{ij}),\quad \tilde{\mu}_{i}^{ij}=T\ln(1-c_jg_{ij}), \tag{36}$$

where g_{ij} is expressed in terms of Meyer's function $f_{ij}=\exp(-\beta v_{ij})-1$ so that

$$g_{ij}=\frac{2f_{ij}}{R_{ij}+1+f_{ij}(c_i+c_j)},$$

$$R_{ij}=\{[(1+(c_i+c_j)f_{ij}]^2-4c_ic_jf_{ij}(f_{ij}+1)\}^{1/2}. \tag{37}$$

In the case of weak interaction, $\beta v_{ij}\ll 1$, Eqs. (36) transform into the results obtained by the mean-field approximation: $\tilde{\Omega}_{ij}=-v_{ij}c_ic_j$; $\tilde{\mu}_{i}^{ij}=v_{ij}c_j$.

3.2. Yang's tetrahedron approximation.[27] As was noted in Introduction, an adequate description of ordering in fcc alloys should take into account at least correlations among all nearest neighbors, i.e., tetrahedron clusters of these neighbors should be included.[8–15,27] Such an analysis was first performed by Yang.[27] He used the ''quasi-chemical'' method, equivalent to the cluster approach discussed here, considered a model with interactions between nearest neighbors, and his basic clusters were tetrahedrons shown in Fig. 1. These tetrahedrons have only one common point, and each lattice site is included in four tetrahedrons. Note that this

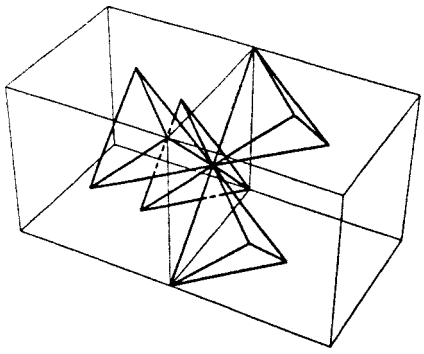


FIG. 1. Decomposition of the fcc lattice into tetrahedron clusters in Yang's quasi-chemical approximation.[27]

choice of basic clusters is different from the configurations of the conventional tetrahedron version of the cluster variation technique to be discussed in Sec. 3.4, where the basic clusters are all eight tetrahedrons including a specific lattice site, so these tetrahedrons have, in addition to one common point, common edges.

With Yang's selection of basic clusters, Eq. (13) or (14) applied to subclusters with $|\alpha|$ or $|\xi|$, which are equated sequentially to 4, 3, 2, and 1, yield the factors $\nu_\alpha=\nu_{|\alpha|}$ and fields ψ^α_ξ (denoted for brevity by $\psi^{|\alpha|}_{|\xi|}$):

$$\nu_4=1, \quad \nu_3=\nu_2=0, \quad \nu_1=-3; \tag{38}$$

$$\psi^4_4=\psi^4_3=\psi^4_2=0, \quad \psi^4_1=\frac{3}{4}\psi^1_1. \tag{39}$$

Thus, interactions within 4-clusters are not renormalized, whereas 3- and 2-clusters do not contribute to observable quantities, thus the cluster variation method in this approximation is again equivalent to the cluster field method. Note that the vanishing of factors ν_3 and ν_2 and corresponding fields ψ^4_3 and ψ^4_2 is the result of the inner subcluster theorem given in Sec. 2.1: Fig. 1 clearly shows that the subclusters formed by a face and an edge of a basic tetrahedron are inner clusters. Note also that Eq. (39) relating fields ψ^4_1 and ψ^1_1 in the 4-cluster and 1-cluster is again consistent with the concept of additivity of field component generated by each ''external'' bond. The number of such bonds for a 4-cluster is 9, and for a 1-cluster (one site) it is 12, which is in agreement with Eq. (39).

3.3. Approximation of Yang's tetrahedrons and pair clusters of non-nearest neighbors. If there are interactions $v_n=v_2, v_3, \ldots, v_m$ with second, third, $\ldots$, mth neighbors, the tetrahedron approximation described in the previous section can be generalized to include these interactions. In the cluster variation technique, this is done by using more complex approximations with larger basic clusters, namely, the tetrahedron–octahedron approximation taking into account v_2, the double tetrahedron-octahedron approximation including v_2 and v_3, and so on.[4,10–13] For most of real alloys and temperatures interesting from the practical viewpoint, however, interactions between non-nearest neighbors are weak and βv_n is much lower than unity.[2] In this connection, the contributions of interactions v_n, where $n\geqslant 2$, are described in terms of the mean-field approximation,[2,3] which is justified if βv_n is small. But if the parameter βv_n is comparable with unity, the pair cluster approximation described in Sec. 3.1 may have a better accuracy than the mean-field approximation.[19–26] At the same time, straightforward equations (32)–(37) make this approximation just as easy, at least in the case of two-component alloys. This approximation, which includes interaction between nearest neighbors in the approximation of Yang's tetrahedrons and interactions v_n for $n\geqslant 2$ in the pair cluster approximation will be dubbed QCTCM. Section 4 will be dedicated largely to investigation of this relatively simple approximation.

Using results of Secs. 3.1 and 3.2, one can write expressions for potentials Ω and μ_p in QCTCM in a form similar to Eqs. (32)–(35):

$$\Omega=\sum_i \Omega_i+\sum_{i<j,n>2}\tilde{\Omega}_{ij,n}+\sum_{\{ijkl\}\in tY}\tilde{\Omega}_{ijkl}, \tag{40}$$

$$\mu_p-\varphi_{pi}=\mu_{pi}+\sum_{j,n>2}\tilde{\mu}^{ij,n}_{pi}+\sum_{\{jkl\}\in tY,i}\tilde{\mu}^{ijkl}_{pi}. \tag{41}$$

Here Ω_i and μ_{pi} are the same as in Eqs. (32)–(34), whereas $\tilde{\Omega}_{ij,n}$ or $\tilde{\mu}^{ij,n}_{pi}$ stands for $\tilde{\Omega}_{ij}$ or $\tilde{\mu}^{ij}_{pi}$ in Eq. (35), but for interaction between nth neighbors. The notation $\{ijkl\}\in tY$ in the last sum of Eq. (40) means that the sum is performed over all Yang's tetrahedrons including sites i,j,k,l of the lattice, and $\{jkl\}\in tY,i$ in Eq. (41) means that summation is performed over four Yang's tetrahedrons that contain site i. The variables $\tilde{\Omega}_{ijkl}$ and $\tilde{\mu}^{ijkl}_{pi}$ for tetrahedron $\{ijkl\}$ are defined similarly to $\tilde{\Omega}_{ij}$ and $\tilde{\mu}^{ij}_{pi}$ for pair clusters in Eq. (35):

$$\tilde{\Omega}_{ijkl}=\Omega_{ijkl}-\Omega_i-\Omega_j-\Omega_k-\Omega_l;$$

$$\tilde{\mu}^{ijkl}_{pi}=\ln(y^{ijkl}_{pi}/y^i_{pi}). \tag{42}$$

3.4. Versions of the cluster variation method. Different approximations in the cluster variation methods have been described in the literature in detail.[4,10–14] In the notation used in this paper, the thermodynamic potentials are expressed by Eqs. (16), (19)–(22), and different versions correspond to different configurations of basic clusters, hence different factors ν_α in Eqs. (16) and (21). Next is a brief discussion of most common versions of the cluster variation method for fcc lattices,[10,13] which will be used in calculations of Sec. 4.

In the tetrahedron approximation, basic clusters are all tetrahedrons including the nearest neighbors. Each triangle of nearest neighbors (which is dubbed minimal) is a part of only one basic cluster, i.e., it is an inner subcluster and, according to the inner subcluster theorem, does not contribute to thermodynamic quantities. Each pair of nearest neighbors (a ''bond'') belongs to two different tetrahedrons, and each site belongs to eight basic clusters and twelve bonds. By solving Eq. (13) with due account of these facts, we obtain for factors $\nu_\alpha=\nu_{|\alpha|}$ in the tetrahedron approximation:

$$\nu_4=1, \quad \nu_3=0, \quad \nu_2=-1, \quad \nu_1=5. \tag{43}$$

In the tetrahedron–octahedron approximation (TOCVM) the basic clusters are all elementary tetrahedrons and octahedrons, such as {1,2,3,4} and {1,3,4,5,6,7} in Fig. 2. All the octahedron subclusters, except minimal triangles, bonds, and

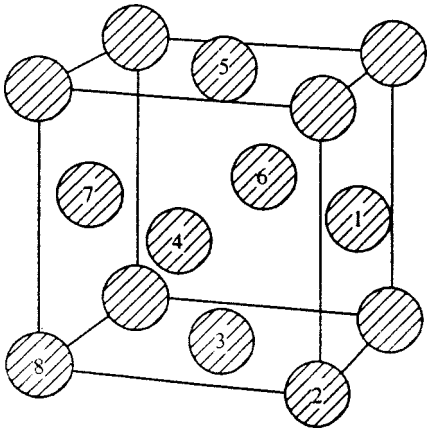


FIG. 2. Sites of the fcc lattice discussed in the paper in connection with different configurations of basic clusters.

single sites, are inner, so their factors ν_α are zero, in accordance with the inner cluster theorem. Each minimal triangle (e.g. {1,2,3} in Fig. 2) belongs to one octahedron and one tetrahedron; each bond (such as {1,3} in Fig. 2) belongs to two octahedrons, two tetrahedrons, and four minimal triangles; each single site belongs to six octahedrons, eight tetrahedrons, twenty-four minimal triangles (three in each tetrahedron) and twelve bonds. Using these relations in solving Eq. (13), we obtain factors $\nu_\alpha = \nu_{|\alpha|}$ in the tetrahedron–octahedron approximation:

$$\nu_6 = \nu_4 = 1, \quad \nu_3 = -1, \quad \nu_2 = 1, \quad \nu_1 = -1. \tag{44}$$

In the double tetrahedron–octahedron approximation, the basic clusters are double tetrahedrons like {1,2,3,4,7,8} and octahedrons like {1,3,4,5,6,7} in Fig. 2, hereafter denoted by dt and o. Using Fig. 2, one can find that the overlapping (i.e., non-inner) subclusters of these basic clusters are tetrahedrons formed of nearest neighbors and denoted by t, tetrahedrons like {2,3,4,8} denoted by $\tilde{t}$, right triangles like {2,4,8} denoted as $\tilde{3}$, minimal triangles, bonds and single sites. Each tetrahedron t belongs to six double tetrahedrons (all of which have one common bond with t), and each tetrahedron $\tilde{t}$ belongs to one octahedron and one double tetrahedron. Triangle $\tilde{3}$ belongs to one octahedron, two double tetrahedron, and two tetrahedrons $\tilde{t}$. Each minimal triangle belongs to one octahedron, six double tetrahedrons, one tetrahedron t, and three tetrahedrons $\tilde{t}$. Each bond belongs to two octahedrons, eleven double tetrahedrons, two tetrahedrons t, ten tetrahedrons $\tilde{t}$, and four minimal triangles, and each single site belongs to six octahedrons, thirty-six double tetrahedrons, eight tetrahedrons t, forty-eight tetrahedrons $\tilde{t}$, twenty-four minimal triangles, and twelve bonds. All these relations and Eq. (13) yield factors ν_α for the double tetrahedron–octahedron approximation:

$$\nu_{dt} = \nu_o = 1, \quad \nu_t = -5, \quad \nu_{\tilde{t}} = -1, \quad \nu_3 = 2,$$
$$\nu_{\tilde{3}} = \nu_2 = 0, \quad \nu_1 = -1. \tag{45}$$

3.5. Approximations employed in the reported work. In our calculations of phase diagrams by the cluster variation method reported in Sec. 4, we have used the tetrahedron–octahedron approximation for ordered phases and the double tetrahedron–octahedron approximation for disordered phases. We selected the more accurate approximation in the latter case because fluctuation effects in the phase transition region (which generate uncertainties largely responsible for calculation errors) are appreciably stronger in the disordered than in ordered phase.[4] Around the points of first-order phase transitions discussed in this paper, fluctuation effects are notably suppressed in the ordered phase owing to the presence of the moderate-size order parameter. Therefore, in order to obtain a similar degree of accuracy in both phases, one should apply a more accurate technique to the disordered phase.

As was noted in Sec. 3.3, the tetrahedron–octahedron approximation is sufficient for taking into account interaction v_2 between second neighbors, and the double tetrahedron–octahedron approximation takes into account interactions v_2 and v_3 between second and third neighbors, but they ignore interactions v_n with higher n. At the same time, taking into account experimental data for v_n,[34] we will consider a model with nonvanishing v_3 and v_4. As was stated in Sec. 3.3, the contributions of interactions v_n beyond large basic clusters will be included in the pair cluster approximation, as in Eqs. (40) and (41), i.e., we will add to expressions like (16) in the cluster variation approximation the components $\tilde{\Omega}_{ij,n}$ or $\tilde{\mu}_{pi}^{ij,n}$ from Eq. (40) or (41) with $n=3$ and 4 for the tetrahedron–octahedron approximation and $n=4$ for the double tetrahedron–octahedron approximation. Thus, unless otherwise stated, we will hereafter define the cluster variation technique in this manner, i.e., the tetrahedron–octahedron approximation for ordered phases, double tetrahedron–octahedron for disordered phases, and, in addition, pair clusters for interactions v_n with $n \geqslant 3$, if such interactions are present and ignored by the basic cluster approximation.

As concerns calculations by the cluster field method described in the next section, we will discuss, in addition to the basic QCTCM approximation treated in Sec. 3.3, an alternative version of this method (QC$\tilde{\text{T}}$CM), which uses basic clusters of the tetrahedron cluster variation method, i.e., each lattice site belongs not to four QCTCM tetrahedrons, but to eight. Since this version relies on the main assumption of the cluster field method, $\psi_\xi^\alpha = 0$ for $|\xi| > 1$, the main self-consistency condition (6) is violated for $|\xi| = 2$, i.e., the mean occupancies g_ξ for $|\xi| = 2$ in the tetrahedrons and pair clusters of nearest neighbors differ. Since the contribution of these clusters to thermodynamic quantities, according to Eq. (43) and unlike Eq. (38), is nonvanishing, there is every reason to suppose that the accuracy of QC$\tilde{\text{T}}$CM should be worse than that of QCTCM. This opinion is supported by calculations of Sec. 4.

4. CALCULATIONS OF PHASE DIAGRAMS WITH $L1_2$ AND $L1_0$ ORDERING BY DIFFERENT CLUSTER APPROXIMATIONS

This section describes results obtained using methods specified in Secs. 2 and 3 in calculating phase diagrams for some models with $L1_2$ or $L1_0$ ordering. In cubic structure

$L1_2$ (which occurs, for example, in alloys like Ni_3Al) minority atoms (Al) occupy predominantly sites of sublattice 1, i.e., vertices of the cell cube shown in Fig. 2, whereas majority atoms (Ni) occupy predominantly sites of sublattice 2, i.e., the centers of cube faces. The mean occupancies c_1 and c_2 of sites (for example, by minority atoms) in these sublattices are related to the mean concentration c of alloy components and order parameter η by the formulas[35]

$$c_1 = c + 3\eta, \quad c_2 = c - \eta, \tag{46}$$

where η may vary between zero and c.

The tetragonal $L1_0$ structure (for example, in TiAl alloys) is formed by alternating [100] lattice planes occupied predominantly by atoms of the first or second sort. In this case, sublattices 1 and 2 are formed by alternating lattice planes, so their occupancies c_1 and c_2 are related to the order parameter η by the formulas

$$c_1 = c + \eta, \quad c_2 = c - \eta, \tag{47}$$

where η, just as in Eq. (46) varies between zero and c.

As was mentioned in Introduction, the main subject of this section is the accuracy of phase diagram calculations based on the simple QCTCM approximation described in Sec. 3.3 in comparison with more complicated calculations based on the cluster variation method. The cluster variation method in this case is the version described in Sec. 3.5, which uses the tetrahedron–octahedron approximation for ordered phases and double tetrahedron–octahedron approximation for disordered phases. Earlier investigations[4,10,11] suggest that the accuracy of this version of the cluster variation method is extremely high, since the difference between calculations by this method and Monte Carlo results were usually within one percent.

As in Refs. 10 and 11, we will analyze only models with interactions v_1 and v_2 between nearest and second-nearest neighbors dubbed for brevity 2-models. In this case, phase transitions to the $L1_2$ or $L1_0$ phase are possible when $v_1 > 0$ and $v_2 < 0$,[11] and if the temperature is replaced by the ''reduced'' temperature $T' = T/v_1$, the phase diagram in the $c - T'$ plane is determined by one parameter $\epsilon = v_2/v_1 < 0$. The available experimental evaluations of ϵ in such alloys[4,25,26,34] yield $-0.3 \lesssim \epsilon \lesssim 0$. For this reason, we have used in our studies of 2-models four values of ϵ equal to $-1/2$, $-1/4$, $-1/8$, and 0. In addition, we have also studied the model including interactions between up to fourth neighbors (the 4-model) at v_n derived from experimental data[34] for the $Ni_{0.927}Al_{0.073}$ alloy at $T = 673$ K. These estimates of interaction constants are $v_1 = 1680$ K, $v_2 = -210$ K, $v_3 = 35$ K, and $v_4 = -207$ K. Note that these values are of the same order of magnitude as both experimental[4,25] and theoretical[2,3] data for various alloys of this kind.

In calculations of temperature $T_{ij}(c)$ at which phases i and j are in equilibrium, we used conventional thermodynamic equations, i.e., equilibrium conditions for chemical potentials μ and potentials Ω of these phases.[11,22] In addition, we also calculated the ordering spinodal $T_{os}(c)$.[1] Lines $T_{os}(c)$ in the $c - T$ plane separate regions of different kinetic mechanisms of phase transformations after quenching of a homogeneous disordered alloy below the ordering temperature, namely, the metastable region of nucleation and nucleus growth for $T > T_{os}(c)$ and the region of exponential growth of concentrational wave amplitudes with time, starting with infinitesimally small fluctuations at $T < T_{os}$.[6,29] The discussed phase transitions to phases $L1_2$ or $L1_0$ have the same ordering spinodal corresponding to ''critical'' concentration waves with the star wave vector $\mathbf{k}_s = (2\pi/a, 0, 0)$, where a is the lattice constant of the fcc structure.[1,35] The shape of this spinodal can be derived from expression (22) for free energy F of an ordered alloy with sublattice occupancies c_1 and c_2 expressed in terms of the mean concentration c and order parameter η using Eqs. (46) and (47). The line $T_{os}(c)$ is determined by the condition of thermodynamic stability against infinitely small η, i.e., the second derivative of F with respect to η should go to zero at $\eta = 0$:

$$[\partial^2 F(T, c, \eta)/\partial \eta^2]_{\eta=0} = 0. \tag{48}$$

In our calculations, we substituted in Eq. (48) the expression for F in the $L1_2$ phase calculated by the QCTCM approximation [Eqs. (22), (40), and (41)]. To the best of our knowledge, previous calculations of $T_{os}(c)$ for ordering transition to phases $L1_2$ and $L1_0$ were based on a very special model with interactions between nearest neighbors.[1] The ordering spinodals given below contain information about important kinetic characteristics of phase transformations based on more realistic alloy models.

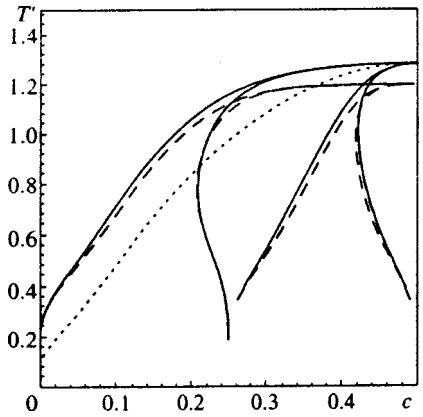


FIG. 3. Phase diagram in the $c - T'$ plane of an fcc binary alloy A_cB_{1-c}, where $T' = T/v_1$ is the reduced temperature, for the following constants of cluster interactions: $v_{n\geqslant 3} = 0$, $\epsilon = v_2/v_1 = -0.5$. The solid lines are curves of equilibrium between phases, $T_{ij}(c)$, calculated by QCTCM as described in Sec. 3.3. The dashed lines are $T_{ij}(c)$ curves calculated by the cluster variation method described in Sec. 3.5. The regions separated by lines $T_{ij}(c)$ correspond to the following phases (from left to right): disordered A_1 phase; coexistence of A_1 and $L1_2$ phases; $L1_2$ phase; coexistence of $L1_2$ and $L1_0$ phases; $L1_0$ phase. The dotted line is the ordering spinodal $T_{os}(c)$ calculated by QCTCM.

The calculation results are plotted in Figs. 3–7. In the models with pair interactions independent of the concentrations, phase diagrams are symmetrical with respect to the line of mean concentration $c = 1/2$, therefore the graphs show only the region $c \leqslant 1/2$. Let us discuss the calculations plotted in the graphs.

First of all, it is clear that the calculations by all models, except the model taking into account only interactions between nearest neighbors (Fig. 7) discussed below, in the

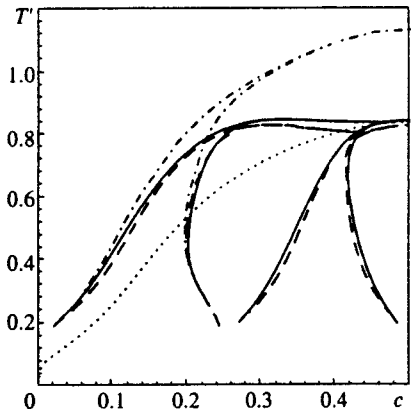


FIG. 4. The same as Fig. 3, but for a model with $\epsilon=-0.25$. The solid, dashed, and dotted lines have the same meaning as in Fig. 3. The dash-dotted lines show $T_{ij}(c)$ curves calculated in the QC$\tilde{\text{T}}$CM approximation described at the end of Sec. 3.5.

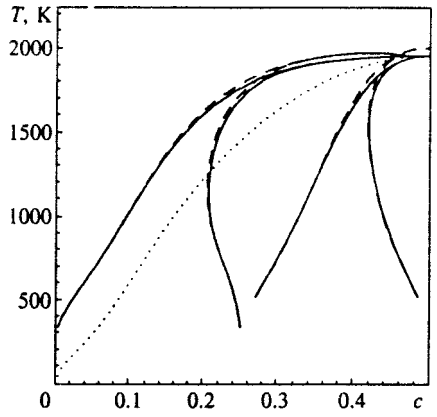


FIG. 6. Phase diagram in the $c-T$ plane for the 4-model described in the text with v_n derived from experimental data.[34] The lines have the same meaning as in Fig. 3.

QCTCM and cluster variation approximations are very close, whereas the latter, as was noted above, should be fairly close to exact results. Thus, in all these models, QCTCM quite adequately describes thermodynamics of phase transformations. This is our main result, which allows us to suppose that QCTCM could be perfectly adequate for solving more complex problems of phase transition kinetics mentioned in Introduction.

At the same time, Fig. 4 demonstrates that the result of QC$\tilde{\text{T}}$CM, which was described in Sec. 3.5, at notable concentrations $c\gtrsim 0.2$ differ considerable from both cluster variation and QCTCM calculations. This indicates that violation of the self-consistency condition (6) by QC$\tilde{\text{T}}$CM, i.e., the condition of equal occupancies in different clusters contributing to thermodynamic quantities, may notably degrade the calculation accuracy. Therefore, one cause of the good accuracy of QCTCM calculations may be the absence of such a self-consistency violation in this technique.

Comparison between graphs in Figs. 5 and 6 illustrates the effect of interactions between non-nearest neighbors on the phase diagram shape. The ratio $\epsilon=v_2/v_1$ in the 4-model in Fig. 6 is very close to $\epsilon=-1/8$ in calculations by the 2-model in Fig. 5. The presence of additional interactions in the 4-model, however, primarily $v_4\simeq v_2$, leads to notably higher temperatures of the phase transition, $T_c(c)/v_1$, and spinodal, $T_{os}(c)/v_1$, as compared to the calculations by the 2-model shown in Fig. 5.

The spinodal temperatures $T_{os}(c)$ in Figs. 3–6 are usually very close to transition points $T_c(c)$, remarkably, not only in the region of ordered phase $L1_2$, but also in the region of $A1$ and $L1_2$ coexistence. Therefore, kinetic effects related to thermodynamic instability in a quenched disordered alloy in the region $T<T_{os}(c)$ and $B2$ ordering discussed in earlier publications[29] can also show up in broad ranges of temperature and concentration. Note also that in 2-models (Figs. 3–5) the ratios $T_{os}(c)/T_c(c)$ drop with increasing $|\epsilon|=|v_2/v_1|$. So, for $\epsilon=-0.125$ (and in QCTCM models with[1,4] $\epsilon=0$) our calculations of $T_{os}(c)/T_c(c)$ in the $L1_2$ phase and near the peak of $T_c(c)$ are close to 0.75. The cluster variation calculations based on the tetrahedron–octahedron scheme with[4] $\epsilon=0$ yield notably higher $T_{os}(c)/T_c(c)\simeq 0.96$. Therefore, our simple calculations with small ϵ by the 2-model probably underestimate the spinodal

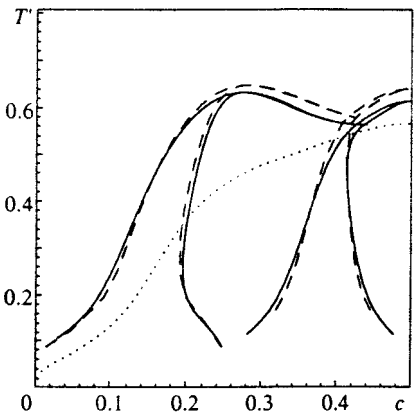


FIG. 5. The same as Fig. 3 for a model with $\epsilon=-0.125$.

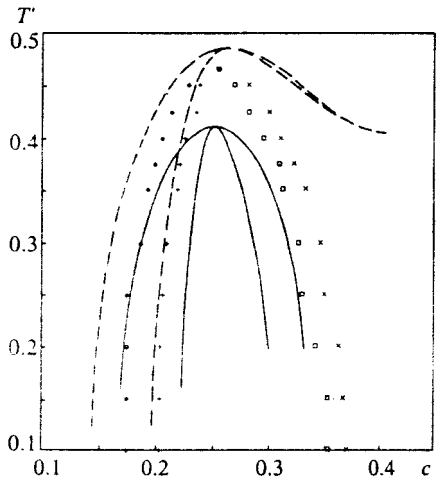


FIG. 7. Phase diagram in the $c-T'$ plane for a model of an fcc alloy with pair interaction only between nearest neighbors describing equilibrium between $A1$ and $L1_2$ phases. The solid lines show $T_{ij}(c)$ calculated by QCTCM, the dashed lines show $T_{ij}(c)$ calculated by TCVM, squares and crosses plot Monte Carlo calculations.[11]

temperature $T_{os}(c)$, which might be caused by fluctuation effects in the phase transition region discussed below.

The model with interaction between only nearest neighbors ($\epsilon=0$) was studied by many researchers.[4,10,11,28] For this reason, Fig. 7 shows only a few calculations of equilibrium lines between the $A1$ and $L1_2$ phases by this model,[11,28] and these data are given as an illustration. The feature of this model is degeneracy (in terms of energy) of several phases with different symmetry properties, in particular, $L1_2$ and DO_{22}, or $L1_0$ and A_2B_2,[11,28] as well as many different configurations with short-range order. This leads to a considerable increase in various fluctuation effects near the phase transitions, hence a poorer accuracy of such simple approximations as QCTCM or TCVM. This, apparently, results in notable differences between results of these methods and Monte Carlo calculations in Fig. 7, as well as similar errors in spinodal calculations discussed by Finel.[4] Nonetheless, it follows from both theoretical calculations[2,3] and experimental data[4,25] that the model taking into account only interactions between nearest neighbors is, seemingly, inadequate for all real alloys that have been studied so far. Therefore the relatively low accuracy of QCTCM in combination with this specific model does not necessarily mean that this method cannot be applied to more realistic alloy models.

5. CONCLUSIONS

In conclusion, let us summarize the main results of this work. A general formulation of cluster methods for calculating thermodynamic quantities in alloys with arbitrary numbers of components and inhomogeneous distributions of atoms is given. The basic variables are effective fields and interactions within a cluster due to its interaction with the environment. This approach seems more simple and wide-ranging than previous versions, can be generalized to various cases, and accommodates as special cases the well-known cluster variation method and the cluster field method described earlier. A simple version of the cluster field method (QCTCM) based on selection of Yang's tetrahedron clusters in an fcc lattice (Fig. 1) and inclusion of interactions with non-nearest neighbors in the pair cluster approximation has been suggested. We have calculated phase diagrams using several models of fcc alloys with $L1_2$ and $L1_0$ ordering using both QCTCM and a more complicated version of the cluster variation method, whose accuracy, in accordance with earlier investigations, should be extremely high. With all realistic interaction models, phase diagrams obtained by the simple QCTCM method are very close to those calculated by the cluster variation technique. This indicates that QCTCM can be applied to topical problems of phase transition kinetics in alloys under discussion, for which the simple mean-field approximation is insufficient, whereas the cluster variation technique is too cumbersome. We have also calculated ordering spinodals $T_{os}(c)$ for these alloys. Our results indicate, in particular, that specific kinetic processes, similar to those associated with simpler phase transitions and discussed previously,[29] can occur when these alloys are quenched into the region $T<T_{os}(c)$. All these results may be used as a base for further development of theoretical models of inhomogeneous states and ordering phase transition kinetics in fcc alloys.

We are grateful to Dr. G. Martin for his interest in this work and hospitality shown to one of the authors (G. D. S.) at Center d'Etude de Saclay, where this investigation began, and to V. V. Kamyshenko for help and useful discussions. This work was supported by the Russian Fund for Fundamental Research (Grant No. 97-02-17842).

*)E-mail:vaks@mbslab.kiae.ru
†)E-mail:G.D.Samolyuk@mbslab.kiae.ru

[1] D. de Fontaine, Solid State Phys. **34**, 74 (1979).
[2] P. E. A. Turchi, in *Intermetallic Compounds*, Vol. 1, *Principles*, J. H. Westbrook and R. L. Fleischer (Eds.), Wiley, New York (1994), p. 21.
[3] A. Zunger, in *Statics and Dynamics of Alloy Phase Transformations*, Vol. 319 of NATO Advanced Study Institute, Series B: Physics, A. Gonis and P. E. A. Turchi (Eds.), Plenum, New York (1994), p. 361.
[4] A. Finel, in *Statics and Dynamics of Alloy Phase Transformations*, Vol. 319 of NATO Advanced Study Institute, Series B: Physics, A. Gonis and P. E. A. Turchi (Eds.), Plenum, New York (1994), p. 495-540.
[5] K. Binder, in *Statics and Dynamics of Alloy Phase Transformations*, Vol. 319 of NATO Advanced Study Institute, Series B: Physics, A. Gonis and P. E. A. Turchi (Eds.), Plenum, New York (1994), p. 467-493.
[6] L. Q. Chen, Y. Z. Wang, and A. G. Khachaturyan, in *Statics and Dynamics of Alloy Phase Transformations*, Vol. 319 of NATO Advanced Study Institute, Series B: Physics, A. Gonis and P. E. A. Turchi (Eds.), Plenum, New York (1994), p. 587-604.
[7] K. D. Belashchenko, V. Yu. Dobretsov, and V. G. Vaks, in *Properties of Complex Inorganic Solids*, Proc. of 1st Intern. Alloy Conf., A. Gonis, A. Meike, and P. E. A. Turchi (Eds.), Plenum, New York (1997), p. 101.
[8] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[9] T. Morita, J. Math. Phys. **13**, 115 (1972).
[10] J. M. Sanchez and D. de Fontaine, Phys. Rev. B **17**, 2926 (1978).
[11] T. Mohri, J. M. Sanchez, and D. de Fontaine, Acta Metall. **33**, 1171 (1985).
[12] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A **128**, 334 (1984).
[13] A. Finel, *Thèse de Doctorat d'Etat*, Université Paris VI, Note Technique ONERA 1987-3 (1987).
[14] F. Ducastelle, in *Cohesion and Structures*, Vol. 3, F. R. de Boer and D. G. Pettifor (Eds.), North Holland, Amsterdam (1991), Ch. 4.
[15] V. G. Vaks and V. V. Kamyshenko, Izv. Akad. Nauk SSSR, Ser. Metally No. 2, 121 (1990).
[16] B. H. Kear, Sci. Am. , No. 12, 99 (1986).
[17] V. G. Vaks, JETP Lett. **63**, 471 (1996).
[18] K. D. Belashchenko and V. G. Vaks, J. Phys. F **10**, 1965 (1998).
[19] V. G. Vaks, N. E. Zeĭn, V. I. Zinenko, and V. G. Orlov, Zh. Éksp. Teor. Fiz. **87**, 2030 (1984) [Sov. Phys. JETP **60**, 1171 (1984)].
[20] V. G. Vaks and V. G. Orlov, Fiz. Tverd. Tela **28**, 3627 (1986) [Sov. Phys. Solid State **28**, 2045 (1986)].
[21] V. G. Vaks and V. G. Orlov, J. Phys. F **18**, 883 (1988).
[22] V. G. Vaks and V. I. Zinenko, J. Phys.: Condens. Matter **1**, 9085 (1989); **3**, 4533 (1991).
[23] V. G. Vaks, N. E. Zein, and V. V. Kamyshenko, J. Phys. F **18**, 1641 (1988).
[24] V. G. Vaks, N. E. Zeĭn, V. V. Kamyshenko, and Yu. V. Tkachenko, Fiz. Tverd. Tela **30**, 477 (1988) [Sov. Phys. Solid State **30**, 270 (1988)].
[25] V. G. Vaks, N. E. Zein, and V. V. Kamyshenko, J. Phys.: Condens. Matter **1**, 2115 (1989).
[26] V. G. Vaks and V. V. Kamyshenko, J. Phys.: Condens. Matter **3**, 1351 (1991).
[27] C. N. Yang, J. Chem. Phys. **13**, 66 (1945).
[28] Y. Y. Li, J. Chem. Phys. **17**, 447 (1949).
[29] V. Yu. Dobretsov, V. G. Vaks, and G. Martin, Phys. Rev. B **54**, 3227 (1996).
[30] K. D. Belashchenko and V. G. Vaks, Zh. Éksp. Teor. Fiz. **112**, 714 (1997) [JETP **85**, 390 (1997)].

[31] J. Dennis and R. Schnabel, *Numerical Methods of Unconditional Optimization and Solution of Nonlinear Equations* [Russian translation], Mir, Moscow (1988).
[32] W. H. Press, S. A. Teukolsky, W. T. Vettering *et al.*, *Numerical Recipes in C*, Camb. Univ. Press (1996), Ch. 10.
[33] A. Finel, V. Mazauric, and F. Ducastelle, Phys. Rev. Lett. **65**, 1016 (1990).
[34] F. Chassagne, M. Bessiere, Y. Calvayrac *et al.*, Acta Metall. **37**, 2329 (1989).
[35] A. G. Khachaturyan, *Theory of Phase Transformations and Structure of Solid Solutions* [in Russian], Nauka, Moscow (1974).

Translation provided by the Russian Editorial office