Effect of scattering media on decays of high-energy particles

A. V. Koshelkin*)

Moscow Institute of Engineering Physics, 115409 Moscow, Russia
(Submitted 14 September 1998)
Zh. Éksp. Teor. Fiz. **116**, 3–10 (July 1999)

The effect of multiple scattering on the decay of high-energy particles has been studied. The self-consistent method for the calculation of decay rates of particles undergoing multiple elastic collisions in an equilibrium medium has been developed. Influence of multiple scattering on the decay rate of a neutral pion in a hadron gas has been studied. © *1999 American Institute of Physics.* [S1063-7761(99)00107-9]

1. INTRODUCTION

The study of nuclear matter generated in the collisions of ultrarelativistic heavy ions makes it necessary to determine the effect of scattering medium on the decay of high-energy particles .[1–7] The photons and various leptons (such as electrons and neutrinos) produced in such matter carry information about the state of the studied medium. One of the main channels in which such particles are generated is the decay in nuclear matter.

The influence of the medium on the particle decay width has been investigated extensively.[1,3,7–11] In those studies the polarization effects in QCD[3,7,8] and QED were examined.[8–11] In investigating decays in a dense nuclear matter, however, one should take into account the polarization effects mentioned above and the multiple scattering of particles in the matter. If the temperature T of the medium is sufficiently low so that polarization effects can be ignored, and if the fluctuations of particle energy due to multiple scattering are on the order of T, then the influence of the medium on the decay rate is largely determined by multiple collisions of a decaying particle.

In this paper we report the results of an experimental study of the decay of elastically multiply scattered particles in an equilibrium medium. The decay rate for such particles has been calculated. If the decaying particle is ultrarelativistic, multiple scattering leads to the strong broadening of its state under certain conditions. On the basis of the developed method for the calculation of decay rates for particles in the media, the decay of neutral pion in an equilibrium hadron gas has been studied. It turned out that under certain conditions multiple scattering results in the considerable increase of the decay rate in the channels $\pi^0 \to 2\gamma$ and $\pi^0 \to \gamma e^+ e^-$.

2. DECAY RATE OF A PARTICLE IN A SCATTERING MEDIUM

Consider a particle of mass M with spin s that undergoes multiple elastic pair collisions in an equilibrium medium. In the quasi-classical approximation, when the particle wavelength is much smaller than its free path, the width $\gamma(x,p)$ of the particle state is given[12] by the standard formula

$$\gamma(x,p) = -2\,\mathrm{Im}\{\Sigma^{\mathrm{ret}}(x,p)\}, \tag{1}$$

where Σ^{ret} is the retarded self-energy in the formalism of the Keldysh graphic technique,[13] $x=(t,\mathbf{r})$ is the 4-coordinate, and $p=(p^0,\mathbf{p})$ is the 4-momentum.

In order to calculate the observed width Γ of the particle decay in the matter, one should sum and average Eq. (1) over the initial and final states of particles and then average it over the observation times τ:

$$\Gamma = \mp\frac{2}{\tau}\mathrm{Im}\left\{\int_0^\tau dt \int d^3r \int \frac{d^4p}{(2\pi)^4} \times \mathrm{Tr}\{\Sigma^{\mathrm{ret}}(x,p)G^{-+}(x,p)\}\right\}, \tag{2}$$

where $G^{-+}(x,p)$ is the Green's function of the decaying particle in the Keldysh graphic technique,[13] the plus and minus signs refer to the Bose and Fermi statistics, respectively, and the normalization volume $V=1$.

Suppose that the effect of the particle scattering on the decay vertex is weak, and the interaction among decay products in the final state is negligible.[1)] Then the retarded self-energy is identical to the self-energy corresponding to the particle decay in vacuum:

$$(\Sigma^{\mathrm{ret}})^{\alpha\beta} = \Sigma^{\alpha\beta}_{\mathrm{vac}}(p), \tag{3}$$

where α and β are spin variables.

By substituting the latter expression into Eq. (2) and using the equation[13,14] relating the Green's function $G^{-+}(x,p)$ to the distribution function $n(t,\mathbf{r},\mathbf{p})$ of particles in the equilibrium medium, we obtain for the width of the decaying particle on the mass surface

$$\Gamma = \frac{1}{\tau}\mathrm{Im}\left\{\int_0^\tau dt \int \frac{d^3p}{E}\mathrm{Tr}[\Sigma^{\alpha\beta}_{\mathrm{vac}}(p)\varrho_{\alpha\beta}] \times \int \frac{d^3r}{(2\pi)^3} n(t,\mathbf{r},\mathbf{p})\right\}, \tag{4}$$

were $\varrho_{\alpha,\beta}$ is the polarization density operator of the decaying particle, $p^0 \equiv E = \sqrt{\mathbf{p}^2 + M^2}$ is the particle energy, and $n(t,\mathbf{r},\mathbf{p})$ is the distribution function of particles that satisfies the Boltzmann kinetic equation.[13,15]

Note that the expression in the brackets in Eq. (4) is a scalar; therefore, it depends only on the masses of the decay-

ing particles and decay products. Moreover, the term $\mathrm{Tr}[\Sigma^{\alpha\beta}_{\mathrm{vac}}(p)\varrho_{\alpha\beta}]$ is proportional to the width of the decaying particle in vacuum. These circumstances allow us to rewrite Eq. (4) in the form

$$\Gamma=\frac{M\Gamma_{\mathrm{vac}}}{\tau}\int_0^{\tau}dt\int_0^{\infty}p^2dp\,\frac{F(E,t)}{E},\tag{5}$$

where Γ_{vac} and the function $F(E,t)$ are given by the formulas

$$\Gamma_{\mathrm{vac}}(p)=\frac{1}{M}\mathrm{Im}\{\mathrm{Tr}\{\Sigma^{\alpha\beta}_{\mathrm{vac}}(p)\varrho_{\alpha\beta}\}\},\tag{6}$$

$$F(E,t)=\int d\Omega_{\mathbf{p}}\int\frac{d^3r}{(2\pi)^3}n(t,\mathbf{r},\mathbf{p}).\tag{7}$$

Thus, the calculation of the width of decaying particle undergoing multiple elastic collisions in an equilibrium scattering medium reduces to the parameter E^{-1} averaging over certain distribution which has to be calculated. Note that $\langle E^{-1}\rangle\simeq M^{-1}$ in the case of decay of nonrelativistic particles, and the effect of scattering on decays of such particles is negligible.

Suppose that the relaxation time in the medium $\tau_{\mathrm{rel}}\ll\min\{\tau,\Gamma^{-1}\}$ is much smaller[2)] than both the observation time τ and reciprocal decay width Γ^{-1}. Then we describe the states of the decaying particle in the medium in terms of the static approximation,[17,18] in which the medium return to its initial state almost immediately after each individual collision. Since the particle distribution function is included in the formula for the decay width in the integral with respect to coordinates and momenta of the particle, we obtain the following expression for the function $F(E;t)$ after integrating the Boltzmann equation over the solid angle in the direction of $\mathbf{p}$ and over $\mathbf{r}$:

$$\frac{\partial F(E;t)}{\partial t}=\int_{E-E_{\max}}^{E+E_{\max}}dE'\int d^3p_1\times\int d\Omega_{\mathbf{p}}\frac{d\sigma(E-E';E;p_1)}{dE'd\Omega_{\mathbf{p}}}vf_{\mathrm{eq}}(p_1)\times\{F(E';t)-F(E;t)\},\tag{8}$$

where $d\sigma(E-E';E,p_1)$ is the cross section of an individual pair scattering event in the medium, $f_{\mathrm{eq}}(p_1)$ is the equilibrium distribution function of particles in the medium, $E_{\max}$ is the maximal energy transferred in an individual collision between two particles in the matter, E and E_1 are energies of the colliding particles, $v=\sqrt{s(s-4M^2)}/2EE_1$, $p=(E,\mathbf{p})$, $p_1=(E_1,\mathbf{p}_1)$, $p'=(E',\mathbf{p}')$, and $s=(p+p_1)^2$.

In the case where the matter is in thermodynamic equilibrium and all collisions between particles are elastic, the energy $|E-E'|$ imparted by a particle in one individual collision is smaller than the medium temperature T. Expanding the scattering integral on the right side of Eq. (8) in the small parameter $|E-E'|/E\ll1$, we then obtain

$$\frac{\partial F(E;t)}{\partial t}=\langle\varepsilon^2\rangle\frac{\partial^2F(E;t)}{\partial E^2},\tag{9}$$

where $\langle\varepsilon^2\rangle$ is the average of the energy transferred per unit time. In deriving the latter equation, we have taken into account the fact that the scattering cross section is the even function of the transferred energy. The parameter $\langle\varepsilon^2\rangle$ is determined by the equation

$$\langle\varepsilon^2\rangle=\int d^3p_1d\Omega_{\mathbf{p}}\int_0^{E_{\max}}dE'\,\frac{d\sigma(E';E,p_1)}{dE'd\Omega_{\mathbf{p}}}vf_{\mathrm{eq}}(p_1)E'^2.\tag{10}$$

This parameter is the function of the instantaneous particle energy. Nonetheless, since the medium is in thermodynamic equilibrium and since only elastic collisions occur, the particle energy is constant, on the average, and it is affected by fluctuations whose amplitude is the function of time. Therefore, let us assume in our approximation that $\langle\varepsilon^2\rangle$ is a function of time, but not of the particle energy. Solving Eq. (9) with the initial condition

$$F(E;t=0)=f_0(E)\,\eta(E-M),\tag{11}$$

we obtain

$$F(E;t)=\frac{\eta(E-M)}{\sqrt{4\pi\int_0^t dt'\langle\varepsilon^2\rangle(t')}}\int_M^{\infty}dE'\times\left\{\exp\left[-\frac{(E-E')^2}{4\int_0^t dt'\langle\varepsilon^2\rangle(t')}\right]+\exp\left[-\frac{(E+E'-2M)^2}{4\int_0^t dt'\langle\varepsilon^2\rangle(t')}\right]\right\}f_0(E'),\tag{12}$$

where $f_0(E)$ is the distribution function which determines the state of the decaying particle at the initial moment, and $\eta(E)$ is the unit step function.[19]

Substituting the calculated function $F(E;t)$ in Eq. (5), we obtain the following expressions for the decay rate W in the scattering medium:

$$W=\Gamma\tau=M\Gamma_{\mathrm{vac}}\int_0^{\tau}\frac{dt}{\sqrt{4\pi\int_0^t dt'\langle\varepsilon^2\rangle(t')}}\int_0^{\infty}\frac{p^2dp}{E(p)}\times\int_M^{\infty}dE'\left\{\exp\left[-\frac{(E-E')^2}{4\int_0^t dt'\langle\varepsilon^2\rangle(t')}\right]+\exp\left[-\frac{(E+E'-2M)^2}{4\int_0^t dt'\langle\varepsilon^2\rangle(t')}\right]\right\}f_0(E').\tag{13}$$

Note that in the case of small occupation numbers the density of particle states in an equilibrium medium has a fairly sharp peak at energies $E\sim T$, where T is the temperature of the medium. Therefore the width of the peak in the integrand in Eq. (12) is determined by the relation between the particle energy fluctuation due to multiple scattering,

$$\int_0^t dt' \langle \varepsilon^2 \rangle (t'),$$

and the temperature T of the medium. If the energy fluctuations are such that

$$\int_0^t dt' \langle \varepsilon^2 \rangle (t') \gg T,$$

the multiple scattering is strong and leads to a considerable broadening of the particle state in the medium, whereas in the opposite limiting case,

$$\int_0^t dt' \langle \varepsilon^2 \rangle (t') \ll T,$$

the effect of the scattering medium is negligible.

Further, let us discuss the application of this method to the calculations of decay rates of specific particles in scattering media.

3. DECAY OF A NEUTRAL PION IN AN EQUILIBRIUM SCATTERING MEDIUM

Consider the effect of elastic multiple scattering on decays of neutral pions, $\pi^0 \to 2\gamma$ and $\pi^0 \to e^+ e^- \gamma$, in an equilibrium pion gas. Investigation of such decays, which are among sources of thermal gamma rays and lepton pairs, are of interest for studies of nuclear media generated in collisions of high-energy heavy ions.[2,17,20] Since the characteristic temperatures of equilibrium pion gases are low ($T \leqslant 200$ MeV) (Refs. 17 and 20) in comparison with the excitation energies of both quark ($\sim T_c \geqslant 250$ MeV) and hadron degrees of freedom ($\sim m_\varrho = 770$ MeV), the effect of scattering on the decay vertex and polarization effects can be ignored.

The calculation of widths of decays $\pi^0 \to 2\gamma$ and $\pi^0 \to e^+ e^- \gamma$ in the approximation of partially conserved axial current[21] gives

$$\frac{d\Gamma_{\rm vac}}{dM_l} = \frac{4e^2}{3\pi M_l^4 M^6} (M^2 - M_l^2)^3 (M_l^2 + 2m^2) \times (M_l^2 - 4m^2)^{1/2} \Gamma_{\pi^0 \to 2\gamma},$$

$$\Gamma_{\pi^0 \to 2\gamma} = 8 \ \text{eV}, \tag{14}$$

where M and m are the pion and electron masses, respectively; $M_l = |p_+ + p_-|$ is the invariant mass of a lepton pair.

In the case where the pion gas temperature $T \sim 200$ MeV (Refs. 2, 17, and 20) we treat the particles of the medium as ultrarelativistic. Then the energy transferred in one collision is smaller than the gas temperature: $E_{\max} \sim T(M/T)^2 \ll T$, and the scattering cross section $d\sigma$ [see Eq. (10)] is the slow function[2,20] of the energy of colliding pions, so $\langle \varepsilon(t) \rangle$ is approximately constant.

Substituting the decay widths in vacuum given by Eq. (14) in the equation for the particle decay rate in the scattering medium and performing all necessary integrations, we obtain

$$W_{\pi^0 \to 2\gamma} = \frac{M}{\pi^{1/2} T} \Gamma_{\pi^0 \to 2\gamma} \left\{ \frac{2\langle \varepsilon^2 \rangle^{1/2} \tau^{3/2}}{3T} + \frac{T^2}{\langle \varepsilon^2 \rangle} \times \left[\exp\left(\frac{\langle \varepsilon^2 \rangle \tau}{T^2} \right) \int_{\sqrt{\langle \varepsilon^2 \rangle \tau}/T}^{\infty} dx \times \exp(-x^2) - \frac{\sqrt{\pi}}{2} + \frac{\sqrt{\langle \varepsilon^2 \rangle \tau}}{T} \right] \right\}, \tag{15}$$

$$\frac{dW}{dM_l} = \frac{4e^2}{3\pi M_l^4 M^6} (M^2 - M_l^2)^3 (M_l^2 + 2m^2) \times (M_l^2 - 4m^2)^{1/2} W_{\pi^0 \to 2\gamma}, \tag{16}$$

where T is the temperature of the medium, and τ is the observation time. In deriving Eqs. (15) and (16) we assumed that the initial state of the decaying pion is described by Boltzmann's distribution function $f_0(E)$ with temperature T, and the pions are ultrarelativistic.

Let us analyze the resulting expression for the pion decay rate in the medium in the main channel $\pi^0 \to 2\gamma$ (the decay channel $\pi^0 \to e^+ e^- \gamma$ is analyzed similarly).

If the fluctuations of particle energy due to multiple scattering are smaller than the temperature of the medium, then $\sqrt{\langle \varepsilon^2 \rangle \tau}/T \ll 1$. Integrating for small values of the lower integration limit in Eq. (15), we obtain

$$W_{\pi^0 \to 2\gamma} = \frac{M\tau}{2T} \Gamma_{\pi^0 \to 2\gamma} \left\{ 1 + \frac{\langle \varepsilon^2 \rangle \tau}{2T^2} + O\left(\left[\frac{\langle \varepsilon^2 \rangle \tau}{T^2} \right]^{3/2} \right) \right\}. \tag{17}$$

It follows from Eq. (17) that multiple scattering of particles in the medium leads to insignificant increase of the decay rate in comparison with the situation of the $\pi^0 \to 2\gamma$ decay in vacuum.

In the opposite limit case of relatively large fluctuations ($\sqrt{\langle \varepsilon^2 \rangle \tau}/T \geqslant 1$), we derive the following expression from Eq. (15):

$$W_{\pi^0 \to 2\gamma} = \frac{M\tau}{\pi^{1/2} T} \Gamma_{\pi^0 \to 2\gamma} \left\{ \frac{2\sqrt{\langle \varepsilon^2 \rangle \tau}}{3T} + O\left(\frac{T}{\sqrt{\langle \varepsilon^2 \rangle \tau}} \right) \right\}. \tag{18}$$

The latter equation indicates that the large fluctuations of energy lead to the strong increase of the decay rate (by the factor $\sqrt{\langle \varepsilon^2 \rangle \tau}/T \gg 1$) due to multiple elastic scattering of pions in the matter.

In a more realistic situation where the pion gas temperature $T = 130$–140 MeV, the decay rate should be calculated numerically using Eqs. (13) and (14). The estimation of $\langle \varepsilon^2 \rangle$ as the product of the frequency of pion collisions[17,20] and the maximum energy transferred in one collision between particles gives $\langle \varepsilon^2 \rangle \simeq \text{const} = 1.1 \times 10^{-4}$ GeV^2/Fm. Performing integration in Eqs. (13) and (14), we obtain

$$\frac{W(\alpha)}{W(\alpha=0)}=\begin{cases}1.25, & \text{if } T=140 \text{ MeV};\\ 1.23, & \text{if } T=135 \text{ MeV};\\ 1.21, & \text{if } T=130 \text{ MeV},\end{cases} \tag{19}$$

where $W(\alpha=0)$ is the decay rate in vacuum.

In conclusion, let us discuss the agreement between the widths of neutral pion states in the scattering medium calculated above and the experimental data on collision of heavy ions S–Au and Pb–Au at energies of 200 GeV/nucleon.[22,23] These data can be compared by estimating the relative yield of electron–positron pairs of small energies, whose main sources are the reactions $\pi^0\to e^+e^-\gamma$ and $\eta\to e^+e^-\gamma$. Note that in the real situation the characteristic temperature T of the hadron medium is within 150 MeV, and the η-meson mass is 547 MeV; therefore, the broadening of its state due to multiple elastic scattering is negligible. Then the ratio between the yields dN_{π^0} of lepton pairs due to the decay $\pi^0\to e^+e^-\gamma$ and dN_η of lepton pairs generated in the reaction $\eta\to e^+e^-\gamma$ provides information about the broadening of pion states in the matter. The experimental data derived from measurements[22,23] give $dN_\eta/dN_{\pi^0}\simeq 0.27$. Calculations give $dN_\eta/dN_{\pi^0}\sim 0.37$ in the absence of scattering effects[17] and $dN_\eta/dN_{\pi^0}\sim 0.30$ with allowance for the pion state broadening due to multiple scattering in the medium with temperature $T=130$ MeV. Thus, the agreement between the calculations and experimental data on the neutral pion decay in the hadron gas is satisfactory.

4. CONCLUSIONS

In the reported work we investigated the effect of multiple scattering on particle decay in an equilibrium medium. The method of calculation of decay rates of particles under such conditions has been developed. The calculated decay rate depends strongly on the temperature of the medium and on the parameters characterizing particle scattering in the medium. We found that in the dense, hot medium in the equilibrium state, when decaying particles are ultrarelativistic and their energies are comparable with the temperature of the medium, multiple elastic scattering leads to the considerable increase in the decay rate of particles in the medium, whereas in the case of low-energy particles the effect of scattering medium is negligible.

The influence of multiple scattering on the decay rate of a neutral pion in an equilibrium pion gas has been studied. The broadening of states of this particle in the experimentally observable range of temperatures has been calculated. A comparison between calculations of the yield of electron–positron pairs in collisions of high-energy heavy ions and experimental data indicates that the agreement between these results is satisfactory.

I wish to thank D. N. Voskresenskii for the discussions.

*)E-mail: koshelkn@gpd.mephi.msk.su

1)The application limits of the model under discussion is determined by a specific situation in which a decay takes place. In particular, when the neutral pion decays in an equilibrium hadron gas (see below), this approximation applies when the medium temperature is lower than the temperature of the phase transition between the hadron gas and quark–gluon plasma, $T_c\geqslant 250$ MeV.

2)The validity of this inequality depends on the dynamics of the nuclear medium generated in collisions of high-energy heavy ions, which expands isentropically,[16] being in a quasi-equilibrium state.

[1] J. Kapusta, P. Lichard, and D. Seibert, Phys. Rev. D **44**, 2774 (1991).
[2] J. Cleymans, V. V. Goloviznin, and K. Redlich, Phys. Rev. D **47**, 173 (1993).
[3] R. Pisarski, Phys. Rev. D **47**, 5589 (1993).
[4] S. K. Ghosh, S. C. Phatak, and P. K. Sahn, Nucl. Phys. A **596**, 670 (1996).
[5] U. Faschingbauer, P. Rehak, J. P. Wurm *et al.*, Nucl. Instrum. Meth. A **377**, 362 (1996).
[6] P. Holl, P. Rehak, F. Cerreto *et al.*, Nucl. Instrum. Methods Phys. Res. A **377**, 367 (1996).
[7] R. Pisarski and M. Tyrgat, Phys. Rev. D **54**, 2989 (1996).
[8] A. B. Migdal, D. N. Voskresenskiĭ, É. E. Sapershteĭn, and M. A. Troitskiĭ, *Pion Degrees of Freedom in Nuclear Matter* [in Russian], Nauka, Moscow (1991).
[9] G. Peressuti and B.-S. Skagerstam, Phys. Lett. **117**, 171 (1982).
[10] R. Baier and R. Kobes, Phys. Rev. D **50**, 5944 (1994).
[11] I. E. Nieves and P. B. Pal, Phys. Rev. D **56**, 365 (1997).
[12] J. Knoll and D. N. Voskresensky, Ann. Phys. **249**, 532 (1996).
[13] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, Pergamon Press, New York–Oxford (1982).
[14] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957); C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).
[15] S. De Groot, W. Van Leeuwen, and Ch. Van Weert, *Relativistic Kinetic Theory*, North Holland, Amsterdam (1980).
[16] J. D. Bjorken, Phys. Rev. D **27**, 140 (1983).
[17] J. Cleymans, K. Redlich, and H. Satz, Z. Phys. C **52**, 517 (1991).
[18] M. Gyulassy, X.-N. Wang, and M. Plumer, Phys. Rev. D **51**, 3236 (1995).
[19] G. Corn and T. Corn, *Reference Book on Mathematics* [in Russian], Nauka, Moscow (1974).
[20] J. Cleymans, V. V. Goloviznin, and K. Redlich, in *Proceedings of Workshop on Pre-Equilibrium Parton Dynamics in Heavy Ion Collisions*, LBL, Berkeley (1993); E-prints archive hep-ph/9310284, Preprint UCT-TP 199/93, Cape Town (1993).
[21] F. Yndураĭn, *Quantum Chromodynamics. An Introduction to the Theory of Quarks and Gluons*, Springer Verlag, Heidelberg (1983).
[22] CERES Collab., G. Agakichiev *et al.*, Phys. Rev. Lett. **76**, 3506 (1996).
[23] A. Drees, Nucl. Phys. A **630**, 449 (1998).

Translation was provided by the Russian Editorial office.

Spin-orbit coupling and the dynamic Jahn-Teller effect in the C_{60}^- system

A. A. Remova*)

Institute of Inorganic Chemistry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 18 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 194–203 (July 1999)

The dynamic Jahn–Teller effect is studied for a charged fullerene molecule C_{60}^- with allowance for spin-orbit coupling. The system of self-consistent equations describing the interaction of an electron and the molecular vibrations in the event of spin-orbit splitting of the electronic level is solved analytically. A novel type of nonlinear vibrations occurring in such a system is described. It is shown that with spin-orbit coupling taken into account, the static Jahn-Teller configurations in the C_{60}^- molecule are unstable even in the limit of strong electron-vibronic coupling and that the symmetry of the atomic configuration of the unperturbed C_{60} molecule is restored under time averaging. © *1999 American Institute of Physics.* [S1063-7761(99)01607-8]

1. INTRODUCTION

One of the problems of the physics of fullerenes, whose symmetry belongs to that of the highest-symmetry point group I_h, is the understanding of the nature of the Jahn–Teller effect in the various excited states of the C_{60} molecule. The vibrational properties of fullerenes have been extensively studied in experiments[1–10] in view of their importance for optical applications. In particular, the behavior of the high-frequency mode A_g is an indication of an excess number of electrons in the molecule, while the H_g modes can shift, broaden, or disappear entirely. The high symmetry of the C_{60} molecule is the reason for the degeneracy of the states of the electron subsystem and the vibrational subsystem of the molecule. An ''excess'' electron in C_{60} occupies the first excited state with the symmetry of t_{1u} and ''distributes'' itself over the molecule surface. It is assumed that the electron-vibronic coupling substantially changes such a state of the C_{60}^- molecule. The interaction of an electron occupying the lowest unoccupied threefold degenerate molecular orbital t_{1u} and the fivefold degenerate vibrational modes H_g gives rise to polaronlike states and to Jahn-Teller distortions of the molecule.[11–13] Here the configurations of the molecules with minimum energy form multiplets. Thus, the symmetry of the ground state lowers, but the degeneracy of the electronic levels is not lifted completely, and the electron subsystem of the C_{60}^- molecule remains very sensitive to perturbations. Typical calculations done in the Born–Oppenheimer approximation allow obtaining the effective potential energy as a function of the static configuration of the ions.[14] In this case the zero-point vibration energy must be much lower than the energy of the Jahn–Teller splitting of an electronic level. It is assumed that in the opposite case we have the dynamic Jahn–Teller effect.

Experimental studies suggest that for an isolated C_{60} molecule the Jahn–Teller effect is dynamic.[4,5,8,10] A possible consequence of dynamic symmetry breaking is the multiphonon structure of the lines observed in Raman and IR spectra[3,4] and in the data on electron spin resonance.[10] Usually, to adjust the ground-state energy, the energy of the zero-point vibrations is added. For instance, in Refs. 11–13 the calculations were begun by determining the positions of the minima in the potential well obtained in the strong coupling limit. The dynamical problem is solved with allowance for electron tunneling in the configuration space between the wells through the potential barriers surrounding these minima. It is difficult to solve this problem systematically, since the effective potential itself is determined by the electron wave function. Electron correlation effects have also been considered.[15] Ihm[16] and Auerbach *et al.*[17] studied the dynamic Jahn–Teller effect in charged fullerenes in terms of the Berry phase with allowance for pseudorotations. Pseudorotations arise as a result of rapid movements of deformed regions on the molecule surface, and the motion of individual atoms is limited by the proximity to their symmetric positions in the neutral C_{60} molecule.

The present paper proposes a novel semiclassical variant of the theory of the dynamic Jahn–Teller effect. We will examine the complete dynamic lifting of the degeneracy of the static Jahn–Teller levels due to spin-orbit coupling. Such coupling generates self-consistent vibrations of the molecules that mix the electronic states with spin flip. In a simplified model of the C_{60}^- molecule that allows for one vibrational multiplet H_g interacting with the electron in state t_{1u} and for spin-orbit coupling, we will derive the exact solutions of a nonlinear system of dynamical equations that describe the temporal evolution of the ground state without resorting to perturbation-theory techniques. We will show that the static ground state is unstable against spin-orbit perturbations even in the strong coupling limit. We will also find an expression for the renormalization of the intramolecular phonon frequency and its nonlinear splitting.

2. MODEL

The model allows for the interaction of π-electrons with vibrations of the C_{60} molecule, since all other electronic lev-

els are located (on the energy scale) at distances that are much larger than the energies of the H_g modes and their perturbation can be ignored. As in Ref. 17, we will consider a one-electron state related to a single vibrational mode H_g of frequency ω_0, and instead of the symmetry group of the frustrum of an icosahedron we will assume that the symmetry is spherical. Then t_{1u} and H_g can be described by the spherical functions $\{Y_{1k}\}_{k=-1}^{1}$ and $\{Y_{2m}\}_{m=-2}^{2}$, respectively.

The main goal of this work is to provide a semiclassical description of the dynamic instability of the ground state of the C_{60}^- molecule, the instability being caused by electron-vibronic and spin-orbit coupling. We will study the redistribution of charge over the molecule surface, electron spin flip, and vibrational motion of the molecule with the passage of time. The Lagrangian L of such a process has the form

$$L=L_{\rm ph}+L_{\rm el}+L_{\rm el,el-vib}+L_{\rm s-o}. \tag{1}$$

Here

$$L_{\rm el}=\frac{i\hbar}{2}\int\left(\psi^*\frac{\partial\psi}{\partial t}-\psi\frac{\partial\psi^*}{\partial t}\right)d\Omega-\int\psi^*E_0\psi d\Omega, \tag{2}$$

$$L_{\rm vib}=\frac{\mu}{2}\int\frac{\partial\eta}{\partial t}\frac{\partial\eta^*}{\partial t}d\Omega-\frac{\kappa}{2}\int\eta\eta^*d\Omega, \tag{3}$$

$$L_{\rm el-vib}=\lambda\int\psi^*(\eta+\eta^*)\psi d\Omega, \tag{4}$$

where μ is the effective mass, κ is the rigidity, $\int\psi^*\psi d\Omega=1$, $d\Omega=\sin\theta d\theta d\phi$, and θ and ϕ are the angles of the spherical system of coordinates. Under the condition

$$\sum_{m=-1}^{1}c_m(t)c_m^*(t)=1, \tag{5}$$

the representation of the wave function ψ of the electronic level t_{1u} with an energy E_0 has the form

$$\psi=\sum_{k=-1}^{1}c_k(t)Y_{1k}(\theta,\phi). \tag{6}$$

Group-theoretical analysis shows that the eight modes H_g and the two modes A_g of the fullerene molecule can interact with an electron in a state with the symmetry of t_{1u}. Here we will consider one such mode, H_g, with the frequency

$$\omega_0=\sqrt{\kappa/\mu}. \tag{7}$$

The rotation of the molecules as a whole is ignored. The electron-vibronic interaction $L_{\rm el-vib}$ in (4) is local and rotation-invariant,[12] with λ the constant of this interaction. In (1), $L_{\rm s-o}$ is the spin-orbit coupling energy. To within an unimportant constant,

$$L_{\rm s-o}=A\langle\hat{\mathbf{l}}\cdot\hat{\mathbf{s}}\rangle. \tag{8}$$

For a spherical molecule of radius R_m we have[18]

$$A\approx\left(\frac{Ze^2}{\hbar c}\right)^2\frac{r_a}{R_m}\frac{me^4}{\hbar^2},$$

where r_a is the radius of the carbon atom. The quantum numbers l and s are fixed, and in averaging in (8) we must allow for the fact that, in contrast to $\hat{\mathbf{s}}$, the projections of $\hat{\mathbf{l}}$ are not equally probable, since the probability $|c_m|^2$ of the projection m occurring is time-dependent. The projections of $\hat{\mathbf{j}}=\hat{\mathbf{l}}+\hat{\mathbf{s}}$ are not equiprobable either:

$$\langle\hat{\mathbf{j}}^2\rangle=\sum_{j,n}j(j+1)b_{j,n}b_{j,n}^*. \tag{9}$$

where $b_{j,n}$ is the probability amplitude of the vector $\hat{\mathbf{j}}=\hat{\mathbf{l}}+\hat{\mathbf{s}}$ having the value j and the projection n. Then, allowing for the quantum mechanical rule of vector summation in terms of Clebsch–Gordan coefficients $\langle l,m;s,M|j,n\rangle$, we obtain

$$b_{j,n}=\frac{1}{\sqrt{2s+1}}\sum_{j=-(|s+l|),M,m}^{s+l}c_m^*\langle l,m;s,M|j,n\rangle. \tag{10}$$

Substituting (10) in (9) yields

$$\langle\hat{\mathbf{j}}^2\rangle=\frac{1}{2s+1}\sum_{j,n,m,m'}j(j+1)c_mc_{m'}^*\langle j,n|l,m;s,M\rangle\times\langle j,n|l,m';s,M'\rangle, \tag{11}$$

$$m,m'=-l,\ldots,0,\ldots,l,\quad j=|s-l|,\ldots,l,\ldots,s+l,$$

$$M,M'=-s,\ldots,,0,\ldots,s,\quad M=n+m,\ M'=n+m'.$$

With allowance for (9)–(11), the expression (8) for $l=1$ and $s=1/2$ becomes

$$L_{\rm s-o}=A\frac{\sqrt{2}}{2}c_0^*(c_1+c_{-1})+\text{c.c.} \tag{12}$$

As shown below, the $\langle\hat{\mathbf{s}}\cdot\hat{\mathbf{l}}\rangle$ coupling is the cause of the perturbations that result in the dynamic splitting of the static Jahn-Teller level with the lowest energy.

We will assume that the vibrational field η of the C_{60}^- molecule is time-dependent and can be described by classical means:

$$\eta=\sum_{k=-2}^{2}\rho_k(t)Y_{2k}(\theta,\phi), \tag{13}$$

where $\rho_k(t)$ is the amplitude of the kth spherical harmonic. If the amplitude of the atomic vibrations is much smaller that the interatomic distances at equilibrium positions, the classical description of a vibronic mode is meaningful. In other words, we will assume that we are dealing with strong electron-phonon coupling and perturbation-theory techniques do not work.

Substituting (5), (6), and (13) in (1)–(4), we find that

$$L_{\rm el}=\frac{i\hbar}{2}\sum_{m=-1}^{1}\left(c_m^*\frac{dc_m}{dt}-c_m\frac{dc_m^*}{dt}\right)-E_0\sum_{m=-1}^{1}c_mc_m^*, \tag{14}$$

$$L_{\rm ph}=\frac{\mu}{2}\sum_{m=-2}^{2}\frac{d\rho_m}{dt}\frac{d\rho_m^*}{dt}-\frac{\kappa}{2}\sum_{m=-2}^{2}\rho_m\rho_m^*, \tag{15}$$

$$L_{\text{el-ph}}=\lambda\sum_{k=-2}^{2}\sum_{m=-1}^{1}\sum_{n=-1}^{1}\rho_k c_m c_n^* F(k,m,n)+\text{c.c.}, \tag{16}$$

where the $\{F(k,m,n)\}_{k=-2,m=-1,n=-1}^{2,1,1}$ are defined in terms of Clebsch–Gordan coefficients:

$$F(k,m,n)=\frac{1}{3^{1/4}}\sqrt{\frac{5}{4\pi}}\langle 1,m;1,\ n|2,\ k\rangle\langle 1,0;1,0|2,0\rangle.$$

3. BASIC EQUATIONS AND THE SOLUTION

Substituting (12) and (14)–(16) in (1) and introducing the dimensionless parameters

$$\tilde{\lambda}=\frac{\lambda}{\mu},\quad \alpha=\left(\frac{5\pi\hbar}{2\tilde{\lambda}^2\mu}\right)^3,\quad \beta=\frac{1}{\sqrt{5\pi}}\tilde{\lambda}\alpha^2,$$

$$\Omega_0^2=\frac{\sqrt{5\pi}\omega_0^2\beta}{\tilde{\lambda}} \tag{17}$$

we arrive at a set of Lagrange equations[19] for the system under consideration:

$$\begin{aligned}
&\frac{d^2v_{-2}}{d\tau^2}=-\Omega_0^2v_{-2}+\sqrt{\frac{3}{2}}\tilde{c}_{-1}\tilde{c}_1^*,\\
&\frac{d^2v_{-1}}{d\tau^2}=-\Omega_0^2v_{-1}+\sqrt{\frac{3}{2}}(\tilde{c}_{-1}\tilde{c}_0^*-\tilde{c}_0\tilde{c}_1^*),\\
&\frac{d^2v_0}{d\tau^2}=-\Omega_0^2v_0+\left(\tilde{c}_0\tilde{c}_0^*-\frac{1}{2}\tilde{c}_{-1}\tilde{c}_{-1}^*-\frac{1}{2}\tilde{c}_1\tilde{c}_1^*\right),\\
&\frac{d^2v_1}{d\tau^2}=-\Omega_0^2v_1-\frac{\sqrt{3}}{2}(\tilde{c}_{-1}^*\tilde{c}_0-\tilde{c}_0^*\tilde{c}_1),\\
&\frac{d^2v_2}{d\tau^2}=-\Omega_0^2v_2+\sqrt{\frac{3}{2}}\tilde{c}_{-1}^*\tilde{c}_1,\\
&i\frac{d\tilde{c}_{-1}}{d\tau}=\frac{1}{2}\tilde{c}_{-1}v_0-\sqrt{\frac{3}{2}}\tilde{c}_1v_{-2}-\frac{\sqrt{3}}{2}\tilde{c}_0v_{-1}-\tilde{A}\tilde{c}_0,\\
&i\frac{d\tilde{c}_0}{d\tau}=\frac{\sqrt{3}}{2}(\tilde{c}_{-1}v_1+\tilde{c}_1v_{-1})-\tilde{c}_0v_0-\tilde{A}(\tilde{c}_1+\tilde{c}_{-1}),\\
&i\frac{d\tilde{c}_1}{d\tau}=\frac{1}{2}\tilde{c}_1v_0+\sqrt{\frac{3}{2}}\tilde{c}_{-1}v_2-\frac{\sqrt{3}}{2}\tilde{c}_0v_1-\tilde{A}\tilde{c}_0,
\end{aligned} \tag{18, 19}$$

$$\sum_{m=-1}^{1}\tilde{c}_m(\tau)\tilde{c}_m^*(\tau)=1, \tag{20}$$

where

$$\tilde{c}_m=c_m\exp\frac{iE_0'\tau}{\hbar},\quad E_0'=\alpha E_0,$$

$$\tilde{A}=A\frac{\sqrt{2}\,\alpha}{2\hbar},\quad v_k=\frac{\rho_k}{\beta},\quad \tau=\frac{t}{\alpha}. \tag{21}$$

Here and in what follows we do not write the complex-conjugate equations. The nonlinear system of dynamic equations (18)–(20) has one exact solution, whose properties are discussed below. We seek this solution in the form

$$\begin{aligned}
&v_{-2}=g\sqrt{\frac{3}{2}}\tilde{c}_{-1}\tilde{c}_1^*,\quad v_{-1}=g\frac{\sqrt{3}}{2}(\tilde{c}_{-1}\tilde{c}_0^*-\tilde{c}_0\tilde{c}_1^*),\\
&v_0=g\left(\tilde{c}_0\tilde{c}_0^*-\frac{1}{2}\tilde{c}_{-1}\tilde{c}_{-1}^*-\frac{1}{2}\tilde{c}_1\tilde{c}_1^*\right),\\
&v_1=-g\frac{\sqrt{3}}{2}(\tilde{c}_{-1}^*\tilde{c}_0-\tilde{c}_0^*\tilde{c}_1),\quad v_2=g\sqrt{\frac{3}{2}}\tilde{c}_{-1}^*\tilde{c}_1.
\end{aligned} \tag{22}$$

Substituting (22) in (19) yields

$$\begin{aligned}
&-i\frac{d\tilde{c}_{-1}}{d\tau}=\frac{3}{4}g\left[2\tilde{c}_{-1}\left(\tilde{c}_1\tilde{c}_1^*+\frac{1}{6}\right)-\tilde{c}_1^*\tilde{c}_0^2\right]-\tilde{A}\tilde{c}_0,\\
&-i\frac{d\tilde{c}_0}{d\tau}=\frac{3}{4}g\left[-2\tilde{c}_1\tilde{c}_{-1}\tilde{c}_0^*+\tilde{c}_0\left(\tilde{c}_0\tilde{c}_0^*+\frac{1}{3}\right)\right]-\tilde{A}(\tilde{c}_{-1}+\tilde{c}_1),\\
&-i\frac{d\tilde{c}_1}{d\tau}=\frac{3}{4}g\left[2\tilde{c}_1\left(\tilde{c}_{-1}\tilde{c}_{-1}^*+\frac{1}{6}\right)-\tilde{c}_{-1}^*\tilde{c}_0^2\right]-\tilde{A}\tilde{c}_0.
\end{aligned} \tag{23}$$

The values of the electron-phonon coupling constant g in (22) and (23) are determined below.

One of the exact solutions of the nonlinear system of equations (20) and (23) for finding $\tilde{c}_m(\tau)$ has the form of harmonic oscillations with frequencies

$$\omega_1=\frac{3g}{4}\sqrt{1-a^*a},\quad \omega_{2,3}=\omega_1\pm\sqrt{2}\tilde{A}, \tag{24}$$

where a is an arbitrary constant such that $a^*a\leqslant 1$. At the same time, the probability amplitudes $\tilde{c}_m(\tau)$ are found to be coupled by the condition

$$\tilde{c}_0^2-2\tilde{c}_{-1}\tilde{c}_1=a\exp\{2ig\tau\}. \tag{25}$$

Under this condition, the system of equations (23) reduces to a linear one and the solution has the form of harmonic oscillations. However, the amplitudes of these oscillations depend on the frequency ω and the superposition principle does not apply. Moreover, such electronic self-oscillating states may be coupled self-consistently to atomic density oscillations only if the resulting solution, according to (22), satisfies the system of equations (18), which is possible if

$$(2\omega_i)^2-\Omega_0^2+\frac{1}{g}=0,\quad i=1,2,3. \tag{26}$$

We will limit ourselves to two important cases in (24): $aa^*=0$ and $aa^*=1$. In the absence of spin-orbit coupling ($A=0$) the first case corresponds to the upper level in the Jahn–Teller static configuration of the C_{60}^- molecule. When $aa^*=1$ holds we are dealing with the lowest stationary level (on the energy scale). In both cases, spin-orbit coupling produces self-consistent electronic transitions and vibration of the atoms in the molecule with frequencies Ω_1 and Ω_2 in dimensionless variables:

$$\Omega_1=\frac{\alpha A}{\hbar},\quad \Omega_2=2\Omega_1. \tag{27}$$

For $aa^*=1$ we have

$$\tilde{c}_1=\frac{\exp\{ig\tau\}}{\sqrt{6}}\left(\frac{\exp\{i\Omega_1\tau\}}{\sqrt{2}}+\frac{\exp\{-i\Omega_1\tau\}}{\sqrt{2}}+i\right),$$

$$\tilde{c}_0=\frac{\exp\{ig\tau\}}{\sqrt{6}}(\exp\{i\Omega_1\tau\}-\exp\{-i\Omega_1\tau\}),$$

$$\tilde{c}_{-1}=\frac{\exp\{ig\tau\}}{\sqrt{6}}\left(\frac{\exp\{i\Omega_1\tau\}}{\sqrt{2}}+\frac{\exp\{-i\Omega_1\tau\}}{\sqrt{2}}-i\right);$$

$\mathrm{Re}\,v_{\pm1}$ and $\mathrm{Im}\,v_{\pm2}$ oscillate with the frequency Ω_1; and $\mathrm{Re}\,v_{\pm2}$, $\mathrm{Im}\,v_{\pm1}$, and $\mathrm{Re}\,v_0$ oscillate with the frequency $2\Omega_1$. These two frequencies appear simultaneously, and the time average of the vibrational field (13), $\overline{\eta}$, vanishes and hence the initial symmetry of the undeformed molecules will be restored (on the average). Similar solutions were obtained for the case $aa^*=0$.

The strong coupling limit requires that $g/\Omega_0\gg1$ and $g/\Omega_{1,2}\gg1$. From (27) it follows that these two conditions are met if either $\Omega_0\ll1$ and $\Omega_{1,2}\ll1$ or $\Omega_0\approx\Omega_{1,2}$. The case $\Omega_0=\Omega_{1,2}$ is highly improbable and we will ignore it. A system with $\Omega_0<\Omega_{1,2}$ undergoes a radical transformation and will also be ignored, since the spin-orbit coupling constant is small and $A/\hbar\omega_0\approx\Omega_{1,2}/\Omega_0\ll1$.

Mathematically, the system of nonlinear equations (19) and (20) has a bifurcation solution. Under any infinitely weak perturbation A the system becomes unstable and self-oscillations set in. The mechanism of excitation of such oscillations is soft, so that we can say that dynamically the system is stable. The phase space of the solutions of the system contains at least two stable limit cycles and two unstable bifurcation points.[20] One of these points fixes the steady state with the lowest energy for the deformed C_{60}^- molecule. Under a spin-orbit perturbation of the system this state becomes unstable.

4. DISCUSSION

This paper studies the electron-vibronic states of a charged C_{60} molecules with allowance for spin-orbit coupling. It is assumed that a real configuration of atoms can be described by a continuous distribution of the atomic density over the spherical surface of the molecule. The spherical functions $\{Y_{1m}\}_{m=-1}^{1}$ are used to represent the wave function of the threefold degenerate electronic state t_{1u} coupled to the atomic vibrations of the fivefold degenerate level H_g. The functions $\{Y_{2k}\}_{k=-2}^{2}$ characterize the corresponding vibronic mode. The electron-phonon coupling is introduced in the ordinary way.[12,18] In addition to electron-phonon coupling, the dynamic ''mixing'' of electronic states due to spin flip for difference projections m of the electron orbital angular momentum $\tilde{\mathbf{l}}$ is also accounted for.

The solution (27) obtained in Sec. 3 and the important results are depicted in Figure 1, which shows the transition from the electronic state t_{1u} (E_0) of the undeformed C_{60}^- molecule to the dynamic Jahn–Teller state (the hatched section); E_1 and E_2 are the split electron energy levels of the atomic configurations of reduced symmetry of the deformed

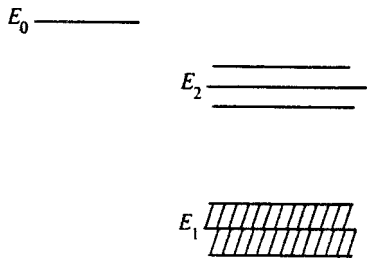

FIG. 1. Dynamic splitting of levels of the Jahn–Teller system C_{60}^-.

C_{60}^- molecule, which arises because of the static Jahn–Teller effect. These two steady states are unstable with respect to spin-orbit coupling. If the degeneracy of the initial state E_0 is threefold, the degeneracies of the levels E_1 and E_2 are twofold and threefold, respectively. As a result of spin-orbit coupling, the degeneracy of the upper Jahn–Teller level, E_2, is lifted in the static limit, too; the lower level E_1 splits only dynamically (according to the Kramers theorem). The hatched section in Fig. 1 corresponds to the electronic states that result from time-dependent ''mixing'' of levels and spin flip. The electronic states vary with time with probability amplitudes $c_1(t)$, $c_{-1}(t)$, and $c_0(t)$ within a certain energy interval. The length of this interval is determined by the spin-orbit coupling constant [see Eq. (8)]. As shown in Sec. 4, the electronic transitions are accompanied by self-consistent stable vibrations of the charge distributed over the molecule surface and of the atoms in the molecule and by electron spin flip. Similar vibrations arise a result of transitions between the split states of the level E_2. The vibrations are nonlinear but harmonic.

The time average of these vibrations v_i is zero and hence the time average of the symmetry of the atomic configuration of the molecule turns out to be the same as for a undeformed molecule, while the symmetry of the electronic states is completely broken.

This treatment is valid if quantum vibrations can be ignored. For this to be true, the zero-point vibration energy must be much lower than the static Jahn–Teller splitting of a level. It is also assumed that static splitting is larger than dynamic splitting, since the new vibrational modes $\Omega_{1,2}$ are classical by assumption. The amplitudes of the vibrations are limited due to the smallness of the spin-orbit coupling constant. This means that the electron-phonon coupling constant is large and the condition for the Born–Oppenheimer approximation is met ($g\approx1/\Omega_0^2\gg1$).

5. CONCLUSION

We have proposed a variant of the semiclassical treatment of the Jahn-Teller effect for the C_{60}^- molecule that facilitates an understanding of the nature of some excited electronic states in fullerenes. There are still many difficulties in directly verifying this effect by experiments. The method developed in the present paper was used to derive expressions for the Jahn–Teller splitting of intramolecular phonon frequencies of H_g and to explain one of the multiphoton singularities in the Raman and IR spectra of fullerene compounds. The consequences of the Jahn–Teller effect also

manifest themselves in the data on electron spin resonance.[10] In particular, the mode H_g may disappear completely, and instead two modes with lower frequencies that differ by a factor of two appear. The occurrence of such singularities cannot be explained if one remains within the framework of the static Jahn–Teller effect or the common dynamic effect.

These novel nonlinear vibrations result from the spin-orbit coupling, so that they must be accompanied by fairly slow electron spin flips. The vibrations occur near the positions of equilibrium of the undeformed molecule, while the static Jahn–Teller configuration of atoms in C_{60}^- is unstable even in the Born–Oppenheimer approximation. More than that, the ground state of an isolated C_{60}^- molecule is time-dependent. The Jahn–Teller effect may become static in the crystalline structure of fullerites or fullerides.[21] In these cases the degeneracy of the Jahn–Teller ground state is lifted completely by the crystalline field.

Thus, the ground state of an isolated electron-vibronic system of the C_{60}^- molecule with spin-orbit coupling is represented by closed cycles in the phase space, which ensure a periodic motion near unstable steady-state solutions. The originating nonlinear vibrations are accompanied by energy transfer from the electron subsystem to the vibrational subsystem when the spin degrees of freedom are excited.

In conclusion we note that allowance for rotations of the molecule as a whole leads to similar effects. Here the Coriolis splitting of levels acts as spin-orbit coupling. Most results are valid in the latter case, too; they can be of use in experimental studies of the Jahn–Teller effect for a freely rotating C_{60}^- molecule.

The work was supported by a grant from the Superconductivity Council of the Russian Academy of Sciences (Project 96107).

*)E-mail: lina@casper.che.nsk.su

[1] T. W. Ebbesen, J. S. Tsai, K. Tanigaki *et al.*, Nature (London) **355**, 620 (1992).
[2] M. G. Mitch, S. J. Chase, and J. S. Lannin, Phys. Rev. Lett. **68**, 883 (1992).
[3] P. J. Horoyski, M. L. W. Thewalt, and T. R. Anthony, Phys. Rev. Lett. **74**, 194 (1995).
[4] P. Zhou, K.-A. Wang, P. C. Eklund, G. Dresselhaus, and M. S. Dresselhaus, Phys. Rev. B **48**, 8412 (1993).
[5] J. Kim and W.-P. Su, Phys. Rev. B **50**, 8832 (1994).
[6] K. Prassides, C. Christides, M. J. Rosseinsky *et al.*, Europhys. Lett. **19**, 629 (1992).
[7] Y. Wang, J. M. Hodden, A. M. Rao, and P. C. Eklund, Phys. Rev. B **50**, 173 (1994).
[8] O. Gunnarsson, H. Handschuh, P. S. Bechthold, B. Kessler, G. Ganteför, and W. Eberhardt, Phys. Rev. Lett. **74**, 1875 (1995).
[9] D. Dubois and K. M. Kadish, J. Am. Chem. Soc. **113**, 4346 (1991).
[10] X. Wei and Z. V. Vardeny, Phys. Rev. B **52**, 2317 (1995).
[11] V. P. Antropov, O. Gunnarsson, and A. I. Liechtenstein, Phys. Rev. B **48**, 7651 (1993).
[12] M. O'Brien, Phys. Rev. B **53**, 3775 (1996).
[13] J. L. Dunn and C. A. Bates, Phys. Rev. B **52**, 5996 (1995).
[14] I. B. Bersuker, *The Jahn-Teller effect and Vibronic Interactions in Modern Chemistry*, Plenum Press, New York (1984).
[15] W. Z. Wang, C. L. Wang, A. R. Bishop, L. Yu, and Z. B. Su, Phys. Rev. B **51**, 10209 (1995).
[16] J. Ihm, Phys. Rev. B **49**, 10726 (1994).
[17] A. Auerbach, N. Manini, and E. Tosatti, Phys. Rev. B **49**, 12998 (1994).
[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977), Chap. XIII.
[19] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976), Chap. I.
[20] V. I. Arnol'd, *Additional Chapters on the Theory of Ordinary Differential Equations* [in Russian], Nauka, Moscow (1978), Chap. VI [English transl.: *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer, Berlin (1983)].
[21] A. A. Remova, V. P. Shpakov, U-Hyon Paek, and V. R. Belosludov, Phys. Rev. B **52**, 13 715 (1995).

Translated by Eugene Yankovsky

Nonsecular contribution to the decay of spin-echo signals of quadrupolar nuclei in magnetically ordered materials

G. N. Abelyashev,*) V. N. Berzhanskiĭ, A. I. Gorbovanov, S. N. Polulyakh, and N. A. Sergeev†)

Simferopol State University, 333007 Simferopol, Ukraine
(Submitted 17 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 204–216 (July 1999)

We analyze the influence of fluctuations of the nonsecular part of the spin Hamiltonian on the decay of ordinary and multiquantum signals of the two-pulse spin echo in a quadrupole spin system with an inhomogeneously broadened spectral line. Expressions are obtained for the rate of decay of an echo in the case of selective excitation of a signal from quadrupole nuclei with arbitrary spin. These expressions are then used to analyze the experimentally observed ordinary and multiquantum echo signals from quadrupole nuclei with spin $I=3/2$ (^{53}Cr, ^{63}Cu, and ^{65}Cu) in ferromagnetic chromium chalcogenide spinels. © *1999 American Institute of Physics.* [S1063-7761(99)02107-1]

1. INTRODUCTION

Nuclear magnetic resonance (NMR) is a method of investigating magnetically ordered materials that allows one to obtain information on a microscopic level about the static and dynamic properties of both the crystalline lattice and the electron spin system.[1,2] The main capabilities of NMR in magnetically ordered materials are well known.[1,2] They are based on the fact that the main interaction probed by NMR is the interaction of the nuclear magnetic moments with the local hyperfine magnetic fields created by the nonzero electronic magnetization. An important feature of NMR in magnetically ordered materials is a natural inhomogeneous broadening of the spectral line. In spin systems with an inhomogeneously broadened spectral line, when using two or more exciting pulses, one observes echo signals, whose formation is described by the Hahn mechanism.[1] In the simplest case of two pulses separated by a time interval τ, the echo signal is formed at the time $t=2\tau$ (the time t is measured from the time of onset of the first exciting pulse).

Nuclei with spin $I>1/2$ can take part in electric quadrupole interactions. As a rule, in magnetically ordered materials quadrupole interactions are weaker than the magnetic interactions and lead to quadrupole splitting of the NMR spectra. In addition, quadrupole interactions can lead to the appearance of multiquantum echo signals, whose formation time differs from $t=2\tau$ (Refs. 3–7).

Multiquantum echo signals were first observed in nonmagnetic compounds.[3] Multiquantum echo signals were later detected in ferromagnets.[4–7] It has been shown, in particular, that the quadrupole satellites are suppressed in NMR spectra recorded with the help of multiquantum echo signals.[4] It was also noted that increasing the time interval between the exciting pulses leads to a more rapid decay of the multiquantum echo in comparison with the ordinary echo at $t=2\tau$ (Ref. 4). The decay of the additional echo signals from the quadrupole nuclei, formed at times between the exciting pulses comparable with the inverse width of the spectral line, was analyzed in Refs. 6 and 7. However, detailed studies of the relaxation properties of the multiquantum echo signals were not performed.

At present, the theory of magnetic relaxation for spin systems with inhomogeneous broadening of the spectral line is well developed for the case where quadrupole interactions are absent.[1,8–10] In this case, the relaxation interactions can be represented as interactions of the resonant spins with an effective magnetic field which is a stochastic function of time. The interaction of the nuclear spin system with a fluctuating magnetic field is treated as a perturbation of the main interaction, which is an interaction with a static magnetic field.

Relative to the static field, the nuclear magnetization and the fluctuating magnetic field can be separated into longitudinal and transverse components. The random process, as a rule, is characterized by two parameters: the correlation time τ_c and the amplitude of the fluctuations σ. Rapid fluctuations ($\sigma\tau_c^{-1}\gg 1$) of the longitudinal component of the effective magnetic field are the reason for the exponential decay of the transverse component of the magnetization. The decay rate is characterized by the transverse relaxation time T_2. Rapid fluctuations of the transverse component of the effective magnetic field lead to an exponential recovery of the longitudinal component of the magnetization with characteristic time T_1, the longitudinal relaxation time. In addition, the transverse fluctuations lead to decay of the transverse component of the magnetization. In the case where all the relaxation processes are due exclusively to the transverse fluctuations, the relation $T_2=2T_1$ is satisfied.

In the case of magnetic relaxation in a quadrupole spin system in the presence of fluctuating magnetic fields it is necessary to take into account fluctuations of the electric quadrupole interactions. The main interaction is represented as a sum of the interaction of the nuclear spins with the static magnetic field and the secular part of the static electric quadrupole interactions. The relaxation interactions are the inter-

actions of the nuclear spins with the fluctuating magnetic and electric fields.

Relative to the main interaction, the relaxation Hamiltonian can be separated into a secular and a nonsecular part. The influence of the secular fluctuations on the decay of the two-pulse echo signals from the quadrupole nuclei was theoretically investigated in Ref. 11 in the spectral diffusion model. The theoretical results of Ref. 11 were successfully applied in Ref. 12 to analyze the experimentally observed decay of two-pulse echo signals from ^{53}Cr (quadrupole) nuclei in the ferromagnet $CdCr_2Se_4$:Ag.

Our aim in the present paper is to analyze the effect of fluctuations of the nonsecular part of the spin Hamiltonian on the decay of a two-pulse echo in an inhomogeneously broadened quadrupolar spin system. The theoretical results obtained in this study are utilized to analyze the experimentally observed decay of ordinary and multiquantum echo signals from copper and chromium nuclei in ferromagnetic chromium chalcogenide spinels.

2. THEORY

We write the Hamiltonian of the main interaction (in units $\hbar=1$) as

$$H_0=-\omega_0 I_z+\omega_Q\left[I_z^2-\frac{I(I+1)}{3}\right], \tag{1}$$

where $\omega_0=\gamma B$, γ is the gyromagnetic ratio, B is the magnetic induction, ω_Q is the magnitude of the quadrupole splitting of the NMR spectrum, and I is the particle spin.

We represent the Hamiltonian describing the fluctuations of the nonsecular part of the magnetic and electric interactions as[1]

$$H_1(t)=-\omega_+(t)I_--\omega_-(t)I_++V_+(t)(I_-I_z+I_zI_-)+V_-(t)(I_+I_z+I_zI_+)+W_+(t)I_-^2+W_-(t)I_+^2. \tag{2}$$

The coefficients $\omega_\pm(t)$ describe the fluctuations of the transverse component of the local magnetic field at the nucleus, and $V_\pm(t)$ and $W_\pm(t)$ are the fluctuations of the nonsecular part of the quadrupole interaction.

Treating the Hamiltonian $H_1(t)$ as a perturbation of the main Hamiltonian H_0, we calculate the response pulses with the help of the density matrix operator $\rho(t)$ in the interaction representation:[9]

$$\rho^*(t)=\exp(iH_0t)\rho(t)\exp(-iH_0t), \tag{3}$$

$$H_1^*(t)=\exp(iH_0t)H_1(t)\exp(-iH_0t). \tag{4}$$

Assuming that $H_1(t)$ varies in time much faster than $\rho^*(t)$, at early times we obtain[9]

$$\rho^*(t)=\rho^*(0)+i\int_0^t[\rho^*(0)H_1^*(t')]dt'-\int_0^t\int_0^{t'}[[\rho^*(0)H_1^*(t'')]H_1^*(t')]dt'dt''. \tag{5}$$

For the transverse magnetization of one spin ensemble with the same law of variation $H_1(t)$, i.e., with one realization of the random process, the mean value of the transverse magnetization is calculated with the help of the density matrix operator

$$\langle M_+(t)\rangle=\mathrm{Tr}\{\rho(t)I_+\}=\alpha_m^+\exp[it(E_{m+1}-E_m)]\times\langle m|\rho^*(t)|m+1\rangle. \tag{6}$$

Here m is the magnetic quantum number (i.e., $I_z|m\rangle=m|m\rangle$); $E_m=\langle m|H_0|m\rangle$ are the eigenvalues of the unperturbed Hamiltonian H_0; $\alpha_m^\pm=\sqrt{I(I+1)-m(m\pm1)}$, i.e., $I_\pm|m\rangle=\alpha_m^\pm|m\pm1\rangle$. In order to find the response of the entire spin system, it is necessary to average expression (6) over the ensemble of ensembles, i.e., over all realizations of the random process. We denote such an average by a bar above the expression.

In the calculation of the transverse magnetization we need to calculate expressions of the type $\overline{\langle m|\rho^*(t)|m'\rangle}$ using the explicit form of $\rho^*(t)$ from Eq. (5). We assume that the random quantities in the relaxation Hamiltonian (2) fluctuate independently of one another:

$$\overline{g_+(t')f_-(t'')}=\overline{g_-(t')f_+(t'')}=\begin{cases}0, & g\neq f,\\ K_f(T), & g=f.\end{cases} \tag{7}$$

Here by $g_\pm(t)$ and $f_\pm(t)$ we mean $\omega_\pm(t)$ and $V_\pm(t)$ or $W_\pm(t)$. The correlation function $K_f(T)$, as usual, depends on $T=|t''-t'|$ for any choice of t' and falls off rapidly with growth of T.

Calculation of $\overline{\langle m|\rho^*(t)|m'\rangle}$ using the ''quantum-mechanical'' expansion of unity $1=\Sigma|m\rangle\langle m|$ and the explicit form of the relaxation Hamiltonian (2) gives rise to factors of the form

$$J_f(a,b,c,d)=\int_0^t\int_0^{t'}\overline{f_+(t')f_-(t'')}\exp[it''(E_a-E_b)-it'(E_c-E_d)]dt'dt''. \tag{8}$$

Transforming from the variable t'' to the variable T and extending the limits of integration over T to the range from $-\infty$ to $+\infty$, we obtain

$$J_f(a,b,c,d)=\int_0^t\int_{-\infty}^{\infty}K_f(T)\exp[iT(E_a-E_b)]\times\exp[it'(E_a-E_b-E_c+E_d)]dT\,dt'. \tag{9}$$

If $E_a-E_b-E_c+E_d\neq0$, then J_f is an oscillating function of time t. The appearance in the expression for $\overline{\langle m|\rho^*(t)|m'\rangle}$ of oscillating terms is equivalent to the appearance of an additional shift of the resonance frequency that results from taking into account the relaxation contribution of the nonsecular terms of the spin Hamiltonian in the Redfield theory.[9] In our analysis of the decay of the echo signal we disregard the oscillating terms. In addition, we will assume the mean values of all oscillating quantities to be zero.

For $a=c$ and $b=d$ from Eq. (9) we obtain

$$J_f(a,b)=\frac{J_f(a,b,a,b)}{t}$$

$$=\int_{-\infty}^{\infty}K_f(T)\exp[iT(E_a-E_b)]dT. \quad (10)$$

The quantity $J_f(a,b)$ has the meaning of a spectral density of the random process at the frequency $\omega_{a,b}=E_a-E_b$.

Omitting cumbersome intermediate manipulations, we present the final result for the mean value of the matrix element of the density matrix operator

$$\overline{\langle m|\rho^*(t)|m'\rangle}=[1-S(m,m')t]\langle m|\rho^*(0)|m'\rangle, \quad (11)$$

where

$$\begin{aligned}S(m,m')=&\alpha_m^+\alpha_{m+1}^-J_\omega(m+1,m)+\alpha_{m-1}^+\alpha_m^-J_\omega\\&\times(m-1,m)+\alpha_{m'}^+\alpha_{m'+1}^-J_\omega(m',m'+1)\\&+\alpha_{m'-1}^+\alpha_{m'}^-J_\omega(m',m'-1)\\&+\alpha_m^+\alpha_{m+1}^-[2m(m+1)+(m+1)^2\\&+m^2]J_V(m+1,m)+\alpha_{m-1}^+\alpha_m^-[2m(m-1)\\&+(m-1)^2+m^2]J_V(m-1,m)\\&+\alpha_{m'}^+\alpha_{m'+1}^-[2m'(m'+1)+(m'+1)^2\\&+m'^2]J_V(m',m'+1)+\alpha_{m'-1}^+\alpha_{m'}^-[2m'(m'\\&-1)+(m'-1)^2+m'^2]J_V(m',m'-1)\\&+\alpha_m^+\alpha_{m+1}^+\alpha_{m+1}^-\alpha_{m+2}^-J_W(m+2,m)\\&+\alpha_{m-1}^+\alpha_{m-2}^+\alpha_m^-\alpha_{m-1}^-J_W(m-2,m)\\&+\alpha_{m'}^+\alpha_{m'+1}^+\alpha_{m'+1}^-\alpha_{m'+2}^-J_W(m',m+2)\\&+\alpha_{m'-1}^+\alpha_{m'-2}^+\alpha_{m'}^-\alpha_{m'-1}^-J_W(m',m'-2).\end{aligned} \quad (12)$$

We are interested in the response of the spin system to a two-pulse input. For the transverse component of the magnetization at the time t after termination of the second exciting pulse, we obtain

$$\begin{aligned}\overline{\langle M_+(t)\rangle}=&\alpha_m^+\langle m|R_2^+|m'\rangle\langle m'|R_1^+\rho(0)R_1^-|m''\rangle\\&\times\langle m''|R_2^-|m+1\rangle\exp[it(E_{m+1}-E_m)\\&+i\tau(E_{m''}-E_{m'})][1-S(m,m+1)t]\\&\times[1-S(m',m'')\tau].\end{aligned} \quad (13)$$

Here the operators $R_j^{\pm}$ describe the evolution of the spin system acted on by the jth exciting pulse ($j=1,2$), τ is the time interval between exciting pulses, $\rho(0)$ is the density matrix at the time of onset of the first exciting pulse [$\rho(0)\propto I_z$].

Apart from the last two factors, expression (13) is identical to the Solomon formula[3] for the two-pulse response of a quadrupole spin system without fluctuating fields. It is known[3–5] that the condition of equality of the argument of the exponent in expression (13) to zero defines the time of formation of the echo signal in an inhomogeneously broadened spin system:

$$t_e=k\tau=\frac{E_{m'}-E_{m''}}{E_{m+1}-E_m}\tau. \quad (14)$$

Here the time t_e is measured from termination of the second exciting pulse.

The relaxation contribution of the nonsecular fluctuations is described by the last two factors in expression (13). Since the expression for the transverse magnetization is based on expression (5), which is valid for small t and τ, expression (13) can be considered as the series expansion of some relaxation function in t and τ to linear terms. However, the specific form of the relaxation function cannot be rigorously derived directly from expression (13). On the other hand, in the Redfield theory,[9] constructed on the basis of the same assumptions as the calculations given above, differential equations of motion of the magnetization vector are obtained in the form of the Bloch equations. It is known that these equations describe the exponential decay of an echo signal. On the basis of what has been said above, we assume that the last two factors in expression (13) are expansions of the exponentials to linear terms. Thus, dropping the averaging notation, we obtain for the amplitude of the echo signal formed at the time $t=(k+1)\tau$ (the time t is measured from the time of onset of the first pulse)

$$M_+(\tau)=M_+(0)\exp\left[-\frac{(k+1)\tau}{T_2}\right], \quad (15)$$

where T_2 is the transverse relaxation time due to fluctuations of the nonsecular part of the spin Hamiltonian

$$T_2^{-1}=\frac{1}{k+1}[kS(m,m+1)+S(m',m'')]. \quad (16)$$

In the derivation of expression (15) we took into account that for an echo signal formed at the time $t=(k+1)\tau$ the total relaxation time is equal to the sum of the time interval τ between pulses and the time interval $k\tau$ between the second pulse and the echo signal.

3. DISCUSSION OF THEORETICAL RESULTS

Expressions (16) and (12) allow us to obtain the relaxation time for an arbitrary echo signal in a system with arbitrary spin. In this case $J_f(m,m')=J_f(m',m)$ for any f, as follows from symmetry of the correlation function relative to time inversion T.

First let us consider the well-known case of decay of an echo signal in a spin system with $I=1/2$. In this case a single echo signal is formed at the time $t=2\tau$ for $m=-1/2$, $m'=1/2$, and $m''=-1/2$. Using (12) and (16), we obtain

$$T_2^{-1}=2J_\omega\left(\frac{1}{2},-\frac{1}{2}\right), \quad (17)$$

in good agreement with the well-known result for the transverse relaxation time due to fluctuations of the transverse component of the local magnetic field.[8,9]

TABLE I. Transverse relaxation rate for quadrupolar nuclei with spin $I=3/2$.

Echo signal	Transverse relaxation rate T_2^{-1}		
	Nonsecular contribution	Secular contribution	
		Gaussian process	Lorentzian process
Echo 2τ $m=\frac{1}{2},\ m'=\frac{3}{2},$ $m''=\frac{1}{2}$ (LF quadrupole satellite)	$4J_\omega(\frac{1}{2},-\frac{1}{2})+6J_\omega(\frac{3}{2},\frac{1}{2})+24J_V(\frac{3}{2},\frac{1}{2})$ $+12J_W(\frac{3}{2},-\frac{1}{2})+12J_W(\frac{1}{2},-\frac{3}{2})$	$\sigma_\Omega^2\tau_{c\Omega}+4\sigma_Q^2\tau_cQ$	$\sigma_\Omega+\sigma_Q$
Echo 2τ $m=-\frac{3}{2},\ m'=-\frac{1}{2},$ $m''=-\frac{3}{2}$ (HF quadrupole satellite)	$4J_\omega(\frac{1}{2},-\frac{1}{2})+6J_\omega(-\frac{1}{2},-\frac{3}{2})$ $+24J_V(-\frac{1}{2},-\frac{3}{2})+12J_W(\frac{3}{2},-\frac{1}{2})$ $+12J_W(\frac{1}{2},-\frac{3}{2})$	$\sigma_\Omega^2\tau_{c\Omega}+4\sigma_Q^2\tau_{cQ}$	$\sigma_\Omega+\sigma_Q$
Echo 2τ $m=-\frac{1}{2},\ m'=\frac{1}{2},$ $m''=-\frac{1}{2}$ (central transition)	$8J_\omega(\frac{1}{2},-\frac{1}{2})+3J_\omega(-\frac{1}{2},-\frac{3}{2})+3J_\omega(\frac{3}{2},\frac{1}{2})$ $+12J_V(-\frac{1}{2},-\frac{3}{2})+12J_V(\frac{3}{2},\frac{1}{2})$ $+12J_W(\frac{2}{3},-\frac{1}{2})+12J_W(\frac{1}{2},-\frac{3}{2})$	$\sigma_\Omega^2\tau_{c\Omega}$	σ_Ω
Echo 4τ $m=-\frac{1}{2},\ m'=\frac{3}{2},$ $m''=-\frac{3}{2}$ (central transition)	$6J_\omega(\frac{1}{2},-\frac{1}{2})+3J_\omega(-\frac{1}{2},-\frac{3}{2})+3J_\omega(\frac{3}{2},\frac{1}{2})$ $+12J_V(-\frac{1}{2},-\frac{3}{2})+12J_V(\frac{3}{2},\frac{1}{2})$ $+12J_W(\frac{3}{2},-\frac{1}{2})+12J_W(\frac{1}{2},-\frac{3}{2})$	$3\sigma_\Omega^2\tau_{c\Omega}$	$\frac{3}{2}\sigma_\Omega$

However, of greatest interest for us is the case of a quadrupole nucleus with spin $I=3/2$ since most of the available experimental results on observation of multiquantum echo signals were obtained for just such nuclei. In the calculation of the transverse relaxation time we will consider echo signals formed at the time $t=2\tau$ at the frequency of the central transition and at the frequencies of the quadrupole satellites. We also take into account the multiquantum echo formed at the frequency of the central transition at the time $t=4\tau$ (Refs. 4 and 5).

Table I gives the magnetic quantum numbers corresponding to each of the echo signals, and expressions for the transverse relaxation time obtained using expressions (12) and (16). In addition, Table I gives expressions for the transverse relaxation rate due to fluctuations of the secular part of the spin Hamiltonian. These latter expressions are based on results of Ref. 10 in the limiting cases of fast Gauss–Markov and fast Lorentz–Markov processes of spectral diffusion. Here we used the following notation: σ_Ω and $\tau_{c\Omega}$ for the amplitude of the fluctuations and the correlation time for the longitudinal component of the local magnetic field; σ_Q and τ_{cQ} for the amplitude of the fluctuations and the correlation time of the secular part of the Hamiltonian of the quadrupole interactions. The restriction to fast processes is dictated by the fact that only in the limiting case $\sigma_\Omega\tau_{c\Omega}^{-1}\gg 1$ or $\sigma_Q\tau_{cQ}^{-1}\gg 1$ will an increase in the time interval between the exciting pulses τ lead to exponential decay of the echo (15). For both relations between the correlation time and the amplitude of the fluctuations the decay of the echo is nonexponential.

It follows from the results presented in Table I that the fluctuations of the nonsecular part of the spin Hamiltonian contribute to the decay of the signals of the nuclear spin echo not only at the frequencies of the quadrupole satellites, but also at the frequency of the central transition. At the same time, there is no contribution from the fluctuations of the secular part of the quadrupole Hamiltonian at the frequency of the central transition.[11]

On the basis of the data in Table I it is possible to introduce a parameter R, equal to the ratio of the transverse relaxation rate for the 4τ echo to the relaxation rate of the 2τ echo at the frequency of the central transition

$$R=T_2^{-1}(4\tau)/T_2^{-1}(2\tau). \tag{18}$$

For fast Gauss–Markov and fast Lorentz–Markov processes of spectral diffusion the parameter R takes the values $R=3$ and $R=1.5$, respectively. In the case where the decay of the echo signals is caused exclusively by fluctuations of the nonsecular part of the quadrupole Hamiltonian, we have $R=1$. If the dominant contribution to T_2^{-1} is governed by the spectral density of the fluctuations of the transverse component of the local magnetic field at the frequency of the central transition $(\pm 1/2\leftrightarrow\mp 1/2)$, then $R=0.75$. In general, for nonsecular fluctuations $0.75\leqslant R\leqslant 1$. Thus, the dimensionless parameter R can be used as a criterion to determine the type of relaxation process.

Also note that to analyze the relaxation contribution due to fluctuations of the nonsecular part of the spin Hamiltonian, it has been traditional to use a correlation function of the form

$$K_f(T)=\sigma_f^2\exp(-|T|/\tau_{cf}), \tag{19}$$

where the subscript f, as before, can take one of three possible values: ω, V, and W; σ_f is the amplitude of the fluctuations, and τ_{cf} is the correlation time of the fth term of the Hamiltonian (2). Substituting expression (19) in Eq. (10), we obtain

$$J_f(m,m') = \frac{2\sigma_f^2 \tau_{cf}}{1+\omega_{mm'}^2 \tau_{cf}^2}, \tag{20}$$

where the frequency $\omega_{mm'} = |E_m - E_{m'}|$ (in units $\hbar = 1$) is determined by the energy levels of the unperturbed Hamiltonian (1).

4. NMR OF ^{63}Cu AND ^{65}Cu IN $CuCr_2S_4$: Sb NUCLEI

To investigate the echo signals from copper nuclei in the ferromagnet copper sulfochromite, we used a pulsed incoherent NMR spectrometer. We examined polycrystalline $CuCr_{1.98}Sb_{0.02}S_4$ samples at $T = 77$ K in the absence of a constant external magnetic field.

It is known[13] that in undoped $CuCr_2S_4$ the spectral lines having their maximum amplitude near the frequencies 100.8 and 107.9 MHz correspond to resonances of the copper isotopes ^{63}Cu and ^{65}Cu. In NMR spectra recorded on the 2τ echo signal in $CuCr_{1.98}Sb_{0.02}S_4$ we detected two spectral lines corresponding to the two copper isotopes. However, the positions of the spectral maximum in the doped compound are shifted toward lower frequencies by 0.150 ± 0.05 MHz, which is due to a weakening of the exchange interactions brought about by the introduction of Sb ions.

Along with the ordinary two-pulse echo signal formed at the time $t = 2\tau$, we detected an additional echo signal in the doped compound at $t = 4\tau$. The maximum amplitude of the 4τ echo signal is observed at frequencies corresponding to the maximum amplitude of the 2τ echo. Both copper isotopes are quadrupole nuclei, and the spin of each of them $I = 3/2$. For such nuclei the echo signal formed at the time $t = 4\tau$ can be considered as a multiquantum echo. The experimentally observed peculiarities of formation of an additional echo signal at 4τ indicates that the given echo signal is indeed a multiquantum signal.[5]

The formation of multiquantum echo signals is possible in the case where the magnitude of the quadrupole splitting of the NMR splitting is nonzero. Ferromagnetic copper sulfochromite has a spinel structure in which the copper ions occupy tetrahedral sites. The local symmetry of these sites is cubic, and the gradient of the electric field for such sites in an ideal lattice should be equal to zero. The experimentally observed lowering of the local symmetry of the tetrahedral sites is apparently due to the influence of the Sb impurity and other defects of the crystalline lattice.

Forgoing a detailed analysis of the NMR spectra of the copper nuclei, let us consider the relaxational properties of ordinary and multiquantum echo signals. To investigate processes of nuclear magnetic relaxation, we measured the dependence of the amplitude of the echo signal on the time interval between the exciting pulses τ. Exponential decay was observed for all the investigated echo signals. The decay rate T_2^{-1} was determined by fitting the experimentally observed dependences of the amplitudes of the echo signals on the time interval using expression (15). Values of T_2^{-1} obtained for the frequencies of the spectral maxima of the 2τ echo are given in Table II.

TABLE II. Transverse relaxation rate in $CuCr_{1.98}Sb_{0.02}S_4$.

Nucleus	T_2^{-1}, μs^{-1}	
	Echo 2τ	Echo 4τ
^{63}Cu	$(1.72 \pm 0.07) \times 10^{-3}$	$(3.57 \pm 0.35) \times 10^{-3}$
^{65}Cu	$(1.74 \pm 0.07) \times 10^{-3}$	$(3.60 \pm 0.35) \times 10^{-3}$

The value of the parameter $R = 2.08 \pm 0.32$, obtained on the basis of the data in Table II, does not allow us to assign a preference to any of the random processes considered here; however, it allows us unambiguously to reject the Lorentz–Markov process of spectral diffusion. To interpret the experimental results, we assume that the decay rate of the echo signals (T_2^{-1}) has both a secular (T_{2s}^{-1}) and a nonsecular (T_{2n}^{-1}) contribution, and that this is valid for the ordinary and for the multiquantum echo:

$$T_2^{-1}(2\tau) = T_{2s}^{-1}(2\tau) + T_{2n}^{-1}(2\tau), \tag{21}$$

$$T_2^{-1}(4\tau) = T_{2s}^{-1}(4\tau) + T_{2n}^{-1}(4\tau). \tag{22}$$

Thus, we obtain the following relation between the nonsecular and secular contributions to the decay of the 2τ echo:

$$\frac{T_{2n}(2\tau)}{T_{2s}(2\tau)} = \frac{R_e - R_n}{R_s - R_e}, \tag{23}$$

where R_s and R_n are the values of the parameter R for the secular and nonsecular fluctuations, and R_e is the experimentally observed value of the parameter R. Assuming that the secular contribution is due to the Gauss–Markov process, and substituting numerical values, we obtain $T_{2n}(2\tau)/T_{2s}(2\tau) \approx 1$. This result implies that the contributions to the decay rate of the two-pulse echo due to fluctuations of the secular and nonsecular parts of the spin Hamiltonian are comparable in magnitude.

5. NMR OF ^{53}Cr IN $CdCr_2S_4$: Ag

Abelyashev *et al.*[12] reported the results of an experimental study of the decay of the 2τ echo signal from ^{53}Cr nuclei in the ferromagnet $Cd_{0.985}Ag_{0.015}Cr_2Se_4$ at $T = 4.2$ K. A theoretical analysis was carried out within the framework of the theory of spectral diffusion based on results of Ref. 11. In particular, it was shown in Ref. 12 that the Gauss–Markov process of spectral diffusion allows one to explain the experimentally observed dependence of the transverse relaxation time on the frequency. The solid curve *1* in Fig. 1 represents the calculated dependence obtained in Ref. 12 by fitting the experimental data.

In the present paper we investigate the decay of a multiquantum 4τ echo from ^{53}Cr nuclei in the same sample that was used in Ref. 12. We have established experimentally that at $T = 4.2$ K the decay of a multiquantum 4τ echo is exponential and that the relaxation time takes different values at different points of the spectrum. The values of transverse relaxation time obtained by fitting the experimentally observed dependences with expression (15) are plotted in Fig. 1 by open and filled circles versus the frequency.[1)]

As follows from the data plotted in Fig. 1, the frequency dependence of the relaxation time T_2 for the 4τ echo differs from the dependence for the 2τ echo. To interpret the experi-

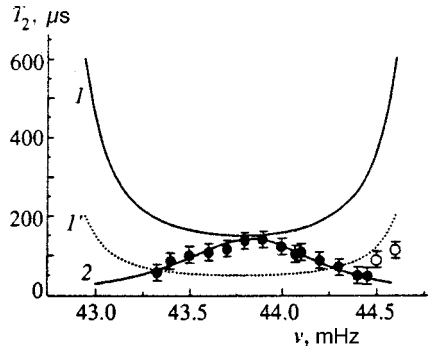

FIG. 1. Frequency dependence of the transverse relaxation time of ^{53}Cr nuclei in $Cd_{0.985}Ag_{0.015}Cr_2Se_4$ at $T=4.2$ K: open and filled circles — experimental points for the 4τ echo; curves *1* and *1′* — Gauss–Markov process of spectral diffusion (*1* — theoretical dependence for the 2τ echo from Ref. 2, *1′* — theoretically expected value for the 4τ echo); *2* — nonsecular contribution to decay of the 4τ echo (experimental points fitted by the theoretical dependence (24).

mentally observed frequency dependence of the relaxation time T_2 measured for the 4τ echo, we assume that the decay of this signal is due to fluctuations of the nonsecular part of the spin Hamiltonian.

Cadmium selenochromite has the structure of a normal spinel, in which the chromium ions occupy octahedral sites. The local symmetry of these sites is trigonal, and the resonance frequency depends on the angle θ—the angle between the electron magnetization vector and the crystallographic direction $\langle 111\rangle$. The NMR signal at different frequencies is created by nuclei of ions with different values of θ.

The coefficients entering into the Hamiltonian (2) also depend on the angle θ, specifically, $\omega_{\pm}\propto\sin 2\theta$, $V_{\pm}\propto\sin 2\theta$, and $W_{\pm}\propto\sin^2\theta$ (Ref. 1). Following Ref. 12, we assume that the source of the fluctuations is orientational inhomogeneity of the electronic magnetization vector. Thus, the amplitude of the fluctuations is determined by the derivative of the corresponding coefficient of the angle θ: $\sigma_f\sim|\partial f/\partial\theta|$, where f takes the values ω, V, and W. Keeping the terms in the Hamiltonian (2) with $\omega_{\pm}$ and $V_{\pm}$, from expressions (16), (12), and (20) we obtain $T_2^{-1}\propto\cos^2 2\theta$. The solid curve *2* in Fig. 1 was obtained using the dependence

$$T_2^{-1}(\theta)=A+B\,\cos^2 2\theta. \tag{24}$$

The coefficients A and B were chosen so as to achieve the best fit of the calculated dependence (24) to the experimentally observed frequency dependence of T_2 for the 4τ echo. The coefficient A in expression (24) takes into account the isotropic contribution to the decay of the echo signal. The relation between the angle θ and the resonance frequency ν is given by the relation

$$\nu=\nu_0+\nu_A(3\cos^2\theta-1) \tag{25}$$

where $\nu_0=44.07\pm0.03$ MHz, and $\nu_A=-0.55\pm0.02$ MHz (Ref. 12). The fitted dependences in Fig. 1 correspond to $A\approx 6.76\times10^{-3}\,\mu s^{-1}$ and $B\approx 23.26\times10^{-3}\,\mu s^{-1}$.

As follows from the data plotted in Fig. 1, in the larger part of the frequency spectrum (the filled circles in Fig. 1) the fluctuations of the nonsecular part of the spin Hamiltonian due to orientational inhomogeneity of the electronic magnetization vector provide a good description of the frequency dependence of the relaxation time.

According to the data in Table I and expression (10), the contribution of the terms of Hamiltonian (2) with coefficients $\omega_{\pm}$ and $V_{\pm}$ is proportional to the spectral density of the random process at the frequencies ω_0 and $\omega_0\pm2\omega_q$ while the contribution of the term with coefficients $W_{\pm}$ is determined by the spectral density of the random process at the frequency $2(\omega_0\pm\omega_q)$. Clearly, a situation is realized in $CdCr_2Se_4$:Ag, where the dominant contribution to decay of the multiquantum echo signals is determined by the spectral density of the random process at the resonance frequencies.

The decay of the echo signals formed at the time $t=2\tau$ is determined by the Gauss–Markov process of spectral diffusion, which is reflected by the solid curve *1* in Fig. 1. It follows from the data of Table I that the transverse relaxation time for the multiquantum 4τ echo in this case should be three times as small. The theoretically expected value of T_2 for the 4τ echo in the case of the Gauss–Markov process is reflected by the dotted curve *1′* in Fig. 1. As follows from the data plotted in Fig. 1, the Gauss–Markov process is responsible for the decay of the 4τ echo only in the high-frequency region of the spectrum (the open circles in Fig. 1).

To interpret the above results, we assume that there are two types of ^{53}Cr nuclei in $CdCr_2Se_4$:Ag. The main contribution to the magnetic relaxation of nuclei of the first type is due to fluctuations of the secular part of the spin Hamiltonian. From these nuclei is observed the 2τ echo, whose properties are discussed in Ref. 12. The multiquantum 4τ echo from nuclei of the first type is observed experimentally only in the high-frequency region (the open circles in Fig. 1). The absence of an echo in the low-frequency region is due to the low intensity of the signal, and in the intermediate-frequency region it is attributable to the inordinately short relaxation time (the dotted curve *1′* in Fig. 1).

Magnetic relaxation of ^{53}Cr nuclei of the second type is due to fluctuations of the nonsecular part of the spin Hamiltonian. The signal from these nuclei is observed as a multiquantum 4τ echo in the intermediate-frequency region. The absence of a signal in the high-frequency region is due to the inordinately short relaxation time (the solid curve *2* in Fig. 1). The absence of a signal in the low-frequency region is due to the short relaxation time and the low intensity of the signal. In our discussion of the theoretical results we noted that in the case where the decay of the echo is due to nonsecular fluctuations, the relaxation time for the 2τ echo is the same or somewhat smaller than the relaxation time for the 4τ echo. It follows from a comparison of the solid curves *1* and *2* in Fig. 1 that the 2τ echo from the second type of nuclei is not observed experimentally because of the small value of T_2.

A possible reason for the appearance of two types of nuclei may be the following. Doping with silver leads to the appearance of Cr^{4+} ions as a result of valence compensation. These ions are not localized, but migrate in some vicinity of the impurity ion Ag^+. For the low concentrations of the silver impurity considered by us the regions of migration of the Cr^{4+} ions which belong to various impurity centers do not

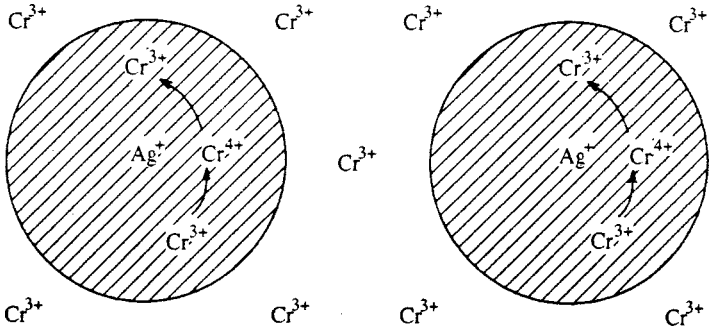

FIG. 2. Cation distribution in $CdCr_2Se_4$:Ag.

overlap each other (Fig. 2). Type-I ^{53}Cr nuclei belong to Cr^{3+} ions which are located outside the regions of migration of the Cr^{4+} ions. Type-II ^{53}Cr nuclei belong to Cr^{3+} ions which are located in the regions of migration of the Cr^{4+} ions. Division of the nuclei into two types according to the results of our study of magnetic relaxation in impurity magnets is similar to the division made in Ref. 14.

6. CONCLUSIONS

The formation of nuclear spin echo signals in magnetically ordered materials with quadrupole nuclei is generally accompanied by a selective excitation, in which each of the echo signals is observed separately from the others. Separate observation is ensured by different formation times of the signals, and for the same formation times separate observation is possible because of formation of the quadrupole spectrum at different frequencies. In our theoretical analysis we did in fact consider the situation of selective excitation since each echo signal was characterized by its own set of magnetic quantum numbers m, m', and m''. A consequence of selective excitation was the fact that decay of the echo is described by one exponential.

In the discussion of the experimental results we demonstrated that in an analysis of the rate of decay of the echo signals it is necessary to take into account both secular and nonsecular contributions to T_2^{-1}. In general, various mechanisms may be responsible for the fluctuations of the secular and nonsecular parts of the spin Hamiltonian. Therefore, the secular and nonsecular fluctuations should, in general, be characterized by coherence times and a different amplitude of the fluctuations.

We have demonstrated here the use of the theoretical results obtained by us for the analysis of experimental data for quadrupole nuclei with spin $I=3/2$. However, these results can be used for quadrupole nuclei with any spin.

We wish to thank T. G. Aminov, E. V. Bushev, and G. G. Shabunin for providing samples of chromium chalcogenide spinels. This work was carried out with the partial support of ISSEP (Grant No. APU072083).

*) E-mail: roton@ccssu.crimea.ua
†) Szczecin University, 70-451 Szczecin, Poland

1) The representation of the experimental results by different symbols is motivated by the results of the theoretical analysis presented below.

[1] A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon Press, Oxford, 1961).
[2] M. I. Kurkin and E. A. Turov, *NMR in Magnetically Ordered Materials and Its Applications* (Nauka, Moscow, 1990).
[3] I. Solomon, Phys. Rev. **110**, 61 (1958).
[4] G. N. Abelyashev, V. N. Berzhanskiĭ, N. A. Sergeev, and Yu. V. Fedotov, Zh. Éksp. Teor. Fiz. **94**(1), 227 (1988) [Sov. Phys. JETP **67**, 127 (1988)].
[5] G. N. Abelyashev, V. N. Berzhanskiĭ, S. N. Polulyakh, N. A. Sergeev, and Yu. V. Fedotov, Zh. Éksp. Teor. Fiz. **100**, 1981 (1991) [Sov. Phys. JETP **73**, 1096 (1991)].
[6] V. O. Golub, V. V. Kotov, A. N. Pogorelyĭ, and Yu. A. Pod'elets, Fiz. Tverd. Tela (Leningrad) **31**, 48 (1989) [Sov. Phys. Solid State **31**, 1864 (1989)].
[7] V. O. Golub, V. V. Kotov, and A. N. Pogorelyĭ, Fiz. Tverd. Tela (St. Petersburg) **40**, 1056 (1998) [Phys. Solid State **40**, 850 (1998)].
[8] H. E. Rorschach, J. Magn. Reson. **67**, 519 (1986).
[9] C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed. (Springer-Verlag, Berlin, 1990).
[10] R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, edited by R. Breslow, J. Halpern FRS, and J. S. Rowlinson FRS (Clarendon Press, Oxford, 1990)
[11] S. N. Polulyakh and N. A. Sergeev, Zh. Éksp. Teor. Fiz. **108**, 14 (1995) [JETP **81**, 7 (1995)].
[12] G. N. Abelyashev, V. N. Berzhansky, Yu. V. Fedotov, S. N. Polulyakh, and N. A. Sergeev, J. Magn. Magn. Mater. **184**, 222 (1998).
[13] H. Yokoyama, R. Watanabe, and S. Chiba, J. Phys. Soc. Jpn. **22**, 659 (1967).
[14] V. N. Berzhanskiĭ, A. I. Gorbovanov, S. N. Polulyakh, and N. V. Pronina, Fiz. Tverd. Tela (St. Petersburg) **40**, 1494 (1998) [Phys. Solid State **40**, 1265 (1998)].

Translated by Paul F. Schippnick

Electron polarization in quantum wells in a strong variable field

V. A. Burdov*)

N. I. Lobachevskiĭ Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia
(Submitted 19 October 1998)
Zh. Éksp. Teor. Fiz. **116**, 217–235 (July 1999)

The wave function of an electron in a symmetric double quantum well placed in a strong time-periodic electric field is found, expressions for quasienergy functions are derived, and the dependence of the dipole moment on the average electric field is analyzed for the case where the average field remains constant. In the case of slow monotonic variation of the ''constant'' component of the electric field, the Schrödinger equation is solved by the WKB method. It is found that the dependence of the dipole moment on the average field is of a clearly nonlinear almost-periodic nature and that in the event of adiabatic monotonic variation of the average field there is a periodic relocation of the electron density from well to well with a small frequency proportional to the rate of variation of the average field. © *1999 American Institute of Physics.* [S1063-7761(99)01507-3]

1. INTRODUCTION

Studies of the electron dynamics in quantum wells activated by variable external fields are of undisputable interest from the standpoint of exposing the potential of modern nanostructures and electron devices based on such structures. The methods developed in such studies make it possible to directly control the wave function of electrons in quantum wells by varying the parameters of a classical external electromagnetic field.[1] In this paper we will examine the effects of electron polarization in a symmetric double quantum well placed in a strong time-periodic external field with a finite average value.

Earlier, in relation to heterostructures, the concept of a double quantum well was used to investigate the phenomenon of dynamic localization of the electron wave function in one well initiated by a sinusoidal external field,[2–4] while in Refs. 5–7 the dipole moment of the system was calculated and the possibility of emission of low-frequency dipole radiation by such a structure was investigated. In Ref. 8 it was found that dynamic localization and the related emission of low-frequency dipole radiation are possible in any periodic field (not necessarily sinusoidal), provided that the field is strong enough. The oscillations of the electron wave packet and the electromagnetic radiation generated by these oscillations were observed in experiments.[9–11] Finally, in Refs. 1 and 12 the dependence of the electron distribution in the wells on the way in which the external periodic field is switched on was studied.

In the above papers the attention was focused on establishing the conditions imposed on the amplitude and frequency of the external field needed for ''locking'' the wave packet in one well, or the regime of low-frequency electromagnetic radiation generation. At the same time, it is still unclear how a constant voltage applied to the structure affects the quantum dynamics of a system driven by high-power laser light. Gorbatsevich *et al.*[1] and Dakhnovskii *et al.*[6] studied the dynamics of the well populations in the presence of a constant external field, and in Ref. 8 it was shown that the constant component has a stabilizing effect on the existence of states in the region of one well. On the whole, however, an analysis of the dependence of the above effects on the external constant field has yet to be done. The present paper is probably the first attempt of such an analysis.

2. EQUATIONS OF EVOLUTION OF THE SYSTEM

We will examine the electron dynamics in the symmetric structure of two quantum wells separated by an impenetrable barrier in the presence of a strong time-periodic electric field $\mathscr{E}$ directed along the axis of the structure. We will assume that below the top of the barrier there are only two energy levels of the unperturbed system, $E_{0,1}=\pm\hbar\Delta/2$ (the zero of energy is chosen exactly midway between the levels), with the ''distance'' $\hbar\Delta$ determined by the tunnel integral through the separating barrier (the integral is exponentially small). We will define the symmetric and antisymmetric wave functions $\chi_{0,1}(\xi)$ corresponding to these energies so that $\chi_0(x)$ is always positive and $\chi_1(x)$ is positive only if $x>0$.

Since the distance from the lower energy levels to the next level E_2 is much larger than the energy splitting $\hbar\Delta$, we will use the two-level approximation throughout the paper, assuming, naturally, that the external field is unable to mix the high-lying levels with the lower levels. Thus, we must impose the restriction $\hbar\Delta \ll V_{01} \ll E_2$, where $V_{01}=e\mathscr{E}x_{01}$ is the matrix element of the perturbation in the dipole approximation, and x_{01} is the coordinate's matrix element. The above inequalities concretize the idea of a strong field: the field must be so strong that the matrix element of the perturbation is much larger than the distance between the levels, $\hbar\Delta$.

Finally, we will require that the frequency ω of the external field be much smaller than $E_2/\hbar$. This will enable us to exclude resonant transitions to higher levels.

Since we are using the two-level approximation, it would seem natural to look for the wave function of a particle in the form of an expansion in the basis consisting of the two stationary states of the unperturbed system. It is clear, however, that this is unlikely to be the most convenient expansion in the presence of a strong variable field. Since below we are interested in the spatial charge distribution, a more appropriate basis is

$$\Psi_L(x)=\frac{\chi_0(x)-\chi_1(x)}{\sqrt{2}},\quad \Psi_R(x)=\frac{\chi_0(x)+\chi_1(x)}{\sqrt{2}},$$

where the functions $\Psi_{L,R}(x)$ are completely localized in the left and the right well, respectively, in accordance with the definition of the stationary-state functions $\chi_{0,1}(x)$.

We can now write the wave function of the system as a linear superposition of the orthonormal vectors $\Psi_{L,R}(x)$ with coefficients that are time-dependent and have yet to be determined:

$$\Psi(x,\tau)=C_L(\tau)\exp\left\{i\int\varepsilon(\tau)d\tau\right\}\Psi_L(x)+C_R(\tau)\times\exp\left\{-i\int\varepsilon(\tau)d\tau\right\}\Psi_R(x). \quad (1)$$

Here we have introduced the dimensionless time variable $\tau=\Delta t$, while $\varepsilon(\tau)$ is the ratio of the perturbation matrix element to the transition energy $\hbar\Delta$ and is a periodic function with a dimensionless period $T=2\pi\Delta/\omega$. We immediately note that $\varepsilon(\tau)$ consists of two parts: the constant part $\overline{\varepsilon}$, which is the value of $\varepsilon(\tau)$ averaged over the period, and the variable part $\tilde{\varepsilon}(\tau)$, whose average is zero. For convenience the exponential phase factors in the expansion coefficient in (1) are written explicitly.

Although $\Psi_L(x)$ and $\Psi_R(x)$ are not eigenfunctions of the operator of a physical quantity, we can attach a definite physical meaning to the coefficients $C_{L,R}(\tau)$ (more precisely, to the squares of their absolute values). If we define the probability of finding the electron in the left or right well as the integral

$$W_{L(R)}=\int_{-\infty(0)}^{0(\infty)}dx|\Psi(x,\tau)|^2,$$

we can easily show that, to within terms whose magnitude is determined by the overlap integral $\int\Psi_L(x)\Psi_R(x)dx$, the probability $W_{L,R}(\tau)$ coincides with the square of theabsolute values of $C_{L,R}(\tau)$. The value of the overlap integral is given by the difference $\chi_0^2(x)-\chi_1^2(x)$, which, as shown in Ref. 13, is small in the parameter $(\hbar\Delta/E_2)^{1/2}$, and according yo the initial assumptions this ratio is the smallest parameter in the problem. In our future calculations we will ignore quantities of this order.

Plugging (1) into the Schrödinger equation yields a pair of equations describing the evolution of the coefficients $C_{L,R}(\tau)$:

$$i\frac{dC_L}{d\tau}=-\frac{C_R}{2}\exp\left\{-2i\int\varepsilon(\tau)d\tau\right\},$$

$$i\frac{dC_R}{d\tau}=-\frac{C_L}{2}\exp\left\{2i\int\varepsilon(\tau)d\tau\right\}. \quad (2)$$

Since $\tilde{\varepsilon}(\tau)$ is a periodic function, we can expand the exponentials in (2) in a Fourier series:

$$\exp\left\{2i\int\varepsilon(\tau)d\tau\right\}=\exp\left\{2i\int\overline{\varepsilon}d\tau\right\}\sum_{n=-\infty}^{\infty}\mu_n\times\exp\{i\psi_n\}\exp\{-in\Omega\tau\},$$

where $\Omega=2\pi/T$, and μ_n and ψ_n are the coefficient and phase of the Fourier expansion, which are related by the formula

$$\mu_n\exp\{i\psi_n\}=\frac{1}{T}\int_0^T d\tau\exp\left\{i\left(n\Omega\tau+2\int\tilde{\varepsilon}(\tau)d\tau\right)\right\}.$$

The explicit expressions for the Fourier coefficients μ_n were obtained in Ref. 8 by the method of stationary phase. Here we will not discuss all the calculations done in Ref. 8—we will only mention the main results.

In accordance with their definition, the coefficients μ_n depend only on the amplitude of the variable part of the perturbation (we denote it by ε_0) and are independent of $\overline{\varepsilon}$. When the amplitude ε_0 is large, which is the case in our problem, the μ_n always decrease with increasing ε_0 either as a power function (as $\varepsilon_0^{-1/3}$ or $\varepsilon_0^{-1/2}$) or exponentially. In both cases the Fourier coefficients μ_n turn out to be small, and this fact substantially simplifies the solution of the problem.

Representing the exponentials in Eqs. (2) by Fourier series, we obtain on the right-hand sides of these equations a set of terms with phases of the form $\psi_n+\int(2\overline{\varepsilon}-n\Omega)d\tau$. Obviously, when the difference $2\overline{\varepsilon}-n\Omega$ is of order Ω, all harmonics in Eqs. (2) are rapidly oscillating functions, and their average effect on the system is essentially nil. But if at $n=l$ the difference $|2\overline{\varepsilon}-n\Omega|$ is much smaller than Ω, the lth term acquires a ''slow'' phase and the effect of this lth harmonic may be resonant (see, e.g., Ref. 14., p. 180). The case of resonant excitation of a two-level system is the most interesting one, and we will discuss it later in this paper.

The condition for resonance,

$$2\overline{\varepsilon}=l\Omega, \quad (3)$$

has a clear physical meaning. The constant component of the perturbation moves the energy levels of stationary states apart by a distance of approximately $2\overline{\varepsilon}$, provided that $2\overline{\varepsilon}\gg 1$. Hence to couple these levels in a resonant manner the system needs exactly the same energy as one photon (or several photons) of the external field has. The lth harmonic of the Fourier expansion is in full agreement with this requirement, and it is this harmonic that couples these two levels.

Note, however, that we are speaking of the lth harmonic of the exponential $\exp(2i\int\tilde{\varepsilon}(\tau)d\tau)$ rather than the lth harmonic of the perturbation. When the amplitude of the function $\tilde{\varepsilon}(\tau)$ is a quantity of order the distance $2\overline{\varepsilon}$ between the levels or is even greater than this distance, the contribution to the resonant transition is provided not only by one-photon processes with the frequency $l\Omega$ but also by various multiphoton processes with a finite difference in the energies of

the emitted and absorbed photons equal to $l\Omega$. Thus, the Fourier coefficient μ_l appears to be the amplitude of all these resonant processes. In the case of small perturbations ε_0 it can easily be verified directly that the coefficient μ_l describes a one-photon process involving the lth harmonic of the perturbation.

Obvious, the system of equations (2) can easily be solved when the average value of the perturbation is close to a resonant value and the condition (3) is met either exactly or at least approximately, i.e., $|2\bar{\varepsilon}-l\Omega|\ll\Omega$. In this case we need only keep the ''slow'' resonant term on the right-hand sides of Eqs. (2) (the lth harmonic of the Fourier series), the term that makes the main contribution to the evolution of the expansion coefficients $C_{L,R}(\tau)$.

Obviously, knowing the expansion coefficients $C_{L,R}(\tau)$ and hence the wave function only in the vicinity of the lth resonance is not enough to be able to analyze their dependence on $\bar{\varepsilon}$ over the entire range of variation of the constant component of the perturbation. Such analysis requires knowing the laws that govern the transition from one resonance to another through the nonresonant region, and this knowledge cannot be supplied by the one-resonance approximation alone.

Of course, one could solve the system (2) near the lth and $(l+1)$st resonances and in the nonresonant region. But then the problem arises of combining all these expressions into one expression that would be valid over the entire range under investigation, which probably is impossible from the mathematical viewpoint.

For this reason we will develop an entirely new method of solution, which will be called the two-resonance approximation. The approach is based on the idea of keeping two neighboring harmonics in (2) rather than one harmonic, with frequencies that are closest to the frequency of the transition between the energy levels. For instance, if $l\Omega\leqslant 2\bar{\varepsilon}\leqslant(l+1)\Omega$, we must keep the lth and $(l+1)$st harmonics. Equations (2) then become

$$i\frac{dC_L}{d\tau}=-\frac{C_R}{2}\left(\mu_l\exp\left\{-i\psi_l-i\int(2\bar{\varepsilon}-l\Omega)d\tau\right\}+\mu_{l+1}\exp\left\{-i\psi_{l+1}-i\int(2\bar{\varepsilon}-(l+1)\Omega)d\tau\right\}\right).$$

$$i\frac{dC_R}{d\tau}=-\frac{C_L}{2}\left(\mu_l\exp\left\{i\psi_l+i\int(2\bar{\varepsilon}-l\Omega)d\tau\right\}+\mu_{l+1}\exp\left\{i\psi_{l+1}+i\int(2\bar{\varepsilon}-(l+1)\Omega)d\tau\right\}\right). \quad (4)$$

Clearly, the solution of this system (if it can be obtained) will have all the necessary properties. At values of $2\bar{\varepsilon}$ close to $l\Omega$ this solution coincides with that obtained in the one-resonance approximation discussed earlier. At values of $2\bar{\varepsilon}$ close to $(l+1)\Omega$ the solution coincides with that obtained in the one-resonance approximation, but near the $(l+1)$st resonance. It is also clear that a more general solution of this form yields universal quasienergy functions of the system over the entire region separating the resonances and a complete picture of the dependence of the quasienergies on $\bar{\varepsilon}$.

It is easy to verify that the error of the two-resonance approximation is of the same order of smallness as the error of the ordinary resonance approximation. As in the one-resonance approximation, in deriving (4) from (2) we discarded the nonresonant harmonics with small Fourier coefficients, $\mu_n\ll 1$, and this formally determines the order of the error.

Note that in a certain sense allowing for the $(l+1)$st harmonic near the lth resonance (or for the lth harmonic near the $(l+1)$st resonance) is superfluous, since these harmonics are nonresonant and can be discarded, as well as the other nonresonant harmonics. The only reason for keeping these harmonics in Eqs. (4) is to obtain equations that describe the dynamics of the system in the entire interval between the neighboring resonances. The fact that these harmonics become nonresonant in turn guarantees that in our calculations there will be no significant buildup of error and that the error will become no larger than the Fourier coefficients μ_n.

3. QUASIENERGY FORMALISM IN THE TWO-RESONANCE APPROXIMATION

We begin by solving Eqs. (4) under the assumption that the average value of the perturbation remains unchanged and hence the system Hamiltonian is strictly periodic in time with a period equal to that of the external field $T=2\pi/\Omega$. It is well known[15] that in this case it is convenient to use the formalism of quasienergies and quasienergy functions with the property $U_\nu(x,\tau+T)=U_\nu(x,\tau)\exp\{-i\nu T\}$, where ν, which is defined to within the frequency Ω of the external field, is called the quasienergy. The quasienergy functions corresponding to different quasienergies form an orthonormal basis, and the particle wave function can be expanded in this basis. Since in a two-level system the number of quasienergies is equal to two (we will denote them by $\nu_\pm$), the basis of the quasienergy functions is also two-dimensional. As a result we arrive at an expansion

$$\Psi(x,\tau)=A_+U_+(x,\tau)+A_-U_-(x,\tau), \quad (5)$$

where $A_\pm$ are the expansion coefficients.

Since the wave function $\Psi(x,\tau)$ and the quasienergy functions $U_\pm(x,\tau)$ are assumed normalized, the squares of the absolute values of the expansion coefficients $A_\pm$ must sum to unity. Moreover, the coefficients in the expansion in the basis of the quasienergy functions are time-independent,[15] in contrast to the expansion in any other base functions as, say, in (1). In accordance with the ordinary rules of quantum mechanics, we can attach a definite physical meaning to the squares of the absolute values of the coefficients $A_\pm$: they define the probability of finding the system in a given quasienergy state with the corresponding quasienergy.

We will now consider to the problem of calculating the quasienergy functions and the quasienergies. The quasienergy function $U_\nu(x,\tau)$ must be a Bloch-type solution of the Schrödinger equation. In accordance with these requirements and with allowance for the structure of the expansion (1), we seek the coefficients $C_{L,R}(\tau)$ in the form

$$C_L(\tau)=f(\tau)\exp\left\{-i\left(\nu+\frac{\delta}{2}\right)\tau\right\},$$

$$C_R(\tau)=g(\tau)\exp\left\{-i\left(\nu-\frac{\delta}{2}\right)\tau\right\}, \quad (6)$$

where $\delta=2\overline{\varepsilon}-l\Omega$. For the sake of definiteness we will assume that the order l of the resonance is even. In this case the functions $f(\tau)$ and $g(\tau)$ (which have yet to be found) must (according to the Floquet theorem) be periodic with a period equal to that of the external field, T, and can be expanded in Fourier series:

$$f(\tau)=\sum_{n=-\infty}^{\infty} f_n \exp\{-in\Omega\tau\},$$

$$g(\tau)=\sum_{n=-\infty}^{\infty} g_n \exp\{-in\Omega\tau\}.$$

Substituting $f(\tau)$ and $g(\tau)$ into (4), we arrive at an infinite-dimensional system of algebraic equations for finding the Fourier coefficients f_n and g_n:

$$\left(\nu+\frac{\delta}{2}+n\Omega\right)f_n+\frac{\mu_l\exp\{-i\psi_l\}}{2}g_n+\frac{\mu_{l+1}\exp\{-i\psi_{l+1}\}}{2}g_{n+1}=0,$$

$$\left(\nu-\frac{\delta}{2}+n\Omega\right)g_n+\frac{\mu_l\exp\{i\psi_l\}}{2}f_n+\frac{\mu_{l+1}\exp\{i\psi_{l+1}\}}{2}f_{n-1}=0. \quad (7)$$

Generally speaking, solving the system of equations (7) is extremely difficult, since many harmonics contribute to the functions $f(\tau)$ and $g(\tau)$. But if we require that the parameters μ_l and μ_{l+1} be small, which is the case in a strong variable field, the sequences of the f_n and g_n decrease rapidly with respect to one of the coefficients, which is the largest, both in the direction of increasing numbers n and in the direction of decreasing numbers n. Here keeping only one expansion coefficient (at most two) and discarding all the others turns out to be sufficient.

By solving the system of equations (7) (see the Appendix) we find two quasienergy values $\nu_\pm=\pm\nu$ belonging to the first Brillouin zone, where

$$\nu=\frac{\delta^2-\sqrt{\delta^2(\Omega-\delta)^2+(2\delta-\Omega)(\mu_l^2(\delta-\Omega)+\mu_{l+1}^2\delta)}}{2(2\delta-\Omega)}. \quad (8)$$

One can easily establish that for $|\delta|\ll\Omega$ and $|\delta-\Omega|\ll\Omega$ the quasienergies (8) become the ordinary quasienergies of the one-resonance approximation (the analog of the Rabi frequency):

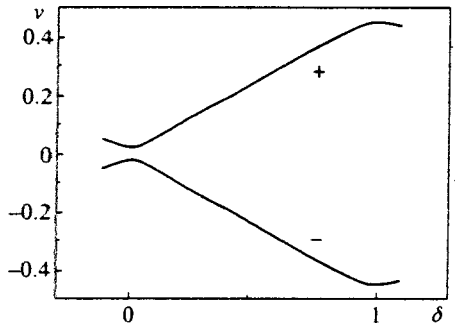

FIG. 1. Dependence of quasienergies on the average field within the first Brillouin zone for the states ''+'' and ''−'' at $\mu_l/\Omega=0.05$ and $\mu_{l+1}/\Omega=0.1$.

$$\nu_\pm=\begin{cases}\pm\dfrac{\sqrt{\delta^2+\mu_l^2}}{2}, & |\delta|\ll\Omega,\\ \pm\dfrac{\Omega-\sqrt{(\delta-\Omega)^2+\mu_{l+1}^2}}{2}, & |\delta-\Omega|\ll\Omega\end{cases}. \quad (9)$$

The complete pattern of the dependence of the quasienergies $\nu_\pm$ on $2\overline{\varepsilon}/\Omega$ corresponding to the general expression (8) is depicted in Fig. 1.

We see that the quasienergy branches are closest at the lth resonance [condition (3)] and are farthest within the first Brillouin zone at values of $\overline{\varepsilon}$ corresponding to the next $(l+1)$st resonance. However, such ''repulsion'' of quasienergy levels within one zone automatically results in their ''attraction'' to the quasienergy levels belonging to the neighboring zones, which are obtained by translation of the first Brillouin zone up or down the quasienergy axis by Ω. Obviously, the minimum distance between the quasienergy branches is achieved at points of resonance and is equal to μ_l or μ_{l+1}.

Note that the fragment of the $\nu_\pm$ vs. $2\overline{\varepsilon}$ dependence depicted in Fig. 1 can easily be continued outside the interval of values of $\overline{\varepsilon}$ for which the calculations were carried out, since the dependence is almost-periodic in $2\overline{\varepsilon}$ with a period equal to 2Ω. For instance, to continue the function $\nu_\pm(2\overline{\varepsilon})$ to the second half of the ''period,'' into the region $(l-1)\Omega\leqslant 2\overline{\varepsilon}\leqslant l\Omega$, it is enough to replace μ_{l+1} by μ_{l-1} and δ by $-\delta$ in (8). After this we can extend the entire ''period'' obtained in this way to any interval on the $2\overline{\varepsilon}$ axis.

Small deviations from periodic behavior will be observed in a small neighborhood of the resonance points $2\overline{\varepsilon}=n\Omega$, where the quasienergy branches are closest to each other or to the zone boundaries. In these neighborhoods the dependence is determined primarily by the coefficients μ_n and, since generally the μ_n are different, there can be no strict periodicity.

Now we will derive the expressions for the quasienergy functions $U_\pm(x.\tau)$. For this we must first determine the coefficients f_0, f_{-1}, g_0, and g_1 from the system of equations (A1) and the normalization condition, which we can write as $|f_0|^2+|f_{-1}|^2+|g_0|^2+|g_1|^2=1$. With allowance for the smallness of the coefficients g_1 and f_{-1} for the branches ''+'' and ''−'', respectively, we have (at $\nu_+=\nu$)

$$g_0^{(+)}=\left(1+\frac{\mu_l^2}{(2\nu+\delta)^2}+\frac{\mu_{l+1}^2}{(2\nu+\delta-2\Omega)^2}\right)^{-1/2}\equiv G,$$

$$f_0^{(+)}=-G\frac{\mu_l\exp\{-i\psi_l\}}{2\nu+\delta},$$

$$f_{-1}^{(+)}=-G\frac{\mu_{l+1}\exp\{-i\psi_l+1\}}{2\nu+\delta-2\Omega}. \tag{10}$$

According to (10), in the interval of values of $\overline{\varepsilon}$ considered here, the coefficient g_0 always plays an important role: it is close to unity almost everywhere in the interval and only at the interval endpoints rapidly decreases to almost zero. The coefficient f_0 is important only near the lth resonance and the coefficient f_-, near the $(l+1)$st resonance.

Reasoning in a similar manner, we arrive at the following expressions in which $\nu_-=-\nu$:

$$f_0^{(-)}=G,\quad g_0^{(-)}=-f_0^{(+)*},\quad g_1^{(-)}=-f_{-1}^{(+)*}. \tag{11}$$

Here the asterisk indicates a complex-conjugate quantity. Now the coefficient f_0 is important in the interval of values of $\overline{\varepsilon}$, while the coefficients g_0 and g_1 are important only at the endpoints of the interval.

We substitute the expressions (10) and (11) for the coefficient into (6) and the result in (1). The quasienergy wave functions $U_\pm(x,\tau)$ are specified as follows:

$$\begin{aligned}U_+(x,\tau)=&\exp\{-i\nu\tau\}[(f_0^{(+)}\\&+f_{-1}^{(+)}\exp\{i\Omega\tau\})\exp\{i\phi_l(\tau)\}\Psi_L(x)\\&+g_0^{(+)}\exp\{-i\phi_l(\tau)\}\Psi_R(x)].\\U_-(x,\tau)=&\exp\{i\nu\tau\}[(g_0^{(-)}+g_1^{(-)}\exp\{-i\Omega\tau\})\\&\times\exp\{-i\phi_l(\tau)\}\Psi_R(x)\\&+f_0^{(-)}\exp\{i\phi_l(\tau)\}\Psi_L(x)].\end{aligned} \tag{12}$$

where $\phi_l(\tau)=\int(\tilde{\varepsilon}(\tau)+l\Omega/2)d\tau$.

Knowing the quasienergy functions (12) will enable us to determine the dipole moment of an electron in the quasienergy states and follow the dependence of the degree of polarization of the system on the magnitude of the constant external field.

4. ELECTRON POLARIZATION IN QUASIENERGY STATES

We will now calculate the dipole moment in quasienergy states. Using the definition of the dipole moment,

$$d^{(\pm)}=-e\langle U_\pm(x,\tau)|\hat{x}|U_\pm(x,\tau)\rangle.$$

the above expressions for the quasienergy functions, and the normalization condition, we arrive at expressions for the dipole moment in the states ''+'' and ''−,'' respectively:

$$d^{(\pm)}=\pm D=\pm ex_{01}(1-2G^2). \tag{13}$$

According to (13), the dependence of the dipole moment D on the double value of the constant component of the external field, $2\overline{\varepsilon}$, is nonmonotonic and, more than that, almost-periodic (as noted earlier, strict periodicity occurs only if all the coefficients μ_n are the same), with the ''period'' determined by the frequency Ω of the variable component of the external field. Here in one ''half-period'' the dipole moment is directed along the constant field and in the other it opposes the field, in both cases reaching maximum values. Thus, in one ''half-period'' these appears an antipolarization effect develops due to the additional action of the strong variable field on the electron.

Let us discuss the Dvs.$2\overline{\varepsilon}$ dependence in detail by using the quasienergy functions (12). To this end we follow the behavior of the degree of localization of one quasienergy function, e.g., $U_+(x,\tau)$, as a function of $\overline{\varepsilon}$. Near the lth resonance the coefficients f_0 and g_0 provide a sizable contribution to $U_+(x,\tau)$. Here, as we move away from the resonance into the region of smaller values of $2\overline{\varepsilon}$, the coefficient g_0 very rapidly decreases to zero on a scale $\delta\sim\mu_l$. while the coefficient f_0 almost equally rapidly becomes almost equal to unity. Thus, for negative values of δ such that $|\delta|\gg\mu_l$, the function $U_+(x,\tau)$ is almost completely localized in the left well.

As we approach the exact equality $\delta=0$, the degree of localization of the quasienergy function in the left well decreases and a fraction of the wave function goes over to the right well. When the condition for resonance is met exactly, the quasienergy function $U_+(x,\tau)$ fills both wells to the same extent. A further increase in $2\overline{\varepsilon}$ causes a further filling of the right well and depletion of the left well. In the limit $\delta\gg\mu_l$, the entire wave function is almost completely localized in the right well.

In the entire range of values of $2\overline{\varepsilon}$ between neighboring resonances the function $U_+(x,\tau)$ undergoes no substantial change as long as $2\overline{\varepsilon}$ is not too close to the next resonance value $(l+1)\Omega$. But if it is close to the resonance value $(l+1)\Omega$, the coefficients g_0 and f_{-1} become important, with the first decreasing and the second increasing as we move closer to resonance. Thus, the weight of the functions $\Psi_R(x)$ and $\Psi_L(x)$ in the expression for $U_+(x,\tau)$ decreases and increases, respectively. At resonance the two wells are filled equally, while as $2\overline{\varepsilon}$ increases, the quasienergy function becomes completely localized in the left well.

If we use (12) and (11), we can easily see that the function $U_-(x,\tau)$ behaves in a manner opposite to that of $U_+(x,\tau)$: when one is localized in the left well, the other is localized in the right well, and vice versa. For this reason the dipole moments of the states ''+'' and ''−'' differ only in sign.

Obviously, each time a resonance is passed, the functions $U_\pm(x,\tau)$ change their localization and completely change the polarization of the given quasienergy state.

Thus, we draw the conclusion that for almost all values of $\overline{\varepsilon}$ the quasienergy functions coincide, to within unimportant phase factors, with the functions $\Psi_{L,R}(x)$ localized in the left and right wells, respectively. The exceptions are narrow resonant regions $|2\overline{\varepsilon}-n\Omega|\sim\mu_n\ll\Omega$ where this assertion is violated and the quasienergy functions become delocalized.

On the other hand, as mentioned earlier, the coefficients in the expansion of the wave function in the basis of the quasienergy functions [Eq. (5)] always remain constant. Thus, if initially we prepare a wave function coinciding with one of the quasienergy functions (by setting one of the coef-

ficients in (5) equal to zero and the other equal to unity), we find that such an expansion is retained at all subsequent moments in time and the wave function of the system always corresponds to the chosen quasienergy state. If, in addition, the value of the external field is such that the system is far from a resonance, the particle wave function becomes localized in one well and remains there for an infinitely long time.

But if the value of $2\overline{\varepsilon}$ lands in a narrow resonant region near an arbitrary nth resonance, the spatial distribution of the charge in pure quasienergy states changes significantly. As noted earlier, a characteristic feature of the resonant values of $2\overline{\varepsilon}$ is that at these points the two quasienergy branches are as close as possible, as shown in Fig. 1. In the zeroth approximation, if we set the coefficient μ_n in the vicinity of the nth resonance to zero, the quasienergy branches cross at points $2\overline{\varepsilon}=n\Omega$, which corresponds to degeneracy of the quasienergies; in this case we speak of a quasienergy resonance. If the finiteness of the μ_n is taken into account, the degeneracy is lifted: the quasienergy levels split, so to speak, which is directly seen from (12), e.g., in the case where $\delta=0$. Obviously, the dipole moment of the system at points of resonance vanishes, a fact that also follows from (13).

Speaking of the antipolarization effect, when the structure is polarized in the direction opposite to the external electric field, we note the analogy between this phenomenon and the effect predicted theoretically by Dakhnovskii and Metiu[16] and Aguado and Platero[17] and detected experimentally by Aguado and Platero[17] and Keay *et al.*,[18] who called it absolute negative resistance: a current flowing through a double-well (three-barrier)[16,17] or multiwell[18] heterostructure was found to be negative when the structure was irradiated by variable laser light, while the constant voltage applied to the structure was kept positive. Obviously, the two effects are similar and are of purely quantum origin, related to electron dynamics in potential wells placed in periodic fields.

Finally, we will briefly discuss the case where all expansion coefficients in (5) are finite and the wave function of the system is a superposition of two quasienergy states. Here the dipole moment is a function of time and can be written

$$d=\overline{d}-4ex_{01}a_+a_-G^2\left\{\frac{\mu_l}{2\nu+\delta}\cos(\overline{\psi}_l+2\nu\tau)\right.$$
$$\left.+\frac{\mu_{l+1}}{2\nu+\delta-2\Omega}\cos(\overline{\psi}_{l+1}+(2\nu-\Omega)\tau)\right\}. \qquad (14)$$

where $\overline{\psi}_n=\psi_n-\alpha$, the parameters a_+, a_-, and α are, respectively, the absolute values and phase difference of the coefficients A_+ and A_-, and $\overline{d}$ is the time average of the dipole moment,

$$\overline{d}=(a_+^2-a_-^2)D, \qquad (15)$$

defined by the population difference of the quasienergy levels.

An analysis of (14) shows that the time dependence of the dipole moment does not always manifest itself, doing so only near resonance. For instance, near the lth resonance the dipole moment is a slowly oscillation function of time with a small frequency $2\nu=\sqrt{\delta^2+\mu_l^2}$, since the coefficient $G^2\mu_l/(2\nu+\delta)$, which determines the amplitude of the first harmonic in (14), is of order unity, while the amplitude of the other harmonic with the frequency $\Omega-2\nu$ is insignificant due to the smallness of the ratio $G^2\mu_{l+1}/(2\nu+\delta-\Omega)\sim\mu_l/\Omega\ll 1$.

Similarly, near the $(l+1)$st resonance the first harmonic, which becomes a rapidly oscillating function $(2\nu\approx\Omega)$. has a small amplitude of order $\mu_l/\Omega\ll 1$, while the second harmonic becomes a slowly oscillating function with a frequency $\sqrt{\mu_{l+1}^2+(\delta-\Omega)^2}\ll\Omega$ and an amplitude of order unity.

In the nonresonant region both harmonics are small and the dipole moment is determined entirely by its average value. Here the double quantum is always found to be polarized if the energy levels are not equally populated, and maximum polarization is achieved when only one level is populated. But if the quasienergy levels are equally populated, the dipole moment vanishes.

5. ADIABATIC VARIATION OF THE AVERAGE FIELD

In Secs. 3 and 4 it was assumed that the average value of the external field is constant and hence the external field is strictly periodic. This enabled us to develop a quasienergy formalism for describing the time evolution of the wave function and to analyze the dependence of the dipole moment of the system on the constant component of the external field.

In particular, we found that there can be a state with a maximum static dipole moment corresponding to the localization of the electron wave function in one well. One must bear in mind, however, that such a state can form not only due to an external field of a certain type acting on the system but also due to a specific choice of the initial condition. By specifying the wave function at the initial moment as one of the quasienergy functions we can ''lock'' the electron in one well by a strong variable field. Any deviations from the given initial condition cause breakdown of the localization regime and to a flow of charge from one well to the other in accordance with (14).

But can we control the localization of the wave function by changing the well populations to the opposite values solely by adjusting the parameters of the external field? For an example we will take the regime of slow monotonic variation of the average field and follow the time evolution of the electron density in the wells and the dipole moment of the system.

To be definite, we will assume that the average value of the perturbation increases adiabatically and select the origin in time when the value of $2\overline{\varepsilon}$ is $l\Omega$. Making the coefficient μ_l so small that δ is much larger than μ_l but still is small compared to Ω, we obtain a solution of Eqs. (2) in the resonance approximation near the lth resonance [see condition (3)]. Discarding in Eqs. (4) all terms corresponding to the $(l+1)$st resonance and expanding the slowly varying function $\overline{\varepsilon}(\tau)$ up to terms linear in τ, we arrive at a system of equations for the coefficients $C_{L,R}(\tau)$:

$$i\frac{dC_L}{d\tau}=-\frac{C_R}{2}\mu_l\exp\{-i(\psi_l-\gamma\tau^2)\},$$

$$i\frac{dC_R}{d\tau}=-\frac{C_L}{2}\mu_l\exp\{i(\psi_l+\gamma\tau^2)\}, \tag{16}$$

where by γ we have denoted the derivative of $\overline{\varepsilon}(\tau)$ with respect to time τ.

We assume that γ is the small parameter of the problem. In particular, it must be much smaller that the characteristic frequency scale of μ_l, which is equal to the smallest distance between the quasienergy branches at $2\overline{\varepsilon}=l\Omega$.

Since now the Hamiltonian operator is not a strictly periodic function of time, it is impossible to introduce the concepts of quasienergy and quasienergy wave function, and so other methods and tools should be used to solve the system of equations (16). We reduce the system (16) of two first-order differential equations to a single second-order equation for, say, the coefficient C_L:

$$\ddot{C}_L+2i\gamma\tau\dot{C}_L+\frac{\mu_l^2}{4}C_L=0. \tag{17}$$

where a dot stands for a time derivative. Allowing for the fact that the coefficient of the first-order derivative is a slowly varying function of time, we use the WKB method to solve this equation.

We represent $C_L(\tau)$ in the form

$$C_L(\tau)=\rho(\tau)\exp\left\{i\int\sigma(\tau)d\tau\right\}$$

and arrive at a system of equations for ρ and σ:

$$2\sigma\dot{\rho}+\rho\dot{\sigma}+2\gamma\tau\dot{\rho}=0,$$

$$\ddot{\rho}-\rho\sigma^2+\frac{\mu_l^2}{4}\rho-2\gamma\tau\rho\sigma=0. \tag{18}$$

In the limit $\gamma=0$, the solution of these equations are constants $\rho(\tau)$ and $\sigma(\tau)=\text{const}$. Allowing for the smallness of γ, we will assume that the functions $\rho(\tau)$ and $\sigma(\tau)$ are slowly varying functions, their first derivatives are first-order quantities, and the second derivative $\ddot{\rho}$ is a second-order quantity.

Ignoring $\ddot{\rho}$ in the second equation in (18), we find two solutions:

$$\sigma^{(\pm)}(\tau)=-\gamma\tau\pm\sqrt{\gamma^2\tau^2+\frac{\mu_l^2}{4}}. \tag{19}$$

Substituting these solutions in the first equation in (18) yields

$$\rho^{(\pm)}(\tau)=\sqrt{\frac{\sqrt{\mu_l^2+4\gamma^2\tau^2}\pm2\gamma\tau}{2\sqrt{\mu_l^2+4\gamma^2\tau^2}}}. \tag{20}$$

Estimates show that it is unnecessary to refine the solutions of Eqs. (18) any further, since the corrections obtained in such a process are determined by integral positive powers of the parameter γ and hence can be discarded.

The general solution of Eq. (17) and hence the initial system (16), which is a linear combination of the particular solutions (19) and (20), is

$$\begin{aligned}C_L(\tau)&=(C_1\rho^{(-)}\exp\{-i\varphi\}+C_2\rho^{(+)}\exp\{i\varphi\})\\&\quad\times\exp\left\{-i\left(\frac{\psi_l}{2}+\frac{\gamma\tau^2}{2}\right)\right\},\\C_R(\tau)&=(C_2\rho^{(-)}\exp\{i\varphi\}-C_1\rho^{(+)}\\&\quad\times\exp\{-i\varphi\})\exp\left\{i\left(\frac{\psi_l}{2}+\frac{\gamma\tau^2}{2}\right)\right\}.\end{aligned} \tag{21}$$

where C_1 and C_2 are arbitrary constants determined by initial conditions, and the phase φ is a function of time:

$$\varphi(\tau)=\frac{\tau}{2}\sqrt{\gamma^2\tau^2+\frac{\mu_l^2}{4}}+\frac{\mu_l^2}{8\gamma}\ln\frac{\sqrt{\mu_l^2+4\gamma^2\tau^2}+2\gamma\tau}{\mu_l}.$$

The expressions (21) for the coefficients $C_{L,R}(\tau)$ fully define the wave function (1) of the system and make it possible to study the dependence of well populations on time for arbitrary initial conditions. For the ''initial''time it is natural to select the time $\tau\to-\infty$, where the functions $\rho^{(\pm)}(\tau)$ attain their stationary values.

Strictly speaking, we must take such values $\tau<0$ at which $|\tau|$ is large compared to $\mu_l/2\gamma$ but at the same time the system is far from the $(l-1)$st resonance, i.e., $\mu_l\ll2\gamma|\tau|\ll\Omega$. Similarly, by infinitely large positive values of τ we mean such values at which the system has left the lth resonance but is still far for the $(l+1)$st resonance, which is ensured by the same inequality. To simplify notation, we will write $\tau\to\pm\infty$, having in mind the large positive and negative values τ in the above sense.

Suppose that at $\tau\to-\infty$ one well, say the left, is completely populated, i.e., $|C_L(-\infty)|=1$, which means that $C_R(-\infty)=0$. Then we have $C_2=0$ and the coefficient C_1 is equal to unity (to within an insignificant phase factor). As noted earlier, the probability of the left or right well being filled is simply the square of the absolute value of the coefficient $C_L(\tau)$ or $C_R(\tau)$. Hence in this case for $W_L(\tau)$ we have

$$W_L(\tau)=\frac{\sqrt{\mu_l^2+4\gamma^2\tau^2}-2\gamma\tau}{2\sqrt{\mu_l^2+4\gamma^2\tau^2}}. \tag{22}$$

We see that as we move toward the condition (3) for resonance, at which $\tau=0$, holds, the probability of the left well being populated monotonically decreases, remaining always larger than the probability of population of the other well. At $\tau=0$ both wells are equally populated, and as $\overline{\varepsilon}$ increases further the occupation of the right well exceeds that of the left and finally amounts to 100%. As $\tau\to+\infty$, the probability W_L vanishes, and this stable distribution of electron density continues indefinitely (in real conditions it is retained as long as $2\overline{\varepsilon}$ does not approach the next resonant value $(l+1)\Omega$, after which the entire charge gradually flows into the left well).

Such behavior of the electron density can be understood if we use (21) and (1) to set up the wave function $\Psi_1(x,\tau)$ for this case. Here it occurs that there is no need to write the expression for the wave function explicitly, since the expression coincides with earlier obtained formula for $U_+(x,\tau)$ in (12) if we replace the constant parameter δ with the time-

dependent function $\delta(\tau)=2\gamma\tau$, the product $\nu\tau$ with $\varphi(\tau)$, and drop the coefficient $f_{-1}^{(+)}$, which is negligible near the lth resonance.

Similarly, when the wave packet is entirely in the right well ($\tau\to-\infty$), the wave function $\Psi_2(x,\tau)$ coincides, to within notation, with the quasienergy function $U_-(x,\tau)$. The time dependence of the wave function $\Psi_2(x,\tau)$ qualitatively follows the dependence of $U_-(x,\tau)$ on the average value of the external field and hence the shift of the point of resonance when $\overline{\varepsilon}$ varies slowly results in the flow of charge from the right well to the left.

A further increase in $\overline{\varepsilon}$ may bring the system into the region of the $(l+1)$st resonance, where the dynamics of the two-level system is also described by equations of the form (16) but with a different value of γ calculated at the time when $2\overline{\varepsilon}=(l+1)\Omega$. Obviously, the solution of these equations is exactly the same as in the vicinity of the lth resonance, and the passage of the point of the $(l+1)$st resonance also gives relocates the wave packet from one well to the other. Hence, if for $2\overline{\varepsilon}<l\Omega$ the wave packet was in the left well, an increase in $\overline{\varepsilon}$ first shifts it to the right well (after $2\overline{\varepsilon}$ becomes greater than $l\Omega$) and then back to the left well, after the system passes the next $(l+1)$st resonance.

Generally speaking, to be sure that the above statement is true we should see whether the wave packet remains in the right well all the time that the value of $\overline{\varepsilon}$ belongs to the nonresonant region. For this we must solve the equations of the two-resonance approximation [Eqs. (4)] with a time-dependent $\overline{\varepsilon}$, which has not been done due to the complexity of such calculations. It is clear, however, that in the nonresonant region all the harmonics in (2), including those left in (4), are found to be ''rapid'' for slow variations of $\overline{\varepsilon}$, and their averaged effect is essentially nil. This means that the coefficients $C_{L,R}(\tau)$ and hence the probabilities $W_{L,R}(\tau)$ will actually retain their values over the entire range of variation of $\overline{\varepsilon}$ between two neighboring resonances.

Thus, a slow variation of the average field in addition to a periodic field of a large amplitude lead to relocation of the electron wave packet from one well to the other as a result of passage of a point of resonance, just as in weak fields there is a transition from one energy level to another under a slow variation of the frequency of the external field.[19]

Finally, if the initial state corresponds to an arbitrary distribution of the electron density in the wells, the wave function is a linear combination of the functions $\Psi_1(x,\tau)$ and $\Psi_2(x,\tau)$ with the coefficients C_1 and C_2 in (21): $\Psi(x,\tau)=C_1\Psi_1(x,\tau)+C_2\Psi_2(x,\tau)$. In accordance with the normalization condition, the sum of the squares of the absolute values of the coefficients C_1 and C_2 is always equal to unity. The functions $\Psi_{1,2}(x,\tau)$ can be called adiabatic quasienergy functions and the quantity $\nu(\tau)=\sqrt{\gamma^2\tau^2+\mu_l^2/4}$, the adiabatic quasienergy near the lth resonance.

It is well known that an adiabatic perturbation does not cause transitions between stationary states of a quantum system. L. D. Landau (see Ref. 14, p. 237) obtained an estimate for the probability of a transition from one stationary state to another initiated by a slowly varying perturbation, and the probability proved to be exponentially small. In our case, due to the strong periodic field, there are no stationary states in the system. Here the quantum states are quasienergy states, and the quantum number characterizing the given state is the quasienergy.

The expressions for the wave functions derived in this paper, which are almost the same as the expressions for the quasienergy functions, indicate that the fact that an adiabatic perturbation does not change the quantum state of a system is true not only for stationary states but can also be generalized to quasienergy states, which effectively replace the states with a definite energy in the presence of a periodic external field.

Obviously, the dipole moment of the system is also described by the expressions (13)–(15), where $\overline{\varepsilon}$ should be replaced by a slowly varying function of time. Correspondingly, in view of the assumed monotonic dependence of $\overline{\varepsilon}$ on τ (for the sake of simplicity and definiteness, we assume that this dependence is strictly linear), the time dependence of the dipole moment in a given adiabatic quasienergy state will follow the dependence on the value of the average field; however, with allowance for the small proportionality coefficient γ, this dependence will be strongly ''elongated'' along the time axis.

As noted earlier, the function $D(2\overline{\varepsilon})$ is almost periodic. Actually, the aperiodicity of this function manifests itself only near the resonant values of $\overline{\varepsilon}$ and is revealed by the fact that the function $D(2\overline{\varepsilon})$ in the vicinity of each point $2\overline{\varepsilon}=n\Omega$ has a its own characteristic step width μ_n. All the μ_n are small quantities, with the result that all steps in the $D(2\overline{\varepsilon})$ dependence are very steep. For this reason we can assume that deviations from periodicity also occur within narrow intervals on the $2\overline{\varepsilon}$ axis near resonance.

Let us write the function $D(2\overline{\varepsilon})$ as a sum of two terms, $D\text{vs.}2\overline{\varepsilon}=D_0(2\overline{\varepsilon})+\tilde{D}(2\overline{\varepsilon})$. One of the terms, $D_0(2\overline{\varepsilon})$, is a strictly periodic function with the period 2Ω. It can be defined by the standard expression for the dipole moment [Eq. (13)], replacing the different coefficients μ_n in this expression by an average value μ, the same for all resonances. The second term, $\tilde{D}(2\overline{\varepsilon})$, represents the deviation of $D(2\overline{\varepsilon})$ from a strictly periodic function, and since this deviation is important only in a near region near a point of resonance, the function $\tilde{D}(2\overline{\varepsilon})$ has the shape of a sequence of equidistant bursts, each of which has its own amplitude and characteristic width defined by the small Fourier coefficient μ_n.

If we now allow for the time dependence of the average field, the dipole moment becomes a function of time. Here, in view of the linear dependence of $\overline{\varepsilon}$ on τ, the function $D_0(\tau)$ is still periodic, but its period is much large than before, $\Omega/\gamma\gg1$, which corresponds to a frequency

$$\lambda=\frac{2\pi\gamma}{\Omega}. \tag{23}$$

In view of its irregularity, the aperiodic part of the dipole moment yields a continuous spectrum, which contains all possible frequencies. The bursts of the function $\tilde{D}(\tau)$ will now have a characteristic width along the time axis equal to μ_l/γ and hence will be smooth (recall that γ is the smallest parameter in the problem), but their width is still much

smaller than the distance $\Omega/2\gamma$ between the bursts. For this reason the amplitudes of the harmonics belonging to the continuous spectrum can be assumed negligible compared to the harmonics of the periodic component.

Thus, the slow linear buildup of the average value of the external field and the permanent periodic field results in a smooth and almost-periodic flow of the electron density from one well to the other. Obviously, such a double-well structure with a periodic flow of charge is a classical dipole capable of emitting electromagnetic energy at its natural frequency. In our case emission of radiation is possible at frequencies that are integral multiples of the frequency (23) of charge oscillations in the wells, since the law of variation of the dipole moment $D_0(\tau)$ as a function of time is periodic but not harmonic.

Two remarks are in order. First, Eq. (23) implies that the fundamental frequency λ is much smaller than the external-field frequency or the transition frequency between the levels of the unperturbed system, which in our notation is equal to unity. As shown in Sec. 4 [formula (14)], small oscillations of the dipole can be generated even for a permanent average value of the external field, but the lowest possible frequency is μ_n. In our case, however, when the average value of the perturbation is a linear of time, the frequency is determined by the rate γ of variation of the average field, which by assumption is much smaller than μ_n.

Second, what is important is that the value of the frequency of the emitted radiation can easily be controlled, i.e., can be varied, if necessary, within broad limits by changing the rate γ of increase of the average value of the external field.

Hence the action on the system of a variable field of frequency Ω accompanied by a slow variation of the average field may lead to low-frequency generation of electromagnetic dipole radiation, one of the main parameters of which, the frequency, can be controlled by the external field itself.

6. CONCLUSION

Let us summarize the main results of our investigation.

In calculating the quasienergies and quasienergy functions for a particle in a strong periodic external field with a finite average value, we used the novel method of the two-resonance approximation. The method is based on allowing for the effect of two neighboring resonances (rather than one resonance) on the quantum dynamics of a system. The two-resonance approximation made it possible to obtain and analyze the dependence of the electron dipole moment on the size of the average value of the external field in almost the entire range of average values, in contrast to the ordinary resonance approximation valid only near a resonance.

In pure quasienergy states the dipole moment of the system is an almost-periodic function of the constant voltage applied to the system. At all points of resonance, $2\overline{\varepsilon}=n\Omega$, the dipole moment is found to be zero, while in the nonresonant regions it reaches its maximum possible positive and negative values due to the complete localization of the electron wave function in one well. When the constant external field applied to the system is positive, the polarization of the double-well structure may oppose the external field, and we have the antipolarization effect.

When the average field varies adiabatically, we used the WKB method of obtain an expression for the electron wave function and found that the problem of the excitation of a system that is in a certain quasienergy state coincides with the problem of the excitation of a system that is in state with a definite energy.[14] In particular, it was found that an adiabatic increase of the average field does not make the system leave the quasienergy state in which it was initially; it only gradually changes the structure of the wave function of the given state. It was also found that an increase in $\overline{\varepsilon}$ that initiates the passage of a point of resonance ''shifts'' the region of localization of the quasienergy wave function from the left well to the right, or vice versa, in view of which the electron density can be made to periodically flow from well to well, provided that the particle wave function is chosen at $\tau\to-\infty$ in the form of one of the quasienergy functions.

APPENDIX

Let us find the expression for the quasienergy in the interval $l\Omega\lesssim 2\overline{\varepsilon}\lesssim(l+1)\Omega$. We will assume that $\delta\ll\Omega$, which corresponds to the lth resonance. Using (7), we can easily see that the coefficients f_0 and g_0 provide the main contribution in this range: their values are of order unity. The coefficients f_{-1} and g_1 are proportional to m_{l+1} (quantities of first order of smallness), the coefficients f_1 and g_{-1} are determined by the product $\mu_l\mu_{l+1}$ (second order of smallness), etc. Near the $(l+1)$st resonance the principal coefficients are either f_0 and g_1 or g_0 and f_{-1} (depending on the quasienergy branch), while the other coefficients are at the most first-order quantities. In the intermediate region between the resonances the main coefficient is f_0 or g_0.

Hence in the interval of variation of $\overline{\varepsilon}$ of interest to us between the lth and $(l+1)$st resonances and near the resonances proper, all coefficients except the four coefficients f_0, f_{-1}, g_0, and g_1 always turn out to be small and can be ignored. For this reason, if in Eqs. (7) we discard all the coefficients except the above four, we arrive at a closed system of equations:

$$\left(\nu+\frac{\delta}{2}-\Omega\right)f_{-1}+\frac{\mu_{l+1}\exp\{-i\psi_{l+1}\}}{2}g_0=0,$$

$$\left(\nu+\frac{\delta}{2}\right)f_0+\frac{\mu_l\exp\{-i\psi_l\}}{2}g_0+\frac{\mu_{l+1}\exp\{-i\psi_{l+1}\}}{2}g_1=0,$$

$$\frac{\mu_{l+1}\exp\{i\psi_{l+1}\}}{2}f_{-1}+\frac{\mu_l\exp\{i\psi_l\}}{2}f_0+\left(\nu-\frac{\delta}{2}\right)g_0=0,$$

$$\frac{\mu_{l+1}\exp\{i\psi_{l+1}\}}{2}f_0+\left(\nu-\frac{\delta}{2}+\Omega\right)g_1=0. \tag{A1}$$

If we nullify the determinant of the system (A1) we obtain two quasienergy values that differ only in sign, with each corresponding to its own quasienergy state:

$$\nu_{\pm}=\pm\frac{1}{\sqrt{2}}\sqrt{\left(\Omega-\frac{\delta}{2}\right)^2+\frac{\mu_{l+1}^2}{2}+\frac{\mu_l^2+\delta^2}{4}-\sqrt{\left[\left(\Omega-\frac{\delta}{2}\right)^2-\frac{\mu_l^2+\delta^2}{4}\right]^2+\Omega^2\mu_{l+1}^2}}. \tag{A2}$$

Plugging (A2) in Eq. (A1), we find that for the state ''+'' the coefficient g_1 is small in the entire interval $l\Omega\lesssim 2\bar{\varepsilon}\lesssim(l+1)\Omega$, while for the state ''−'' the coefficient f_{-1} is always small. We will use this fact to simplify (A2), since it is rather cumbersome in the present form.

Let us consider the state ''+.'' If we ignore g_1 and drop the last equation in (A1), we arrive at the equation

$$\nu^3+\nu^2\left(\frac{\delta}{2}-\Omega\right)-\frac{\nu}{4}(\delta^2+\mu_l^2+\mu_{l+1}^2)-\left(\frac{\delta}{2}-\Omega\right)\frac{\delta^2+\mu_l^2}{4}-\frac{\mu_{l+1}^2\delta}{8}=0. \tag{A3}$$

A simple analysis of (A2) shows that the deviations $\nu(\delta)$ from the linear dependence $\nu_+=\delta/2$ are always small and reach maximum values near resonances, where they amount to quantities of order μ_l and μ_{l+1}, respectively.

Hence, putting $\nu=\delta/2+\eta$, where $\eta\ll 1$ (of order μ_l or μ_{l+1}), in (A3) and ignoring third-order quantities, $\eta^3\sim\eta\mu_l^2\sim\eta\mu_{l+1}^2$, we arrive an ordinary quadratic equation for η:

$$\eta^2(\Omega-2\delta)+\eta\delta(\Omega-\delta)+\frac{\mu_l^2}{4}(\delta-\Omega)+\frac{\mu_{l+1}^2}{4}\delta=0.$$

Solving it and allowing for the fact that for the state ''+'' the quasienergy differs only in sign, we arrive at (8).

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

[1] A. A. Gorbatsevich, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **107**, 1320 (1995) [JETP **80**, 734 (1995)].
[2] F. Grossmann, T. Dittrich, P. Jung, and P. Hänggi, Phys. Rev. Lett. **67**, 516 (1991); Z. Phys. B **84**, 315 (1991); F. Grossmann, T. Dittrich, and P. Hänggi, Physica B **175**, 293 (1991); F. Grossmann and P. Hänggi, Europhys. Lett. **18**, 571 (1992).
[3] J. M. Gomez Llorente and J. Plata, Phys. Rev. A **45**, R6958 (1992).
[4] Y. Kayanuma, Phys. Rev. A **50**, 843 (1994).
[5] Y. Dakhnovskii and R. Bavli, Phys. Rev. B **48**, 11020 (1993).
[6] Y. Dakhnovskii and R. Bavli, Phys. Rev. B **48**, 11010 (1993); Y. Dakhnovskii, R. Bavli, and Metiu, Phys. Rev. B **53**, 4657 (1996).
[7] F. I. Gauthey, C. H. Keitel, P. L. Knight, and A. Maquet, Phys. Rev. A **55**, 615 (1997).
[8] V. A. Burdov, Zh. Éksp. Teor. Fiz. **112**, 1209 (1997) [JETP **85**, 657 (1997)].
[9] K. Leo, J. Shah, E. O. Göbel, Th. C. Damen, S. Schmitt-Rink, W. Schäfer, and K. Köhler, Phys. Rev. Lett. **66**, 201 (1991).
[10] I. Brener, P. C. M. Planken, M. C. Nuss, L. Pfeiffer, D. E. Leaird, and A. M. Weiner, Appl. Phys. Lett. **63**, 2213 (1993).
[11] M. S. C. Luo, S. L. Chuang, P. C. M. Planken, I. Brener, and M. C. Nuss, Phys. Rev. B **48**, 11043 (1993).
[12] R. Bavli and H. Metiu, Phys. Rev. A **47**, 3299 (1993).
[13] V. A. Burdov, Teoret. Mat. Fiz. **108**, 69 (1996) [Theor. Math. Phys. **108**, 904 (1996)].
[14] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[15] J. H. Shirley, Phys. Rev. **138B**, 979 (1965); Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **51**, 1492 (1966) [Sov. Phys. JETP **24**, 1006 (1967)]; V. I. Ritus, Zh. Éksp. Teor. Fiz. **51**, 1544 (1966) [Sov. Phys. JETP **24**, 1041 (1967)].
[16] Y. Dakhnovskii and H. Metiu, Phys. Rev. B **51**, 4193 (1995).
[17] R. Aguado and G. Platero, Phys. Rev. B **55**, 12860 (1997).
[18] B. J. Keay, S. Zeuner, S. J. Allen Jr., K. D. Maranowski, A. C. Gossard, U. Bhattacharya, and M. J. W. Rodwell, Phys. Rev. Lett. **75**, 4102 (1995).
[19] M. E. Crenshaw, Phys. Rev. A **54**, 3559 (1996).

Translated by Eugene Yankovsky

Perturbation of atomic energy levels by a metal surface

M. I. Chibisov*) and A. V. Roĭtman

Russian Research Center ''Kurchatov Institute,'' Institute of Nuclear Synthesis, 123182 Moscow, Russia
(Submitted 8 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 236–256 (July 1999)

The energy level shifts of one-electron atomic particles H, He^+, Li^{++}, etc. which interact with a metal surface have been investigated. In the approximation of image charges, an operator describing perturbations of atomic levels has been obtained. By numerically solving the Schrödinger equation, we have calculated energy levels of $H(1s)$, $H^*(n=2)$, and $C^{5+}(n)$ as functions of the distance between an atom and surface. Asymptotic behavior of atomic levels at large distances from the surface has been studied. The linear Stark effect for excited states, which was earlier mentioned by A. V. Chaplik, has been found and investigated in detail.
 [S1063-7761(99)01707-2]

1. INTRODUCTION

Thermonuclear research has stimulated an interest in studies of collisions of slow ($v<10^8$ cm/s), multiply charged ions with metal surfaces. Numerous investigations[1–20] have revealed that there is a strong probability that slow, multiply charged ions can be neutralized before colliding with a surface (above the surface), and that this neutralization generates ''hollow atoms'' with vacant internal shells and occupied outer shells. An inverse process, namely, destruction of highly excited Rydberg atoms near a metal surface, was studied in experiments,[21–24] where Rydberg states were generated by laser radiation. The theoretical description of such effects requires shifts of atomic energy levels due to the presence of metal because the rates of electron exchange between atoms and metal surfaces essentially depend on these shifts.[17]

This paper reports on calculations of energy level shifts in one-electron (hydrogen-like) ions H, He^+, Li^{++}, ..., $A^{(Z-1)+}$ in the approximation of image charge. The objective is to determine the perturbation operator responsible for these shifts. Earlier[8,14–17] the problem of hydrogen levels near a metal surface was investigated, but the problem of a correct form of the perturbation operator has not been solved. In order to determine this operator, we will analyze in the next section the image charge approach in the context of calculation of atomic energy levels. This analysis is indispensable because one can hardly find in the classical electrostatic problems[25–28] a system similar to an atom, when the mass of one particle (electron) is smaller than that of another particle (nucleus) by a factor of almost 2000.

Asymptotic shifts of atomic levels in the limit of large separations between atoms and metal surfaces has been studied. In the case of a hydrogen-like ion with a nonvanishing dipole moment, we have found a linear Stark effect (previously mentioned by Chaplik[17]), which is similar to the effect of a uniform electric field or the interaction with an ion. The magnitude of this effect is also finite when a neutral hydrogen atom interacts with a metal surface, even though the atom is acted upon by its neutral image. The correct interpretation of this effect requires a perturbation operator written in a correct form.

2. OPERATOR OF ATOMIC LEVEL PERTURBATION

Shifts of atomic levels are determined by a perturbation generated by a metal and acting only on the electron if isotopic effects, which are proportional to $m/M \ll 1$, where m and M are the electron and nucleus masses, respectively, can be ignored. Therefore, potentials of interactions between the metal, on the one hand, and the electron and nucleus, on the other, should be calculated separately. The total energy of the atom in this case is the sum of the energies of the electron and nucleus. The potential energy of a classical point-like particle is the work done in the process of its movement in vacuum from infinity to a certain point near the metal surface. The potential energy of one isolated particle (electron or nucleus) near a metal surface is well known.[25–28] The problem is in finding the electron's potential energy in the presence of the nucleus, as well as the potential energy of the nucleus in the presence of an electron. Obviously, the potential energy of each particle is notably affected by the presence of another particle owing to the infinite polarizability of the metal.

The potential energy of a system of charged particles depends on relative coordinates of particles and their positions with respect to the metal surface, and is independent of the trajectories along which they move to the metal. The method of calculating the potential energy which is the most obvious is considering the motion of an atom as a whole to the metal surface with a constant relative radius-vector that connects the nucleus and electron, so that both the electron and nucleus move in the direction perpendicular to the surface.

For a system of two real particles, an electron e and nucleus Z, their images $\tilde{e}$ and $\tilde{Z}$ are introduced in the approximation of image charges,[25–28] and we have a system of four particles e, $\tilde{e}$, Z, and $\tilde{Z}$ instead of the atom + metal system. The potential energy $V_{e\tilde{Z}}$ of interaction between the real electron and the nucleus image is equal to the work:

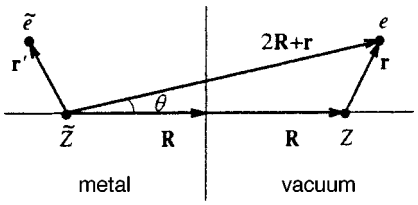

FIG. 1. Coordinates of an electron e, nucleus Z, and their images for a hydrogen-like ion near a metal surface.

$$V_{e\tilde{Z}}=-\int_{S_e}^{\infty}\mathbf{F}_{e\tilde{Z}}d\mathbf{S}_e=-\int_{S_e}^{\infty}F_{e\tilde{Z}}\cos\theta\,dS_e,$$
$$\cos\theta=\frac{2S_e-z}{|2\mathbf{R}+\mathbf{r}|},\quad F_{e\tilde{Z}}=\frac{e^2\tilde{Z}}{|2\mathbf{R}+\mathbf{r}|^2}, \tag{1}$$

where S_e is the distance between the electron and surface (here the vector $\mathbf{S}_e$ is turned perpendicular to the surface and is directed towards the electron), $F_{e\tilde{Z}}$ is the magnitude of the repulsive force between the electron and the image of the nucleus, θ is the angle between the force and electron displacement $d\mathbf{S}_e$ along vector $\mathbf{S}_e$, i.e., the angle between the vectors $2\mathbf{R}+\mathbf{r}$ and $\mathbf{S}_e$; and z is the component of the vector $\mathbf{r}$ perpendicular to the surface, which defines the direction of the z-axis (Fig. 1). In the approximation of image charges,[25–28] the force $F_{e\tilde{Z}}$ is identical to the Coulomb repulsive force between two charges e and $\tilde{Z}$.

The integral in Eq. (1) can be transformed to

$$V_{e\tilde{Z}}=-\frac{e^2\tilde{Z}}{4}\int_{S_e}^{\infty}\frac{d[(2S_e'-z)^2]}{[(2S_e'-z)^2+h^2]^{3/2}},$$

after which it can be calculated analytically. Here h is the component of vector $\mathbf{r}$ perpendicular to vector $\mathbf{R}$. Both h and z are independent of the integration variable S_e'. Note that this transformation does not apply if vector $\mathbf{r}$ is not constant when the atom is driven towards the surface. The interaction energy between the real nucleus and electron image, $V_{Z\tilde{e}}$ can be written in a similar form, and after elementary integration we have both these energies:

$$V_{e\tilde{Z}}=\frac{Ze^2}{2|2\mathbf{R}+\mathbf{r}|},\quad V_{Z\tilde{e}}=\frac{Ze^2}{2|2\mathbf{R}-\mathbf{r}'|}. \tag{2}$$

In general, integration in Eq. (1) should be performed separately for specific values of vector $\mathbf{r}$. Since the calculation scheme under discussion is applicable for arbitrary $\mathbf{r}$ values, Eq. (2) and subsequent formulas (3)–(6) are valid for arbitrary $\mathbf{r}$.

Adding to Eq. (2) the energies of interaction between the electron and nucleus, on the one hand, and their images, on the other,

$$V_{e\tilde{e}}=-\frac{e^2}{2|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|},\quad V_{Z\tilde{Z}}=-\frac{Z^2e^2}{4R}, \tag{3}$$

which are calculated similarly, we obtain the total energies of interaction between the metal, on the one hand, and the electron, $V_{\rm el}(\mathbf{r},\mathbf{R})$, and nucleus, $V_{\rm nuc}(\mathbf{r},\mathbf{R})$, on the other:

$$V_{\rm el}\equiv V_{e\tilde{e}}+V_{e\tilde{Z}}=-\frac{e^2}{2|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|}+\frac{Ze^2}{2|2\mathbf{R}+\mathbf{r}|}, \tag{4}$$

$$V_{\rm nuc}\equiv V_{Z\tilde{Z}}+V_{Z\tilde{e}}=-\frac{Z^2e^2}{4R}+\frac{Ze^2}{2|2\mathbf{R}-\mathbf{r}'|}. \tag{5}$$

Each term in Eqs. (4) and (5) is half of the similar term for interaction between two real particles, because in calculating the work needed to drive a particle one should perform integration over the distance through which it has been driven to the surface, but not over the distance between the particle and its image. If an atom is driven to a surface, its image also moves from the metal depth to its surface. When the distance between an atom and surface changes by ΔR, the distance between the atom and its image varies by $2\Delta R$.

The full energy of interaction between the atom and metal surface should be written in the form

$$V_{\rm tot}=V_{\rm el}+V_{\rm nuc}$$
$$=-\frac{Z^2e^2}{4R}-\frac{e^2}{2|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|}+\frac{Ze^2}{2|2\mathbf{R}+\mathbf{r}|}+\frac{Ze^2}{2|2\mathbf{R}-\mathbf{r}'|}. \tag{6}$$

Owing to the symmetry between an object and its mirror image, the following equality applies: $|2\mathbf{R}+\mathbf{r}|=|2\mathbf{R}-\mathbf{r}'|$; therefore, interaction energies $V_{e\tilde{Z}}$ and $V_{Z\tilde{e}}$ have identical magnitudes and signs: $V_{e\tilde{Z}}=V_{Z\tilde{e}}$, but these are interactions between the metal and different particles, namely, the electron and nucleus, respectively. An important point is that these cross terms describe interactions of the electron and nucleus with the metal, namely, with electrons and ions of the metal. The atomic electron generates on the metal surface a positive charge composed of particles inherent to the metal. The nucleus interacts with this charge, and the term $V_{Z\tilde{e}}$ describes this interaction. Using the symmetry between the particles and their mirror images, one can formally express the function $V_{Z\tilde{e}}(\mathbf{R},\mathbf{r}')$ in terms of the electron coordinates, but this expression would not mean that it is the energy of the atomic electron. The term $V_{z\tilde{e}}$ is calculated by integrating the force acting between the metal and atomic nucleus. In the same way, the term $V_{e\tilde{Z}}$ is calculated by integrating the force acting between the metal and atomic electron; therefore, it is a part of the electron potential energy, although it is a function of the nuclear coordinates.

Let us discuss an alternative way of the motion of an atom from infinity to the surface and demonstrate that the potential energies of the electron and nucleus, given by Eqs. (4) and (5), are the same. First, let the nucleus move towards the surface and be fixed at finite distance R from the surface (Fig. 1). Then we have only the interaction between the nucleus and its image, $V_{Z\tilde{Z}}$, given by the second part of Eq. (3), and the electric field outside the metal is that generated by the dipole $\{Z,\tilde{Z}\}$. If the electron is moved to the surface at a very high velocity, $v\gg 10^8$ cm/s, there is not enough time to generate the electron image charge on the metal surface. In this case (and only in this case) the electron interacts with the dipole field $\{Z,\tilde{Z}\}$ and its interaction with the metal is determined by the interaction with the nucleus at rest. This

interaction is $e^2Z/|2\mathbf{R}+\mathbf{r}|$, i.e., it coincides with the last term in Eq. (6). By analogy with the physics of atomic collisions, this approximation for interaction between an electron and metal can be called static potential approximation. In the case of adiabatically slow motion of the electron to the metal surface, this approximation is very rough because the metal polarizability is infinite.

If the electron approaches the metal + fixed nucleus system adiabatically slowly, there is enough time to generate on the metal surface the charge due to the electron image. In this case, two charges simultaneously approach the surface, instead of one. Concurrently with the real electron, its image moves to the surface from the metal bulk. Then the work done to drive the electron to the surface not only changes its potential energy, but also changes the potential energy of the nucleus due to the interaction between the nucleus and electron image. The total work is $e^2Z/|2\mathbf{R}+\mathbf{r}|$, as in the case of fast electron motion, but it should be divided between two components, the electronic and nuclear ones. Owing to the symmetry of the mirror image discussed above, these two components are equal; i.e., the electronic and nuclear potential energies each equal half the total work. This means that Eqs. (4) and (5) are also valid, as was expected, in the case of sequential approach of the nucleus and then the electron to the surface.

Equations (4) and (5) describe the potential energies of two parts of one system, namely, the electron and nucleus of one atom that interacts with the metal surface. They can be easily generalized to the case of a larger number of particles. Consider one simple case of such a generalization. Imagine a nucleus with an atomic number Z as a system of Z protons concentrated at one point (the presence of neutrons does not affect its interaction with the metal). According to Eqs. (3), (4), and (5), the energies of attraction between each proton and its image, and with images of other protons, are equal to $-1/4R$. Thus, the energy of attraction of one proton to all Z proton images is $-Ze/4R$, and the total energy of attraction between the nucleus and metal is the sum of all proton energies. The summation reduces to multiplication of one proton energy by the total number of protons, i.e., the total energy is $-Z^2e^2/4R$, as was expected. This calculation justifies Eqs. (4) and (5).

Classical equations describing the motion of the atom's center of mass and interatomic motion can be written in the general form

$$(M+m)\ddot{\mathbf{R}}_{\mathrm{CM}}=\mathbf{F}_{Z\tilde{e}}+\mathbf{F}_{Z\tilde{Z}}+\mathbf{F}_{e\tilde{Z}}+\mathbf{F}_{e\tilde{e}},$$

$$\mu\ddot{\mathbf{r}}=\mathbf{F}_{eZ}+\frac{M}{M+m}(\mathbf{F}_{e\tilde{Z}}+\mathbf{F}_{e\tilde{e}})-\frac{m}{M+m}(\mathbf{F}_{Z\tilde{Z}}+\mathbf{F}_{Z\tilde{e}}),$$

where $\mathbf{F}_i$ is the force due to the ith interaction introduced previously, and $\mu=mM/(M+m)$. Ignoring the small terms of the order of $m/M\ll 1$, we obtain the equation for the interatomic motion:

$$m\ddot{\mathbf{r}}=-\mathbf{F}_{eZ}-\mathbf{F}_{e\tilde{Z}}+\mathbf{F}_{e\tilde{e}}+O(m/M),$$

the right-hand side of which is the sum of forces acting only on the electron. These classical equations clearly indicate that, in calculating interatomic energies to within the ratio $m/M\sim 5.45\times 10^{-4}$, one can ignore the forces acting on the nucleus, whereas all forces should be taken into account in calculating the total energy of the atom.

In earlier works[8,14–17] the nucleus potential energy was equated to the energy of its attraction to its image, $-Z^2e^2/4R$, both in the presence and in the absence of an electron, whereas the electron energy consisted of three terms: $V_{e\tilde{e}}$, $V_{e\tilde{Z}}$, and $V_{Z\tilde{e}}$. In other words, the interaction[8,14–17] between the nucleus and electron image, $V_{Z\tilde{e}}(\mathbf{r}',\mathbf{R})$, was attributed to the electronic component. It follows from the analysis given above that this incorrect interpretation can be attributed to the highly nontrivial subtlety of the issue under discussion: owing to the infinite metal polarizability, the interaction of any charged particle with the metal surface is essentially modified by the presence of other particles near the surface.

3. WAVE EQUATION FOR A HYDROGEN-LIKE ION PERTURBED BY A METAL SURFACE

In accordance with Eq. (4), the wave equation for determination of electron levels in the adiabatic approximation, i.e., at low atom velocities, is

$$[\hat{H}_{\mathrm{at}}+\hat{V}_{e\tilde{e}}(\mathbf{r},\mathbf{R})+V_{e\tilde{Z}}(\mathbf{r},\mathbf{R})-E(R)]\psi(R,\mathbf{r})=0, \tag{7}$$

where $\hat{H}_{\mathrm{at}}$ is the Hamiltonian of the isolated atom:

$$\hat{H}_{\mathrm{at}}=-\frac{\nabla^2}{2}-\frac{Z}{r}. \tag{8}$$

Below we will use the system of atomic units, unless stated otherwise, such that $e^2=\hbar=m=1$.

We treat the heavy atomic nucleus as a classical point-like particle, whereas the atom's electron has an extended wave function, and hence an extended distribution of its charge density, while the image of this density generated in the metal also has an extended distribution:

$$\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')=\mathrm{image}\{\varrho_e(\mathbf{r})\},\quad \varrho_e(\mathbf{r})\equiv|\psi(\mathbf{r})|^2. \tag{9}$$

Here we have normalization conditions

$$\int \varrho_e(\mathbf{r})d^3r=\int|\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')|d^3r'=1.$$

The simple functional relation (9) uniquely determines the image charge distribution inside the metal, given the real charge distribution outside the metal volume. The total electric field component parallel to the metal surface equals zero if each element of the real charge, $dq(\mathbf{r})=\varrho_e(\mathbf{r})d^3r$, at point $\mathbf{r}$ has its counterpart of the equal magnitude but opposite sign, $d\tilde{q}(\mathbf{r}')=\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')d^3r'$, at point $\mathbf{r}'$, which is a mirror reflection of point $\mathbf{r}$.[25–28] Calculation of the interaction energy between real and image charges, however, requires a more complicated mathematical procedure. The interaction between the nucleus and its image in the metal was calculated in the previous section as a sum of interactions between all protons and all their images. In the case of an extended charge distribution, summation is replaced with integration. The potential $\hat{V}_{e\tilde{e}}(\mathbf{R},\mathbf{r})$ in Eq. (7) generated at point $\mathbf{r}$ outside the metal by the entire charge of the electron image is

$$\hat{V}_{e\tilde{e}}(\mathbf{R},\mathbf{r}) = -\int \frac{\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')d^3r'}{2\,|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|}. \tag{10}$$

It thus follows that Eq. (7) is, in general, a nonlinear integral-differential equation. The image charge density in Eqs. (7) and (10) is uniquely related to the real charge density [Eq. (9)], and the latter equals the absolute value of the wave function squared. The total energy of interaction between the electron and its image is

$$W_{e\tilde{e}}(R) = \int \hat{V}_{e\tilde{e}}(R,\mathbf{r})\varrho_e(\mathbf{r})d^3r$$

$$= -\int\int \frac{\varrho_e(\mathbf{r})\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')d^3r'd^3r}{2\,|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|}, \tag{11}$$

so that each charge element $dq_e(\mathbf{r})$ of the real electron interacts with the full charge of the electron image (as should be in accordance with the electrostatic laws[25–28]), but not only with its own image $d\tilde{q}_{\tilde{e}}(\mathbf{r}')$. The vectors $\mathbf{r}$ and $\mathbf{r}'$ in Eq. (11) should be treated as independent variables unrelated by the laws of mirror reflection. The approximation based on the wave equation (7) and Eq. (10) is a usual Hartree approximation, which is commonly used in calculations of parameters of atomic and molecular systems.[25] In our specific case, this approximation is applied to the system of an atom + its image with the reflection conditions (9).

The interaction between the atomic nucleus and metal is determined by Eq. (5) integrated over the charge density distribution inside the metal due to the electron image, $\tilde{\varrho}_e(\mathbf{r}')$:

$$\overline{V}_{\rm nuc}(R) = -\frac{Z^2}{4R} + \frac{Z}{2}\int \frac{\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')d\mathbf{r}'}{|2\mathbf{R}-\mathbf{r}'|}. \tag{12}$$

The integral-differential equation (7) is unlikely to be solved analytically. In the case of multiply charged ions, $Z \gg 1$, discussed in this paper, the total interaction energy between the electron and its image, $W_{e\tilde{e}}$, is a small parameter since it is independent of Z. Then we can use in the zeroth-order approximation, instead of Eq. (7), the equation

$$\left(-\frac{\Delta}{2} - \frac{Z}{r} + \frac{Z_{\rm eff}}{r_2} - E_0(R)\right)\psi_0 = 0, \tag{13}$$

where $r_2 = |2\mathbf{R}+\mathbf{r}|$ is the distance between the real electron and the image of the nucleus, and $Z_{\rm eff}$ is defined as

$$Z_{\rm eff} \equiv Z/2. \tag{14}$$

The functions ψ_0 and $E_0(R)$ are zeroth-order parameters in the perturbation theory applied to this specific case. In the first order of the perturbation theory, the electron energy is

$$E(R) \simeq E_0(R) + W^0_{e\tilde{e}}(R) + \cdots. \tag{15}$$

Here $W^0_{e\tilde{e}}(R)$ is given by Eq. (11), where charge densities due to the real electron, ϱ^0_e, and its image, $\tilde{\varrho}^0_e$, calculated in the zeroth-order approximation should be substituted.

In the elliptic coordinates ξ, η, and φ:

$$\xi = \frac{r_2+r}{2R}, \quad \eta = \frac{r_2-r}{2R}, \quad 1 \leqslant \xi \leqslant \infty,$$

$$-1 \leqslant \eta \leqslant 1, \quad 0 \leqslant \varphi \leqslant 2\pi, \tag{16}$$

where φ is the angle of rotation around the $Z\tilde{Z}$ axis, the variables can be separated in Eq. (13), and we have the system of two ordinary differential equations[29]

$$\psi_0(\xi,\eta,\varphi) = \frac{N\,U(\xi)V(\eta)}{\sqrt{(\xi^2-1)(1-\eta^2)}}\,\frac{e^{im\varphi}}{\sqrt{2\pi}}, \tag{17}$$

$$\frac{d^2U}{d\xi^2} + \left[\frac{2R^2E_0\xi^2 + 2(Z-Z_{\rm eff})R\xi + A}{\xi^2-1} - \frac{m^2-1}{(\xi^2-1)^2}\right]U = 0, \tag{18}$$

$$\frac{d^2V}{d\eta^2} + \left[\frac{-2R^2E_0\eta^2 + 2(Z+Z_{\rm eff})R\eta - A}{1-\eta^2} - \frac{m^2-1}{(1-\eta^2)^2}\right]V = 0, \tag{19}$$

where A is the separation constant, and N is the normalization constant.

In order to obtain a wave function ψ_0 that would behave in a regular manner at infinity and at the atomic nucleus, we should satisfy the boundary conditions[29]

$$U(\xi=1)=0, \quad U(\xi=\infty)=0, \quad V(\eta=1)=0. \tag{20}$$

In the case of two real Coulomb centers, the wave function should also tend to zero near the second center, so the condition $V(\eta=-1)=0$ should also be satisfied.[29] In the case of interaction with a metal surface, the second Coulomb center with charge $Z_{\rm eff}$ is fictitious. The real screening charge is distributed over the metal surface.[25–28] On the metal surface at $\eta=0$ and inside the metal, $\eta<0$, we set the wave function to zero:

$$V(\eta \leqslant 0) = 0. \tag{21}$$

This condition forbids the atom's electron to penetrate inside the metal. If the electron energy is in a band gap of the metal, condition (21) holds. But if $E(R)$ coincides with a band of electron states in the metal, and this energy is below the Fermi level in the metal, $E(R)<\varepsilon_{\rm F}$, then the atom's electrons cannot move inside the metal in accordance with the Pauli principle because all the states below the Fermi energy are occupied by metal's electrons. Suppose that the metal temperature is zero or, at least, much smaller than the Fermi energy. Formally, one should substitute in Eq. (7) the metal potential for the electron image potential, then the solution of this equation should be nonvanishing but equal to the wave function of a metal electron of the same energy. Since these two functions are identical, we determine the full wave function symmetrized with respect to exchanges between the electrons of the atom and metal in the form of the Slater determinant, which is zero because two columns of this determinant are identical. Hence follows condition (21) for this case. The validity of boundary condition (21) will be discussed in detail in Sec. 6.

We have solved the system of equations (17)–(19) with boundary conditions (20) and (21) using the predictor-corrector numerical technique[30] for a hydrogen atom in the ground state, H($1s$), and in the lowest excited state, H($n=2$), and also for a carbon ion with charge five, $C^{5+}(n=9)$, with one electron excited to level $n=9$. The electron

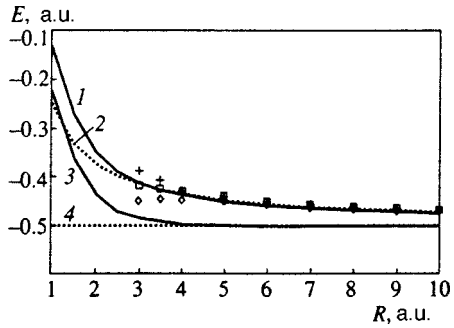

FIG. 2. Energy of the hydrogen ground state H($1s$) as a function of the distance to the surface. The results of the zeroth-order approximation, $E_0(R)$ [without interaction between the electron and its image, $W_{e\tilde{e}}(R)$]: (*1*) our numerical calculations; (*2*) asymptotic expression (34) for $Z=1$: $E_0(R)=-0.5+1/4R$. The results of the first-order approximation, $E(R)=E_0(R)+W_{e\tilde{e}}(R)$: (*3*) our numerical calculations; (*4*) asymptotic value from Eq. (41): $E(R)=-0.5$. The results reported by other authors: crosses from Ref. 8, rhombi from Ref. 16, and squares from Ref. 14.

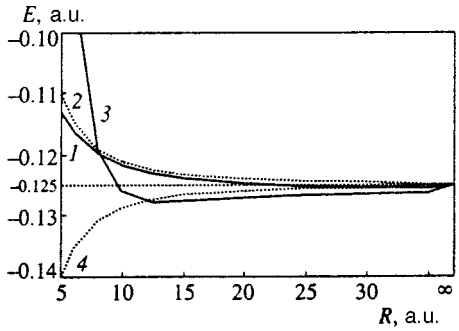

FIG. 4. Energies $E(R)$ of the excited $n=2$ states of the hydrogen atom in the first-order approximation, $E(R)=E_0(R)+W_{e\tilde{e}}(R)$, as functions of the distance to the surface. Term $n_1=1$, $n_2=0$: (*1*) our numerical calculations; (*2*) asymptotic expression (41b): $E(R)=-0.125+3/8R^2-3/2R^3+\cdots$ Term $n_1=0$, $n_2=1$: (*3*) our numerical calculations; (*4*) asymptotic expression (41b): $E(R)=-0.125-3/8R^2-3/2R^3+\cdots$ The orbit size in the lowest excited term $n=2$ in the unperturbed atom is about 8 a.u. As $R\to\infty$, the energy of the $n=2$ terms tends to -0.125 a.u.

binding energy of this level (the ionization potential), which equals 6.047 eV, is very close to the work function for most metals, so a $C^{5+}(n=9)$ ion can be formed with a high probability as a result of a collision of a bare C^{6+} nucleus with a metal surface through capture of a band electron in the metal. The probability of this process essentially depends on the location of $E(R)$ levels (which is the reason for calculating these levels). The energy eigenvalue $E_0(R)$ and separation constant A were derived from the condition of compatibility between Eqs. (18) and (19), i.e., by finding the values of $E_0(R)$ and A at which both the solution of Eq. (18) and that of Eq. (19) satisfy boundary conditions (20) and (21) (Ref. 29). The calculations are plotted in Figs. 2–6 and will be discussed in the following sections.

In order to calculate the electron energy in the first order of the perturbation theory using Eq. (15), which takes into account the interaction between the electron and its image, one should determine the image charge density as a function of coordinates. The reflection of each point outside the metal with elliptic coordinates $(\xi,\ \eta,\ \phi)$ is a point with elliptic coordinates $(\xi,\ -\eta,\ \phi)$ inside the metal. In accordance with the image charge approximation,[25–28] the densities of the real electron charge and of its image at these two points should be equal:

$$\tilde{\varrho}_{\tilde{e}}(\xi,-\eta,\phi)=\varrho_e(\xi,\eta,\phi)\quad(0\leqslant\eta\leqslant1). \tag{22}$$

This condition unambiguously determines the electron image charge density; therefore, we can write

$$\varrho_e(\xi,\eta)=\frac{N^2}{2\pi}\frac{U^2(\xi)V^2(\eta)}{(\xi^2-1)(1-\eta^2)}\quad(0\leqslant\eta\leqslant1), \tag{23}$$

$$\tilde{\varrho}_{\tilde{e}}(\xi',\eta')=\frac{N^2}{2\pi}\frac{U^2(\xi')V^2(-\eta')}{(\xi'^2-1)(1-\eta'^2)}\quad(-1\leqslant\eta'\leqslant0). \tag{24}$$

In accordance with the boundary condition (21), the function $V(\eta)$ is nonzero only at the positive values of its argument.

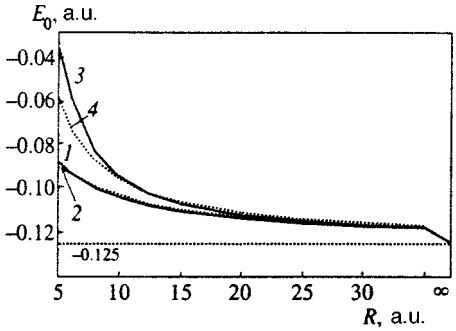

FIG. 3. Energies $E_0(R)$ of the excited state $n=2$ of the hydrogen atom in the zeroth-order approximation [without $W_{e\tilde{e}}(R)$] as functions of the distance to the surface. Term $n_1=1$, $n_2=0$: (*1*) our numerical calculations; (*2*) asymptotic expression from Eq. (34): $E_0^{as}(R)=-0.125+1/4R-3/8R^2+3/8R^3+\cdots$ Term $n_1=0$, $n_2=1$: (*3*) our numerical calculations; (*4*) asymptotic expression (34): $E_0(R)=-0.125+1/4R+3/8R^2+3/8R^3+\cdots$ The size of the lowest excited orbit $n=2$ in the unperturbed hydrogen atom is about 8 a.u. As $R\to\infty$, the energies of $n=2$ terms tend to -0.125 a.u.

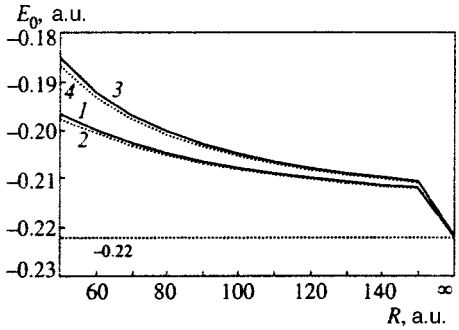

FIG. 5. Energies $E_0(R)$ of the extreme components of the excited $n=9$ state of carbon ion $C^{5+}(Z=6)$ in the zeroth-order approximation [without $W_{e\tilde{e}}(R)$] as functions of the distance R to the surface. Term $n_1=8$, $n_2=0$: (*1*) our numerical calculations; (*2*) asymptotic expression (34): $E_0(R)=-0.22(2)+1.5/R-13.5/R^2+128.25/R^3+\cdots$ Term $n_1=0$, $n_2=8$: (*3*) our numerical calculations; (*4*) asymptotic expression (34): $E_0(R)=-0.22(2)+1.5/R-13.5/R^2+128.25/R^3+\cdots$ The orbit size in the $n=9$ term of the carbon ion is ~27 a.u. As $R\to\infty$, the energies of these terms tend to $-2/9=-0.22(2)$ a.u.

The real electron charge density ϱ_e in this case is nonzero only outside the metal, and the charge density of its image, $\tilde{\varrho}_{\tilde{e}}$, is nonzero only inside the metal.

The volume element expressed in the elliptic coordinates defined by Eq. (16) is $d\tau = R^3(\xi^2-\eta^2)d\xi\, d\eta\, d\varphi$, and the interaction between the electron and its image is described by the formula

$$W^0_{e\tilde{e}}(R) = -\frac{N^4R^6}{4\pi}\int_0^1 d\eta\int_0^1 d\eta'\int_1^\infty d\xi\int_1^\infty d\xi'\int_0^{2\pi} d\phi \times \frac{U^2(\xi)U^2(\xi')V^2(\eta)V^2(\eta')(\xi^2-\eta^2)(\xi'^2-\eta'^2)}{(\xi^2-1)(1-\eta^2)(\xi'^2-1)(1-\eta'^2)\, r_{12}(\xi,\xi',\eta,\eta',\phi)}, \tag{25}$$

where

$$r_{12} = R[\xi^2+\xi'^2+\eta^2+\eta'^2-2(1-\xi\eta\xi'\eta') + 2\cos(\phi)\sqrt{(\xi^2-1)(1-\eta^2)(\xi'^2-1)(1-\eta'^2)}\,]^{1/2} \tag{26}$$

is the distance between two arbitrary points (ξ, η, φ) and (ξ', η', φ') located inside and outside the metal, respectively, and these points are not related by the rules of mirror reflection in the metal surface; $\phi=\varphi-\varphi'$, and the variable substitution $\eta'\to-\eta'$ has been performed in the integrand in Eq. (25). The function $W^0_{e\tilde{e}}(R)$ was calculated by formula (25) with functions U and V found by solving numerically the system of equations (18) and (19).

4. ASYMPTOTIC EXPANSION

At large distances R between the atom and metal surface, the energy $E(R)$ of the atom's electron can be expanded in inverse powers of R. The distance R should be much larger than the average size of the electron orbit in question. The interaction between the atom and metal surface is weak at such distances and can be treated by the perturbation theory.

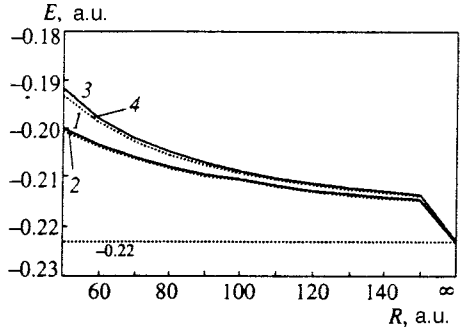

FIG. 6. Energies $E(R)$ of the extreme components of the excited $n=9$ state of carbon ion C^{5+} ($Z=6$) in the first-order approximation $E(R)=E_0(R)+W_{e\tilde{e}}(R)$. Term $n_1=8$, $n_2=0$: (*1*) our numerical calculations; (*2*) asymptotic expression (41): $E(R)=-0.22(2)+1.25/R-9/R^2+45/R^3+\cdots$ Term $n_1=0$, $n_2=8$: (*3*) our numerical calculations; (*4*) asymptotic expression (41): $E(R)=-0.22(2)+1.25/R+9/R^2+45/R^3+\cdots$ The orbit size in the $n=9$ term of the carbon ion is $\simeq 27$ a.u. As $R\to\infty$, the energies of these terms tend to $-2/9=-0.22(2)$ a.u.

At large R the elliptic coordinates near the atom can be approximated by parabolic coordinates[29] μ, ν, and φ:

$$\mu = r(1+\cos\theta), \quad \nu = r(1-\cos\theta), \tag{27}$$

where θ is the angle between the vectors $\mathbf{r}$ and $\mathbf{R}$ (the azimuthal angle φ is the same for both coordinate systems), and the atom is described in terms of the zero-order wave functions ψ_{00} in the parabolic coordinates[31]

$$\psi_{00} = \frac{\sqrt{2}Z^{3/2}}{n^2} f_{n_1m}\left(\frac{Z\mu}{n}\right) f_{n_2m}\left(\frac{Z\nu}{n}\right)\frac{e^{im\phi}}{\sqrt{2\pi}}, \tag{28}$$

where

$$f_{pm}(\rho) = \frac{1}{|m|!}\sqrt{\frac{(p+|m|)!}{p!}}F(-p,|m|+1,\rho)\, e^{-\rho/2}\rho^{|m|/2}, \tag{29}$$

where F is the degenerate hypergeometric function.[25] An important point for further analysis is that the electron in these states has constant dipole, d, and quadrupole, Q, moments[1)] (Refs. 29 and 31):

$$d \equiv -\int |\psi_{00}(\mathbf{r})|^2 r\cos\theta\, d^3r = -\frac{3n(n_1-n_2)}{2Z}, \tag{30}$$

$$Q \equiv -\int |\psi_{00}(\mathbf{r})|^2 r^2(3\cos^2\theta-1)d^3r = -\frac{n}{2Z^2}\{(m+2) \times (m+3)[n+5(n_1+n_2)]+15(m+3)(n_1^2-n_1 + n_2^2-n_2)+10(n_1^3-3n_1^2+2n_1+n_2^3-3n_2^2+2n_2) - 3n[6n_1n_2+(m+1)(3n-2m-1)]\}. \tag{31}$$

In Eq. (31) m is the absolute value of the magnetic quantum number, i.e., $|m|$.

In the first order of the perturbation theory, we have

$$E(R) \simeq E_{00}+(V_{\rm el})_{00} = E_{00}+\langle\psi^*_{00}(\mathbf{r})|\hat{V}_{e\tilde{e}}+\hat{V}_{e\tilde{z}}|\psi_{00}(\mathbf{r})\rangle, \tag{32}$$

where the energy of the unperturbed atom $E_{00}=-Z^2/2n^2$. The interaction between the atom and metal, i.e., the interaction between the atom and its image on the metal surface, has features that distinguish it from the interaction between an atom and a real particle, i.e., another atom or a molecule. For this reason, a detailed analysis of the resulting asymptotic expansion is given below.

The interaction between the real electron and the nucleus image, $\hat{V}_{e\tilde{z}}$, does not explicitly depend on the electron image coordinates. At large R this interaction can be expanded as follows:

$$\hat{V}_{e\tilde{z}} = \frac{Z}{2}\frac{1}{|2\mathbf{R}+\mathbf{r}|} = \frac{Z}{2}\left(\frac{1}{2R}-\frac{r\cos\theta}{(2R)^2}+\frac{r^2}{(2R)^3}\frac{3\cos^2\theta-1}{2}+\cdots\right). \tag{33}$$

Multiplication of Eq. (33) by the probability density $\varrho_e(\mathbf{r})$ for the real electron and integration yield an expansion of energy $E_0(R)$ in the zeroth-order approximation (without the interaction between the electron and its image, see the previous section):

$$E_0(R)=E_{00}+\langle\psi_{00}^*(\mathbf{r})|\hat{V}_{e\tilde{z}}|\psi_{00}(\mathbf{r})\rangle$$
$$=-\frac{Z^2}{2n^2}+\frac{Z}{4R}+\frac{Zd}{8R^2}-\frac{ZQ}{32R^3}+\cdots. \tag{34}$$

The interaction between the electron and its image has the following expansion:

$$-\frac{1}{2|2\mathbf{R}+\mathbf{r}-\mathbf{r}'|}\simeq-\frac{1}{4R}+\frac{r\cos\theta}{8R^2}-\frac{r^2(3\cos^2\theta-1)}{32R^3}$$
$$-\cdots-\frac{r'\cos\theta'}{8R^2}-\frac{r'^2(3\cos^2\theta'-1)}{32R^3}$$
$$-\cdots-\frac{\mathbf{r}\mathbf{r}'-3rr'\cos\theta\cos\theta'}{16R^3}+\cdots, \tag{35}$$

where θ and θ' are the polar angles of the vectors $\mathbf{r}$ and $\mathbf{r}'$. Expansion (35) is a function of the coordinates of the real electron and its image. In order to calculate the matrix element of this operator, one should determine the electron image charge distribution $\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')$, given the charge distribution of the real electron in the parabolic coordinate system. Accordingly, we consider a second parabolic coordinate system with the origin at the nucleus image:

$$\mu'=r'(1+\cos\theta'),\quad \nu'=r'(1-\cos\theta'), \tag{36}$$

such that the angle θ' is measured with respect to the vector $\mathbf{R}$, as in the case of the first parabolic coordinate system introduced for the real electron, and the azimuthal angle φ is the same for both coordinate systems. The mirror reflection transforms lines $\mu=\text{const}_1$ and $\nu=\text{const}_2$ to lines $\nu'=\text{const}_1$ and $\mu'=\text{const}_2$. This means that the point with coordinates $(\mu,\ \nu,\ \varphi)$ in the first coordinate system is reflected into the point with coordinates $(\nu,\ \mu,\ \varphi)$ in the second coordinate system. For the wave functions (28) and (29) in the parabolic coordinates, the permutation $\mu\leftrightarrow\nu$ is equivalent to the exchange of indices $n_1\leftrightarrow n_2$. This means that the charge distribution in the electron image is determined by the same function as that of the real electron, but with the interchanged parabolic quantum numbers:

$$\tilde{\varrho}_{\tilde{e};(n,n1,n2,|m|)}(\mu',\nu',\varphi')=\varrho_{e;(n,n2,n1,|m|)}(\mu',\nu',\varphi'), \tag{37}$$

where ϱ_e is the absolute value of the wave function (31) squared.

Since the charges of the electron and nucleus change their signs as a result of the mirror reflection, there is the plus sign in front of the integrals in definitions of the image dipole and quadrupole moments:

$$\tilde{d}\equiv\int\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')r'\cos\theta'\,d^3r'=d, \tag{38}$$

$$\tilde{Q}\equiv\int\tilde{\varrho}_{\tilde{e}}(\mathbf{r}')r'^2(3\cos^2\theta'-1)d^3r'=-Q, \tag{39}$$

whereas the sign in front of the integrals in definitions for the real electron is minus [see Eqs. (30) and (31)].

Since the integral in Eq. (38) is antisymmetrical with respect to the permutation $n_1\leftrightarrow n_2$ [see Eq. (30)], we can see that $\tilde{d}=d$, i.e., the dipole moment of the electron whose state is determined by the parabolic wave functions (28) does not change after its mirror reflection from the metal surface.[25] Note that in the states described by Eqs. (28) and (29) only the z-component of the dipole moment is nonzero: $d_z\neq0$, but $d_x=d_y=0$. In general, the parallel components of the dipole moment, if they do not equal zero, change their signs as a result of the mirror reflection: $\tilde{d}_x=-d_x$, $\tilde{d}_y=-d_y$ (Ref. 25). The quadrupole moment changes its sign after reflection: $\tilde{Q}=-Q$, since the integral in Eq. (39) is symmetrical with respect to the permutation $n_1\leftrightarrow n_2$ [Eq. (31)].

By substituting expansion (35) in the integral in Eq. (11), we obtain an expansion of the full interaction between the electron and its image

$$W_{e\tilde{e}}(R)\simeq-\frac{1}{4R}-\frac{2d}{8R^2}+\frac{2Q-4d^2}{32R^3}+\cdots\quad(R\to\infty). \tag{40}$$

Adding Eqs. (34) and (40), we obtain an expansion for the electron energy

$$E(R)=-\frac{Z^2}{2n^2}+\frac{Z-1}{4R}-\frac{3n(n_1-n_2)}{8R^2}\left(\frac{Z-2}{2Z}\right)$$
$$-\frac{(Z-2)Q+4d^2}{32R^3}+\cdots\quad(R\to\infty). \tag{41}$$

Let us discuss the results. Recall that in the expansion of the interaction between the two charge distributions the main term of the Coulomb interaction is due to the direct intercharge interaction ($\propto R^{-1}$); then follow the terms describing the interactions between the real charge and image dipole, and the image charge and real dipole ($\propto R^{-2}$), etc. First, let us consider the interaction between the electron and its image [Eq. (40)]. Only this interaction depends on the charge distributions due to the electron and its image.

The first term on the right-hand side of Eq. (40), $-1/4R$, is due to the Coulomb attraction between the total charges of the electron and its image. The second term in Eq. (40) is due to two interactions: between the real charge and image dipole, and between the image charge and real dipole. It is clear that since both the magnitudes and signs of these terms are equal, they are added, but are not canceled out. The image charge has the sign opposite to that of the real charge; therefore, the electric fields generated by these charges have the same direction on the connecting line. For a negative charge, both these fields are directed from the metal to the charge, and for a positive charge both fields are directed towards the metal surface. In both cases, these field really exist only in the space outside the metal. In the electric field $\mathbf{F}$, the dipole energy is $-\mathbf{dF}$; therefore, both terms due to the charge-dipole interaction are identical: their magnitudes and

signs are equal. The same result can be obtained for the two interactions between the charge and the quadrupole, which also are added.[2)]

The first term in the expansion of the total electron energy (41) is the electronic energy of the unperturbed atom. The second term in Eq. (41) is due to the interaction between the real electron and image $\tilde{q}$ of the total charge of the ion, $q \equiv Z-1$. Since $\tilde{q}=-q$ and $q>0$, this interaction is repulsive and drives the electron level in the ion towards the continuum states.

Of the greatest interest is the third term in Eq. (41), which is due to the sum of the charge–dipole interaction. It can be expressed as a sum of two terms:

$$\frac{(Z-2)d}{8R^2}=\frac{q\,d}{8R^2}-\frac{d}{8R^2} \quad (q\equiv Z-1), \tag{41a}$$

where the first term is due to the interaction between the real dipole and the full ion charge image, and the second is due to the interaction between the real electron and image dipole. The third term in Eq. (41) can be interpreted as a linear Stark effect of a hydrogen-like ion in an electric field generated by the ion image. An interesting point, however, is that this effect does not vanish in the case of the neutral hydrogen atom, $Z=1$. It follows from Eq. (41a) that at $Z=1$ only the first term turns to zero, whereas the second term does not vanish and, moreover, is independent of Z. For the neutral hydrogen atom, $Z=1$, Eq. (41) takes the form

$$E(R)=-\frac{1}{2n^2}+\frac{3n(n_1-n_2)}{16R^2} -\frac{9n^2(n_1-n_2)^2-Q(n,n_1,n_2)}{32R^3}+\cdots \quad (R\to\infty), \tag{41b}$$

where $Q(n,n_1,n_2)$ is given by Eq. (31). The second term on the right-hand side of Eq. (41b) is the same as for a hydrogen atom in a uniform electric field with the strength proportional to R^{-2} or for interaction with a positively charged ion at distance R. The possibility of this effect for a neutral atom near a metal surface was first indicated by Chaplik,[17] who studied tunneling of an electron from a highly excited atom into a metal. The numerical factor in the second term on the right of Eq. (41b) is a factor of 2 smaller than the respective factor in Chaplik's formula,[17] since the perturbation operator [Eq. (3) in Ref. 17] is twice as large as that in our equations.

An interesting question is why the linear Stark effect takes place when a neutral atom interacts with its own neutral image, similarly to the case of an atom in a homogeneous electric field or when it interacts with a charged particle such as an ion. When an excited hydrogen atom interacts with a real particle equipped with a constant dipole moment, the second term on the right of Eq. (41b) also takes place, but it is a function of the dipole moment of the second particle, which is independent of the hydrogen dipole moment. In this case, the interaction between the full charge of the hydrogen electron and the dipole (and the other moments of the second particle) is independent of all quantum parameters of the hydrogen atom; therefore, the states of the latter are not mixed up. On the other hand, the image dipole generated by interaction with metal is fully determined by the real dipole. Any change in the real dipole also changes the image dipole. Because of this strong correlation between the real and image dipoles, the asymptotic expansion of the energy of an atom interacting with a metal surface [Eq. (41)] has a character radically different from that of two interacting real particles, namely, all coefficients in expansion (41) depend only on quantum parameters of the real atom since it is acted upon by its own image generated on the metal surface. As a result, the interaction between the zeroth-order moment or the total charge of the real electron (which is the same for all real states $|n,n_1,n_2,m\rangle$, i.e., independent of the quantum numbers n, n_1, n_2, and m) and the image dipole moment depends on the quantum numbers n, n_1, n_2, and m because the parameters of the image dipole depend on these numbers. This interaction brings about transitions between different hydrogen states. Thus, the correlation between the real and image moments is the cause of the linear Stark effect in a neutral hydrogen atom near a metal surface.

Equation (41) clearly shows that the level splitting due to the Stark effect has a nontrivial dependence on the atomic number Z. In the helium ion $He^+(n)$ $(Z=2)$ there is no such effect since the factor in front of the third term on the right-hand side of Eq. (41) vanishes. For the neutral hydrogen atom $(Z=1)$ and hydrogen-like ions Li^{2+}, Be^{3+}, ... $(Z=3, 4, \ldots)$ the factor in front of the third term in Eq. (41) is nonzero, so the effect should take place. Note that in the neutral hydrogen atom Stark components with $n_1>n_2$ have higher energies than components with $n_1<n_2$, whereas in ions with $Z>2$ the arrangement of levels is different: components with $n_1>n_2$ are lower than those with $n_1<n_2$.

The existence of the Stark effect in hydrogen-like ions placed close to metal surfaces is important for applications. In many experiments, beams of atoms in Rydberg states moving near metal surfaces are used. Interaction with metal leads to transitions between Rydberg sublevels. The charge–dipole interaction responsible for these transitions [the third term in Eq. (41)] decreases with the distance to the surface fairly slowly, $\propto R^{-2}$, and can lead to a notable mixing between Rydberg states (see the next section).

The fourth term on the right-hand side of Eq. (41) is for the same reason (correlation between real and image dipoles) different from a similar term for the case of interaction between two real (therefore independent) atoms. The dipole–dipole interaction, which is proportional to the product of two dipole moments, $\mathbf{d}_1\mathbf{d}_2-3(\mathbf{nd}_1)(\mathbf{nd}_2)$, is proportional in Eq. (41) to the atom's dipole moment squared, d^2. The total charge–quadrupole interaction, $(Z-2)Q/32R^3$, is also a function of only one quadrupole moment Q. This interaction is zero for He^+ $(Z=2)$ and nonvanishing for the neutral hydrogen atom and Li^{2+}, Be^{3+}, ... ions, just as in the case of the dipole–charge interaction.

The asymptotic expansion of the energy of interaction between the nucleus and the metal surface averaged over the spatial charge distribution in the electron image is

$$E_{\rm nuc}(R)=-\frac{Z^2}{4R}+\langle\psi_{00}^*|\hat{V}_{z\tilde{e}}|\psi_{00}\rangle$$

$$=-\frac{Z(Z-1)}{4R}+\frac{Zd}{8R^2}-\frac{ZQ}{32R^3}+\cdots \quad (42)$$

The first term on the right describes the Coulomb interaction between the nuclear charge Z and that of the ion image $-(Z-1)$. The two following terms are due to the interaction of the nucleus with image dipole and quadrupole moments.

The expansion of the energy of interaction between the entire atom, which consists of the electron and the nucleus, and the metal is

$$E_{\rm tot}(R)=E(R)+E_{\rm nuc}(R)$$

$$=-\frac{Z^2}{2n^2}-\frac{(Z-1)^2}{4R}+\frac{(Z-1)d}{4R^2}$$

$$-\frac{(Z-1)Q+2d^2}{16R^3}+\cdots \quad (43)$$

It is clear that for the neutral atom $(Z=1)$ the second and third terms on the right of Eq. (43) turn to zero. The only nonzero term in Eq. (43) in this case is that due to the dipole–dipole interaction, $-d^2/8R^3$ if the dipole moment $d\neq 0$.

5. MIXING OF HYDROGEN-LIKE RYDBERG STATES NEAR A METAL SURFACE

Many experiments deal with beams of metastable hydrogen atoms H$(2s)$. Near a metal surface, the hydrogen atom can transfer from state $2s$ to state $2p$, which rapidly relaxes to the ground state $1s$ with emission of a photon. Let us investigate in detail the rate of $2s\leftrightarrow 2p$ transitions in the hydrogen atom near the metal surface.

Let a metastable hydrogen atom H$(2s)$ start its motion along the metal surface at large distance R at the time $t=0$, so that we could use the asymptotic expressions for the energies of atomic levels given in the previous section. The wave functions of atomic steady states near the metal surface are described by the functions in the parabolic coordinates $\psi_{n_1,n_2,m}(\mu,\nu,\varphi)$. The coordinate wave functions in spherical coordinates, $\psi_{2s}(\mathbf{r})$ and $\psi_{2p}(\mathbf{r})$, for the states with the principal quantum number $n=2$ are expressed as linear combinations of the coordinate functions in the parabolic coordinates:

$$\psi_{2s}=\frac{1}{\sqrt{2}}(\psi_{010}+\psi_{100}),\quad \psi_{2p}=\frac{1}{\sqrt{2}}(\psi_{010}-\psi_{100}). \quad (44)$$

If the atom is in the state $2s$ at $t=0$, then at the time $t\geqslant 0$ its state is described by the function

$$\Psi(\mathbf{r},t)=\frac{1}{\sqrt{2}}\left[\psi_{100}\exp\left(-i\int_0^t E_{10}\,dt'\right)\right.$$

$$\left.+\psi_{010}\exp\left(-i\int_0^t E_{01}dt'\right)\right], \quad (45)$$

where E_{10} and E_{01} are the energies of states $|n_1=1,n_2=0\rangle$ and $|n_1=0,n_2=1\rangle$, respectively [see Eq. (41)]. The probabilities of detecting the atom in states $2s$ and $2p$ at $t\geqslant 0$ are

$$P_{2p}(t)=|\langle\psi_{2p}|\Psi(\mathbf{r},t)\rangle|^2=\sin^2\frac{\Delta(t)}{2}, \quad (46)$$

$$P_{2s}(t)=|\langle\psi_{2s}|\Psi(\mathbf{r},t)\rangle|^2=\cos^2\frac{\Delta(t)}{2}, \quad (47)$$

where

$$\Delta(t)\equiv\int_0^t(E_{10}-E_{01})dt'. \quad (48)$$

Probabilities (46) and (47) vary with time as a result of interference between states $|1,0,0\rangle$ and $|0,1,0\rangle$, which have different energies. According to Eq. (41), for $n=2$ the difference between energies of Stark components is $E_{10}-E_{01}=3/4R^2$. If the atom travels along the surface, $R=\text{const}(t)$ and $\Delta(t)=3t/4R^2=3L/4R^2v$, where L is the distance through which the atom has moved along the surface and v is its velocity. For the complete $2s\rightarrow 2p$ transition we need $\Delta=\pi$, and the distance through which the atom should travel in order to complete its transition is $L_\pi=(4\pi/3)vR^2$ (in atomic units). Equation (41) yields correct parameters of the Stark splitting as long as it is larger than the fine structure splitting. For $n=2$ the splitting ΔE_0 between levels $2p_{3/2}$ and $2p_{1/2}$ in an unperturbed hydrogen atom is $\Delta E_0=0.365\ \text{cm}^{-1}=4.53\times10^{-5}$ eV$=1.66\times10^{-6}$ a.u. (Ref. 32). The distance at which the Stark splitting is equal to ΔE_0 is

$$R_0=\sqrt{\frac{3}{4\Delta E_0}}\simeq 3.5\times10^{-6}\ \text{cm}\simeq 670\ \text{a.u.}$$

At $R>R_0$ the Stark effect becomes quadratic, $E_{10}-E_{01}\propto R^{-4}$. If the atom velocity is $v=10^8$ cm/s and the distance to the surface is $R=R_0$, then the required travel distance is $L_\pi=5\times10^{-5}$ cm. The time of this travel is $\tau_\pi=L_\pi/v=5\times10^{-11}$ s, which is much shorter than the radiation lifetime of the $2p$-state: $\tau_{2p-1s}\simeq 1.6\times10^{-9}$ s. We can conclude that an atom in the $2s$-state moving a distance of $\sim 600a_0$ from the surface should travel through half a micrometer (5×10^{-5} cm) to complete its transition to the $2p$-state. If an atom travels along the surface through a distance much larger than half a micrometer, then $\Delta[t(L)]\gg\pi$ and the atom undergoes multiple $2s\rightarrow 2p\rightarrow 2s$ transitions.

6. DISCUSSION

The boundary condition (21) used in our calculations is approximate. Calculations based on this condition can be accurate if the surface region makes a small contribution to the energies of levels under discussion. This condition is valid *a fortiori* if the electron orbit size is much larger than the surface region thickness, which is usually within two to three Bohr radii in different metals. Such is the situation, for example, in case of collisions between excited atomic particles and a metal surface. For example, the sizes of electron orbits of $C^{5+}(n=9)$ and $\mathbf{H}(n=2)$ are $\sim 27a_0$ and $\sim 8a_0$, respectively. Note that in this case condition (21) leads to a correct description of atomic levels even though the electron orbit is

strongly deformed by the interaction with metal. If the orbit size is of order of the surface layer thickness, which is the case in the hydrogen ground state, H($1s$), the distance between the atom and surface should be larger than the orbit size. In this case, the electron wave function is exponentially small near the surface because of the large barrier that separates the atom and metal, and the surface layer has little effect on the energy of the atomic electron. If the electron energy is higher than the Fermi level, condition (21) is also satisfied at large distances between the atom and metal for the same reason.

The approximation of image charge applies to the case of large distances between the atom and metal. Usually the condition $R \gg R_D$ is discussed; here R_D is the Debye radius, which equals the Bohr radius a_0 times a factor of 1 or 2. A more restrictive condition, however, is the smallness of the perturbation of the metal conduction electrons, namely, the additional density of the image charge on the surface should be much smaller than the charge density in the unperturbed metal. The density of degenerate conduction electrons can be estimated using the formula[33–35] $n_0(\varepsilon_F) = 2\sqrt{2}\varepsilon_F^{3/2}/3\pi^2$. For an isolated charge Z the image charge density on the surface[25–28] is $\Delta n(\varrho, R) = -(ZR/2\pi R_D)/(\varrho^2 + R^2)^{3/2}$, where ϱ is the distance measured along the surface from the perpendicular line drawn through the charge. Using the maximum value of this density at $\varrho = 0$, we obtain the condition

$$R \gg R_0\sqrt{Z}, \quad R_0 \equiv 1.5\varepsilon_F^{-5/8}. \tag{49}$$

For most metals, R_0 is within $\simeq (4-10)a_0$. For cesium (or any metal whose surface more than 60% is coated with cesium) the Fermi energy $\varepsilon_F = 1.5$ eV $= 0.055$ a.u., so we have $R_0 = 9.2a_0$. This value is six times the Debye radius for metallic cesium, $R_D^{\mathrm{Cs}} \simeq 1.54a_0$. For the carbon ion with $Z = 6$ the distance from the cesium surface should be larger than $22.5a_0$, which is a factor of 14.6 larger than R_D^{Cs}.

If $R \sim R_0\sqrt{Z}$, the perturbation in the electron density in the metal cannot be ignored. The potential barrier between the atom and surface is lowered, and electrons from the metal can transfer to the atomic ion more easily. This effect was previously studied in detail,[2] and it turned out that, owing to this neutralization effect, the acceleration of the atomic ion near the surface caused by its attraction to the image charge is terminated at distances much larger than the Debye radius and equal to approximately $2R_0\sqrt{Z}$.

The result of our calculations are plotted in Figs. 2–6. Figure 2 shows the ground-state level $1s$ of hydrogen in comparison with calculations by other authors. Our calculations for the $1s$ level are close to the unperturbed value $E_{00} = 13.6$ eV at distances $R \geqslant 4a_0$. On the other hand, the $1s$ energy level calculated by other authors[8,14–16] follows the function $E_{1s}(R) \simeq E_{00}(\infty) + 1/4R$ at large distances and rises to 1.90 eV above E_{00} at $R = 4a_0$. It follows from our calculations [Eq. (41)] that deviation of the level $E(R) - E_{00} \simeq 1/4R$ is possible for an atomic ion with unit charge, for example, He^+. But for a neutral atom, $Z = 1$, the Coulomb term $(Z-1)/4R$ is zero. The image of a neutral particle in a metal is a neutral particle, and interaction between two neutral particles cannot contribute a Coulomb term to the level energy. In the earlier calculations,[8,14–16] a fraction of the potential energy of the nucleus due to its interaction with the electron image, $V_{Z\tilde{e}}$, was attributed to the electron energy. At larger distances, this interaction has an asymptotic behavior $1/4R$, and it is precisely this term that causes the Coulomb rise of the hydrogen ground level proportional to $1/4R$ found in Refs. 8 and 14–16 for $Z = 1$.

Figure 2 also shows the results of our calculations of the energy $E_0(R)$ for the hydrogen ground state $1s$ in the zeroth-order approximation. This energy does not include interaction between the electron and its image and was calculated by numerically solving the system of equations (18) and (19). The asymptotic expansion of energy (34) at $Z = 1$ contains the Coulomb term $1/4R$ due to the repulsion between the electron and proton image. For $R > 4a_0$, our calculations of $E_0(R)$ are in good agreement with calculations[8,14–16] for the total $1s$ level. Previously[8] the energy of the $1s$ level was calculated using the electric potential on the metal surface, whereas our calculations are based on condition (21). The agreement between these calculations indicates that the surface layer contributes little to the electron energy of the hydrogen atom in the $1s$ state at distances from the surface larger than $4a_0$.

Figures 3 and 4 show energies of excited states of the hydrogen atom ($n = 2$). One can see that the calculations in the zeroth order and first order of the perturbation theory (shown in Figs. 3 and 4, respectively) differ considerably. The energies of the electron interaction with its image and with the nucleus image are values of the same order for both the ground state $1s$ and the excited states with $n = 2$ of the hydrogen atom.

Figures 4–6 show the results of our calculations for the excited state of the $C^{5+}(n = 9)$ ion. Figure 5 shows electronic energies calculated in the zeroth-order approximation, and Fig. 6 shows these energies calculated in the first order of the perturbation theory. The differences between the energies plotted in these graphs are small because the energy of electron interaction with its image is small in comparison with its interaction with the nucleus image. The numerical calculations are compared in these graphs with asymptotic expansions in the limit of large distances from the surface. It is clear that the asymptotic expressions are close to numerical calculations at distances down to the orbit size; hence asymptotic formulas can be used in calculating energies of other C^{5+} levels, as well as levels of other A^{Z+} ions. The levels in Figs. 4–6 are driven upwards, and this tendency is due to the repulsion between the electron and nucleus image, whose energy at large distances is $(Z-1)/4R$. Since the second Coulomb center, which is the nucleus image, has an effective charge $Z_{\mathrm{eff}} = Z/2$ in Eq. (13), in the limit $R \to 0$ the total charge of two centers is equal to $-Z/2$ and all states are bound. From the levels that are above the Fermi energy, an electron can tunnel into the metal.[17,19] At the same time, an atomic ion approaching the metal surface cannot capture a resonant electron to these levels. This capture of an electron from the conduction band requires that additional energy be imparted to the electron.[36]

We have calculated the energies of the C^{5+} ion for $R \geqslant 50a_0$ (Figs. 4–6). According to the condition (52) and

the calculations of Ref. 2, the distances to the metal surface should satisfy the condition $R \geqslant 2R_0\sqrt{Z}$, which means in the case of cesium (see above) that $R > 45a_0$.

The Rydberg states of hydrogen-like ions and of the neutral hydrogen atom are strongly mixed near a metal surface. At large distances from the surface, the amplitude of this mixing is approximately the same as the mixing caused by a charged atomic ion,[17] when the splitting between Stark components is proportional to R^{-2}. This effect is caused by interactions of two types: 1) the interaction between the constant dipole moment of the real atom with the image of the total ion charge generated on the metal surface; 2) the interaction between the real electron charge and the dipole moment of the atom image. For the neutral hydrogen atom, the former interaction is zero, but the latter is nonvanishing and responsible for the Stark effect under discussion. The sign of each interaction is opposite for all $Z > 1$, and the magnitudes depend on Z. Since for the He^+ ion the magnitudes of these two interactions are equal, the Stark effect in this ion is zero.

We express our profound gratitude to A. V. Chaplik for very helpful discussions. The work was supported by the Russian Fund for Fundamental Research (Project No. 96-15-96815) within the program of support for leading scientific schools.

*)E-mail: chib@qq.nfi.kiae.su

1)The origin of the coordinate system is set at the nucleus.

2)It is remarkable that in the asymptotic expansion of the total interaction operator (6) for a neutral atom ($Z=1$) all moments of odd orders, in particular, the dipole moment, turn to zero. The reason is that Eq. (6) describes the total interaction between the metal, on the one hand, and the nucleus and electron, on the other, and the latter two have opposite charge signs. Therefore, the total interaction of the electron and the proton with the dipole image turns to zero in operator (6), i.e., before averaging over the spatial charge distributions.

[1] H. Winter, Europhys. Lett. **18**, 207 (1992).
[2] J. Burgdörfer and F. Meyer, Phys. Rev. A **47**, R20 (1993).
[3] F. W. Meyer, S. H. Overbury, C. C. Havener, P. A. Zeijlmans van Emmichoven, and D. M. Zehner, Phys. Rev. Lett. **67**, 723 (1991).
[4] F. W. Meyer, S. H. Overbury, C. C. Havener, P. A. Zeijlmans van Emmichoven, J. Burgdörfer, and D. M. Zehner, Phys. Rev. A **44**, 7214 (1991).
[5] J. Burgdörfer, P. Lerner, and F. W. Meyer, Phys. Rev. A **44**, 5674 (1991).
[6] F. W. Meyer, C. C. Havener, and P. A. Zeijlmans van Emmichoven, Phys. Rev. A **48**, 4476 (1993).
[7] F. W. Meyer, L. Folkerts, I. G. Hughes, S. H. Overbury, D. M. Zehner, P. A. Zeijlmans van Emmichoven, and J. Burgdörfer, Phys. Rev. A **48**, 4479 (1993).
[8] S. A. Deutscher, X. Yang, and J. Burgdörfer, Phys. Rev. A **55**, 466 (1997).
[9] J. Das, L. Folkerts, and R. Morgenstern, Phys. Rev. A **45**, 4669 (1992).
[10] J. Das and R. Morgenstern, Phys. Rev. A **47**, R755 (1993).
[11] R. Köhrbrück, N. Stolterfoht, S. Schippers, S. Hustedt, W. Heiland, D. Lecler, J. Kemmler, and J. Bleck-Neuhaus, Phys. Rev. A **48**, 3731 (1993).
[12] J. Bleck-Neuhaus, A. Saal, R. Page, P. Biermann, R. Köhrbrück, and N. Stolterfoht, Phys. Rev. A **49**, R1539 (1994).
[13] J.-P. Briand, G. Giardino, G. Borsoni, M. Froment, M. Eddrief, C. Sebenne, S. Bardin, D. Schneider, J. Jin, H. Khemliche, Z. Xie, and M. Prior, Phys. Rev. A **54**, 4136 (1996).
[14] P. Nordlander and J. C. Tully, Phys. Rev. Lett. **61**, 990 (1988); Surf. Sci. **211/212**, 207 (1989); Phys. Rev. B **42**, 5564 (1990).
[15] D. Teillet-Billy and J. P. Gauyacq, Surf. Sci. **239**, 343 (1990).
[16] A. G. Borisov, D. Teillet-Billy, and J. P. Gauyacq, Nucl. Instrum. Methods Phys. Res. B **78**, 49 (1993).
[17] A. V. Chaplik, Zh.Éksp. Teor. Fiz. **54**, 332 (1968) [Sov. Phys. JETP **27**, 178 (1968)].
[18] R. K. Janev, J. Phys. B **7**, 1506 (1974).
[19] R. K. Janev and N. N. Nedeljković, J. Phys. B: At. Mol. Phys. **18**, 915 (1985).
[20] M. I. Chibisov, Dokl. Akad. Nauk SSSR **304**, 854 (1989) [Sov. Phys. Dokl. **34**, 138 (1989)].
[21] S. E. Kupriyanov, JETP Lett. **5**, 197 (1967).
[22] C. Fabre, M. Gross, J. M. Raimond, and S. Haroche, J. Phys. B: At. Mol. Phys. **16**, L671 (1983).
[23] V. Sandogher, C. I. Sukenik, E. A. Hinds, and S. Haroche, Phys. Rev. Lett. **68**, 3432 (1992).
[24] K. Ganesan and K. T. Taylor, J. Phys. B: At. Mol. Opt. Phys. **29**, 1293 (1996).
[25] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford–New York (1984).
[26] W. R. Smythe, *Static and Dynamic Electricity*, Second Edition, New York–Toronto–London (1950).
[27] J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, INC, New York–London (1962).
[28] I. E. Tamm, *Principles of Electricity Theory* [in Russian], Nauka, Moscow (1957).
[29] I. V. Komarov, L. I. Ponomarev, and S. Yu. Slavyanov, *Spherical and Coulomb Spherical Functions* [in Russian], Nauka, Moscow (1976).
[30] R. W. Hamming, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York (1973).
[31] L. D. Landau and E. M. Lifshitz, *Quantum mechanics: Non-Relativistic Theory*, Pergamon Press, Oxford–New York (1977).
[32] A. A. Radtsig and B. M. Smirnov, *Parameters of Atoms and Atomic Ions* [in Russian], Énergoatomizdat, Moscow (1986).
[33] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, Oxford–New York (1978).
[34] D. Pines, *Elementary Excitations in Solids*, W. A. Benjamin, New York (1963).
[35] Yu. V. Gott, *Interaction between Particles and Matter in Plasma Research* [in Russian], Atomizdat, Moscow (1978).
[36] M. I. Chibisov, *Auger Recombination Rate of Atomic Ions Near a Metal Surface*, Preprint IAE-5908/6, Moscow (1995).

Translation was provided by the Russian Editorial office.

The $a\gamma\gamma\gamma$ vertex and three-photon axion decay in an external magnetic field

V. V. Skobelev

Moscow State Industrial University, 109280, Moscow, Russia
(Submitted 6 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 26–34 (July 1999)

The axion vertex $a\gamma\gamma\gamma$, the probability of three-photon axion decay in an external magnetic field, and the cross section of the crossing process $a\gamma\rightarrow 2\gamma$, which *CP* invariance forbids in vacuum, are calculated for the first time. It is shown that in superstrong magnetic fields $B\gg F_0=m^2/|e|=4.41\cdot 10^{13}$ G the probability of three-photon decay is greater than the probability of two-photon decay. The astrophysical aspects of the questions examined are discussed.
 [S1063-7761(99)00307-8]

1. INTRODUCTION

The necessity of the existence of an axion as a pseudoscalar goldstone boson, explaining the *CP* invariance of strong interactions,[1] is now generally accepted. To give the axion real standing laboratory experiments are now being conducted or planned,[2,3] while previously the axion parameters were estimated only on the basis of astrophysical and cosmological considerations. Thus, the first laboratory estimate of the upper limit of the axion–photon coupling constant in an experiment based on the Primakov effect with conversion of solar axions into photons in a coherent interaction with a crystal lattice was given in Ref. 2 as $g_{a\gamma\gamma}<2.7\cdot 10^{-9}\,\mathrm{GeV}^{-1}$. The limit on $g_{a\gamma\gamma}$ was obtained irrespectively of the axion mass m_a right up to the upper limit $\sim$1 keV. An experiment on the conversion of solar axions into x-ray photons in a system of strong magnetic fields could be realized in the very near future.[3]

The most reliable data on the values of the axion parameters are still based on cosmological and astrophysical results. For example, the possible axion mass range is predicated on existing ideas about the rate at which stars lose energy (upper mass limit) remaining unchanged and on the axion contribution to the nonbaryonic component of the invisible mass of the universe (lower limit):[4]

$$10^{-6}\ \mathrm{eV}\leqslant m_a\leqslant 10^{-3}\ \mathrm{eV}. \tag{1}$$

The main axion production channel in the interior regions of stars and collapsed objects could be the Primakov effect[5] and possibly a synchrotron mechanism due to the presence of strong magnetic fields. However, the only decay channel is considered to be the two-phonon channel $a\rightarrow 2\gamma$ with the decay probability per unit time[7]

$$W_{2\gamma}=\frac{g_{a\gamma\gamma}^2 m_a^4}{64\pi q_0},\quad g_{a\gamma\gamma}=\frac{e^2 c_\gamma}{2\pi f}, \tag{2}$$

where $e^2=\alpha=1/137$, q_0 is the axion energy, c_γ is a constant depending on the model and is of the order of 1, and f is the energy scale for breaking of the Peccei–Quinn symmetry.[1] However, strong magnetic fields of the order of and greater than the characteristic Schwinger field[8] $F_0=m^2/|e|=4.41\times 10^{13}$ G (m — electron mass) can exist in extreme astrophysical situations (the Big Bang, neutron stars), where the axion component of the invisible mass was mainly formed. Such fields can fundamentally change the characteristics of decays by influencing the propagator factors of the electron loop, which makes the main contribution to the amplitude because of the very small mass of the charged fermion. Specifically, three-photon decay $a\rightarrow 3\gamma$ (see Fig. 1), which *CP* invariance forbids in vacuum, becomes possible. The structure of the axion vertex in the diagrams is determined by the adopted form of the axion–electron interaction Lagrangian[9]

$$\mathscr{L}_{ae}=\frac{c_e}{2f}(\overline{\Psi}\gamma_\mu\gamma^5\Psi)\frac{\partial a}{\partial x_\mu}, \tag{3}$$

where $c_e\sim 1$ is a constant.

In the present paper the amplitude and probability of three-photon axion decay in a magnetic field and the cross section of the crossing process $a\gamma\rightarrow 2\gamma$ are calculated for the first time, and it is shown that in fields with induction $B\gg F_0$ the probability of three-photon decay of relativistic axions is greater than the probability of two-photon decay.

The exposition is organized as follows. The form of the Green's function for the Dirac equation in a magnetic field, including the asymptotic limit of superstrong fields, is presented together with a discussion of the algebra of $\tilde{\gamma}$ matrices in two-dimensional space, in Sec. 2. The form of the $(a\gamma\gamma\gamma)$ vertex in a magnetic field is obtained in Sec. 3, and the probability of the decay $a\rightarrow 3\gamma$ and the cross section of the inelastic process $a\gamma\rightarrow\gamma\gamma$ are obtained in Sec. 4. The results obtained, including the astrophysical aspects, are analyzed in the last section.

2. GREEN'S FUNCTION FOR THE DIRAC EQUATION IN AN EXTERNAL MAGNETIC FIELD AND THE ALGEBRA OF 2×2 $\tilde{\gamma}$ MATRICES

The solution of the singular Dirac equation

$$(i\hat{\partial}-e\hat{A}-m)G(x,y)=\delta(x-y) \tag{4}$$

in an external constant and uniform magnetic field with induction B in the special gauge of the A potential can be written as

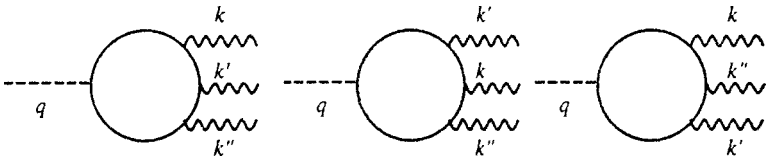

FIG. 1.

$$G(x,y)=\exp\left[-\frac{i\gamma}{2}(x_1+y_1)(x_2-y_2)\right]G(x-y), \tag{5}$$

where $\gamma=|eB|$ and the 3 axis is oriented along the vector **B**. It is convenient to use for the Fourier transform $G(p)$ of the function $G(x-y)$, which depends only on the coordinate difference, the representation

$$G(p)=\frac{1}{\gamma\eta}\int_0^1 dt\left(\frac{1+t}{1-t}\right)^{\eta}\exp(-\delta t)\left\{(\hat{p}_{\parallel}+m)\times\left[\Pi_-(1-\delta t)-\frac{\eta}{1+t}\right]-\eta\hat{p}_{\perp}\right\} \tag{6}$$

obtained in Ref. 10. Here

$$\eta=(p_{\parallel}^2-m^2)/2\gamma,\quad p_{\parallel}^2=p_0^2-p_3^2,$$
$$\delta=p_{\perp}^2/\gamma,\quad p_{\perp}^2=p_1^2+p_2^2,\quad \hat{p}_{\parallel}=\gamma^0p_0+\gamma^3p_3, \tag{7}$$
$$\hat{p}_{\perp}=\gamma^1p_1+\gamma^2p_2,$$

and $\Pi_-=(1-i\gamma_1\gamma_2)/2$ is the operator projecting spin on a direction opposite to the field. The expansion of the function (6) in the field in an invariant form in terms of the field tensor F has the form

$$G(p)=G_0-\frac{ie}{2}G_0\left(\gamma F\frac{\partial}{\partial p}\right)G_0+\ldots,$$
$$G_0=\frac{\hat{p}+m}{p^2-m^2}. \tag{8}$$

In the strong-field limit $B\gg F_0$, the integrand in the expression (6) must be formally expanded into powers of η (in the inverse field). This expansion is valid if the integrals over the two-dimensional momentum in the subspace (0, 3) on the electron mass m converges. The leading term, employed below, of the expansion is

$$G(p)=2e^{-\delta}\Pi_-\frac{\hat{p}_{\parallel}+m}{p_{\parallel}^2-m^2}. \tag{9}$$

As noted in Ref. 11, this procedure actually ''two-dimensionalizes'' the mathematical apparatus of the theory. Specifically, for loop diagrams, the expression remaining after Gaussian-type integrals over $p_{\perp}$ are performed reduces to a two-dimensional expression in the subspace (0, 3), since vertex factors of the form $\Pi_-\gamma_\alpha\Pi_-$ and $\Pi_-\gamma_\alpha\gamma^5\Pi_-$ are different from zero only for $\alpha=0$ and 3, $\gamma^0\gamma^3$ playing the role of the matrix γ^5 since $\Pi_-\gamma^5\Pi_-=\Pi_-\gamma^0\gamma^3$. Finally, the presence of the operator Π_- decreases the dimension of the γ matrices to 2×2, after which Π_- can be dropped, denoting the 2×2 matrices by the symbol $\tilde{\gamma}_\alpha$ ($\alpha=0,3$), while $\tilde{\gamma}^5=\tilde{\gamma}^0\tilde{\gamma}^3$. The function

$$G_s(p)=\frac{\check{p}+m}{p_{\parallel}^2-m^2},\quad \check{p}=\tilde{\gamma}^0p_0+\tilde{\gamma}^3p_3, \tag{10}$$

plays the role of the Green's function in two-dimensional space. It enters in the matrix elements of the corresponding diagrams (see Sec. 3).

It can be shown[11] that an expanded variant of Furry's theorem holds in two-dimensional space — the matrix elements of loop diagrams with an odd number of vertices vanish irrespective of their P classification (vector of pseudovector). As a result, diagrams with an even number of vertices are linear functions of the field,[12] whereas diagrams with odd number of vertices (even beyond the two-dimensional approximation) approach a constant, indicating that the former predominate in fields $B\gg F_0$.

We note also that the maximum divergence in two-dimensional expressions of the vacuum diagrams is logarithmic, and in all cases which we have considered previously it cancels in accordance with the condition for applicability of the expansion (9) (the only exception arises in the calculation of the electron mass operator,[13] where the method works to logarithmic accuracy in the field).

The two-dimensionalization of the mathematical apparatus can be explained physically by the suppression of the transverse excitations of virtual electrons in fields $B\gg F_0$ with the Landau ground state making the dominant contribution.

The algebraic relations used below for the $\tilde{\gamma}$ matrices follow from their basic reduction property:

$$\tilde{\gamma}^\alpha\tilde{\gamma}^\beta=\tilde{g}^{\alpha\beta}+\tilde{\gamma}^5\varepsilon^{\alpha\beta},\quad (\tilde{\gamma}^5)=1, \tag{11}$$

where $\tilde{g}^{\alpha\beta}=(1,-1)$ is the metric tensor in the subspace (0, 3), and $\varepsilon^{\alpha\beta}$ is the absolutely antisymmetric tensor in (0, 3) with the values $\varepsilon^{03}=-\varepsilon^{30}=1$ and $\varepsilon^{00}=\varepsilon^{33}=0$. The identities

$$\varepsilon^{\alpha\beta}\varepsilon^{\rho\sigma}=\tilde{g}^{\alpha\sigma}\tilde{g}^{\beta\rho}-\tilde{g}^{\alpha}\tilde{g}^{\beta\sigma}, \tag{12a}$$

$$\tilde{g}^{\alpha\beta}\varepsilon^{\rho\sigma}+\varepsilon^{\alpha\beta}\tilde{g}^{\rho\sigma}=\tilde{g}^{\beta\rho}\varepsilon^{\alpha\sigma}+\varepsilon^{\rho\beta}\tilde{g}^{\alpha\sigma}, \tag{12b}$$

$$\tilde{\gamma}_a(\tilde{\gamma}_{\alpha_1}\ldots\tilde{\gamma}_{\alpha_{2n+1}})\tilde{\gamma}^\alpha=0 \tag{12c}$$

greatly simplifies the procedure for calculating the traces:

$$\frac{1}{2}\mathrm{Tr}(\tilde{\gamma}^\alpha\tilde{\gamma}^\beta\tilde{\gamma}^\rho\tilde{\gamma}^\sigma)=\tilde{g}^{\alpha\beta}\tilde{g}^{\rho\sigma}+\tilde{g}^{\alpha\sigma}\tilde{g}^{\beta\rho}-\tilde{g}^{\alpha\rho}\tilde{g}^{\beta\sigma},$$

$$\frac{1}{2}\mathrm{Tr}(\tilde{\gamma}^5\tilde{\gamma}^\alpha\tilde{\gamma}^\beta\tilde{\gamma}^\rho\tilde{\gamma}^\sigma)=\tilde{g}^{\alpha\beta}\varepsilon^{\rho\sigma}+\varepsilon^{\alpha\beta}\tilde{g}^{\alpha\beta}\tilde{g}^{\rho\sigma}, \tag{12d}$$

and so on.

3. THE $a\gamma\gamma\gamma$ VERTEX IN A CONSTANT AND UNIFORM MAGNETIC FIELD

The $a\gamma\gamma\gamma$ vertex corresponds to the three diagrams in Fig. 1, and the complete interaction Lagrangian is a sum of the expression (3) and the electrodynamic part $e(\overline{\Psi}\gamma^{\alpha}\Psi)A_{\alpha}$ with the operators represented in the Furry picture and the electron loop making the dominant contribution. Noting that the phase factor in Eq. (5) in a loop with an even number of vertices cancels, we shall determine the matrix element of the process $a\to 3\gamma$ in terms of the S-matrix element as

$$\langle f|S|i\rangle=\frac{(2\pi)^4\delta(q-k-k'-k'')}{2q_0 2k_0 2k_0' 2k_0''}M, \tag{13}$$

$$M=M_{\alpha\alpha'\alpha''}e^{*\alpha}e^{*\alpha'}e^{*\alpha''}, \tag{13a}$$

where q is the momentum of the axion, k, k', k'' and e, e', e'' are the momenta and polarization vectors of the photons, and the pseudotensor $M_{\alpha\alpha'\alpha''}$ has the form

$$M_{\alpha\alpha'\alpha''}=\frac{-iec_e(4\pi)^{3/2}}{6f}\int d^4y\int d^4x\int d^4x' \times\exp[i(kx+k'x'-qy)]\mathrm{Tr}[\gamma_{\alpha}G(x-x')\gamma_{\alpha'}G(x')\gamma_{\alpha''}G(-y)\hat{q}\gamma^5 G(y-x)] + \text{two photon permutations}. \tag{14}$$

For fields $B\ll F_0$, in the low-energy approximation with respect to the momenta of the external lines (as compared with the electron mass m), the nonzero contribution according to the ''standard'' Furry theorem corresponds to an odd number of interactions with the field in the loop. From gauge and Lorentz invariance considerations the pseudotensor $M_{\alpha\alpha'\alpha''}$ is, to within numerical factors, in first orders in the field

$$M_{\alpha\alpha'\alpha''}=\frac{e^4c_e}{fm^4}\left[\{F\otimes e\otimes k^3\}_{\alpha\alpha'\alpha''}+\frac{1}{F_0^2}\times\{F^3\otimes e\otimes k^3\}_{\alpha\alpha'\alpha''}+\dots\right], \tag{15}$$

where in the brackets $e\equiv e_{\mu\nu\alpha\beta}$ is the absolutely antisymmetric tensor, and the direct-product symbols in the braces denote gauge-invariant combinations of the form

$$(k''\tilde{F})_{\alpha''}[k_{\alpha'}k'_{\alpha}-g_{\alpha\alpha'}(kk')],$$

$$(k''F)_{\alpha''}e_{\mu\nu\alpha\alpha'}k^{\mu}k'^{\nu};$$

. . . .

in the first term and

$$(k''\tilde{F})_{\alpha''}(k'F)_{\alpha'}(kF)_{\alpha},$$

$$(k''\tilde{F}F^2)_{\alpha''}[k_{\alpha'}k'_{\alpha}-g_{\alpha\alpha'}(kk')],$$

. . . .

in the second term, and so on ($\tilde{F}$ — dual tensor). In this approximation there is no need to calculate the tensor $M_{\alpha\alpha'\alpha''}$ exactly, since the corresponding decay probability $W_{3\gamma}$ in any case is small compared with the probability $W_{2\gamma}$ (see Sec. 4).

The situation is completely different in the strong-field limit $B\gg F_0$. Here detailed calculations are required. Substituting into Eq. (14) the expression

$$G(z)=\frac{1}{(2\pi)^4}\int d^4p\,G(p)\exp[-i(pz)]$$

with $G(p)$ from Eq. (9) and integrating the loop over all variables except for the two-dimensional momentum using the two-dimensional representation of the Green's function (10) in accordance with the method described in Sec. 2, we obtain

$$M_{\alpha\alpha'\alpha''}=\frac{2e^3c_e\gamma}{3\sqrt{\pi}f}J_{\alpha\alpha'\alpha''}, \tag{16}$$

$$J_{\alpha\alpha'\alpha''}=I_{\alpha\alpha'\alpha''}(k,k',k'')+I_{\alpha'\alpha\alpha''}(k',k,k'')+I_{\alpha\alpha''\alpha'}(k,k'',k'), \tag{16a}$$

$$I_{\alpha\alpha'\alpha''}(k,k',k'')=\frac{i}{\pi}\int d^2p\cdot\frac{1}{2}\mathrm{Tr}[\tilde{\gamma}^5\check{q}G_s(p)\tilde{\gamma}_{\alpha}G_s\times(p+k)\tilde{\gamma}_{\alpha'}G_s(p+k+k')\tilde{\gamma}_{\alpha''}G_s\times(p+q)];$$

$$q=k+k'+k''; \quad \alpha,\alpha',\alpha''=0,3. \tag{16b}$$

The symmetrized expression $J_{\alpha\alpha'\alpha''}$ vanishes to first order of the expansion of $I_{\alpha\alpha'\alpha''}$ in terms of the momenta, as should happen because of gauge invariance.

The exact result for the two-dimensional rank-2 pseudotensor $I_{\alpha\alpha'\alpha''}$ in the form of a triple integral over the Feynman parameters is

$$
\begin{aligned}
&I_{\alpha\alpha'\alpha''}(k,k',k'')\\
&=2\int_0^1 dx_1\int_0^{x_1} dx_2\int_0^{x_2} dx_3\Bigg\{-\frac{m^4}{(m^2-\tilde m^2)^3}[q_\alpha\varepsilon_{\alpha'\alpha''}\\
&+(q\varepsilon)_\alpha\tilde g_{\alpha'\alpha''}]+\frac{m^2}{(m^2-\tilde m^2)^2}(q\varepsilon)_{\alpha'}\tilde g_{\alpha\alpha''}-\frac{m^2}{(m^2-\tilde m^2)^3}\\
&\times[-\varepsilon_{\alpha\alpha'}\kappa^{(2)}_{\alpha''}(\kappa^{(3)}q)-\tilde g_{\alpha\alpha'}(\kappa^{(2)}\varepsilon)_{\alpha''}(\kappa^{(3)}q)\\
&-\tilde g_{\alpha\alpha'}\kappa^{(2)}_{\alpha''}(\kappa^{(3)}\varepsilon q)-\varepsilon_{\alpha\alpha'}(\kappa^{(2)}\varepsilon)_{\alpha''}(\kappa^{(3)}\varepsilon q)\\
&+(q\varepsilon)_\alpha\kappa^{(1)}_{\alpha'}\kappa^{(2)}_{\alpha''}+q_\alpha(\kappa^{(1)}\varepsilon)_{\alpha'}\kappa^{(2)}_{\alpha''}+q_\alpha\kappa^{(1)}_{\alpha'}(\kappa^{(2)}\varepsilon)_{\alpha''}\\
&+(q\varepsilon)_\alpha(\kappa^{(1)}\varepsilon)_{\alpha'}(\kappa^{(2)}\varepsilon)_{\alpha''}-q^2(\kappa^{(1)}\varepsilon)_\alpha\tilde g_{\alpha'\alpha''}\\
&+q^2\kappa^{(1)}_\alpha\varepsilon_{\alpha'\alpha''}+2(q\varepsilon\kappa)\kappa^{(1)}_\alpha\tilde g_{\alpha'\alpha''}-2(q\varepsilon\kappa)\\
&\times(\kappa^{(1)}\varepsilon)_\alpha\varepsilon_{\alpha'\alpha''}+(q\varepsilon\kappa)\tilde g_{\alpha\alpha'}(\kappa^{(2)}+\kappa^{(3)})_{\alpha''}\\
&+(q\kappa)\varepsilon_{\alpha\alpha'}(\kappa^{(2)}+\kappa^{(3)})_{\alpha''}+(q\kappa)\tilde g_{\alpha\alpha'}(k''\varepsilon)_{\alpha''}\\
&+(q\varepsilon\kappa)\varepsilon_{\alpha\alpha'}(k''\varepsilon)_{\alpha''}]\\
&-\frac{1}{(m^2-\tilde m^2)^3}[(q\varepsilon\kappa)(\kappa^{(1)}_\alpha\kappa^{(2)}_{\alpha'}\kappa^{(3)}_{\alpha''}\\
&+\kappa^{(1)}_\alpha(\kappa^{(2)})\varepsilon_{\alpha'}(\kappa^{(3)}\varepsilon)_{\alpha''}+(\kappa^{(1)}\varepsilon)_\alpha\kappa^{(2)}_{\alpha'}(\kappa^{(3)}\varepsilon)_{\alpha''}\\
&+(\kappa^{(1)}\varepsilon)_\alpha(\kappa^{(2)}\varepsilon)_{\alpha'}\kappa^{(3)}_{\alpha''})-(q\kappa)((\kappa^{(1)}\varepsilon)_\alpha\kappa^{(2)}_{\alpha'}\kappa^{(3)}_{\alpha''}\\
&+\kappa^{(1)}_\alpha(\kappa^{(2)}\varepsilon)_{\alpha'}\kappa^{(3)}_{\alpha''}+\kappa^{(1)}_\alpha\kappa^{(2)}_{\alpha'}(\kappa^{(3)}\varepsilon)_{\alpha''}\\
&+(\kappa^{(1)}\varepsilon)_\alpha(\kappa^{(2)}\varepsilon)_{\alpha'}(\kappa^{(3)}\varepsilon)_{\alpha''})]\Bigg\}. \qquad (17)
\end{aligned}
$$

In Eq. (17) and below in this section all contractions and scalar products are two-dimensional in the space (0, 3). For brevity, the following notation has been introduced:

$$
\begin{aligned}
\tilde m^2&=k^2(1-x_2+x_3)(x_2-x_3)+(q+k'')^2(1-x_1+x_2)\\
&\times(x_1-x_2)+q^2x_1(1-x_1)-2k(q-k'')(x_2-x_3)\\
&\times(x_1-x_2)-2(kq)(x_2-x_3)(1-x_1)-2q(q-k'')\\
&\times(x_1-x_2)(1-x_1), \qquad (17a)\\
\kappa&=k(x_2-x_3)+(q-k'')(x_1-x_2)+q(1-x_1),\\
\kappa^{(1)}&=\kappa-k,\quad \kappa^{(2)}=\kappa-q+k'',\quad \kappa^{(3)}=\kappa-q.
\end{aligned}
$$

In principle the expressions (16) and (17) hold in the entire region $\gamma\gg q_0^2$, m^2, but in practice it is possible only to study the low-energy approximation, where all momenta are much less than the electron mass. In this case the simple result

$$
J_{\alpha\alpha'\alpha''}=\frac{4q^2}{15m^6}(k\varepsilon)_\alpha(k'\varepsilon)_{\alpha'}(k''\varepsilon)_{\alpha''} \qquad (18)
$$

is obtained from Eqs. (17) and (16a) after complicated calculations. This expression is explicitly gauge-invariant

$$
J_{\alpha\alpha'\alpha''}k^\alpha=J_{\alpha\alpha'\alpha''}k'^{\alpha'}=J_{\alpha\alpha'\alpha''}k''^{\alpha''}=0. \qquad (18a)
$$

We note that $J_{\alpha\alpha'\alpha''}$ is proportional to the fifth power of the momenta, though in principle the condition (18a) does not rule out a cubic combination, which, however, vanishes identically in a specific calculation.

If the state of linear polarization of a photon is characterized by the vector

$$
e_\alpha=(k\varepsilon)_\alpha/\sqrt{k^2} \qquad (19)
$$

and the orthogonal vector, then the latter does not contribute at all because of the two-dimensionality of the contractions. Thus for ''nonsterile'' polarization states the matrix element is

$$
\begin{aligned}
M&=M_{\alpha\alpha'\alpha''}\frac{(k\varepsilon)^\alpha}{\sqrt{k^2}}\frac{(k'\varepsilon)^{\alpha'}}{\sqrt{k'^2}}\frac{(k''\varepsilon)^{\alpha''}}{\sqrt{k''^2}}\\
&=-\frac{8e^3c_e(B/F_0)}{45\sqrt{\pi}m^4f}q^2k_\perp k'_\perp k''_\perp. \qquad (20)
\end{aligned}
$$

Here the relation (12a) was used and the fact that $\sqrt{k^2}=k_\perp$, where $k_\perp$ is the component of the photon momentum transverse to the field, is taken into account. As indicated in Sec. 2, M increases linearly with the field.

4. THREE-PHOTON AXION DECAY AND THE INELASTIC PROCESS $\gamma a\to\gamma\gamma$ IN A MAGNETIC FIELD

In the region $F\ll F_0$ (F — field amplitude) it is easy to obtain, using the expressions (13) and (15), to within numerical factors and taking account of the significant powers of the small parameters the probability of three-photon axion decay summed over the polarization states of the photons

$$
\begin{aligned}
W_{3\gamma}&\simeq\frac{\alpha^3c_e^2m_a^2m^2}{f^2q_0}\Bigg[a_1\left(\frac{F}{F_0}\right)^2\left(\frac{m_a}{m}\right)^6+a_2\chi^2\left(\frac{m_a}{m}\right)^4\\
&+a_3\chi^4\left(\frac{m_a}{m}\right)^2+a_4\chi^6\Bigg], \qquad (21)
\end{aligned}
$$

where a_i are numerical coefficients and

$$
\chi=\frac{\sqrt{e^2(qF^2q)}}{m^3}\ll 1. \qquad (22)
$$

Comparing the expressions (21) and (2) shows that only the contribution of the last term in Eq. (21) can compete with the probability of two-photon decay, but it is less than the corresponding field corrections to $W_{2\gamma}$.

If the tensor (15), just as in the two-dimensional variant (18), is proportional to the fifth power of the momenta, then this result only becomes stronger, since additional small factors will appear in the expression (21).

We now consider the case of superstrong fields, where the matrix element is determined by the expression (20). The integral over k'' removes the δ function, and to integrate over k' it is convenient to orient the axis of the spherical coordinate system along the vector $\mathbf{p}=\mathbf{q}-\mathbf{k}$. The calculations yield the following expression for the probability of three-photon axion decay per unit time as a distribution over the momentum of a single photon:

$$W_{3\gamma}=\frac{4\alpha^3 c_e^2 (q^2)^2}{45^3(2\pi)^5 f^2 m^8 q_0}\left(\frac{B}{F_0}\right)^2 \int \frac{d^3k}{2k_0}k_\perp^2 \times[(p^2)^2+p^2p_\perp^2+p_\perp^4]. \quad (23)$$

Here the index $\perp$ once again denotes the vector components that are perpendicular to the field, q^2 and p^2 are squares in the subspace (0, 3), and the expression for $W_{3\gamma}$ is additionally divided by 3! because the photons are indistinguishable.

Next, orienting the axis of the spherical coordinate system along the vector **q**, the limits of integration over the photon energy are determined by the relation

$$0\leqslant k_0\leqslant \frac{m_a^2}{2(q_0-|\mathbf{q}|\cos\theta)}, \quad (24)$$

and the final result is

$$W_{3\gamma}=\frac{\alpha^3 c_e^2 m_a^2 (q_\perp^2+m_a^2)^2}{84\cdot 90^3(2\pi)^4 f^2 m^8 q_0}\left(\frac{B}{F_0}\right)^2 (4q_\perp^6+6q_\perp^4 m_a^2 + 8q_\perp^2 m_a^4+3m_a^6). \quad (25)$$

In the limiting cases of an axion at rest ($q_\perp=0$) and an ultrarelativistic axion ($q_\perp \gg m_a$) the corresponding probabilities are

$$W_{3\gamma}=\frac{\alpha^3 c_e^2 m_a^{12}(B/F_0)^2}{28\cdot 90^3(2\pi)^4 f^2 m^8 q_0}, \quad (25a)$$

$$W_{3\gamma}=\frac{\alpha^3 c_e^2 m_a^2 q_\perp^{10}(B/F_0)^2}{21\cdot 90^3(2\pi)^4 f^2 m^8 q_0}, \quad (25b)$$

The cross section of the inelastic process $a\gamma\to\gamma\gamma$ can be easily found using the expressions (20) and (23) and is

$$\sigma=\frac{\alpha^3 c_e^2 k_\perp^2 (m_a^2+q_\perp^2)^2 (B/F_0)^2}{9\cdot 15^3(2\pi)^2 f^2 m^8 (q_0k_0-\mathbf{q}\cdot\mathbf{k})}[(p^2)^2+p^2p_\perp^2+p_\perp^4], \quad (26)$$

where $p=q+k$.

5. DISCUSSION

It was shown in Ref. 14 that for $B\gg F_0$ the two-photon decay probability $W_{2\gamma}$ no longer depends on the field, differing from the probability (2) in the absence of a field by the formal substitution $c_\gamma\to c_e$. In other words, a superstrong field restores the isotropy of the space with respect to the decay $a\to 2\gamma$, as a result of which the relativistic factor $q_\perp^n$ is absent in the expression for $W_{2\gamma}$, in contrast to Eq. (25b). This and the proportionality of $W_{3\gamma}$ to the square of the field together make it possible for the three-photon axion decay channel to predominate over the two-photon channel. On the basis of these remarks and Eqs. (2) and (25b) we obtain

$$\frac{W_{3\gamma}}{W_{2\gamma}}\simeq 2.5\cdot 10^{-9}\left(\frac{q_\perp}{m}\right)^{10}\left(\frac{m}{m_a}\right)^2\left(\frac{B}{F_0}\right)^2. \quad (27)$$

As one can see, $W_{3\gamma}/W_{2\gamma}\gtrsim 1$ for hard axions ($q_\perp \lesssim m$) in the assumed mass range (1). For example, $W_{3\gamma}/W_{2\gamma}\sim 10$ for $q_\perp=0.1m$, $m_a=10^{-3}$ eV, and $B=10F_0$, not to mention smaller values of m_a from the ''window'' (1) or large values of the magnetic induction.

Of course, here there is no contradiction with perturbation theory, since the probabilities $W_{(2n+1)\gamma}$ and $W_{2n\gamma}$ of decay into odd and even numbers of photons, respectively, decrease with increasing n, but because of the specific nature of the decays in superstrong fields $W_{(2n+1)\gamma}>W_{2n\gamma}$ can hold for all n for the indicated values of the parameters.

As is well known,[9] the axion lifetime with respect to two-photon decay is greater than the age of the Universe, and this also holds in superstrong magnetic fields.[14] For this reason, the process $a\to 2\gamma$ plays no role on the astrophysical level. Using Eq. (25b), we represent the lifetime with respect to three-photon decay as

$$\tau_{3\gamma}\simeq\left(\frac{f/c_e}{10^{10}\,\mathrm{GeV}}\right)^2\left(\frac{10^{-3}\,\mathrm{eV}}{m_a}\right)^2\left(\frac{m}{q_\perp}\right)^9\left(\frac{F_0}{B}\right)^2\cdot 3\cdot 10^{33}\,\mathrm{s}. \quad (28)$$

If $\tau_{3\gamma}$ is to be comparable to the time of existence of superstrong magnetic fields during the epoch of the Big Bang, then $B\gtrsim 10^{15}F_0$ for hard axions $q_\perp\sim m$, which is hardly possible. Therefore the astrophysical aspects of the three-photon axion decay process are quite problematic. Careful estimates show that this is also true for the other channel, whose cross section is given by Eq. (26), for ''vanishing'' of an axion.

I thank Yu. O. Yakovlev for technical assistance.

[1] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977). R. D. Peccei, *CP Violation*, edited by C. Jarlskog (World Scientific Publishers, Singapore, 1989).
[2] F. T. Avignone III, D. Abriola, R. L. Brodzinski *et al.*, Yad. Fiz. **61**, 1237 (1998) [Phys. At. Nucl. **61**, 1137 (1998)]
[3] M. Minova, Yad. Fiz. **61**, 1217 (1998) [Phys. At. Nucl. **61**, 1117 (1998)].
[4] G. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (University of Chicago Press, Chicago, 1996).
[5] K. van Bibber, P. N. McIntyre, D. E. Morris, and G. G. Raffelt, Phys. Rev. D **39**, 2089 (1989).
[6] A. V. Borisov and V. Yu. Grishina, Zh. Éksp. Teor. Fiz. **106**, 1553 (1994) [JETP **79**, 837 (1994)]; V. V. Skobelev, Zh. Éksp. Teor. Fiz. **112**, 25 (1997) [JETP **85**, 13 (1997)].
[7] G. G. Raffelt, Phys. Rev. D **33**, 897 (1986).
[8] C. Thompson and R. C. Duncan, Astrophys. J. **408**, 194 (1993); M. Bocquet *et al.*, Astron. Astrophys. J. **301**, 757 (1995).
[9] G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
[10] Yu. M. Loskutov and V. V. Skobelev, Yad. Fiz. **31**, 1279 (1980) [Sov. J. Nucl. Phys. **31**, 661 (1980)].
[11] V. V. Skobelev, Zh. Éksp. Teor. Fiz. **71**, 1263 (1976) [Sov. Phys. JETP **44**, 660 (1976)].
[12] Yu. M. Loskutov and V. V. Skobelev, Phys. Lett. A **62**, 53 (1977).
[13] Yu. M. Loskutov and V. V. Skobelev, Vestn. MGU, Ser. Fiz., Astr., No. **6**, 111 (1977).
[14] V. V. Skobelev, Yad. Fiz. **61**, 2236 (1998) [Phys. At. Nucl. **61**, 2123 (1998)].

Translated by M. E. Alferieff

Probing the field-induced variation of the chemical potential in $Bi_2Sr_2CaCu_2O_y$ via magneto-thermopower measurements

S. A. Sergeenkov

SUPRAS, Institute of Physics, University of Liege, B-4000 Liege, Belgium; Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, Moscow Region, Russia

M. Ausloos

SUPRAS, Institute of Physics, University of Liege, B-4000 Liege, Belgium

(Submitted 21 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 257–262 (July 1999)

Approximating the shape of the magneto-thermoelectric power (TEP) $\Delta S(T,H)$ measured in $Bi_2Sr_2CaCu_2O_y$ by an asymmetric linear triangle of the form $\Delta S(T,H)\simeq S_p(H)\pm B^{\pm}(H)(T_c-T)$ with positive $B^-(H)$ and $B^+(H)$ defined below and above T_c, we observe that $B^+(H)\simeq 2B^-(H)$. To account for this asymmetry, we explicitly introduce the field-dependent chemical potential $\mu(H)$ of holes into the Ginzburg–Landau theory and calculate both an average $\Delta S_{\rm av}(T,H)$ and fluctuation contribution $\Delta S_{\rm fl}(T,H)$ to the total magneto-TEP $\Delta S(T,H)$. As a result, we find a rather simple relationship between the field-induced variation of the chemical potential in this material and the above-mentioned magneto-TEP data around T_c, viz. $\Delta\mu(H)\propto S_p(H)$. © *1999 American Institute of Physics.* [S1063-7761(99)01807-7]

As is well-known,[1,2] the variation of the chemical potential μ of free carriers in an applied magnetic field H provides direct information about the magnetization structure inside a superconducting sample. Namely, the field-induced change of the chemical potential in superconducting state reads[3] $\Delta\mu(H)\equiv\mu(H)-\mu(0)=-M(H)H/n$, where $M(H)$ is the field-induced magnetization and n is the carrier number density. At the same time, due to the existence of the so-called compensation effect,[4] it is rather difficult to observe field-induced modulations of μ in bulk samples, since in equilibrium any field-induced variations of μ will be completely canceled by similar variations caused by the magnetostrictive changes of the volume. However, this compensation does not occur in thin films[1,2] and oriented powders.[5] Thus we can expect to see significant changes of $\mu(H)$ in layered (anisotropic) structures as well. On the other hand, in view of their carrier-sensitive nature, thermopower (TEP) measurements seem to be the most adequate tool for probing the field-induced changes of the chemical potentials. Indeed, TEP results have already proved to be useful for providing reasonable estimates for such important physical parameters as the Fermi energy, Debye temperature, interlayer spacing etc.[6,7] Studying the observable magneto-TEP $\Delta S(T,H)=S(T,H)-S(T,0)$ also provides important insights into different aspects of the material in the mixed state[7–9] (when $H_{c1}\ll H\ll H_{c2}$). When experimental results are presented in the form of $\Delta S(T,H)$ one observes that its temperature dependence has a Λ-like shape asymmetric around T_c where it reaches its magnetic field-dependent peak value $S_p(H)\equiv\Delta S(T_c,H)$. Then for small fields, approximating the shape of $\Delta S(T,H)$ by the asymmetric linear triangle of the form[8]

$$\Delta S(T,H)\simeq S_p(H)\pm B^{\pm}(H)(T_c-T), \tag{1}$$

with positive slopes $B^-(H)$ and $B^+(H)$ defined for $T<T_c$ and $T>T_c$, respectively, one finds (see Fig. 1) that $B^+(H)\simeq 2B^-(H)$ in the vicinity of T_c.

In the present paper, using the Ginzburg–Landau theory and utilizing some typical magneto-TEP data[7,8] on textured $Bi_2Sr_2CaCu_2O_y$, we discuss the mixed-state behavior of the magneto-TEP [and in particular the origin of the asymmetry given by Eq. (1)] in terms of the corresponding behavior of the chemical potential in applied magnetic field.

It is well-known[7–9] that for external fields H such that $H_{c1}\ll H\ll H_{c2}$ and for the Ginzburg–Landau parameter $\kappa\gg 1$, the magneto-TEP $\Delta S(T,H)$ is proportional to the strength of the external field. To describe the observed behavior of the magneto-TEP both below and above T_c, we can roughly present it in a two-term contribution form[7]

$$\Delta S(T,H)=\Delta S_{\rm av}(T,H)+\Delta S_{\rm fl}(T,H), \tag{2}$$

where the average term $\Delta S_{\rm av}(T,H)$ is assumed to be nonzero only below T_c (since in the normal state the TEP of high-T_c superconductors (HTSCs) is found to be very small[8,9]) while the fluctuation term $\Delta S_{\rm fl}(T,H)$ should contribute to the observable $\Delta S(T,H)$ for $T\simeq T_c$. In what follows, we shall discuss these two contributions separately within the mean-field theory approximation.

MEAN VALUE OF THE MAGNETO-TEP: $\Delta S_{\rm av}(T,H)$

Assuming that the net result of the magnetic field is to modify the chemical potential (Fermi energy) μ of quasiparticles, we can write the generalized GL free energy functional $\mathcal{G}$ of a superconducting sample in the mixed state as

$$\mathcal{G}[\psi]=a(T)|\psi|^2+\frac{\beta}{2}|\psi|^4-\mu|\psi|^2. \tag{3}$$

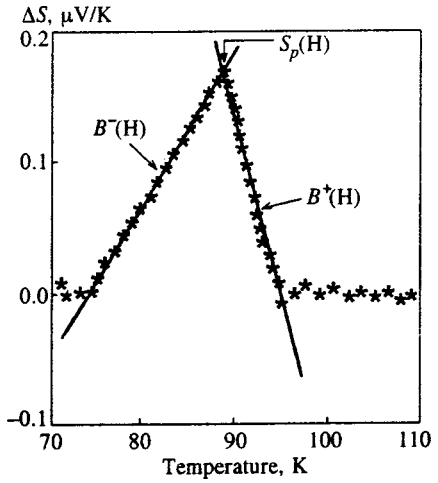

FIG. 1. A typical pattern of the observed[8] magneto-TEP $\Delta S(T,H)$ of $Bi_2Sr_2CaCu_2O_y$ at $H=0.12$ T. The best fit to the data points according to Eq. (1) yields $S_p(H)=0.16\pm0.01$ μV/K for the peak, and $B^-(H)=0.012\pm0.001$ μV/K^2 and $B^+(H)=0.027\pm0.003$ μV/K^2 for the slopes.

Here $\psi=|\psi|e^{i\phi}$ is the superconducting order parameter, $\mu(H)$ stands for the field-dependent in-plane chemical potential of the quasiparticles; $a(T,H)=\alpha(H)(T-T_c)$ and the GL parameters $\alpha(H)$ and $\beta(H)$ are related to the critical temperature T_c, the zero-temperature BCS gap $\Delta_0=1.76k_BT_c$, the out-of-plane chemical potential (Fermi energy) is $\mu_c(H)$, and the total particle number density n as $\alpha(H)=\beta(H)n/T_c=2\Delta_0k_B/\mu_c(H)$. In fact, in layered superconductors, $\mu=\mu_c/\gamma^2\simeq m^*_{ab}(J_cd/2\hbar)^2$, where d and J_c are the interlayer distance and coupling energy within the Lawrence–Doniach model, and $\gamma=m^*_c/m^*_{ab}$ is the mass anisotropy ratio. The magnetic field is applied normally to the *ab*-plane where the strongest magneto-TEP effects are expected.[9] In what follows, we ignore the field dependence of the critical temperature since for all fields under discussion $T_c(H)=T_c(0)(1-H/H_{c2})\simeq T_c(0)\equiv T_c$.

As usual, the equilibrium state of such a system is determined from the minimum-energy condition $\partial\mathcal{G}/\partial|\psi|=0$ which yields for $T<T_c$

$$|\psi_0|^2=\frac{\alpha(H)(T_c-T)+\mu(H)}{\beta(H)}. \tag{4}$$

Substituting $|\psi_0|^2$ into Eq. (3) we obtain for the average free energy density

$$\Omega(T,H)\equiv\mathcal{G}[\psi_0]=-\frac{[\alpha(H)(T_c-T)+\mu(H)]^2}{2\beta(H)}. \tag{5}$$

In turn, the magneto-TEP $\Delta S(T,H)$ can be related to the corresponding difference of the transport entropies,[7,8] $\Delta\sigma\equiv\partial\Delta\Omega/\partial T$, as $\Delta S(T,H)=\Delta\sigma(T,H)/en$, where e is the charge of the quasiparticles. Finally the mean value of the mixed-state magneto-TEP reads (below T_c)

$$\Delta S_{\rm av}(T,H)=S_{p,{\rm av}}(H)-B_{\rm av}(H)(T_c-T), \tag{6}$$

with

$$S_{p,{\rm av}}(H)=\frac{\Delta\mu(H)}{eT_c}, \tag{7}$$

and

$$B_{\rm av}(H)=\frac{8\Delta_0k_B\Delta\mu(H)}{eT_c\gamma^2\mu^2(0)}. \tag{8}$$

Before we proceed to compare the above theoretical findings with the available experimental data, we first have to estimate the corresponding fluctuation contributions to the observable magneto-TEP, both above and below T_c.

MEAN-FIELD GAUSSIAN FLUCTUATIONS OF THE MAGNETO-TEP: $\Delta S_{\rm fl}(T,H)$

The influence of superconducting fluctuations on the transport properties of HTSCs (including TEP and electrical conductivity) has been extensively studied for the past few years (see, e.g., Refs. 10–14 and further references therein). In particular, it was found that the fluctuation-induced behavior may extend to temperatures more than 10 K higher than the corresponding T_c. Let us consider now the region near T_c and discuss the Gaussian fluctuations of the mixed-state magneto-TEP $\Delta S_{\rm fl}(T,H)$. Recall that according to the theory of Gaussian fluctuations,[15] the fluctuations of any observable, which is conjugate to the order parameter ψ (such as heat capacity, susceptibility, etc.) can be presented in terms of the statistical average of the square of the fluctuation amplitude $\langle(\delta\psi)^2\rangle$ with $\delta\psi=\psi-\psi_0$. Then the TEP above $(+)$ and below $(-)$ T_c have the form

$$S^{\pm}_{\rm fl}(T,H)=A\langle(\delta\psi)^2\rangle_{\pm}=\frac{A}{Z}\int d|\psi|(\delta\psi)^2e^{-\Sigma[\psi]}, \tag{9}$$

where $Z=\int d|\psi|e^{-\Sigma[\psi]}$ is the partition function with $\Sigma[\psi]\equiv(\mathcal{G}[\psi]-\mathcal{G}[\psi_0])/k_BT$, and A is a coefficient to be defined below. Expanding the free energy density functional $\mathcal{G}[\psi]$

$$\mathcal{G}[\psi]\approx\mathcal{G}[\psi_0]+\frac{1}{2}\left[\frac{\partial^2\mathcal{G}}{\partial\psi^2}\right]_{|\psi|=|\psi_0|}(\delta\psi)^2, \tag{10}$$

around the mean value of the order parameter ψ_0, which is defined as a stable solution of equation $\partial\mathcal{G}/\partial|\psi|=0$, we can explicitly calculate the Gaussian integrals. Because $|\psi_0|^2$ is given by Eq. (4) below T_c and vanishes at $T\geqslant T_c$, we obtain finally

$$S^-_{\rm fl}(T,H)=\frac{Ak_BT_c}{4\alpha(H)(T_c-T)+4\mu(H)}, \quad T\leqslant T_c \tag{11}$$

and

$$S^+_{\rm fl}(T,H)=\frac{Ak_BT_c}{2\alpha(H)(T-T_c)-2\mu(H)}, \quad T\geqslant T_c. \tag{12}$$

As we shall see below, for the experimental range of parameters under discussion. $\mu(H)/\alpha(H)\gg|T_c-T|$. Hence, with a good accuracy we can linearize Eqs. (11) and (12) and obtain for the fluctuation contribution to the magneto-TEP

$$\Delta S^{\pm}_{\rm fl}(T,H)\simeq S^{\pm}_{p,{\rm fl}}(H)\pm B^{\pm}_{\rm fl}(H)(T_c-T), \tag{13}$$

where

$$S^{-}_{p,\mathrm{fl}}(H)=-\frac{Ak_BT_c\Delta\mu(H)}{4\mu^2(0)},\quad S^{+}_{p,\mathrm{fl}}(H)=-2S^{-}_{p,\mathrm{fl}}(H),\quad (14)$$

and

$$B^{-}_{\mathrm{fl}}(H)=-\frac{3Ak_B^2T_c\Delta_0\Delta\mu(H)}{\gamma^2\mu^4(0)},\quad B^{+}_{\mathrm{fl}}(H)=-2B^{-}_{\mathrm{fl}}(H). \quad (15)$$

Furthermore, it is quite reasonable to assume $S^{-}_{p}=S^{+}_{p}\equiv S_p$, where $S^{-}_{p}=S_{p,\mathrm{av}}+S^{-}_{p,\mathrm{fl}}$ and $S^{+}_{p}=S^{+}_{p,\mathrm{fl}}$. Then the above equations yield an explicit expression for the constant parameter A, namely $A=4\mu^2(0)/3ek_BT_c^2$. This in turn leads to the following expressions for the fluctuation contribution to peaks and slopes through their average counterparts [see Eqs. (7) and (8)]: $S^{+}_{p,\mathrm{fl}}(H)=(2/3)S_{p,\mathrm{av}}(H)$, $S^{-}_{p,\mathrm{fl}}(H)=-(1/3)S_{p,\mathrm{av}}(H)$, $B^{-}_{\mathrm{fl}}(H)=-(1/2)B_{\mathrm{av}}(H)$, and $B^{+}_{\mathrm{fl}}(H)=B_{\mathrm{av}}(H)$. Finally, the total contribution to the observable magneto-TEP reads [cf. Eq. (1)]

$$\Delta S(T,H)=S_p(H)\pm B^{\pm}(H)(T_c-T), \quad (16)$$

where

$$S_p(H)=\frac{2\Delta\mu(H)}{3eT_c},\quad B^{+}(H)\equiv B^{+}_{\mathrm{fl}}(H)=2B^{-}(H), \quad (17)$$

and

$$B^{-}(H)\equiv B_{\mathrm{av}}(H)+B^{-}_{\mathrm{fl}}(H)=\frac{4\Delta_0k_B\Delta\mu(H)}{eT_c\gamma^2\mu^2(0)}. \quad (18)$$

Let us compare now these theoretical expressions with typical experimental data[8] on textured $Bi_2Sr_2CaCu_2O_y$ for the slopes $B^{\pm}(H)$ and the peak $S_p(H)$ values for $H=0.12$T (see Fig. 1): $S_p=0.16\pm0.01\,\mu$V/K, $B^{-}=0.012\pm0.001\,\mu$V/K^2 and $B^{+}=0.027\pm0.003\,\mu$V/K^2. First we notice that the calculated slopes $B^{+}(H)$ above T_c are twice their counterparts below T_c, i.e., $B^{+}(H)=2B^{-}(H)$ in a good agreement with the observations. Using $\gamma\simeq55$ and $d=1.2$ nm for the anisotropy ratio and interlayer distance in this material,[9,13,16] we obtain reasonable estimates of the field-induced changes of the in-plane chemical potential (Fermi energy) $\Delta\mu(H)$ [along with its zero-field value $\mu(0)$] and the interlayer coupling energy J_c, namely, $\mu(0)\simeq1.6$ meV, $\Delta\mu(H)\simeq0.02$ meV, and $J_c\simeq4$ meV. Furthermore, relating the field-induced variation of the in-plane chemical potential to the change of the corresponding magnetization $M(H)$, viz.,

$$\Delta\mu(H)=-\frac{M(H)H}{n_h}, \quad (19)$$

where $M(H)$ for $H_{c1}\ll H\ll H_{c2}$ has the form[3] (recall that the lower critical field for this material is $H_{c1}=(\phi_0/4\pi\lambda^2_{ab})\ln\kappa\simeq40G$ with $\lambda_{ab}\simeq250$ nm, $\xi_{ab}\simeq1$ nm, and $\kappa\simeq250$)

$$\mu_0M(H)=\frac{2\phi_0}{\sqrt{3}\lambda^2_{ab}}-\left\{\ln\left[\frac{3\phi_0}{4\pi\lambda^2_{ab}(H-H_{c1})}\right]\right\}^{-2}-H, \quad (20)$$

we obtain $n_h\simeq2.5\cdot10^{27}\,\mathrm{m}^{-3}$ for the hole number density in this material, in reasonable agreement with the other estimates of this parameter.[17] Figure 2 shows $\Delta\mu(H)$ calculated according to Eq. (19) with the experimental data points deduced [via Eq. (17)] from the magneto-TEP measurements on the same sample.[7] As is seen, the data are in a good agreement with the model predictions. Finally, using the above parameters (along with the critical temperature), we find that $\mu(H)/\alpha(H)\simeq100$ K which justifies the use of the linearized Eq. (13) since, as is seen in Fig. 1, the observed magneto-TEP practically vanishes for $|T_c-T|\geqslant15$ K.

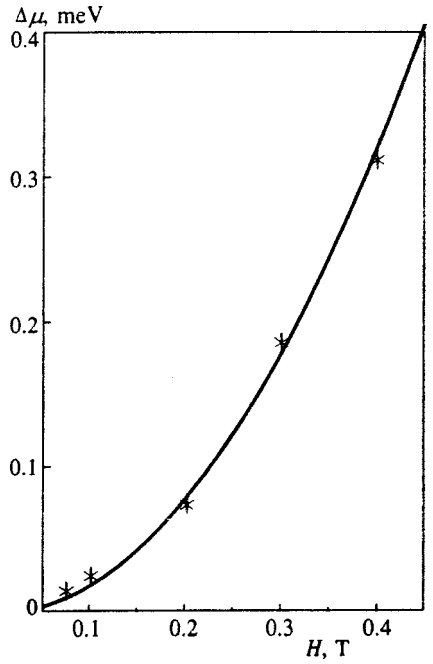

FIG. 2. The change of the chemical potential $\Delta\mu(H)$ in applied magnetic field calculated according to Eq. (19). The experimental points are deduced from the magneto-TEP data[7] on $Bi_2Sr_2CaCu_2O_y$ and related to $\Delta\mu(H)$ via Eq. (17).

To summarize, to probe the variation of the chemical potential $\Delta\mu(H)$ of quasiparticles in anisotropic materials in an applied magnetic field, we calculated the mixed-state magneto-thermopower $\Delta S(T,H)$ in the presence of field-modulated charge effects near T_c. Using the available magneto-TEP experimental data on textured $Bi_2Sr_2CaCu_2O_y$, the field-induced behavior of the in-plane $\Delta\mu(H)$ was obtained along with reasonable estimates for its zero-field value (Fermi energy) $\mu(0)$, interlayer coupling energy J_c, and the hole number density n_h in this material.

We thank A. Varlamov for very useful discussions on the subject. Part of this work has been financially supported by the Action de Recherche Concertées (ARC) 94-99/174. S.A.S. acknowledges the financial support from FNRS.

[1] V. I. Nizhankovskii, R. N. Sheftal', and S. G. Zybtsev, JETP Lett. **55**, 238 (1992).
[2] V. I. Nizhankovskii and S. G. Zybtsev, Phys. Rev. B **50**, 1111 (1994).
[3] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, Elsevier, Amsterdam (1988).
[4] N. E. Alekseevskii and V. I. Nizhankovskii, Zh. Éksp. Teor. Fiz. **88**, 1771 (1985) [Sov. Phys. JETP **61**, 1051 (1985)].
[5] C. M. Fowler, B. L. Freeman, W. L. Hults, J. C. King, F. M. Muller, and J. L. Smith, Phys. Rev. Lett. **68**, 534 (1992).

[6] V. Gridin, S. Sergeenkov, R. Doyle, P. de Villiers, and M. Ausloos, Phys. Rev. B **47**, 14594 (1993).
[7] S. Sergeenkov, M. Ausloos, H. Bougrine, R. Cloots, and V. Gridin, Phys. Rev. B **48**, 16680 (1993).
[8] V. Gridin, P. Pernambuco-Wise, C. G. Trendall, W. R. Datars, and J. D. Garrett, Phys. Rev. B **40**, 8814 (1989).
[9] N. V. Zavaritskii, A. V. Samoilov, and A. A. Yurgens, JETP Lett. **55**, 127 (1992).
[10] L. Reggiani, R. Vaglio, and A. A. Varlamov, Phys. Rev. B **44**, 9541 (1991).
[11] A. A. Varlamov, D. V. Livanov, and F. Federici, JETP Lett. **65**, 196 (1997).
[12] M. Houssa, H. Bougrine, S. Stassen, R. Cloots, and M. Ausloos, Phys. Rev. B **54**, R6885 (1996).
[13] M. Houssa, M. Ausloos, R. Cloots, and H. Bougrine, Phys. Rev. B **56**, 802 (1997).
[14] A. A. Varlamov and M. Ausloos, in *Fluctuation Phenomena in High Temperature Superconductors*, edited by M. Ausloos and A. A. Varlamov, vol. 32 in the NATO ASI Partnership Sub-Series, Kluwer. Dordrecht (1997), p. 3.
[15] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Clarendon Press, Oxford (1968).
[16] W. C. Lee, R. A. Klemm, and D. C. Johnston, Phys. Rev. Lett. **63**, 1012 (1989).
[17] Xin-Fen Chen, G. X. Tessema, and M. J. Skove, Phys. Rev. B **48**, 13141 (1993).

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

Collapse of resonance in quasi-one-dimensional quantum channels

C. S. Kim

Department of Physics, Chonnam National University, Kwangju 500-757, Korea

A. M. Satanin*)

Nizhniĭ Novgorod State University, 603091 Nizhniĭ Novgorod, Russia

Yong S. Joe, and R. M. Cosby

Department of Physics, Ball State University, Muncie, Indiana 47306-0505, U.S.A.

(Submitted 29 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 263–275 (July 1999)

We study the resonance structure of the conductance (transmissivity) of a quasi-one-dimensional channel that contains an attractive impurity of finite dimensions and derive an exact expression for the scattering matrix. We show that an impurity of finite dimensions may cause a set of Fano resonances to appear in the transmissivity. We also find that due to the coherent interaction the Fano resonances can collapse and discrete levels may appear in the continuum. Finally, we establish the wave function of the discrete levels and study the channel transmissivity in the critical regime. © *1999 American Institute of Physics.* [S1063-7761(99)01907-1]

1. INTRODUCTION

Lately nanotechnology methods have been used to produce quasi-one-dimensional channels, or electron waveguides.[1] A remarkable property of such channels is the quantization of conductance.[2,3] This means, for instance, that the dependence of conductivity on the electron Fermi energy acquires steps. In such channels artificial scatterers or impurities can be created, which makes it possible to control the transmissivity of the waveguide.[4] The effect of short-range impurities on the transmissivity was studied earlier and it was found that such impurities lead to quantum erosion of transmissivity. Chu and Sorbello,[5] Bagwell,[6] and Tekman and Ciraci[7] found that short-range impurities generate dips in the transmissivity, while impurities of finite dimensions in the transverse direction may lead to the appearance of asymmetric resonances.[8]

In the present paper we study the scattering of electron waves in a quasi-one-dimensional waveguide by an impurity of finite dimensions. We derive an exact expression for the scattering matrix. If the impurity is short-range in the longitudinal direction, it gives rise to an asymmetric resonance (Fano resonance) related to the existence of a virtual level in the low-lying bands continuum.[9] For the first time we study a situation where the impurity generates a large number of levels. New coherent effects may arise in the process: the interaction of levels immersed in the continuum, and collapse of resonances. The physics of Fano resonances differs substantially from that of ordinary Breit–Wigner resonances. We show that Fano resonance may disappear at certain (critical) parameters of the system. Here the collapse of resonances is accompanied by the appearance of discrete levels in the continuum, for which levels we find the wave function and show that it is normalizable.

2. MODEL OF IMPURITY AND THE SCATTERING MATRIX

We examine the scattering of electron waves in a quantum waveguide of width W aligned with the x axis. Suppose that the confinement potential acting in the transverse (lateral) direction is described by a function $V_c(y)$ and the impurity potential by a function $V(x,y)$. The electron wave function is found by solving the Schrödinger equation

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

where m is the effective electron mass.

Let us derive the exact expression for the scattering matrix of an electron in a waveguide containing an impurity of finite size. Here we use the model of a two-dimensional well, examined earlier in Refs. 10 and 11 (Fig. 1). The impurity potential can be written

$$V(x,y-Y_s)=-V_{\rm att}\theta\left(\frac{L_a}{2}-|x|\right)\theta\left(\frac{W_a}{2}-|y-Y_s|\right), \tag{2}$$

where $\theta(x)=0$ for $x<0$ and $\theta(x)=1$ for $x>1$, $X_s=0$ and Y_s are the coordinates of the center of the well, and $V_{\rm att}$ is its depth.

In calculating the scattering matrix it is convenient to use two different bases: the wave functions $\varphi_n(y)$ of a perfect waveguide ($V(x,y)=0$), which can be found by solving the equation

$$\left\{-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2}+V_c(y)\right\}\varphi_n(y)=E_n\varphi_n(y) \tag{3}$$

(E_n is the energy of traverse motion), and the wave function in the potential $V_{\rm tr}(y)=-V_{\rm att}g(y)+V_c(y)$; we denote the

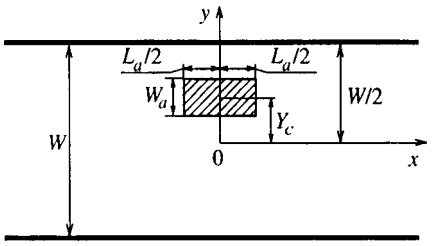

FIG. 1. Schematic of a quantum waveguide containing an attractive impurity.

wave functions and energy levels of an electron in the potential $V_{\rm tr}(y)$ by $\chi_j(y)$ and $E_j^{\rm tr}$, respectively. The base functions $\varphi_n(y)$ and $\chi_j(y)$ are related by a matrix **U** defined as follows:

$$U_{n,j}=\int \varphi_n(y)\chi_j(y)dy. \tag{4}$$

The solution of Eq. (1) can be written

$$\psi(x,y)=\sum \left[A_n \exp\left\{ik_n\left(x+\frac{L_a}{2}\right)\right\} + B_n \exp\left\{-ik_n\left(x+\frac{L_a}{2}\right)\right\}\right]\varphi_n(y) \quad \text{for } x<-L_a/2, \tag{5}$$

$$\psi(x,y)=\sum \left[a_j \exp\{iq_j x\}+b_j \exp\{-iq_j x\}\right]\chi_j(y) \quad \text{for } -L_a/2<x<L_a/2, \tag{6}$$

$$\psi(x,y)=\sum C_n \exp\{ik_n(x-L_a/2)\varphi_n(y)\} \quad \text{for } x>L_a/2, \tag{7}$$

where the quantities $k_n=\sqrt{2m(E-E_n)/\hbar^2}$ and $q_j=\sqrt{2m(E-E_j^{\rm tr})/\hbar^2}$ act as the wave vectors of the particle outside and inside the region occupied by the impurity, respectively.

Note that solutions with real k_n and q_j correspond to propagating waves, while solutions with imaginary $k_n=i|k_n|$ and $q_j=i|q_j|$ represent nonuniform waves. The boundary conditions give rise to equations for the wave amplitudes, which can be conveniently written in matrix form:

$$\begin{aligned}
&\mathbf{d}^{-1}\mathbf{a}+\mathbf{d}\mathbf{b}=\mathbf{U}(\mathbf{A}+\mathbf{B}),\\
&\mathbf{q}(\mathbf{d}^{-1}\mathbf{a}-\mathbf{d}\mathbf{b})=\mathbf{U}\mathbf{k}(\mathbf{A}-\mathbf{B}),\\
&\mathbf{d}\mathbf{a}+\mathbf{d}^{-1}\mathbf{b}=\mathbf{U}\mathbf{C},\\
&\mathbf{q}(\mathbf{d}\mathbf{a}-\mathbf{d}^{-1}\mathbf{b})=\mathbf{U}\mathbf{k}\mathbf{C},
\end{aligned} \tag{8}$$

where we used matrices with elements

$$(\mathbf{k})_{n,n'}=k_n\delta_{nn'}, \quad (\mathbf{q})_{n,n'}=q_n\delta_{nn'}, \quad (\mathbf{d})_{n,n'}=\exp\{i\theta_n\}\delta_{nn'}, \tag{9}$$

and $\theta_n=q_nL_a/2$ are the phases of the waves; the wave amplitudes **a**, **b**, **A**, **B**, and **C** are considered infinite vectors.

After excluding the intermediate amplitude **a** and **b** from (8), we can find the elements of the scattering matrix. We introduce the amplitude transmission matrix by $\mathbf{C}=\mathbf{t}\mathbf{A}$. Then Eqs. (8) yield

$$\mathbf{M}\mathbf{C}=\mathbf{A}. \tag{10}$$

The matrix **M** has the form

$$\mathbf{M}=\tfrac{1}{4}\mathbf{U}^{-1}[(1+\hat{\mathbf{k}}^{-1}\mathbf{q})\mathbf{D}^{-1}(1+\mathbf{q}^{-1}\hat{\mathbf{k}})+(1-\hat{\mathbf{k}}^{-1}\mathbf{q}) \times\mathbf{D}(1-\mathbf{q}^{-1}\hat{\mathbf{k}})]\mathbf{U}, \tag{11}$$

where $\hat{\mathbf{k}}=\mathbf{U}\mathbf{k}\mathbf{U}^{-1}$, and $\mathbf{D}=\mathbf{d}^2$. The way in which matrix **M** is written clearly shows that **M** is similar to the corresponding amplitude in the one-dimensional case.[12] It is convenient to write the transmission matrix as

$$\mathbf{t}=\mathbf{M}^{-1}=\frac{\mathbf{M}_C^T}{\det(\mathbf{M})}, \tag{12}$$

where $\mathbf{M}_C$ is the cofactor (adjoint) of M.

It is well known that the poles of the scattering matrix $t_{n,n'}(E)$, obtained as a result of analytic continuation in the energy E, determine the bound states or resonances in the system. The total conductance G of the electron waveguide is expressed in terms of the channel transmissivity T by the Büttiker–Landauer formula[13,14]

$$G=\frac{2e^2}{h}T, \quad T=\sum_{n,n'}\frac{k_n}{k_{n'}}|t_{n,n'}|^2, \tag{13}$$

where the sum is over all open channels. Equation (12) suggests that the analytic properties of the transmission amplitude as a function of energy are fully determined (in the given problem) by the structure of the matrix **M**. Here the poles of **t** can be found from

$$\det(\mathbf{M})=0, \tag{14}$$

and the zeros follow from the equations

$$[\mathbf{M}_C]_{n,m}=0, \quad n,m=1,2,\ldots, \tag{15}$$

The inversion symmetry of the adopted impurity potential makes it possible to factorize **M**:

$$\mathbf{M}=\mathbf{U}^{-1}\mathbf{M}_a\mathbf{M}_s\mathbf{U}, \tag{16}$$

where we have introduced the matrices

$$\mathbf{M}_s=\tfrac{1}{2}[-(\mathbf{d}-\mathbf{d}^{-1})+(\mathbf{d}+\mathbf{d}^{-1})\mathbf{q}^{-1}\hat{\mathbf{k}}], \tag{17}$$

$$\mathbf{M}_a=\tfrac{1}{2}[-(\mathbf{d}-\mathbf{d}^{-1})+\hat{\mathbf{k}}^{-1}\mathbf{q}(\mathbf{d}+\mathbf{d}^{-1})]. \tag{18}$$

Hence we can write Eq. (14) as two equations:

$$\det(\mathbf{M}_s)=0 \tag{19}$$

for poles in the case of symmetric states, and

$$\det(\mathbf{M}_a)=0 \tag{20}$$

for poles in the case of antisymmetric states.

3. BREIT–WIGNER AND FANO RESONANCES IN A QUANTUM CHANNEL

We begin with the limit where $W_a = W$. From (4) it follows that $\mathbf{U} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This means that in this limit the problem becomes one-dimensional, since $\mathbf{M}$ is diagonal. Equation (11) can be written

$$(\mathbf{M})_{n,n'} = \frac{1}{4}\left[\left(1+\frac{q_n}{k_n}\right)\left(1+\frac{k_n}{q_n}\right)\exp\{-2i\theta_n\} + \left(1-\frac{q_n}{k_n}\right) \times \left(1-\frac{k_n}{q_n}\right)\exp\{2i\theta_n\}\right]\delta_{nn'}, \tag{21}$$

where $\theta_n = q_n L_a/2$. This equation can easily be used to find the transmission matrix $t_{n,n'}(E) = t_n(E)\,\delta_{nn'}$. It is convenient to write the amplitude $t_n(E)$ as

$$t_n(E) = \frac{ik_n/q_n}{(\sin\theta_n + i(k_n/q_n)\cos\theta_n)(-\cos\theta_n + i(k_n/q_n)\sin\theta_n)}. \tag{22}$$

For $E > E_n$, the wave vectors k_n and q_n are real. Suppose that the well is deep: $V_{\text{att}} \gg E_n$ and $k_n/q_n \ll 1$. In this case the zeros of the denominator are approximately determined by

$$\sin\theta_{n,j} = 0, \quad \theta_{n,j} = \pi j, \quad j = 1,2,\ldots, \tag{23}$$

or by

$$\cos\theta_{n,j} = 0, \quad \theta_{n,j} = \frac{\pi}{2}(2j+1), \quad j = 0,1,\ldots. \tag{24}$$

The first equation has the solution

$$E_{n,j} = E_n - V_{\text{att}} + \left(\frac{2\pi}{L_a}\right)^2 \frac{\hbar^2 j^2}{2m}. \tag{25}$$

Near $E_{n,j}$ we expand the denominator of $t_{n,n'}$ approximately (for symmetric states), assuming that $E = E_{n,j} + \varepsilon$ in (22):

$$\sin\theta_n + i\,(k_n/q_n)\cos\theta_n \approx \theta'_{nj}(\varepsilon + i\Gamma_{nj})\cos\theta_n, \tag{26}$$

where

$$\Gamma_{nj} = \frac{k_n(E_{n,j})}{q_n(E_{n,j})\,\theta'_{nj}}, \quad \theta'_{nj} = \left.\frac{\partial\theta_n(E)}{\partial E}\right|_{E=E_{n,j}}, \quad \cos\theta_n \approx \pm 1.$$

An expansion near the second solution of Eq. (24) yields a similar expression for the denominator of the amplitude. Hence for $E > E_n$ the poles of the scattering amplitude lie in the complex plane, and near the poles the amplitude has the structure of a Breit–Wigner resonance:

$$t_{n,n}(E) = \frac{i\Gamma_{nj}}{E - E_{nj} + i\Gamma_{nj}}. \tag{27}$$

We now continue the amplitude into the complex energy plane. If the energy is in the interval $E_n - V_{\text{att}} < E < E_n$ (for the sake of definiteness we assume that $V_{\text{att}} < E_n - E_{n-1}$), $k_n = i|k_n|$ is imaginary and q_n is real. In this case the amplitude has poles on the real energy axis, and their positions are determined by (19) and (20) for symmetric and antisymmetric states, respectively:

$$\tan\theta_n = -i\frac{k_n}{q_n}, \quad \cot\theta_n = i\frac{k_n}{q_n}. \tag{28}$$

Thus, as Eq. (28) implies, we can find two types of pole. For $E < E_n$ the amplitude $t_n(E)$ has poles at real energies. When the energy is below the subband with $n \geqslant 1$, these poles land in the continuum of the lowest subbands. Below we will see that in the quasi-one-dimensional regime, when $W_a < W$ holds, the interaction between the discrete levels and the continuum is the reason for the formation for Fano resonances. At the same time, due to the interaction with the states of the continuum, the Breit–Wigner resonances shift in the complex plane.

Now we examine the situation where the size of the impurity, L_a, is smaller than the electron wavelength along the channel, $L_a \ll q_n^{-1}$. Expanding Eq. (11) in the parameter $L_a q_n^{-1}$, we easily find that

$$\mathbf{M} = (i\mathbf{k})^{-1}[i\mathbf{k} + \mathbf{v}], \tag{29}$$

where $\mathbf{v} = (L_a/2)\mathbf{U}^{-1}\mathbf{q}^2\mathbf{U}$. This case can be analyzed both by perturbation techniques and by numerical methods (the details of calculations can be found in Ref. 8). Here are the conclusions important for our analysis below. If we allow only for the diagonal elements of $\mathbf{v}$, we arrive at a set of poles that split off each subband of size quantization. Allowance for off-diagonal elements leads to the interaction of the levels that have split off the subbands with $n > 1$ with the continuum of the low-lying bands and to a shift of the levels in the complex plane. However, in contrast to the one-dimensional case, the scattering amplitude has zeros (in addition to poles), and in the weak coupling regime a zero and a pole lie close to each other in the complex plane. In particular, for $E_1 < E < E_2$, near a zero and pole the amplitude can be represented as

$$t_{11}(E) \sim \frac{E - E_0}{E - E_R + i\Gamma}, \tag{30}$$

where E_0, E_R, and Γ are the parameters of a Fano resonance.[9] According to Nöckel and Stone,[15] for a system with inversion symmetry, the transmissivity $T_{11} = |t_{11}|^2$ can be written

$$T_{11}(E) = \frac{1}{1+q^2}\frac{(\varepsilon+q)^2}{\varepsilon^2+1}, \tag{31}$$

where $\varepsilon = (E - E_R)/\Gamma$, and $q = (E_R - E_0)/\Gamma$. Depending on the parameters of the system, the dimensionless asymmetry parameter q, which is the ratio of the distance $|E_0 - E_R|$ between zero and pole to the resonance width Γ can be much larger or much smaller than unity. In the limit $q \ll 1$ formula (31) yields a dip, while in the opposite case $q \gg 1$ it yields a peak in the transmissivity, with the transmissivity probability at the peak being exactly unity.

Thus, when the electron is scattered by a short-range attractive impurity, the scattering amplitude near virtual levels has the structure of a Fano resonance.

4. COLLAPSE OF FANO RESONANCES

It is well known that the interaction of discrete levels with the continuum in multichannel atomic systems gives rise to a universal resonance curve for the lifetime of quasibound states and excitations.[9] In Ref. 8 we showed that this happens in the case of the transmission amplitude in a quantum channel with a short-range impurity. An impurity of finite dimensions generates a large number of levels in the continuum. One should expect nontrivial interactions between levels and nontrivial behavior of asymmetric resonances. In atomic systems the interaction between levels results in an overlap of resonances. Earlier Mies[16] discussed such effects by using a phenomenological approach. Our goal is to study the effect of the interaction of resonances on the tunneling of electrons though multilevel configurations. We limit ourselves to the energy interval $E_1<E<E_2$, since earlier we found[8] that it is in this energy interval that total reflection of electrons is possible when the particle energy coincides with the energy of the zero of a Fano resonance.

Qualitative reasoning suggests that a deep impurity can yield dramatically new effects. Clearly, a two-dimensional well generates a set of levels in the interval $E_1<E<E_2$, and in the case of a well of finite depth these levels lie in the interval $(E_2-V_{\rm att},E_2)$. If we ignore the interaction between the electronic states belonging to different channels, Eq. (28) implies that the levels corresponding to symmetric and antisymmetric states are determined, respectively, by the equations

$$\tan\theta_n=\frac{|k_n|}{q_n},\quad \cot\theta_n=-\frac{|k_n|}{q_n}. \tag{32}$$

Since in this case we have $E_2^{\rm tr}<E_2$, the levels lie in the interval $(E_2^{\rm tr},E_2)$ and their interaction with the continuum gives rise to a set of Fano resonances. Obviously, these resonances can interact with each other. Such interaction is also present for Breit–Wigner resonances, with the resonances repelling each other. What makes the case of Fano resonances so different is that the zeros, while moving in the complex plane, may collide with poles. Below we will see that this leads to two interesting consequences. First, the resonances narrow, which means that the regime $q\ll 1$ in (31) may be replaced by $q\gg 1$. This is accompanied by the appearance of a peak in transmissivity, with the transmissivity probability at the peak being exactly unity. Second, there are critical values of the parameters of the system at which the resonances disappear entirely, accompanied by the appearance of discrete levels in the continuum.

To illustrate the predicted effect, we analyze the simplest situation where the well size in the transverse direction, W_a, is much smaller than L_a:

$$V_{\rm tr}(y)=-\frac{\hbar^2 u}{m}\,\delta(y-Y_s)+V_c(y), \tag{33}$$

with $u=mV_{\rm att}W_a/\hbar^2$. For such a model the energy of transverse motion can be found by solving the equation

$$\sin\kappa W=2\,\frac{u}{\kappa}\sin\kappa Y_s\sin\kappa(W-Y_s), \tag{34}$$

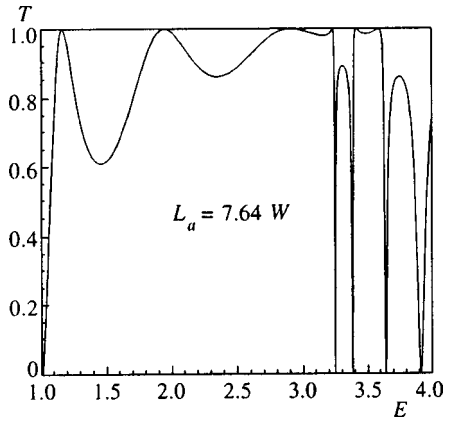

FIG. 2. Transmissivity of a channel with a finite impurity (the impurity parameters are $V_{\rm att}=6.37E_1$ and $W_a=0.1W$). The Fano resonances are clearly visible for $3.19E_1<E<4.0E_1$ and the Breit–Wigner resonances, for $1.0E_1<E<3.19E_1$.

where $\kappa=\sqrt{2mE/\hbar^2}$. From (34) it follows that when the coupling parameter u is zero, we have the solution $E_j^{\rm tr}=E_j$, with $E_j=\pi^2\hbar^2/(2mW^2)$, but when u is finite, we have a set of levels with energies $E_j^{\rm tr}<E_j$. Below we list the results of numerical calculations for which the the impurity parameters were taken to be $W_a=0.1$ W and $V_{\rm att}=6.37E_1$. For the sake of orientation we give only the first three levels of transverse motion determined by the numerical solution of Eq. (34): $E_1^{\rm tr}=-0.60E_1$, $E_2^{\rm tr}=3.19E_1$, and $E_3^{\rm tr}=8.99E_1$. Hence, we should expect Fano resonances in the energy interval $3.19E_1<E<4.0E_1$ and Breit–Wigner resonances, in the interval $1.0E_1<E<3.19E_1$. Figure 2 depicts the transmissivity of a channel with an attractive impurity. The Fano-resonance structure in the transmissivity is clearly visible in the interval $E_2^{\rm tr}<E<E_2$, while the Breit–Wigner resonances are clearly visible in the interval $E_1<E<E_2^{\rm tr}$.

Now we study the transmissivity analytically, and for $E_2^{\rm tr}<E<E_2$ we keep to resonantly interacting channels. In the region $|x|<L_a/2$ we leave two propagating modes with wave vectors $q_1=\sqrt{2m(E-E_1^{\rm tr})/\hbar^2}$ and $q_2=\sqrt{2m(E-E_2^{\rm tr})/\hbar^2}$. In the region $|x|>L_a/2$ (outside the impurity) we leave a propagating solution with $k_1=\sqrt{2m(E-E_1)/\hbar^2}$ and a nonuniform solution with $k_2=i|k_2|$, $|k_2|=\sqrt{2m(E_2-E)/\hbar^2}$.

An approximate expression for the scattering matrix can be obtained from (12). The transmission amplitude t_{11} between the open channels can be written

$$t_{11}=\frac{[\mathbf{M}_C]_{11}}{\det(\mathbf{M}_s)\det(\mathbf{M}_a)}. \tag{35}$$

The poles of symmetric states can be found from

$$\det(\mathbf{M}_s)$$
$$=i\left(-i\sin\theta_1+\frac{k_1}{q_1}\cos\theta_1\right)\left(-\sin\theta_2+\frac{|k_2|}{q_2}\cos\theta_2\right)$$
$$\times U_{11}U_{22}-i\left(-i\sin\theta_2+\frac{k_1}{q_2}\cos\theta_2\right)$$

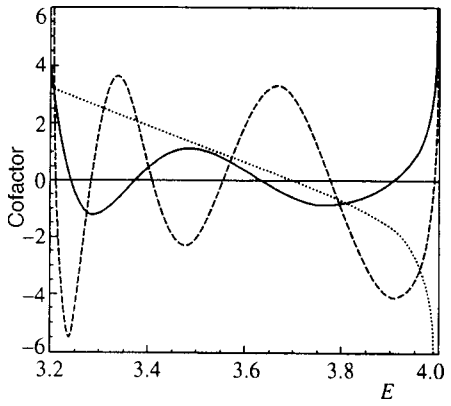

FIG. 3. Energy dependence of the cofactor for the characteristic impurity sizes: $L=1.27W$ (dotted curve), $L=7.64W$ (solid curve), and $L=11.46W$ (dashed curve). The zeros of the cofactor determine the zeros of the transmission amplitude. The other parameters of the impurity are the same as in Fig. 2.

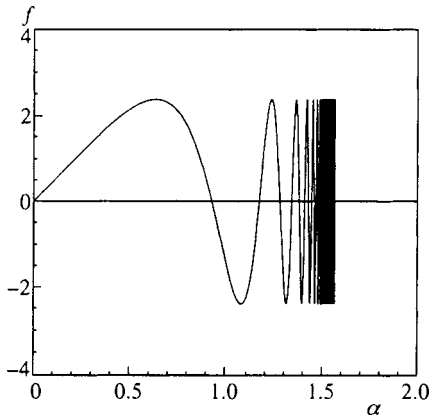

FIG. 4. The characteristic function. The critical parameters are determined by the zeros of the characteristic function.

$$\times\left(-\sin\theta_1+\frac{|k_2|}{q_1}\cos\theta_1\right)U_{12}^2=0, \tag{36}$$

and those of antisymmetric states, from

$$\begin{aligned}\det(\mathbf{M}_a)&=i\left(-i\sin\theta_1+\frac{q_1}{k_1}\cos\theta_1\right)\left(\sin\theta_2+\frac{q_2}{|k_2|}\cos\theta_2\right)\\&\times(U^{-1})_{11}(U^{-1})_{22}-i\left(-i\sin\theta_2+\frac{q_2}{k_1}\cos\theta_2\right)\\&\times\left(\sin\theta_1+\frac{q_1}{|k_2|}\cos\theta_1\right)(U^{-1})_{12}(U^{-1})_{21}=0.\end{aligned} \tag{37}$$

An analysis of the equations for the poles shows that generally these equations have two sets of complex-valued solutions, which is similar to the one-dimensional case. However, in the quasi-one-dimensional case the amplitudes have zeros, which can be found by solving the equation $[\mathbf{M}_C]_{11}=0$ or, in greater detail,

$$\begin{aligned}&q_2\sin 2\theta_2\left(-\tan\theta_2+\frac{|k_2|}{q_2}\right)\left(\cot\theta_2+\frac{|k_2|}{q_2}\right)U_{22}(U^{-1})_{22}\\&-q_1\sin 2\theta_1\left(-\tan\theta_1+\frac{|k_2|}{q_1}\right)\left(\cot\theta_1+\frac{|k_2|}{q_1}\right)\\&\times U_{12}(U^{-1})_{21}=0.\end{aligned} \tag{38}$$

Since the coefficients of these equations are real for $E_2^{\rm tr}<E<E_2$, the equations have real-valued solutions. Here the transmission amplitude vanishes. Figure 3 depicts the energy dependence of the cofactor for three characteristic impurity sizes: $L=1.27W$, $7.64W$, and $11.46W$. Hence Eqs. (36) and (37) determine the poles and the solutions of Eq. (38) determine the zeros (or dips) in the transmission. For the energy interval $E_1<E<E_2^{\rm tr}$ we must rewrite (38), assuming that $q_2=i|q_2|$. Clearly, in this case the zeros are shifted into the complex plane.

An important observation is that Eq. (36) [or (37)] can have a solution for a real energy simultaneously with (38). In the symmetric case this occurs when

$$\tan\theta_2=\frac{|k_2|}{q_2}, \tag{39}$$

$$\tan\theta_1=\frac{|k_2|}{q_1}. \tag{40}$$

Thus, Eqs. (39) and (40) should have a simultaneous solution in the energy interval $E_2^{\rm tr}<E<E_2$. Here we have a two-parameter spectral problem. For the spectral parameters it is convenient to take the energy E and the impurity size L_a, or the spectral pair (E,L_a). Note that for an arbitrary value of the parameter L_a, Eq. (39) determines the levels in a one-dimensional well,[13] but these levels lie below E_2. Hence the levels acquire a finite width, since they interact with the continuous spectrum of the states of the subband $n=1$. The level width may vanish when Eqs. (39) and (40) have a common solution. Thus, when the system parameters coincide with the values of the spectral parameter $[E(j),L_a(j)]$, discrete levels appear in the continuum of states of the quantum channel.

Using Eqs. (39) and (40), we can easily show that the spectral parameters are given by the expressions

$$E(j)=E_2^{\rm tr}+(E_2-E_2^{\rm tr})\cos^2\alpha(j), \tag{41}$$

$$L_a(j)=\frac{2}{\pi}W\frac{\alpha(j)}{\cos\alpha(j)}\left[\frac{E_1}{E_2-E_2^{\rm tr}}\right]^{1/2},\quad j=1,2,\ldots, \tag{42}$$

where the parameters $\alpha(j)$ are the solutions of

$$f(\alpha)=\sqrt{\delta+\cos^2\alpha}\,\tan\left[\frac{\alpha}{\cos\alpha}\sqrt{\delta+\cos^2\alpha}\right]-\sin\alpha=0, \tag{43}$$

with $\delta=(E_2-E_1)/(E_2-E_2^{\rm tr})$. Figure 4 depicts the characteristic function $f(\alpha)$. We see that Eqs. (39) and (40) have (simultaneously) an infinite set of solutions. In the same way, antisymmetric solutions are possible if

$$\cot\theta_2=-\frac{|k_2|}{q_2}, \tag{44}$$

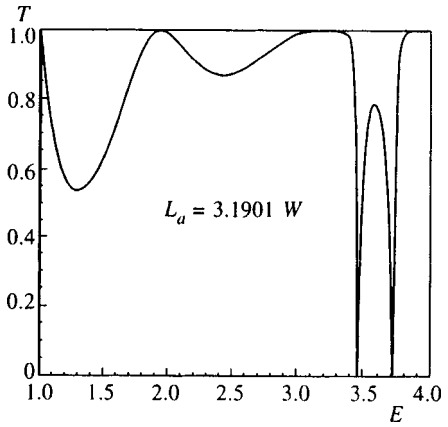

FIG. 5. Energy dependence of the transmissivity for the critical value $L_c \equiv L_a(3) = 3.1901W$ of the impurity size (the other parameters of the impurity are the same as in Fig. 2). The resonance disappears at $E = E_c(3) = 3.255E_1$, but the transmissivity remains finite.

$$\cot\theta_1 = -\frac{|k_2|}{q_1}. \tag{45}$$

From unitarity it follows that the zeros $E_2^{\rm tr} < E < E_2$ of the cofactor must coincide with the poles. Equation (38) shows that it is valid when Eqs. (39) and (40) [or Eqs. (44) and (45)] are valid. This implies that the Fano resonance disappear at certain (critical) values of the parameters $L_a(j)$, values at which $\Gamma = 0$ and $E_0 = E_p$ hold for the parameters of a resonance in Eq. (30).

To illustrate the effect of disappearance of resonances, or collapse of resonances, we turn to Fig. 5, which depicts the energy dependence of the transmissivity for one of the critical impurity sizes ($L_c \equiv L_a(3) = 3.1901W$). The critical value L_c was calculated by Eqs. (39) and (40) with the depth of the well fixed, $V_{\rm att} = 6.37E_1$. Figure 6 depicts the energy dependence of the transmissivity near the critical energy for several values of the impurity size: $L_a = 3.06W$, $3.12W$, $3.1901W$, and $3.25W$. The results show

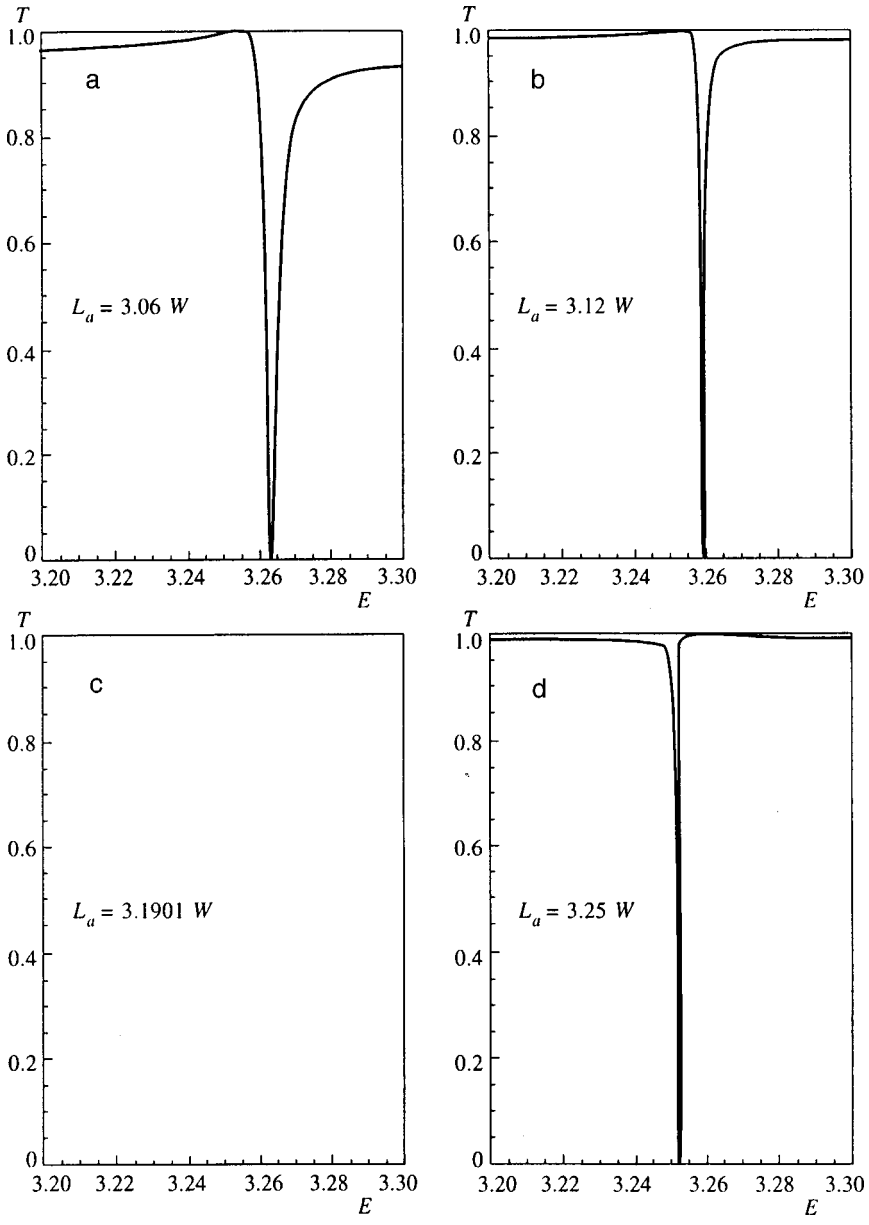

FIG. 6. Energy dependence of the transmissivity for several values of the impurity size ($L_a = 3.06W, 3.12W, 3.1901W$, and $3.25W$) near the critical size $L_c = 3.1901W$ (the other parameters of the impurity are the same as in Fig. 2). At the critical length the resonance disappears (in Fig. 6c the transmissivity T is unity for the selected energy interval). The resonance reappears when the impurity size passes the critical value L_c (Fig. 6d).

that when L_a reaches the critical value L_c, the Fano resonance disappears but the transmissivity remains finite at the critical energy value $E_c(3)=3.255E_1$. Figure 6d shows that resonances appear again when the impurity size passes the critical value. Clearly, the resonance collapses because of the coherent interaction of counterpropagating waves inside the region occupied by the impurity.

For orientation we point out that if a two-dimensional channel of width W contains an impurity, for the impurity parameters used above the critical size is $L_a(0)=1.10W$, and the minimum critical energy is $E_c=E(0)=3.48E_1$, where $E_1=\pi^2\hbar^2/2mW^2$.

To make the picture clearer we write the wave function of the discrete levels explicitly and show that the function can be normalized. The symmetric solution of Eq. (1) can be written as

$$\psi_c(x,y)=a_1\cos(q_1x)\chi_1(y)+a_2\cos(q_2x)\chi_2(y) \quad \text{for } |x|<\frac{L_a}{2}, \tag{46}$$

$$\psi_c(x,y)=C_1\exp\left\{|k_2|\left(x-\frac{L_a}{2}\right)\right\}\varphi_2(y) \quad \text{for } x>\frac{L_a}{2}. \tag{47}$$

The boundary conditions imply that the solution determined by (46) and (47) is valid if Eqs. (39) and (40) are valid.

Similarly, for the antisymmetric solutions we have

$$\psi_c(x,y)=b_1\sin(q_1x)\chi_1(y)+b_2\sin(q_2x)\chi_2(y) \quad \text{for } |x|<\frac{L_a}{2}, \tag{48}$$

$$\psi_c(x,y)=D_1\exp\left\{|k_2|\left(x-\frac{L_a}{2}\right)\right\}\varphi_2(y) \quad \text{for } x>\frac{L_a}{2}, \tag{49}$$

which are valid if (44) and (45) are valid. There exists a simple physical interpretation of the localized solution—it is a standing wave with an amplitude a_1 between two Fano ''mirrors'' and a nonuniform mode in channel $n=2$. Since outside the impurity the wave function decreases exponentially, it can be normalized. On the other hand, for the same critical parameter the channel may contain a propagating solution of the form (5)–(7). Thus, we have found that two types of state may coexist at the same energy: one localized state and one propagating state.

5. CONCLUSION

We have analyzed the passage of a particle through a quantum waveguide containing an attractive impurity. By finding the exact solution for the scattering matrix we have shown that the impurity generates a set of quasibound states, which manifest themselves as resonance–antiresonance pairs in the transmissivity. When the impurity is short-range, a single Fano resonance can be observed in the transmissivity. For an impurity of finite dimensions a new coherent effect occurs, where at certain (critical) values of the parameters of the system the resonances collapse. As a result, for the critical parameters of the system the continuum acquires discrete levels. The transmissivity changes dramatically in the process, which can be verified in experiments involving low-dimensional channels with impurities. By way of an example, we point to an interesting paper by Yamada and Yamamoto,[4] who proposed a method for producing artificial impurities in a quantum channel.

The present work was supported by the Ministry of Education of Korea (Grants Nos. BSRI-96-2431 and BSRI-97-2431) and the International Center for Advanced Studies in Nizhniĭ Novgorod, Russia. One of the authors (A.M.S.) is grateful to KOSEF and the Russian Fund for Fundamental Research (Grant No. 97-02-16923a) for support.

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

[1] *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk (Eds.), Academic Press, Boston (1989).
[2] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).
[3] D. A. Wharam, T. J. Thorton, R. Newbury *et al.*, J. Phys. **21**, L209 (1988).
[4] Syoji Yamada and Masafumi Yamamoto, Appl. Phys. Lett. **79**, 8391 (1996).
[5] C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
[6] P. F. Bagwell, Phys. Rev. B **41**, 10354 (1990).
[7] E. Tekman and S. Ciraci, Phys. Rev. B **42**, 9098 (1990).
[8] C. S. Kim and A. M. Satanin, Zh. Éksp. Teor. Fiz. **115**, 211 (1999) [JETP **88**, 118 (1999)].
[9] U. Fano, Phys. Rev. **124**, 1866 (1961).
[10] Yong S. Joe and R. M. Cosby, Appl. Phys. Lett. **81**, 6217 (1997).
[11] Yong S. Joe and R. M. Cosby, Solid State Commun. **101**, 731 (1997).
[12] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[13] R. Landauer, Philos. Mag. **21**, 863 (1970).
[14] M. Büttiker, Phys. Rev. B **35**, 4123 (1987).
[15] J. U. Nöckel and A. D. Stone, Phys. Rev. B **50**, 17415 (1994).
[16] F. H. Mies, Phys. Rev. **175**, 164 (1968).

Translated by Eugene Yankovsky

Nonlinear percolation conductivity and negative differential resistivity in microcrystalline PbTe layers with an adjustable potential well

V. D. Okunev*) and N. N. Pafomov

Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, 340114 Donetsk, Ukraine
(Submitted 20 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 276–298 (July 1999)

The effect of electric fields on the electrical conductivity of PbTe films with block sizes smaller than the Debye screening length is studied. As the temperature is varied, a readjustment of the potential well is observed due to thermal spread of barriers with height $\varphi < kT$ and the expansion of higher barriers. Spatial ensembles, which consist of several blocks that increase rapidly with temperature, are established for each T. This process leads to an increase in the height of the potential barriers as the linear size of these ensembles increases. This determines the potential well in these films and their nonlinear properties, which originate in the nonlinear percolation conductivity of a microscopic crystalline system with intergranular barriers. A comparison with the experimental data of Shklovskiĭ shows that the scale length of the spatial inhomogeneity $a = 3.7 \times 10^{-6}$ cm at $T = 4.2$ K corresponds to the average block size. The value of a increases with temperature, reaching 5×10^{-4} cm at $T = 240$ K. This mechanism for electrical conductivity is compared with the hopping conductivity with a variable hopping length. The negative differential resistance in the structures examined here is found to be electrothermal in nature. © *1999 American Institute of Physics.* [S1063-7761(99)02007-7]

1. INTRODUCTION

The basic reasons for the essentially continuous rise in the interest of researchers in systems with disordered structures in recent decades are well known. On one hand, research on microcrystalline, nanocrystalline, granulated, amorphous, glassy, and other similar materials is stimulated by the possibilities which been discovered in the course of studying the effect of inhomogeneities and, as a rule, of the localized states associated with them, on the electrical and optical properties of these systems.[1–5] Studies of different aspects of the influence of the structural features of disordered systems on the localization of electronic states play an important role in the development of the physics of the condensed state. On the other hand, interest in disordered systems is stimulated by advances in practical application of related materials. Experience shows that extending the spectrum of the structural states of matter can yield qualitatively new results, even when working with traditional materials. Finally, this area has recently been rather strongly, although indirectly, influenced by research on quantum-well effects in media with correlated siting of clusters with ordered structure.[6–9]

Materials with a microcrystalline structure are of independent scientific interest with stable domains of application, and they have been well studied. Since the theoretical models for their behavior are regarded as well developed, these objects are often used as models for studying various phenomena in other types of disordered systems.[3,10–19] This may apply to models of structure and to mechanisms for electrical conductivity or optical absorption.

Many of the standard models are obtained by comparing the objects under study with microcrystalline samples which have rather large grains whose size greatly exceeds the Debye screening length. At the same time, the most important generalizations often have to be carried out for much smaller grain sizes and, therefore, much smaller inhomogeneity scale lengths, where such comparisons are not adequately justified. This refers directly, for example, to amorphous and glassy materials, whose structure and properties have long been modelled using microcrystalline systems. The popularity of the various microcrystalline or cluster models changes with time, but the need for them has remained fairly stable.

The experience gained by studying the effects observed in microcrystalline films and in bulk samples is useful, both for creating new models, and for choosing analogs for other forms of materials with disordered structures. It can also be of interest for studying structures with extremely small sized active elements and a high packing density in crystals with perfect long-range order.

Of the published data on the structure and properties of semiconductors with microcrystalline structure, the most complete information is on polycrystalline silicon. Progress in research on other materials, including narrow-gap materials, has been more modest. Published reports of research on polycrystalline semiconductors of the IV–VI group mostly concern special cases[20–25] and usually do not deal with the general problems typical of microcrystalline systems. As a rule, they contain little data demonstrating a relationship between the electrical parameters of films and their structure, and do not take full account of the influence of the potential well peculiar to microcrystalline systems on charge carrier transport. The reasons for this are quite understandable. The

IV–VI semiconductors are a very interesting group of materials[26–28] and have been a source of enthusiasm for several generations of researchers. At the same time, because of the small band gap of the major representatives of these materials and their high concentration of intrinsic charge carriers,[26,27] they are considered to be of little interest in studies of the phenomena observed in microcrystalline systems. When films of the narrow-band IV–VI semiconductors are mentioned, their advantages are customarily taken to include the closeness of the actual parameters to those of the corresponding single crystals, while the specific problems intrinsic to microcrystalline systems are usually ignored. In addition, even for films with grain sizes $D \sim 10^{-6}$ cm, the small barrier heights in PbTe and other materials make it possible to ignore the features of microcrystalline systems resulting from the presence of regions of surface space charge near the grain boundaries. In particular, this applies to determining the main parameters of films using Hall coefficient measurements, at and below room temperature. For example, in the literature there are reports that the mobility of charge carriers in microcrystalline PbTe films with grain sizes $D \simeq 250$ Å at $T = 77$ K is close to the carrier mobility in the best single crystal specimens,[29] although it is clearly known that charge carrier mean free path in them greatly exceeds the average grain size. When there is an explicit dependence of the electronic properties of a film on the grain size,[30,31] data of this kind should be interpreted more carefully.

As for their nonlinear conductivity, despite the possible observation of negative differential resistivity, this group of materials has been of almost no interest to researchers and a very limited number of papers[32–35] are devoted to this topic. At the same time, it is evident that narrow-band semiconductors may turn out to be useful in measurements at low temperatures, where it is difficult to work with wide-gap materials and there are strict limits on the parameters and methods that can be used for these materials.

As a rule, the theoretical models apply to microcrystalline samples whose grain size substantially exceeds the Debye screening length.[36] Within the volume of these grains, the internal electric field strength is zero, and the free carrier concentration is the same as the carrier concentration in a single crystal with a similar impurity concentration. At the same time, as noted above, it is often necessary to deal with polycrystalline samples with much smaller grain sizes, where the results obtained for large grains may not seem adequately justified and, in principle, incorrect.

In this paper we present some data on the low-temperature nonlinear conductivity of microcrystalline PbTe films on glass substrates with inhomogeneities smaller than the Debye screening length. Possible mechanisms for the nonlinear properties and negative differential resistivity of these samples are discussed.

2. SUMMARY OF THE CHARACTERISTICS OF THE SAMPLES AND EXPERIMENTAL PROCEDURE

Polycrystalline PbTe layers on glass substrates are distinguished by the large variety of their properties and observed effects, which are caused primarily by the possible variation in the average grain size over 4-5 orders of magnitude.[30,31,37] The electronic properties of the samples are then related directly to the structure of the films. The classical and quantum-well effects observed during studies of the electrical conductivity of these structures are of interest for the physics of microcrystalline systems, as well as for specific practical developments. As for the technology, it is not difficult to reproduce these results: it is sufficient to ensure the specified temperature and growth rate. Recall, also, the importance of the film thickness, since the structure of microcrystalline PbTe films and, therefore, the sample properties are determined, to a substantial extent, by the elastic stresses at the interface boundary, which, in turn, are related to the degree of mismatch between the lattice parameters of materials which come into contact and by the thickness of the films.

The polycrystalline PbTe films studied here were grown by dc cathode sputtering of a stoichiometric target in an argon atmosphere with deposition of the sputtered material onto substrates of cover glass with a thickness of approximately 100 μm. The rate of film growth was held constant at about 4 Å/s. The target material was a cylinder of height 0.2 cm and diameter 2 cm, cut from an ingot of single crystal *p*-type PbTe with a resistivity of about 0.1 $\Omega \cdot$cm at room temperature. The current density at the target was 5×10^{-4} A/cm^2 and the power dissipated by the target was 6 W. The target was cooled by liquid nitrogen; the growth (substrate) temperature was held near room temperature, at 293 ± 2 K. Sputtered film electrodes of $Pt_{0.8}Mo_{0.2}$ alloy, obtained by sputtering of platinum and molybdenum together, were used as ohmic contacts and ensured linear current-voltage characteristics of films with small grains ($D < 10^{-4}$ cm). All the electrical measurements were made for direct current; the conductivity type of the films was determined by the thermal emf method. The structural state of the samples was monitored by x-ray techniques: the diffraction pattern obtained in a Debye chamber was analyzed.

Polycrystalline PbTe films grown on glass substrates turn out to have a structural disequilibrium,[30,31] and their parameters can vary substantially during storage under ordinary conditions (annealing at room temperature). The structural transformations observed during aging are caused by two main factors: the tendency of a polycrystalline system to change its surface energy (this can be achieved by an increase in the grain size) and the competing process associated with the existence of elastic stresses owing to the presence of the film-substrate interface. The direction of the structural changes depends on the average grain size D in the original films, which, in turn, is determined by the growth temperature and thickness of the grown layers.[31] If $D < 10^{-6}$ cm, then the grains become larger. When $D > 10^{-6}$ cm, fragmentation involving the breakup of large crystallites into blocks is observed. As a result, regardless of the film thickness and the average initial value of D, the grain size approaches $D^* \simeq 10^{-6}$ cm in the course of aging. It turns out that the unit cell parameter in layers with $D \simeq 10^{-6}$ cm corresponds to the lattice parameter of single crystal PbTe.[37]

In films with grain sizes $D > 10^{-6}$ cm, the electrical

properties of the samples are caused mainly by scattering of charge carriers at the grain boundaries. Some specific mechanisms for scattering at the surface of a crystal have been examined elsewhere.[38] For $D<10^{-6}$ cm, the total surface area of the grains contained in unit volume is $S\propto 1/D$; therefore, the density of localized states at their boundaries, $N_s=\gamma/D$, is fairly high ($N_s\simeq 8\times 10^{20}$ for $D=1\times 10^{-6}$ and $N_s\simeq 4\times 10^{21}$ cm$^{-3}\cdot$ eV^{-1} for $D=2\times 10^{-7}$ cm),[30] and jump conductivity is observed in samples of this type at low temperatures with the participation of these states, which can be interpreted as conduction along the grain boundaries. Within the interval 4.2 K$<T<$50 K, hopping conductivity with a variable hopping length is observed and in the range 50 K$<T<$200 K, hopping conductivity between nearest neighbors with a constant activation energy is observed.[30] For $D\sim 10^{-7}$ cm, the lattice parameter and band gap are found to depend on the average grain size.[30,37]

During the electrical conductivity measurements, most of the films of micron and submicron thickness ($d<5$ μm) obeyed Ohm's law or the deviations from linearity in the current-voltage characteristics were small. The current-voltage characteristics of the finely crystalline films were linear over wide ranges of variation in the temperature and electric field strengths. A significant nonlinearity in the conductivity was observed only in large crystalline samples with grain sizes $D>10^{-4}$ cm and at temperatures below 300 K. A similar structure and the corresponding properties could be reproduced only in rather thick ($d=7-10$ μm) films. Because of the significant anisotropy in their growth, the average grain size for these samples was always at least a few times greater than the thickness of layers with a polycrystalline structure.[31]

In accordance with an analysis of the first data on the nonlinear conductivity of microcrystalline PbTe layers, it was already clear, in general outline, that the nonlinear properties of coarse crystalline films are related to the existence of potential barriers at the grain boundaries, but the specific mechanism for this phenomenon was not understood for a long time. This situation was cleared up only after detailed studies of the electrical properties of samples at low temperatures, which involved different structural states of microcrystalline PbTe films, were carried out regularly over a period of five years.

These measurements always indicated that, as the temperature was lowered, the current-voltage characteristics of samples with $D=(2-5)\times 10^{-3}$ cm became nonlinear. The critical electric field, at which a nonlinear dependence of the current on the voltage began to be observed, fell as the temperature was reduced and turned out to be very low. For example, in freshly prepared films at $T\simeq 80$ K, a nonlinear conductivity was observed in fields above 50 V/cm. Further increases in the electric field led to the appearance of segments with a stable S-type negative differential resistivity in the current-voltage characteristic, but this, unfortunately, was accompanied by substantial, irreversible hysteresis in the current as a function of voltage. X-ray structural analysis showed that the hysteresis in the current-voltage characteristics is caused by irreversible changes in the structure of the films associated with their fragmentation,[31] i.e., the breakup of large grains into smaller ones. This fragmentation, which is mostly stimulated by stresses at the film-substrate interface, ultimately leads to a division of each of the original grains into 10^6-10^7 blocks with sizes on the order of 10^{-6} cm and fairly perfect interface boundaries. Because of the change in the film properties during aging and the impossibility of making a correct measurement of the characteristics of freshly prepared samples, we were forced to delay a study of their behavior in strong electric fields until no further irreversible changes in their parameters could be observed for measurements at moderate currents.

The measurements were made on polycrystalline films with *p*-type conductivity (the equilibrium hole concentration in the bulk of the large grains was $p_0=5\times 10^{16}$ cm^{-3}) (Ref. 30) and a thickness of 7–10 μm, grown on glass substrates at room temperature. The average grain size in as-grown samples was $(2-5)\times 10^{-3}$ cm (Ref 31). The deviations of the current-voltage characteristics of diodes fabricated from these films from Ohm's law showed up most distinctly. Samples with the following geometry were used: their length was 0.07 cm, their cross-sectional area was $(4.2-5.5)\times 10^{-4}$ cm^2 with contacts, as in Refs. 30 and 31, in the form of film electrodes made by sputtering a $Pt_{0.8}Mo_{0.2}$ alloy.

The figures shown here apply to samples obtained from a 9-μm-thick film. The average grain size in a freshly prepared film was $D=5\times 10^{-3}$ cm, which is 5.5 times the thickness of the polycrystalline layer in Ref. 31. X-ray structural analysis showed that the average block size D^* in the samples was about 2×10^{-6} cm after aging for three years (annealing at room temperature under natural conditions, which causes the reduction in grain size). The sample geometry and measurement technique are illustrated in Fig. 1.

We have shown previously that the electrical properties of polycrystalline PbTe layers in weak electric fields are closely related to their potential well and that the latter is, in turn, related to the structure, primarily, the average grain size.[30] Perhaps, this applies to an even greater degree to the nonlinear conductivity. It would be impossible to overlook that the current-voltage characteristics of diodes fabricated from coarse crystalline films for studying their behavior in strong electric fields are related in a most direct fashion to changes in the film structures as they age. Fragmentation (note again that here we mean the breakup of large crystallites into fine blocks), accompanied by the appearance of new boundaries and an increase in the number of potential barriers in the way of the charge carriers, causes a substantial rise in the threshold voltage U_t (U_t is defined at the cutoff point, where the differential resistance $dU/dI=0$.) and a drop in the threshold current I_t. The threshold power $W_t=U_tI_t$ decreases. For some of the samples, W_t was found to decrease by several factors. This is quite clear from the data shown in Fig. 2.

In principle, control measurements taken during the aging of the films showed that the relationship between the current-voltage characteristics of the samples and the potential well, which is determined by the film structure and changes as they age, is rather obvious. Within 6–8 months after fabrication of the films, it was possible to obtain current-voltage characteristics with segments of stable nega-

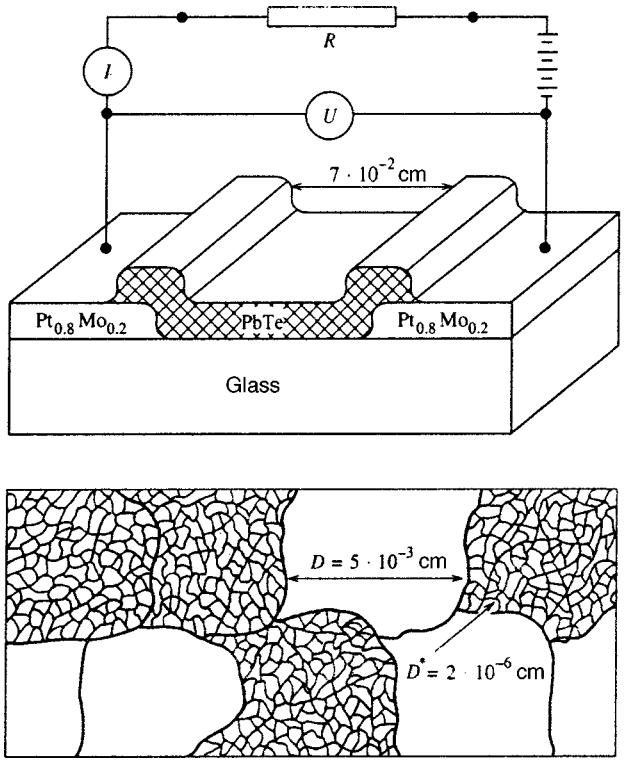

FIG. 1. Geometry of the microcrystalline PbTe samples with an *S*-shaped current-voltage characteristic and the measurement setup.

tive resistance without any special problems. With time, as the average block size decreased and the number of blocks increased, a continuous rise in the resistance of the samples, which is accompanied by a drop in the threshold current and a rise in the threshold voltage, is observed. The threshold power (for fixed temperature) decreased. Two such states are shown in Fig. 2. If the current-voltage characteristics of the samples were taken with sufficiently high load resistances,

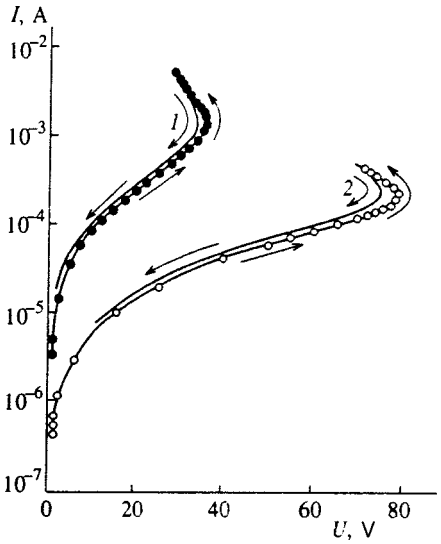

FIG. 2. Current-voltage characteristics ($T=77$ K) of a PbTe film sample 11 months (curves *1*) and 33 months (curves *2*) after fabrication. The film thickness is 9×10^{-4} cm, the distance between electrodes is 7×10^{-2} cm, and the cross-sectional area is 5×10^{-4} cm^2.

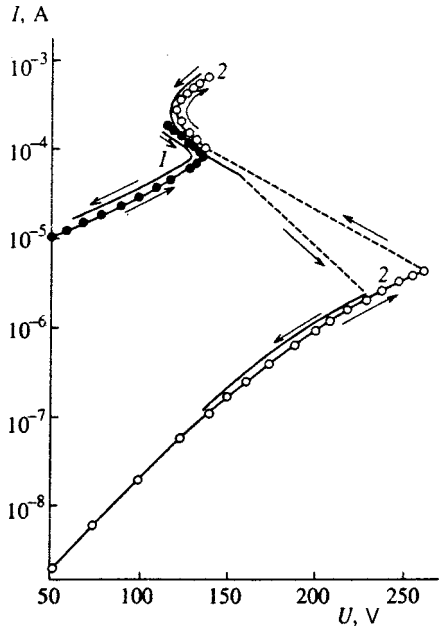

FIG. 3. Current-voltage characteristic ($T=25$ K) before (curves *1*) and after (curves *2*) several measurement cycles at high currents.

i.e., close to a current source regime, then the hysteresis phenomena observed between forward and reverse currents are small (curves *1* and *2* in Fig. 2) or are essentially not observed at all. There were also no signs of the formation of current filaments.

At high currents, a segment with a negative differential resistance is followed by another segment with a positive differential resistance. The appearance of this segment should not be regarded as something unusual, since its presence is even predicted by a purely thermal breakdown theory.[39] However, high currents, which facilitate further fragmentation, produced irreversible changes in the current-voltage characteristic, and led to a rise in U_t and a drop in I_t. Figure 3 (curve *2*) shows a current-voltage characteristic ($T=25$ K) for the same sample as in Fig. 2, but after several high-current measurement cycles. The initial state before the high-current measurements, which corresponds to the current-voltage characteristic of curve *2* in Fig. 2, was obtained at $T=77$ K and is shown as curve *1*. It can be seen that the changes in the parameters of the current-voltage characteristic are substantial. Thus, during measurements at currents beyond the threshold, we usually restricted the analysis to relatively low currents through the sample, without reaching the segment with a positive differential resistance.

With decreasing resistance of the load in series with the power supply and sample, a switching effect is observed (from a high resistance to a low resistance state and back again when the current through the sample is being lowered) along the load curve. The hysteresis phenomena become much more substantial in this case (Fig. 3, curve *2*).

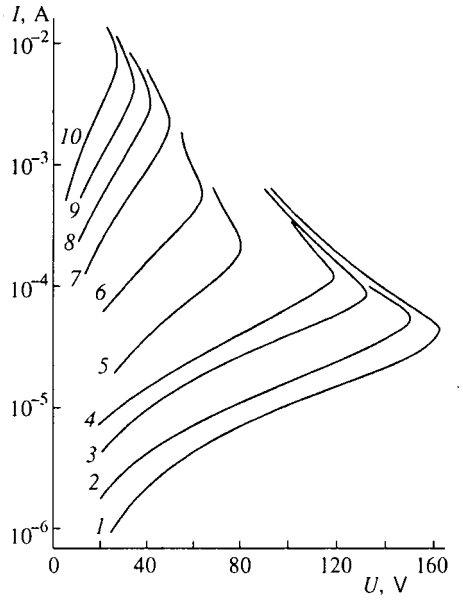

FIG. 4. Current-voltage characteristics of a PbTe sample after aging for 33 months at room temperature for different measurement temperatures (K): (*1*) 8, (*2*) 20, (*3*) 30, (*4*) 40, (*5*) 77, (*6*) 100, (*7*) 140, (*8*) 170, (*9*) 200, and (*10*) 240.

3. CURRENT-VOLTAGE CHARACTERISTICS OF SAMPLES AT DIFFERENT TEMPERATURES AND THEIR NONLINEAR CONDUCTIVITY IN THE SUBTHRESHOLD REGION

The state corresponding to curve *2* in Fig. 2 is already quite stable and a complete cycle of studies of the effect of electric fields on the electrical conductivity of the film can be conducted using it. Figure 4 shows a series of static current-voltage characteristics of the sample which were measured at different temperatures and which correspond to this structural state, in which the film parameters are essentially constant in time for ordinary storage conditions. Segments with a stable negative differential resistance show up at temperatures below 250 K. In the entire range of temperatures studied, down to $T=4.2$ K, no signs of current filament formation were observed. For operation at moderate currents, there are hardly any hysteresis effects. At $T>77$ K, a sharp increase in the threshold current is observed, and the segments with a negative differential resistance become ever less conspicuous as the temperature is raised. The absolute magnitude of the derivative dU/dI decreases rapidly with increasing T, and for $T>250$ K formation of a negative differential resistance is suppressed altogether.

The subthreshold current-voltage characteristics at temperatures below 250 K are highly nonlinear. The available data on polycrystalline semiconductors indicates that the most probable mechanisms for the nonlinearity are related to the presence of potential barriers at grain (block) boundaries. In contrast with the wide-gap semiconductors with microcrystalline structures, for narrow-gap materials that also have small-sized blocks the direct relationship between electrical conductivity and the presence of potential barriers at the boundaries of the inhomogeneities is not too obvious and must be clarified.

According to the effective medium theory, for weak and strong electric fields, the electrical conductivity of a polycrystalline film is determined by the sum of two terms:

$$\sigma=\sigma_H+\sigma_G, \tag{1}$$

where σ_H is the electrical conductivity across the grain boundaries, and σ_G is that for the barrier transport mechanism. In general, σ_H can also be the sum of two terms, the hopping term and the band term, when low-resistance layers exist at the grain boundaries.[28] We shall not consider the latter mechanism here, since it is not observed in the films we have studied.

Since potential barriers for electrons or holes develop as a result of the localization of the latter at the grain boundaries, there are no free carriers at the boundaries and, in this case, σ_H is obviously caused by hopping conductivity involving the localized states. The average hopping length for hopping conductivity with a variable hopping length is a function of temperature:[3,5]

$$l(T)=\alpha^{-1}\xi(T)=\alpha^{-1}(T_0/T)^{1/4}, \tag{2}$$

where T_0 is a well-known parameter in Mott's law, and α^{-1} is the length of the localized states. Given the possible values of α^{-1} and T_0 (Ref. 30), we find that the average jump length in the films studied here is, at most, 80 Å. Therefore, according the data of Ref. 30, we can ignore hopping conductivity in films with grain sizes greater than 10^{-6} cm and the mechanism for the nonlinearity for the current-voltage characteristics must be found in the intercrystallite potential barriers at the grain (block) boundaries.

Theoretical discussions of phenomena at interfaces usually consider the contact of a semiconductor with a semiconductor,[40–44] which actually consists of two Schottky barriers connected in opposition.[36,45] In a theory that provides a satisfactory description of the electrical conductivity of polycrystalline semiconductors with grain sizes greatly exceeding the Debye screening length $L_D=(kT\chi\chi_0/q^2p_0)^{1/2}$, where χ is the dielectric constant, χ_0 is the permitivity of free space, and q is the electronic charge, they are treated as a set of intercrystalline barriers (bicrystals) with identical properties. It should be noted that the transition from a bicrystalline to a polycrystalline material does not change the results essentially in this case. Many years of experimental studies have shown that this is entirely acceptable when working with weak electric fields. Here an average barrier height φ appears in the calculations.

The situation changes fundamentally for microcrystalline samples in strong electric fields. Here, also, the already inhomogeneous character of the polycrystalline system caused primarily by the spread in the height of the barriers at the grain boundaries can greatly change the resulting equations. The scatter in the barrier heights is related to a change in the form and size of the microcrystallites, their mutual orientation, and the degree to which the localized states are filled at the grain boundaries. As a number of theoretical and experimental studies[46–51] have shown, the randomness of the potential well of inhomogeneous semiconductors can lead to a fundamentally new effect – nonlinear percolation conductivity in strong electric fields. Nonlinear percolation conduc-

tivity has been observed experimentally[51] in photosensitive layers of PbS, an IV–VI material with properties close to those of PbTe. However, in contrast with the results discussed here, the data presented in Ref. 51 apply to ordinary polycrystalline samples with micron-sized grains.

This effect involves the onset of deviations from Ohm's law much sooner in randomly inhomogeneous systems than in systems with more or less periodic variations in the potential of equal amplitude. In addition, while there is an external similarity in the functional dependence of the current on the voltage for the Schottky effect and for nonlinear percolation conductivity, the nonlinearity of the current-voltage characteristic in the latter case is very much greater than for a homogeneous sample with a Schottky barrier, since the local electric field strength in inhomogeneous samples can be considerably higher than the measured average.

The behavior of a single barrier in a strong electric field is, to a great extent, determined by the Schottky effect. According to the theory of the Schottky effect caused by lowering a potential barrier in an electric field,[36] the current through the sample is

$$I = AT^2 \exp(-\varphi/kT) \exp[(q^3/4\pi\chi\chi_0)^{1/2} F^{1/2}/kT], \qquad (3)$$

where A is the Richardson constant, φ is the barrier height, and F is the electric field strength. However, experiments indicate that the nonlinearity coefficient for an inhomogeneous polycrystalline system can be an order of magnitude higher than $\beta = (q^3/4\pi\chi\chi_0)^{1/2}$ in Eq. (3) or the Poole-Frenkel coefficient $\beta = (q^3/\pi\chi\chi_0)^{1/2}$ (Ref. 50).

According to the landmark paper of Shklovskiĭ,[47] the current in a randomly inhomogeneous system is

$$I = I_0 \exp[(CqFaV_0^{\nu})^{1/(1+\nu)}/kT], \qquad (4)$$

where V_0 is the amplitude of the fluctuations in the potential well, $\nu \approx 0.9$, and $C = 0.25$; a is the scale length of the spatial inhomogeneity.

Equation (4) is valid for the range of electric fields determined by the inequality $V_0 \gg qFa \gg kT(kT/V_0)^{\nu}$. The condition $V_0 \gg kT$ must also be satisfied. If the critical index ν is replaced by unity, then Eq. (4) can be rewritten in the form

$$I = I_0 \exp(\beta F^{1/2}/kT), \qquad (5)$$

where $\beta = (CqaV_0)^{1/2}$. Later this theory was developed directly for microcrystalline systems with intergranular barriers.[46,48] The authors of Refs. 46 and 48, while confirming the validity of Shklovskiĭ's basic assumptions and conclusions, nevertheless find that these relations provide a better description of the behavior of polycrystalline semiconductors in strong electric fields. However, despite this critique[46,48] of Shklovskiĭ's results, as regards their application to inhomogeneous polycrystalline systems in high fields, on comparing the theory and experiment we were compelled to restrict the analysis to Shklovskiĭ's work, not solely because of its clarity and evident advantages for analyzing experimental data. In the calculations in Refs. 46 and 48 it was assumed that $L_D \ll a$. For our samples, this condition is clearly not satisfied: even at $T = 4.2$ K, the Debye screening length is 7×10^{-6} cm, or three times the average block size. Shklovskiĭ's theory[47] does not contain such strict requirements for the potential well and is, we believe, more universal. In accordance with Ref. 49, the theory and experiment were compared at rather high electric fields.

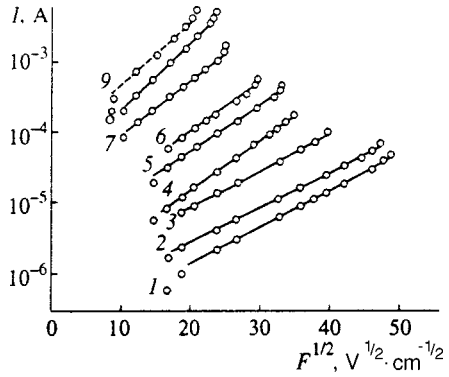

FIG. 5. Subthreshold portions of the current-voltage characteristics of a diode at different temperatures (K): (*1*) 8, (*2*) 20, (*3*) 40, (*4*) 77, (*5*) 90, (*6*) 100, (*7*) 150, (*8*) 200, and (*9*) 240.

As the theory predicts, the current-voltage characteristics of these samples deviate anomalously early from Ohm's law. The critical value of F is 0.3 V/cm at $T = 4.2$ K, 0.7 V/cm at $T = 15$ K, and about 5 V/cm at $T = 77$ K. It should be recalled that in as-grown samples, these values were one or two orders of magnitude higher because of the enhanced inhomogeneity of the microcrystalline films as they fragment due to the breakup of coarse grains into finer blocks.

In the region where their conductivity depends on the electric field strength, the current-voltage characteristics of these samples actually correspond to a dependence of the form $\ln I \propto F^{1/2}$ over a wide temperature range (Fig. 5). As Fig. 5 shows, the segments of the current-voltage characteristic which are consistent with the nonlinear percolation conductivity theory shift toward higher electric fields as the temperature is lowered. The deviation of the experimental $I = I(F^{1/2})$ curves from the theoretical dependence (5) is caused by ohmic segments in low fields and the development of a negative differential resistance in high fields. The experimental values of β, consistent with the theory for low temperatures, are 5–7 times the values calculated for the Poole-Frenkel effect $[(1.0-1.9) \times 10^{-5}$ eV$^{1/2}\cdot$cm$^{1/2}$ in the temperature interval 4.2–300 K]. As the temperature is raised, because of the rapid rise in β with temperature, this difference increases by more than two orders of magnitude (Fig. 6, curve *2*).

The average barrier height φ, which determines the conductivity of a microcrystalline medium at a given temperature and, therefore, V_0 in Eq. (5) which corresponds roughly to φ, can be determined from the temperature variation in the electrical conductivity σ in low electric fields shown in Fig. 7, assuming that

$$\sigma \propto \exp(-\varphi/kT). \qquad (6)$$

The electrical conductivity of highly inhomogeneous media has been examined in Ref. 5. In terms of the basic concepts of the percolation theory, φ characterizes the average height of the barriers, which form a critical sublattice in the ran-

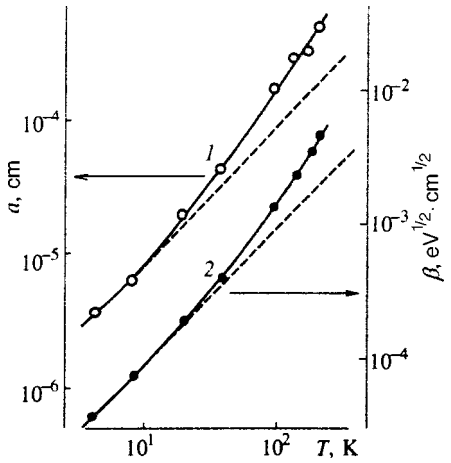

FIG. 6. Temperature variations in the spatial inhomogeneity scale length a (1) and nonlinearity coefficient β (2).

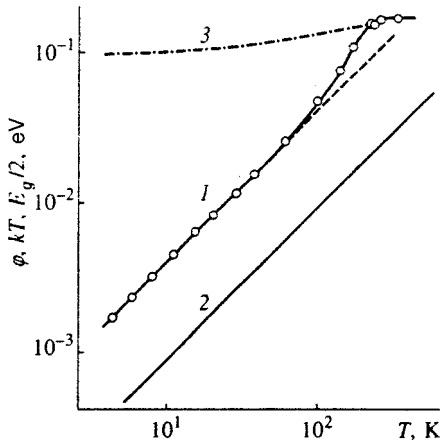

FIG. 8. Average height of the intercrystallite barriers, φ (1), kT (2), and $E_g/2$ (3) as functions of temperature.

domly inhomogeneous medium under consideration. The effect of the intrinsic conductivity at low temperatures can be ignored.[30,31] This is also quite clear from the data shown in Fig. 8, on comparing the behavior of $\varphi=\varphi(T)$ with the temperature dependence of the optical band gap. The hopping conductivity is not observed in these samples. This last feature is well illustrated by the curves shown in Fig. 7: the conductivity of these samples decreases with aging over the entire temperature range, while hopping conductivity should increase as $\sigma\propto \exp(-D^{*1/3})$ when the average block size D^* is reduced.[30]

The dependence of φ on T in Fig. 8 (curve *1*) shows that the barrier height increases almost linearly with T at low temperatures, i.e.,

$$\varphi\propto T^t, \tag{7}$$

where $t\simeq 0.96$–0.97, while at higher temperatures ($T\geqslant 240$ K) $\varphi=\varphi(T)$ has a saturation dependence. At $T>240$ K, the temperature dependence of the conductivity is obviously caused by the temperature dependence of the intrinsic conductivity.

The Schottky barriers formed when metals are in contact with single crystal PbTe have an effect on the electrical conductivity of the samples up to somewhat higher temperatures. The properties of contacts between single crystal *n*- and *p*-type PbTe and various metals (In, Cu, Ag, Au) have been studied in detail.[52] It was shown[52] that the height of the Schottky barrier is independent of the work function of the metal and varies in the range 0.175–0.200 eV, or 0.5–0.6 times E_g for PbTe at room temperature. As the temperature is lowered, the barrier height decreases in accordance with the reduction in the band gap.

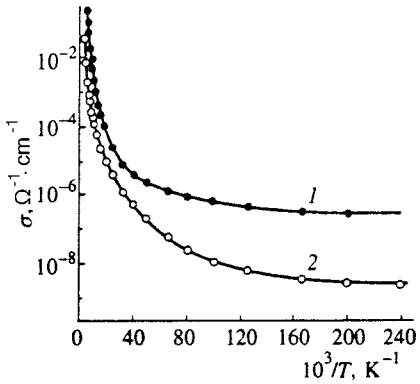

FIG. 7. Temperature variations in the electrical conductivity of a microcrystalline PbTe film in a weak electric field 8 months (curve *1*) and 33 months (curve *2*) after fabrication.

4. MECHANISM FOR READJUSTMENT OF THE POTENTIAL WELL AS THE SAMPLE TEMPERATURE IS VARIED

Assuming $V_0=\varphi$ and knowing β, one can use Eq. (5) to find the scale length a of the spatial inhomogeneity. At $T=4.2$ K, a is 3.7×10^{-6} cm, which is still substantially greater than the possible tunnelling length for charge carriers in PbTe. The quantity $a=3.7\times10^{-6}$ cm is already close to the average block size $D^*=2\times10^{-6}$ cm found by x-ray structural analysis[31] and indicates that Shklovskiĭ's theory[47] applies to our results. The inequality $V_0\gg qFa\gg kT(kT/V_0)^{\nu}$ holds solidly in the entire range of temperatures studied here.

However, because a increases with temperature, an initially quite unexpected situation arises. Thus, for example, at 8 K the value of a increases to 6.4×10^{-6} cm and reaches 5×10^{-4} at 240 K. The dependence of a on T is plotted in Fig. 6 (curve *1*). This behavior of the system originates in the relationship between the potential φ and the scale length a of the inhomogeneity. The relationship between the potential and the inhomogeneity scale length is apparently typical of any system on a mesoscopic scale, where the scale length varies from a few nanometers to several tens of nanometers. This situation is fundamentally independent of the particular type of structure, whether we are speaking of nanocrystals or systems of quantum dots.[9]

It should be noted that the results presented here can be used for films with arbitrary grain (block) size or inhomogeneity scale length to determine the temperature above which the conductivity of the samples no longer depends on the electric field strength.

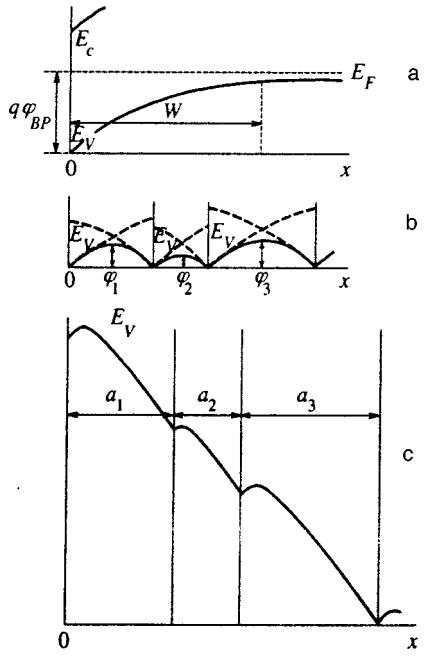

FIG. 9. Band energy diagrams of a *p*-type polycrystalline semiconductor: a—$a \gg W$, b $a < W$, c as in b, but in a strong electric field.

If the grain size is sufficiently large and $D \gg L_D$, then the solution of the Poisson equation for the one-dimensional case,

$$\frac{\partial^2 V}{\partial x^2} = -\frac{q}{\chi\chi_0} N_A \tag{8}$$

is[36]

$$V(x) = -\frac{qN_A}{\chi\chi_0}\left(Wx - \frac{1}{2}x^2 \right) - \varphi_{bp}, \tag{9}$$

where $V(x)$ is the potential, φ_{bp} is the potential barrier height, W is the width of the space charge layer, N_A is the acceptor concentration, and x is the coordinate. This situation is illustrated in Fig. 9a. A *p*-type semiconductor is being considered, as in our particular case. On the thermodynamic equilibrium band energy diagram shown in Fig. 9a, E_F is the Fermi energy, E_c is the energy corresponding to the bottom of the conduction band, and E_v is the energy corresponding to the ceiling of the valence band. If the scale length a of the spatial inhomogeneity is small and the second boundary at $x = a$ does not affect the potential distribution, then for $a \ll W$ Eq. (9) yields

$$V(a) \simeq -\frac{qN_A}{\chi\chi_0} Wa - \varphi_{bp}. \tag{10}$$

The height of the potential barrier,

$$\varphi(0) = -\varphi_{bp} - V(a) \simeq \frac{qN_A}{\chi\chi_0} Wa, \tag{11}$$

is proportional to a in this case. If we assume the existence of a symmetric barrier at $x = a$ (Fig. 9b), then the solution of the Poisson equation (8) must be taken in the form

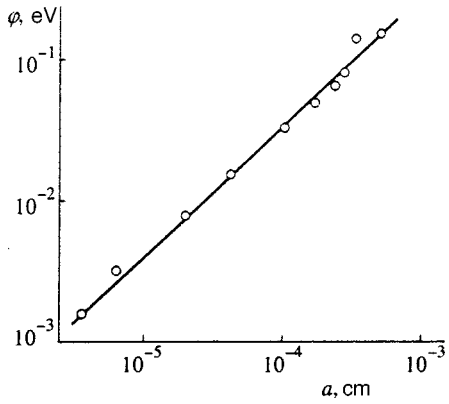

FIG. 10. Relationship between the average intercrystallite barrier height φ and the spatial inhomogeneity scale length a.

$$V(x) = -\frac{qN_A}{\chi\chi_0}\left[Wx - \frac{1}{2}x^2 \right] - \frac{qN_A}{\chi\chi_0} \times \left[W(a-x) - \frac{1}{2}(a-x)^2 \right] - \varphi_{bp}. \tag{12}$$

We obtain an expression for the height of the potential barrier which differs from Eq. (11):

$$\varphi(0) = \varphi(a) = V(0) - V\left(\frac{a}{2}\right) = V(a) - V\left(\frac{a}{2}\right) = \frac{qN_A}{\chi\chi_0}\frac{a^2}{4}. \tag{13}$$

An analogous result was obtained in Ref. 41, where the kernel was approximated by a sphere of radius r_0. The space charge is concentrated in a layer from $r_1 < r_0$ to r_0; i.e., $r_0 - r_1$ has the significance of a screening length. The height of the potential barrier is given by

$$\varphi(r_0) = \frac{2\pi qN_A}{3\chi\chi_0}\left(r_0^2 + \frac{2r_1^3}{r_0} - 3r_1^2 \right). \tag{14}$$

For $r_1 = 0$ we have, as in Eq. (13), a quadratic dependence of the potential barrier height on the grain size,

$$\varphi(r_0) = \frac{2\pi qN_A}{3\chi\chi_0} r_0^2. \tag{15}$$

As the data of Fig. 10 show, however, in the entire range of variation of a the relation between the height of the potential barrier and the scale length of the spatial inhomogeneity is close to a linear dependence, which can be approximated by a power law of the form

$$\varphi \propto a^{\eta}, \tag{16}$$

where $\eta \simeq 0.94$. The height of the barrier is determined by the inhomogeneity scale length $a = a(T)$ and, as the data of Fig. 8 imply, is a function of temperature. Clearly, the temperature dependence of the band gap E_g cannot be the reason for the dependence (16) with $\eta \simeq 1$. On the other hand, since E_g in PbTe increases as the temperature rises,[27,53] one should rather expect Eq. (16) with $\eta > 2$. The reason for this assumption can be understood if we note that all the measurements of the sample parameters were made in strong electric fields. This is clearly evident for the one-dimensional case

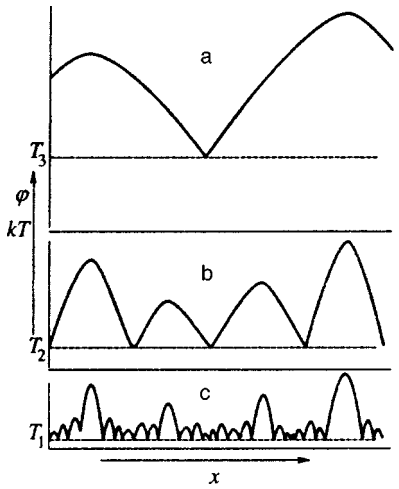

FIG. 11. Scheme for readjustment of the potential well as the temperature changes: $T_3 > T_2 > T_1$.

(Fig. 9c). In an electric field the potential distribution is highly distorted toward one of the barriers compared to the equilibrium case. If, for example, we consider an inhomogeneity of size $a = a_1$, then it is evident from Fig. 9c that the barrier at $x = 0$ makes the main contribution and then, as Eq. (11) implies, $\varphi \propto a$. In addition, since the measurements are made in strong electric fields, there is an inevitable reduction in the barrier height in the electric field for each individual barrier due to the Schottky effect. This is qualitatively understandable from the change in the potential distribution in Fig. 9c, while the quantitative changes should obey Eq. (3).

As for the mechanism for the rapid increase in the spatial inhomogeneity scale length a as the temperature rises, here the main contribution is from the ''vanishing'' or, more precisely, the thermal smearing out of small barriers with heights $\varphi_{\min} < kT$. Judging from the relationship between φ and kT, i.e., given that $\varphi \simeq 4kT$, we may conclude that during current flow, not only the barriers with heights $\varphi_{\min} < kT$, but also the barriers whose average height $\varphi < 4kT$, do not participate. The same applies to barriers of height $\varphi > 4kT$. The primary contribution is from barriers of height $\varphi \approx 4kT$, which form a critical sublattice and determine the electrical conductivity of this randomly inhomogeneous medium with an adjustable potential well. In the case $a < W$, space charge layers of the barriers whose height $\varphi_{\max} \gg kT$ propagate to regions with vanishing barriers. It is not by chance that the relationship between φ and T in Fig. 8 is extremely close to linear, especially at low temperatures. The total surface area of the blocks united in these unique ensembles (or clusters) decreases with temperature and, as free charge carriers are trapped by surface states, the specific charge per unit surface area increases.

As they propagate in the space which previously belonged to lower barriers, the barriers at the boundaries of the large blocks simultaneously increase in amplitude. This process continues until W exceeds $a/2$. When the condition $W < a/2$ is satisfied, the dependence of φ on T should saturate, subsequently following the dependence of E_g on T, which in our case occurs at $T > 230$ K (Fig. 8, curve *1*). The calculated value of $W = (2\chi\chi_0\varphi/qp_0)^{1/2} = 1.7 \times 10^{-4}$ cm for $T = 230$ K, in good agreement with the value $a/2 = 2.1 \times 10^{-4}$ cm found by analyzing the electrical measurement data in accordance with the Shklovskiĭ theory. We regard this fact as one of the most weighty arguments in favor of using the Shklovskiĭ theory for analyzing our experimental data. The dependence of χ on T was taken into account when calculating W.[55] Since the band gap E_g in PbTe increases considerably more slowly than kT [from 0.19 eV at $T = 4.2$ K to 0.32 eV at $T = 300$ K (Refs.27 and 53)], the increase in E_g with temperature only facilitates the growth of ''high'' barriers. As a result, the change in the height distribution function of the barriers changes the paths of the current flow in the system in such a way that the nonlinear conductivity includes temporary configurational ensembles which determine the spatial inhomogeneity scale length and within which the number of blocks increases constantly with temperature. This process is illustrated schematically in Fig. 11. The height of the barriers increases in this case much more rapidly than the increase in E_g with T, while the reduction in their number leads to a continuous rise in the nonlinearity coefficient β with T.

5. ELECTRICAL CONDUCTIVITY IN STRUCTURES WITH AN ADJUSTABLE POTENTIAL WELL AND HOPPING CONDUCTIVITY WITH A VARIABLE HOPPING LENGTH

Given the rise in the activation energy for electrical conductivity in our film samples with increasing temperature, it is meaningful to compare the observed mechanism for electrical conductivity in microcrystalline systems that have an adjustable potential well with hopping conductivity when the hopping length is variable. This is especially appropriate, since the theory of hopping conductivity based on percolation theory[5] also deals with a critical sublattice that changes with temperature and plays a key role in deriving the basic equations.

If, in accordance with the known activation dependence for the specific resistivity of a semiconductor of the form $\rho = \rho_0 \exp(\Delta E/kT)$, where $\Delta E = \partial \ln \rho / \partial (kT)^{-1}$, we define a local activation energy for hopping conductivity with a variable hopping length of the form $\delta\varepsilon = \partial \ln \rho / \partial (kT)^{-1}$ (Ref. 54), then for the temperature dependence of the local activation energy $\delta\varepsilon$ we obtain

$$\delta\varepsilon = \frac{1}{4} k T_0^{1/4} T^{3/4} = \frac{1}{4} (T_0/T)^{1/4} kT. \tag{17}$$

However, despite an outward similarity in the temperature variations of the local activation energies $\delta\varepsilon = \delta\varepsilon(T)$ and $\varphi = \varphi(T)$, it is obvious that there is, in fact, no analogy here at all and we are dealing with completely different conductivity mechanisms. The mechanism under consideration here, as for the overwhelming majority of other microcrystalline systems with semiconductor conductivity, is caused by superbarrier transport of charge carriers, and, in our case, the existence of a temperature dependence for the local activation energy (according to Eq. (7), $\varphi \propto T^{0.96}$) is the result of a potential well that adjusts itself with temperature and the dependence of the barrier height on the linear size a of the resulting ensembles (clusters). The hopping mechanism is

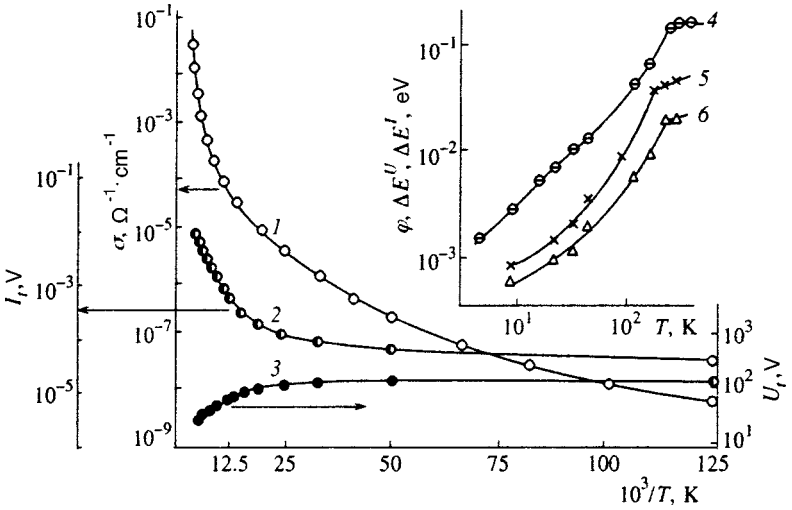

FIG. 12. Electrical resistivity σ in weak electric fields (*1*), threshold current I_t (*2*), and threshold voltage U_t (*3*) as functions of temperature. The inset shows the average intercrystallite barrier height φ (*4*) and activation energies ΔE^I (*5*) and ΔE^U (*6*) as functions of temperature.

based on tunnelling of carriers between two localized states which differ in energy by the energy of a phonon.

The difference between these mechanisms becomes especially evident when the behavior of the inhomogeneity scale length a is compared with that of the characteristic size of the critical sublattice in the theory of hopping conductivity,[5,56]

$$L_0 = l(T)\xi_c^{\nu} \simeq \alpha^{-1}\xi_c^{1+\nu}, \tag{18}$$

where $l(T) = \alpha^{-1}\xi_c(T) = \alpha^{-1}(T_0/T)^{1/4}$ is the average jump length, α^{-1} is the radius of the localized state, and $\nu \simeq 0.9$ is the exponent on the correlation length. This length (L_0) plays an important role in the theory of all effects associated with hopping conductivity, including in the behavior of hopping conductivity in strong electric fields.[56,57] This shows that, in contrast with the inhomogeneity scale length a, which increases linearly or superlinearly with temperature in the microcrystalline system under consideration, L_0 has an reciprocal temperature dependence and decreases with rising temperature according to the power law

$$L_0 \propto \xi_c^{1+\nu} \propto T^{-0.47}. \tag{19}$$

6. NEGATIVE DIFFERENTIAL RESISTANCE

An idea of the mechanism for the negative differential resistance can be obtained from the well-tested procedure of comparing the dependences of the threshold voltage U_t and current I_t on T with the temperature dependence of the electrical conductivity σ in a weak electric field (see Fig. 12). The changes in the threshold parameters U_t and I_t are small in the interval 8–40 K. Raising the temperature above 40 K substantially increases the rate at which U_t and I_t change with temperature. This pertains largely to the temperature dependence of the threshold current I_t. Figure 12 shows that the threshold voltage U_t and threshold current I_t vary much more slowly with T than does $\sigma = \sigma(T)$. A quantitative criterion can be found if we write the temperature dependences of σ, U_t, and I_t in the conventional forms

$$\sigma \propto \exp(-\varphi/kT), \quad U_t \propto \exp(\Delta E^U/kT),$$

$$I_t \propto \exp(-\Delta E^I/kT),$$

where φ, ΔE^U, and ΔE^I, in turn, depend on the temperature and, therefore, on local activation energies with the corresponding temperature dependences. The temperature dependences of φ, ΔE^U, and ΔE^I are shown in the inset in Fig. 12 for comparison.

The similarity in the functions $\Delta E^U = \Delta E^U(T)$ and $\varphi = \varphi(T)$ indicates an explicit relationship between the mechanisms for the negative differential resistance and thermal processes in our samples. The activation energy ΔE^U, however, varies in the range $(0.12–0.20)\varphi$ in the entire temperature interval. At the same time, for a purely thermal mechanism for the negative differential resistance, it should be equal to 0.5φ (Ref. 39). Similarly, instead of equalling 0.5φ, ΔE^I for the thermal mechanism varies in the range $(0.21–0.31)\varphi$. Given the above arguments, it must be assumed that in this case the mechanism for the negative differential resistance is electrothermal, with a field component determined by the nonlinear percolation conductivity of the microcrystalline samples. It is obvious from the smallness of ΔE^U that electron processes make a large contribution. It is greatest in the temperature range 20–30 K, where ΔE^U and ΔE^I form a minimal fraction of φ. At lower temperatures, the resistances of the barriers are low and at high temperatures, the threshold currents are too high. Given that the activation energies ΔE^U for U_t as a function of T vary little over the entire temperature range studied here [they vary in the range $(0.12–0.20)\varphi$], the contribution of thermal processes to the development of a negative differential resistance in this temperature range is essentially constant. The data of Fig. 8 show that a negative differential resistance can be observed in these films at temperatures up to 240–250 K, beyond which thermal generation of charge carriers of both signs and a transition to intrinsic conductivity will destroy the state with a negative differential resistivity and liquidate the nonlinear conductivity in these samples.

The data of Fig. 3 can serve as important evidence of an electrothermal mechanism for the negative differential resistance. The curves in that figure refer to entirely different structural states, and their resistances measured at 25 K and in weak electric fields, differ by four orders of magnitude. At the same time, the curves essentially merge in the segment

with a negative differential resistance at currents $\sim 10^{-4}$ A; i.e., the resistance of this sample is the same in both structural states. This means that for the same scattered power (which is large in this segment, according to the data shown here), which produces high enough temperatures in the active region, the scale length of the spatial inhomogeneity that determines the magnitude of the resistance and therefore the resistance of the sample, is the same regardless of the situation at low temperatures. Here it is also important that all the remaining parameters, and primarily those related to the sample geometry, remain constant.

7. CONCLUSIONS

We have cited some evidence of the unusual behavior of polycrystalline samples in strong electric fields, when the inhomogeneity scale length (grain and block sizes) is much shorter than the Debye screening length. The data presented here show that this unusual behavior of polycrystalline samples in the nonlinear percolation conductivity regime is caused by a potential well that readjusts itself with temperature. In this case, the height of the potential barrier and the inhomogeneity scale length are functions of temperature. These effects could be observed because of a successful combination of the structure of the test samples (small blocks with perfect interfaces formed during fragmentation of coarse grain films and capable of joining together easily into ensembles consisting of a different number of these blocks) and the experimental conditions. We have also found good agreement between the experimental data and the theory of Shklovskiĭ. Two major indications of this agreement should be noted. First, as shown in Sec. 4, at low (≈ 4.2 K) temperatures the inhomogeneity scale length $a = 3.7 \times 10^{-6}$ cm found by analyzing the experimental data in accordance with Shklovskiĭ's theory actually does approach its minimum value, which equals the average block size ($D^* \simeq 2 \times 10^{-6}$ cm) after fragmentation of coarse crystalline ($D \simeq 5 \times 10^{-3}$ cm) films. The other unique ''reference'' point is the inhomogeneity scale length determined by using the formulas proposed by Shklovskiĭ at high temperatures, when the temperature dependence of the average potential barrier height (Fig. 8) begins to saturate. This occurs when, as the temperature rises, the linear size of the spatial ensembles, a, becomes comparable to the width W of the space charge layer in single-crystal samples as the number of blocks increases. The calculated $W = 1.7 \times 10^{-4}$ cm is in good agreement with the value $a/2 = 2.1 \times 10^{-4}$ cm at $T = 230$ K found by analyzing electrical measurement data in accordance with Shklovskiĭ's theory. Subsequently the barrier height essentially ceases to rise with temperature, and the system behavior becomes similar to the well-studied behavior of coarse crystalline samples with grain sizes much greater than the Debye screening length. In narrow-gap semiconductors this phenomenon can be observed only at low temperatures, but, as the gaps of polycrystalline semiconductors increase, the upper bound can evidently be shifted toward higher temperatures.

As the data presented here imply, the negative differential resistance is caused by an electrothermal mechanism. This is indicated by a quantitative analysis of the temperature variations in U_t and I_t in Fig. 12. A further weighty proof of the importance of thermal processes in the development of the negative differential resistance is the closeness of the two current-voltage characteristics for different structural states of the same sample shown in Fig. 3 for high ($\sim 10^{-4}$ A) currents. We believe that these current-voltage characteristics are a good illustration of the feasibility of the mechanism proposed here for the readjustment of the potential well.

We thank Z. A. Samoĭlenko for the x-ray structural analyses of the samples.

*)E-mail: okunev@host.dipt.donetsk.ua

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] V. L. Bonch-Bruevich, I. P. Zvyagin, R. Kaĭper, A. G. Mironov, R. Enderleĭn, and B. Esser, *Electron theory of Disordered Semiconductors* [Russian translation], Nauka, Moscow (1981).
[3] N. Mott and E. Davis, *Electronic Processes in Noncrystalline Substances* [Russian translation], Mir, Moscow (1982).
[4] N. F. Mott, *Metal-Insulator Transitions* [Russian translation], Nauka, Moscow (1979).
[5] B. I. Shklovskiĭ and A. L. Efron, *Electronic Properties of Doped Semiconductors* [in Russian], Nauka, Moscow (1979).
[6] V. N. Lutskii, Phys. Status Solidi A **1**, 199 (1970).
[7] Zh. I. Alferov, Fiz. Tekh. Poluprovodn. **32**, 1 (1998) [Semiconductors **32**, 1 (1998)].
[8] N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, P. S. Kop'ev, Zh. I. Alferov, and D. Bimberg, Fiz. Tekh. Poluprovodn. **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].
[9] M. G. Mil'vidskiĭ and V. V. Chaldyshev, Fiz. Tekh. Poluprovodn. **32**, 513 (1998) [Semiconductors **32**, 457 (1998)].
[10] H. Günterodt and H. Beck, *Metallic Glasses* [Russian translation], Mir, Moscow (1983).
[11] A. Feltz, *Amorphous and Glassy Nonorganic Solids* [Russian translation], Mir, Moscow (1986).
[12] A. S. Bakaĭ, *Polycluster Amorphous Materials* [in Russian], Energoatomizdat, Moscow (1987).
[13] Ping Sheng and J. Klafter, Phys. Rev. B **27**, 2583 (1983).
[14] Yu. F. Komnik, Fiz. Nizkikh Temperatur **8**, 115 (1982) [Sov. J. Low Temp. Phys. **8**, 57 (1982)].
[15] B. Abeles, Ping Sheng, M. D. Coutts, and Y. Arie, Adv. Phys. **24**, 407 (1975).
[16] A. M. Glukhov, N. Ya. Fogel', and A. A. Shablo, Fiz. Tverd. Tela. **28**, 1043 (1986) [Sov. Phys. Solid State **28**, 583 (1986)].
[17] M. Mostefa and G. Olivier, Solid State Commun. Solid State Commun. **63**, 219 (1987).
[18] V. D. Okunev and Z. A. Samoĭlenko, JETP Lett. **43**, 28 (1986).
[19] V. D. Okunev and Z. A. Samoĭlenko, JETP Lett. **53**, 44 (1991).
[20] A. P. Gorchakov, Yu. A. Zarif'yants, and O. I. Tananaev, Izv. Akad. Nauk SSSR, Neorg. Mater. **18**, 1275 (1982).
[21] Yu. A. Boĭkov and V. A. Kutasov, Fiz. Tverd. Tela **24**, 3192 (1982) [Sov. Phys. Solid State **24**, 1813 (1982)].
[22] Z. M. Dashevskiĭ and M. P. Rulenko, Fiz. Tekh. Poluprovodn. **27**, 662 (1993) [Semiconductors **27**, 366 (1993)].
[23] A. V. Burlak, V. V. Zotov, A. V. Ignatov, A. V. Tyurin, and V. G. Tsukerman, Fiz. Tekh. Poluprovodn. **26**, 548 (1992) [Sov. Phys. Semiconductors **26**, 311 (1992)].
[24] L. N. Neustroev and V. V. Osipov, Fiz. Tekh. Poluprovodn. **18**, 359 (1984) [Sov. Phys. Semicond. **18**, 224 (1984)].
[25] L. N. Neustroev and V. V. Osipov, Fiz. Tekh. Poluprovodn. **20**, 59 (1986) [Sov. Phys. Semicond. **20**, 34 (1986)].
[26] Yu. I. Ravich, B. A. Efimova, and I. A. Smirnov, *Methods for Studying Semiconductors as Applied to the Lead Chalcogenides PbTe, PbSe, and PbS* [in Russian], Nauka, Moscow (1968).
[27] N. Kh. Abrikosov and L. E. Shelimov, *Semiconducting materials based on $A^{IV}B^{VI}$ Compounds* [in Russian], Nauka, Moscow (1975).

[28] S. A. Azimov and Sh. B. Atakulov, *Kinetic Phenomena in Polycrystalline films of Lead and Bismuth Chalcogenides* [in Russian], Izd. ''Fan,'' Tashkent (1985).
[29] S. V. Plyatsko, Fiz. Tekh. Poluprovodn. **32**, 257 (1998) [Semiconductors **32**, 231 (1998)].
[30] Yu. A. Bratashevskiĭ, V. D. Okunev, N. N. Pafomov, and Z. A. Samoĭlenko, Fiz. Tverd. Tela **27**, 723 (1985) [Sov. Phys. Solid State **27**, 447 (1985)].
[31] Yu. A. Bratashevskiĭ, V. D. Okunev, and Z. A. Samoĭlenko, Izv. Akad. Nauk SSSR, Neorg. Mater. **21**, 1124 (1985).
[32] H. St-Onge and J. N. Walpole, Phys. Rev. B **6**, 2337 (1972).
[33] Yu. G. Troyan, F. F. Sizov, and V. M. Lakeenkov, Ukr. Fiz. Zh. **32**, 467 (1987).
[34] B. A. Akimov, A. V. Albul, and E. V. Bogdanov, Neorg. Mater. **28**, 2377 (1992).
[35] B. A. Akimov, A. V. Albul, E. V. Bogdanov, and V. Yu. Il'in, Fiz. Tekh. Poluprovodn. **26**, 1300 (1992) [Sov. Phys. Semicond. **26**, 726 (1992)].
[36] S. Sze, *Physics of Semiconductor Devices* [Russian translation], Vol. 1, Mir, Moscow (1984).
[37] Yu. A. Bratashevskiĭ, V. D. Okunev, and Z. A. Samoĭlenko, Izv. Akad. Nauk SSSR, Neorg. Mater. **22**, 1568 (1986).
[38] V. F. Gantmakher and I. B. Levinson, *Scattering of Charge Carriers in Metals and Semiconductors* [in Russian], Nauka, Moscow (1984).
[39] B. Yu. Lototskiĭ and L. K. Chirkin, Fiz. Tverd. Tela **8**, 1967 (1966) [Sov. Phys. Solid State **8**, 1564 (1966)].
[40] G. J. Korsh and R. S. Muller, Solid State Electron. **21**, 1045 (1978).
[41] V. G. Kovka, R. P. Kormirenko, Yu. V. Kornyushin, Yu. P. Medvedev, and O. V. Tretyak, Fiz. Tekh. Poluprovodn. **16**, 2176 (1982) [Sov. Phys. Semiconductors **16**, 1404 (1982)].
[42] Y. F. Matare, J. Appl. Phys. **56**, 2605 (1984).
[43] E. I. Gol'dman and A. G. Zhdan, Fiz. Tekh. Poluprovodn. **12**, 833 (1978) [Sov. Phys. Semicond. **10**, 491 (1978)].
[44] E. I. Gol'dman, A. G. Zhdan, Yu. V. Markin, and P. S. Shul'zhenko, Fiz. Tekh. Poluprovodn. **17**, 390 (1983) [Sov. Phys. Semicond. **17**, 242 (1983)].
[45] A. Milnes and D. Voigt, *Heterojunctions and Metal-semiconductor Junctions* [Russian translation], Mir, Moscow (1975).
[46] A. Ya. Vinnikov, A. M. Meshkov, and V. N. Savushkin, Fiz. Tverd. Tela **22**, 2989 (1980) [Sov. Phys. Solid State **22**, 1745 (1980)].
[47] B. I. Shklovskiĭ, Fiz. Tekh. Poluprovodn. **13**, 93 (1979) [Sov. Phys. Semicond. **13**, 53 (1979)].
[48] A. Ya. Vinnikov, A. M. Meshkov, and V. N. Savushkin, Fiz. Tverd. Tela **24**, 1352 (1982) [Sov. Phys. Solid State **24**, 766 (1982)].
[49] E. I. Levin, Fiz. Tekh. Poluprovodn. **18**, 255 (1984) [Sov. Phys. Semicond. **18**, 158 (1984)].
[50] A. Ya. Vinnikov, A. M. Meshkov, and V. N. Savushkin, Pis'ma Zh. Tekh. Fiz. **6**, 726 (1980) [Sov. Tech. Phys. Lett. **6**, 312 (1980)].
[51] Sh. B. Atakulov and K. É. Onarkulov, Fiz. Tekh. Poluprovodn. **19**, 1324 (1985) [Sov. Phys. Semicond. **19**, 811 (1985)].
[52] F. F. Sizov, V. V. Teterkin, Yu. G. Troyan, and V. Yu. Chopik, Optoélektronika is poluprovodn. tekh., No. 12, 47 (1987).
[53] R. N. Tauber, A. A. Machonis, and I. B. Cadoff, J. Appl. Phys. **37**, 4855 (1966).
[54] V. D. Okunev and N. N. Pakhomov, Fiz. Tekh. Poluprovodn. **20**, 1302 (1986) [Sov. Phys. Semicond. **20**, 822 (1986)].
[55] G. M. T. Foley and D. N. Langenberg, Phys. Rev. B **15**, 4830 (1977).
[56] B. I. Shklovskiĭ and A. L. Éfros, Usp. Fiz. Nauk **117**, 401 (1975) [Sov. Phys. Usp. **18**, 845 (1975)].
[57] B. I. Shklovskiĭ, Fiz. Tekh. Poluprovodn. **10**, 1440 (1976) [Sov. Phys. Semicond. **10**, 855 (1976)].

Translated by D. H. McNeill

Hard soliton excitation regime: Stability investigation

E. A. Kuznetsov*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 29 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 299–317 (July 1999)

The problem of the stability of one-dimensional solitons in the hard regime of soliton excitation, where the matrix element of the four-wave interaction has an additional smallness, is studied. It is that shown for optical solitons striction can weaken the Kerr nonlinearity. It is shown that solitons with a finite amplitude discontinuity at the critical soliton velocity, equal to the minimum phase velocity of linear waves, are unstable while solitons with a soft transition remain stable with respect to one-dimensional perurbations. Two- and three-dimensional solitons near threshold are unstable with respect to modulation perturbations. © *1999 American Institute of Physics.* [S1063-7761(99)02207-6]

1. INTRODUCTION

This paper is devoted to a study of hard and soft soliton excitation regimes (in other words — sub- and supercritical bifurcations) accompanying a change in the soliton velocity **v** — one of the main characteristics of a soliton. It is well known that if the velocity **v** of a moving object is such that

$$\omega_k = \mathbf{k}\cdot\mathbf{v}, \tag{1.1}$$

where $\omega=\omega_k$ is the dispersion law for linear waves and **k** is the wave vector, then such an object will lose energy vis Cherenkov radiation. This also applies fully to solitons as localized stationary formations. They cannot exist if the resonance condition (1.1) is satisfied. Hence follows the first, and simplest, selection rule for solitons: The soliton velocity must be either less than the minimum phase velocity of linear waves or greater than the maximum phase velocity. The boundary separating the region of existence of solitons from the resonance region (1.1) determines the critical soliton velocity $v_{\rm cr}$. As is easy to see, this velocity is the same as the group velocity of linear waves at the point where the straight line $\omega=kv$ is tangent to the dispersion curve $\omega=\omega_k$ (in the multidimensional case — the point of tangency of the plane $\omega=\mathbf{k}\cdot\mathbf{v}$ to the dispersion surface). If the tangency occurs from below, then the critical velocity determines the maximum soliton velocity for this parameter range and, conversely, for tangency from above $v_{\rm cr}$ is the same as the minimum phase velocity. Two regimes are possible in crossing this boundary: soft or hard excitation, in other words, super- or subcritical bifurcation.

It has been determined previously[1–5] that near the critical velocity solitons in the soft excitation regime behave in a universal manner. This universality is manifested, in the first place, as a square-root velocity dependence of the soliton amplitude, as is typical for second-order phase transitions, making this phenomenon similar to second-order phase transitions. However, an important distinction from phase transitions is manifested in the shape of a soliton. The shape is also universal: As the velocity approaches the critical value, the soliton shape acquires the form of an envelope soliton for the nonlinear Schrödinger equation (NLSE). This behavior occurs for any spatial dimension.

This isue was first considered for gravity-capillary surface waves on deep water. One-dimensional soliton solutions were first found numerically in Ref. 1. Later, a bifurcation — a transition from periodic solutions to a soliton solution — was studied in Refs. 2 and 3 using normal forms. The stationary NLSE for gravity-capillary wave solitons was derived in Ref. 4. In Ref. 5 it was shown that this mechanism can be extended to optical solitons. It followed from this work, essentially, that this mechanism occurs for waves of arbitrary nature.

The question of whether the bifurcation is super- or subcritical depends on the character of the nonlinear interaction. The soft transition regime occurs with focusing nonlinearity when the product $\omega'' T<0$, where $\omega''=\partial^2\omega/\partial k^2$ is the second derivative of the frequency with respect to the wave vector, evaluated at the point of tangency $k=k_0$, and T is the value of the matrix element $T_{k_1k_2k_3k_4}$ of the four-wave interaction for $k_i=k_0$. If $\omega'' T>0$, which corresponds to a defocusing nonlinearity, there are no solitons — localized solutions — with amplitude vanishing smoothly as $v\to v_{\rm cr}$. In the theory of phase transitions this corresponds to a first-order phase transition, and in the theory of turbulence, using Landau's terminology,[6] it corresponds to a hard regime of excitation. In the present case the transition through the critical velocity is accompanied by a discontinuity in the soliton amplitude. The magnitude of the jump is determined by the next higher-order terms in the expansion of the Hamiltonian. Just as for first-order phase transitions, in the situation of the general position universality of soliton behavior is no longer assured. However, when the hardness of this transition is small, attention can be confined to the next approximation in the expansion of the Hamiltonian and all other terms can be neglected. In phase transitions this corresponds to a first-order phase transition close to a second-order transition, which occurs, for example, near a tricritical point. As shown in Ref. 7, this situation arises for one-dimensional internal-

wave solitons propagating along a density jump in liquids. According to Ref. 7 the matrix element T in this case vanishes for density ratio $\rho_1/\rho_2=(21-8\sqrt{5})/11$. Near this point the authors of Ref. 7 were able to perform a complete bifurcation analysis for solitons using the method of normal forms.

In the present paper, the Hamiltonian description (concerning this subject, see the recent review in Ref. 8, as well as Ref. 5) is used to obtain a closed description of solitons near the critical velocity, and the stability of solitons with respect to modulation disturbances is studied on the basis of a nonstationary generalized nonlinear Schrödinger equation. It should be noted that in contrast to the method of normal forms, which is extensively used in Refs. 2, 3, 7, and 9 to study bifurcations of solitons, the Hamiltonian approach is fundamental for investigating soliton stability. In the method of normal forms, the introduction of envelopes is not unique, as a result of which, after averaging, the Hamiltonian equations of motion lose their original Hamiltonian structure.

As will be shown below, the expansion of the Hamiltonian for solitons with hard bifurcation near threshold differs substantially from the expansion for the soft regime. Terms containing the so-called Lifshits invariant[10] which plays an important role for first-order phase transitions, are present in this expansion. In the context of solitons, the Lifshits invariant gives rise to a nonlinear coordinate dependence of the soliton phase (in optics it is called a chirp).

It is shown below that only one-dimensional solitons are stable in the soft transition regime. Two- and three-dimensional solitons are subject to modulation instability. For hard transitions this same instability also occurs for two- and three-dimensional solitons.

Using integral estimates of the Sobolev type in their multiplicative variant (Galiardo–Neirenberg inequalities) it is shown that only one-dimensional solitons with weak attraction are stable in the Lyapunov sense. For these solitons, for a fixed number of particles (intensity), the Hamiltonian is bounded from below and, correspondingly, the solitons themselves realize its minimum, which by virtue of Lyapunov's theorem implies their stability. It should be noted that the existence of stable localized structures — solitons — due to a relatively weak four-wave interaction against the background of strong attraction ($\sim|\psi|^6$), which leads to collapse (see, for example, Ref. 11), is related to the phenomenon of weak localization.[12]

Solitons with a finite discontinuity in amplitude at the transition point are strictly unstable with respect to small perturbation, where there is no contribution from the Lifshits invariant. The stability criterion for these solutions has the form of the Vakhitov–Kolokolov criterion[13] for the NLSE:

$$\partial N_s/\partial\lambda^2>0, \tag{1.2}$$

where N_s is the number of particles in the soliton solution and $\varepsilon=-\lambda^2<0$ is the energy of the bound state — a soliton. For a positive derivative, i.e., when the addition of a single particle decreases the energy ε, the soliton is stable. In the opposite case, where level expulsion occurs as N increases, the soliton is unstable. When the contribution from the Lifshits invariant is nonzero, the instability criterion (1.2) for solitons with weak repulsion is only a necessary criterion. However, solutions of this type are unstable with respect to finite perturbations.

The criterion (1.2) makes solitons with weak attraction stable in the entire range of λ^2. This is in complete agreement with the Lyapunov stability analysis.

This paper is organized as follows. The nonstationary generalized nonlinear Schrodinger equation is derived in Sec. 2. Here the effect of the spatial dimension on soliton stability in the soft excitation regime is discussed. Specifically, it is shown that stable solitons can exist only in the one-dimensional case. For this reason, in what follows we shall focus on only one-dimensional solitons. In Sec. 3 the exact analytic expressions for a one-dimensional soliton solution with an amplitude jump at the transition point are found. The next section is devoted to the Lyapunov stability of solitons. Using exact integral estimates, it is established that the Hamiltonian has a lower limit only for solitons with weak attraction ($T\omega''<0$). A soliton realizes the minimum of H and is Lyapunov-stable, i.e., not only with respect to small disturbances but also with respect to finite disturbances. In this case this criterion can, and must, be viewed as an energy principle. Specifically, hence it is easy to see that merging of solitons is energetically favorable (note that merging is impossible for the integrable NLSE). As a rule, this process will be accompanied by radiation (see Ref. 14), since all the conservation laws can be satisfied only for very exceptional dispersion dependences of the energy of the solitons on the momentum and number of particles.

The question of linear stability is investigated in Sec. 5. The Hamiltonian approach is also very helpful here — it greatly simplifies the entire derivation of the analog of the Vakhitov–Kolokolov criterion.

The final section is devoted to an application to optical solitons in optical fibers. Specifically, it is shown that the striction mechanism can decrease the four-wave matrix element, due to the Kerr effect, and in principle can change its sign.

2. BASIC EQUATIONS

Let us consider a nonlinear medium where waves can propagate. We shall assume that the medium is purely conservative, and its nonlinear oscillations can be described by the Hamiltonian

$$H=\int \omega_k|a_k|^2 d\mathbf{k}+H_{\rm int}, \tag{2.1}$$

where ω_k is the dispersion law of low-amplitude waves, a_k are the amplitudes of the waves, and the Hamiltonian $H_{\rm int}$ describes the nonlinear interaction of the waves.

The equations of motion of the medium can be written in terms of the amplitudes a_k in the standard manner

$$\frac{\partial a_k}{\partial t}+i\omega_k a_k=-i\frac{\delta H_{\rm int}}{\delta a_k^*}, \tag{2.2}$$

so that in the absence of an interaction the system consists of a set of noninteracting oscillators (waves):

$$a_k(t)=a_k(0)e^{-i\omega_k t}.$$

The equation (2.2) describes dynamics in the k representation. The inverse Fourier transform

$$\psi(\mathbf{x},t)=\frac{1}{(2\pi)^{d/2}}\int a_k(t)e^{i\mathbf{k}\cdot\mathbf{r}}d\mathbf{k} \tag{2.3}$$

must be performed to return to the x representation.

Ordinarily, the function $\psi(\mathbf{x},t)$ is related with the characteristics of the medium (fluctuations of the density and velocity of the medium, electric and magnetic fields, and so on) by linear transformation (see, for example, Ref. 8). It is important that if $\psi(\mathbf{x},t)$ is a periodic function of the coordinates, then its spectrum $a_k(t)$ consists of a set of δ functions. For localized distributions $\psi(\mathbf{x},t)\to 0$ as $|\mathbf{x}|\to\infty$, the Fourier amplitude $a_k(t)$, being a localized function of $\mathbf{k}$, does not contain δ-function singularities.

Let us now consider the solution of Eq. (2.2) in the form of a soliton propagating with constant velocity $\mathbf{v}$:

$$\psi(\mathbf{x},t)=\psi(\mathbf{x}-\mathbf{v}t).$$

In this case the entire dependence of a_k on the time t is contained in the oscillating exponent:

$$a_k(t)=c_k e^{-i\mathbf{k}\cdot\mathbf{v}t},$$

where by virtue of Eq. (2.2) the amplitude a_k will satisfy the equation

$$(\omega_k-\mathbf{k}\cdot\mathbf{v})_k=-\frac{\partial H}{\partial c_k^*}\equiv f_k. \tag{2.4}$$

The difference $\omega_k-\mathbf{k}\cdot\mathbf{v}$ appearing in this equation will be positive for all $\mathbf{k}$ if the soliton velocity is less than the minimum phase velocity

$$|\mathbf{v}|<\min(\omega_k/k). \tag{2.5}$$

Conversely, the difference will be negative for all k if the soliton velocity is greater than the maximum phase velocity:

$$|\mathbf{v}|>\max(\omega_k/k). \tag{2.6}$$

We shall show that a soliton solution is possible if the condition (2.5) [or (2.6)] is satisfied. Let us assume the opposite to be true — let the condition (2.5) be satisfied, i.e., the equation

$$\omega_k=\mathbf{k}\cdot\mathbf{v} \tag{2.7}$$

possesses a solution. For simplicity, we shall assume that it is unique: $\mathbf{k}=\mathbf{k}_0$. Then, since $x\delta(x)=0$, the homogeneous linear equation

$$(\omega_k-\mathbf{k}\cdot\mathbf{v})C_k=0$$

possesses a nontrivial solution in the form of a monochromatic wave

$$C_k=A\,\delta(\mathbf{k}-\mathbf{k}_0).$$

In this case Eq. (2.4) admits (in accordance with the Fredholm alternative) the representation

$$c_k=A\,\delta(\mathbf{k}-\mathbf{k}_0)+\frac{f_k-f_{k_0}}{\omega_k-\mathbf{k}\cdot\mathbf{v}}. \tag{2.8}$$

This equation, in contrast to Eq. (2.4), contains a free parameter — the complex amplitude A. It can be solved, for example, by iterations, taking $A\,\delta(\mathbf{k}-\mathbf{k}_0)$ as the zeroth term. It is important that as a result of iteration harmonics which are multiples of $\mathbf{k}=\mathbf{k}_0$ will appear in the solution because of the nonlinearity. The solution so obtained will consist of a collection of δ functions. Correspondingly, in the x representation the solution will be a periodic function of the coordinates, i.e., it will be nonlocalized. Hence follows the first selection rule for solitons: The difference $\omega_k-\mathbf{k}\cdot\mathbf{v}$ must be sign-definite, which is equivalent to the requirements (2.5) or (2.6) of the absence of Cherenkov radiation.

In this entire scheme, however, there is an important exception. Having represented Eq. (2.4) in the form (2.8), we have actually assumed that the singularity in the expression

$$\frac{f_k}{\omega_k-\mathbf{k}\cdot\mathbf{v}} \tag{2.9}$$

is nonremovable. This may not be the case — the singularity in the denominator in Eq. (2.9) could cancel with the numerator, i.e., it could be removable.[5] For example, this happens for the classic soliton of the KdV equation, for equations which are generalizations of the KdV equation,[15] for a combination of the one-dimensional NLSE and the MKdV equations[5,16] generated by the same Zakharov–Shabat operator,[17] and so on. In all of these cases cancellation occurs as a result of the k dependence of the matrix elements. However, even in this case, after the resonance (2.7) is removed, the selection rule for solitons remains the same — the part remaining in the denominator must be sign-definite.

In what follows the singularities in Eq. (2.9) will be assumed nonremovable in the forbidden region, and the behavior of the soliton solution as the soliton velocity approaches the critical value will be studied. For definiteness, it will be assumed that the plane $\omega=\mathbf{k}\cdot\mathbf{v}$ is tangent to the dispersion surface $\omega=\omega_k$ from below, i.e., the criterion (2.5) holds. Let tangency occur at the point $\mathbf{k}=\mathbf{k}_0$. Then, instead of Eq. (2.8), in the allowed region

$$c_k=\frac{f_k}{\omega_k-\mathbf{k}\cdot\mathbf{v}}.$$

As the velocity v approaches the critical value $v_{\rm cr}$, the denominator in this expression becomes small near the point of tangency, so that c_k possesses a sharp peak at this point

$$c_k=\left[\frac{1}{2}\omega_{\alpha\beta}\kappa_\alpha\kappa_\beta+k_0(v_{\rm cr}-v)\right]^{-1}f_k. \tag{2.10}$$

Here $\omega_{\alpha\beta}=\partial^2\omega/\partial k_\alpha\partial k_\beta$ is a symmetric, positive-definite, tensor of the second derivatives, evaluated at $\mathbf{k}=\mathbf{k}_0$, and $\kappa=\mathbf{k}-\mathbf{k}_0$.

It is evident from Eq. (2.10) that as v approaches the critical velocity the width of the peak narrows as $\propto\sqrt{v_{\rm cr}-v}$, and the distribution corresponding to the main peak $\mathbf{k}=\mathbf{k}_0$ approaches a monochromatic wave. On account of the nonlinearity the spectrum contains harmonics which are multiples of $\mathbf{k}=\mathbf{k}_0$. If it is assumed that the amplitude of the soliton vanishes smoothly as $v\to v_{\rm cr}$ (which would corre-

spond to a second-order phase transition), then the solution $\psi(\mathbf{x})$ (or, equivalently, c_k) can be sought as an expansion in terms of harmonics:

$$\psi(\mathbf{x}')=\sum_{h=-\infty}^{\infty}\psi_n(\mathbf{X})e^{ih\mathbf{k}_0\cdot\mathbf{x}'},\quad \mathbf{x}'=\mathbf{x}-\mathbf{v}t. \tag{2.11}$$

Here the small parameter

$$\lambda=\sqrt{1-v/v_{\rm cr}} \tag{2.12}$$

and the ''slow'' coordinate $\mathbf{X}=\lambda\mathbf{x}'$ are formally introduced, so that $\psi_n(\mathbf{X})$ is the amplitude of the envelope of nth harmonic. The assumption that the soliton amplitude vanishes continuously at $v=v_{\rm cr}$ means that the leading term of the series in Eq. (2.11) corresponds to the first harmonic, and all other harmonics are small in the parameter λ. This is the condition under which the nonlinear Schrödinger equation is derived (see, for example, Refs. 5, 18, and 19). In the case at hand, in leading order in λ we arrive at the stationary NLSE (compare with Ref. 5)

$$-k_0v_{\rm cr}\lambda^2\psi_1+\frac{1}{2}\omega_{\alpha\beta}\frac{\partial^2\psi_1}{\partial X_\alpha\partial X_\beta}+B|\psi_1|^2\psi_1=0, \tag{2.13}$$

where B is related with the matrix element $\tilde{T}_{k_0k_1k_2k_3}$ of the four-wave interaction as

$$B=-(2\pi)^d\tilde{T}_{k_0k_0k_0k_0}. \tag{2.14}$$

In this approximation the leading term in the interaction Hamiltonian has the form

$$H_{\rm int}=\frac{\tilde{T}_{k_0k_0k_0k_0}}{2}\int c_k^*c_{k_1}^*c_{k_2}c_{k_3}\delta_{\mathbf{k}+\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}_3}d\mathbf{k}d\mathbf{k}_1d\mathbf{k}_2d\mathbf{k}_3$$

$$=-\frac{B}{2}\int|\psi_1|^4d\mathbf{x}, \tag{2.15}$$

and the overtilda signifies that renormalization of the vertex by the three-wave interaction is taken into account in the matrix element — in the case at hand the interaction with the zeroth and second harmonics. As we have already noted, $\omega_{\alpha\beta}$ in Eq. (2.13) is a symmetric positive-definite tensor. For this reason, performing a rotation to its principal axes and carrying out the corresponding extensions along each axis, Eq. (2.15) can be put into the standard form

$$-\lambda^2\psi+\Delta\psi-\mu|\psi|^2\psi=0, \tag{2.16}$$

where $\mu=\mathrm{sign}(\tilde{T}\omega_{\alpha\alpha})$.

Hence it follows, in the first place, that solitons (focusing nonlinearity) are possible only if μ is negative (the product $\tilde{T}\omega_{\alpha\alpha}$ is negative) and, in the second place, that the amplitude of the solitons is proportional to

$$\lambda=\sqrt{1-v/v_{\rm cr}},$$

i.e., the amplitude vanishes according to a square-root law, the size of the soliton increasing as the velocity approaches the critical value inversely as this factor.

The number of particles (or intensity) in a soliton solution as a function of λ has the form

$$N=\int|\psi|^2d\mathbf{x}=\lambda^{2-d}\int|g(\xi)|^2d\xi, \tag{2.17}$$

where d is the dimension of the space and $g(\xi)$ satisfies the equation

$$-g+\Delta g+|g|^2g=0.$$

In the one-dimensional case $g=\sqrt{2}\,\mathrm{sech}\,\xi$ and, correspondingly, $N=4\lambda$. In the two-dimensional case N is independent of λ on the entire family of solitons, while in the three-dimensional case N decreases with increasing λ. The dependence of N on λ^2 is determining from the standpoint of soliton stability. It is obvious that the most dangerous disturbances will be those having wave numbers close to $\mathbf{k}=\mathbf{k}_0$ moving together with the soliton, i.e., modulation-type disturbances. To include the time dependence in the averaged equations the amplitudes ψ_n in the expansion (2.11) must be assumed to depend not only on the ''slow'' coordinate X but also on the slow time $T=\lambda^2t$. Then a multiscale expansion gives the nonstationary analog of the NLSE

$$i\psi_t-\lambda^2\psi+\Delta\psi-\mu|\psi|^2\psi=0 \tag{2.18}$$

instead of the stationary NLSE (2.16). The soliton stability problem for this equation has been well studied (see, for example, Refs. 5 and 20). We shall recall the basic assumptions.

The equation (2.18) as an equation for envelopes inherits the canonical Hamiltonian form (2.2)

$$i\frac{\partial\psi}{\partial t}=\frac{\delta\tilde{H}}{\delta\psi^*}, \tag{2.19}$$

where the Hamiltonian

$$H=\lambda^2N+\int(|\nabla\psi|^2-|\psi|^4)d\mathbf{r}\quad(\mu=-1) \tag{2.20}$$

arises as a result of averaging the initial Hamiltonian. The equation (2.18) preserves, besides H, the total number N of particles (adiabatic invariant), so that solitons are stationary points of the energy functional $E=H-\lambda^2N$ for a fixed number of particles:

$$\delta(E+\lambda^2N)=0.$$

Hence it can be shown following Ref. 20 that a soliton realizes the minimum energy only in the one-dimensional case, while for $d\geqslant2$ solitons are a hyperbolic point. This means that solitons are stable only in the one-dimensional case, while in the two-dimensional (critical) and three-dimensional cases solitons are unstable and can be considered as separatrix solutions separating collapsing solutions from diffracting solutions.[21]

This is probably the simplest method for explaining the well-known empirical fact that solitons, as a rule, exist only in one-dimensional systems. For multidimensional systems stable solitons are rare and can appear as a result of only topological constraints or a mechanism that removes Cherenkov singularities (which is discussed in the present paper). The latter, as is easy to understood, is due to the existence of a certain class of symmetry.

In the present work we shall focus our attention mainly on solitons arising in the hard excitation regime, which is observed when the coefficient μ in Eq. (2.16) is positive. In this case Eq. (2.16) no longer possesses stationary localized (vanishing at infinity) solutions. In order for them to exist it is necessary to take account of the next higher-order terms in the expansion of the Hamiltonian relative to the parameter $\Delta k/k_0$, where Δk is the width of the main peak. If the jump in the soliton amplitude at $\nu=\nu_{cr}$ is large (of the order of 1), then the entire series must used and it is no longer possible to count on a systematic theory based on an expansion of the Hamiltonian. Only if the matrix element $\tilde{T}_{k_0k_0k_0k_0}=\tilde{T}_0$ is small, i.e., there is an additional smallness compared with the supercriticality (and as a result the jump is also small), it is sufficient in this case to retain several of the next terms in the expansion. We shall study only one-dimensional solitons, since, as we have seen above, multidimensional solitons are unstable for a soft transition. This same tendency also remains in hard regimes. It is easy to see that in this situation two terms make the main contribution to the interaction Hamiltonian. The first term is a correction to the local four-wave Hamiltonian $-B/2\int|\psi|^4dx$. It arises because a term linear in $\kappa_i=k_i-k_0$ is retained in the expansion of the matrix element $\tilde{T}_{k_1k_2k_3k_4}$

$$\tilde{T}_{k_1k_2k_3k_4}=\tilde{T}_0+\frac{\partial\tilde{T}}{\partial k_1}(\kappa_1+\kappa_2)+\frac{\partial\tilde{T}^*}{\partial k_1}(\kappa_3+\kappa_4)$$
$$=\tilde{T}_0+\mathrm{Re}\left(\frac{\partial\tilde{T}}{\partial k_1}\right)(\kappa_1+\kappa_2+\kappa_3+\kappa_4).$$

Here

$$\frac{\partial\tilde{T}}{\partial k_1}\equiv\left.\frac{\partial\tilde{T}_{k_1k_2k_3k_4}}{\partial k_1}\right|_{k_i=k_0}.$$

As a result, the Hamiltonian of the four-wave interaction in the envelope approximation can be written in the x representation as

$$H^{(4)}=\pi\int\left\{\tilde{T}_0|\psi|^4+2i\,\mathrm{Re}\left(\frac{\partial\tilde{T}}{\partial k_1}\right)(\psi_x^*\psi-\psi_x\psi^*)|\psi|^2\right\}dx. \tag{2.21}$$

The expression $i(\psi_x^*\psi-\psi_x\psi^*)$ in this integral is well known in the theory of phase transitions (see Ref. 10) — it is the so-called Lifshits invariant.

The second term — local in ψ — is a six-wave interaction

$$H^{(6)}=-C\int|\psi|^6dx. \tag{2.22}$$

As will be evident from what follows, the interaction constant C can be both negative and positive — the combination of both contributions (2.21) and (2.22) will be important.

The total Hamiltonian in dimensionless variables will depend on three constants μ, β, and C

$$H=\lambda^2N+\int\left[|\psi_x|^2+\frac{\mu}{2}|\psi|^4+i\beta(\psi_x^*\psi-\psi_x\psi^*)|\psi|^2-C|\psi|^6\right]dx. \tag{2.23}$$

The constant μ is assumed to be small, and the constants β and C do not contain any additional smallness.

The equations of motion for ψ that correspond to this Hamiltonian can be written according to Eq. (2.19) as

$$i\psi_t-\lambda^2\psi+\psi_{xx}-\mu|\psi|^2\psi+3C|\psi|^4\psi+4i\beta|\psi|^2\psi_x=0. \tag{2.24}$$

Besides the energy $E=H-\lambda^2N$ and the number of particles, this equation also conserves the total momentum

$$P=\frac{i}{2}\int(\psi_x^*\psi-\psi_x\psi^*)dx.$$

3. SOLITON SOLUTIONS

The stationary (independent of t) soliton solutions of Eq. (2.24) will be determined from the following ordinary differential equation:[1)]

$$-\lambda^2\psi+\psi_{xx}-\mu|\psi|^2\psi+3C|\psi|^4\psi+4i\beta|\psi|^2\psi_x=0. \tag{3.1}$$

This equation can be integrated easily, if the amplitude $r=|\psi|$ and phase $\varphi=\arg\psi$ are introduced instead of ψ: $\psi=re^{i\varphi}$. Next, substituting ψ in Eq. (3.1) and separating real and imaginary parts we obtain for the imaginary part the equation

$$\varphi_x=-\beta r^2. \tag{3.2}$$

After eliminating the phase, the equation for r reduces to Newton's equation

$$2r_{xx}=-\partial U/\partial r \tag{3.3}$$

with the potential

$$U=-\lambda^2r^2-\frac{\mu}{2}r^4+C_1r^6,$$

where the interaction constant C is renormalized as $C_1=C+\beta^2$. Then Eqs. (3.2) and (3.3) can be integrated using the energy integral:

$$r^2=\frac{4\lambda^2}{\sqrt{16\lambda^2C_1+\mu^2}\cosh(2\lambda x)-\mu}, \tag{3.4}$$

$$\varphi=-\frac{\beta^2}{\sqrt{C_1}}\tan^{-1}\left[\frac{\sqrt{16\lambda^2C_1+\mu^2}\,e^{2\lambda x}-\mu}{4\lambda\sqrt{C_1}}\right]. \tag{3.5}$$

This soliton-type solution exists only if $C_1>0$.[2)] It is interesting to note that the renormalization of the interaction constant C is due to the β term in the Hamiltonian. This can be seen directly from Eq. (2.23), rewriting H in terms of the amplitude and phase as

$$H=\lambda^2N+\int\left[r_x^2+r^2(\varphi_x+\beta r^2)^2+\frac{\mu}{2}r^4-(C+\beta^2)r^6\right]dx. \tag{3.6}$$

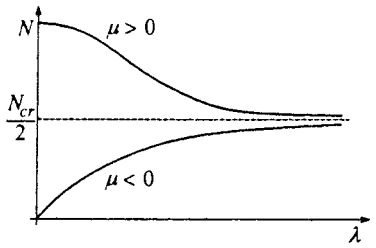

FIG. 1. N versus λ for soliton solutions: The top curve corresponds to weak repulsion and the bottom curve corresponds to weak attraction.

It is also easy to see that the soliton solution (3.4) is a stationary point of H. Indeed, the variation of H with respect to φ leads to Eq. (3.2), and Newton's equation (3.3) arises as a result of varying H with respect to r.

The solutions (3.4) and (3.5) with $\lambda=0$ and $\mu>0$ degenerate into a soliton with power-law decay[9]

$$r_{\lim}^2=\frac{2\mu}{\mu^2x^2+4C_1},$$

$$\varphi_{\lim}=-\frac{\beta}{\sqrt{C_1}}\tan^{-1}\frac{\mu x}{2\sqrt{\beta^2+C}}. \tag{3.7}$$

Thus, as the velocity passes through v_{cr}, the soliton undergoes a jump. The amplitude of the soliton has its maximum value at the jump

$$A^2=\mu/2C_1.$$

This value of the jump is easily found from Eq. (3.6).

As λ increases, the amplitude of the soliton grows according to a square-root law, and the soliton size decreases as λ^{-1}.

An important feature of the solution (3.4) is the existence of a nonlinear coordinate dependence (called a chirp in optics) of the phase φ. The maximum change in phase (from $-\infty$ to $+\infty$ in x)

$$\Delta\varphi=-\frac{\beta\pi}{\sqrt{\beta^2+C}}$$

is reached at the jump for $v=v_{cr}$. It can be both greater and less than π, depending on the sign of the constant C.

The solution (3.4) can also be used for negative but small values of μ. In this case, as should be, the soliton solution softly splits off zero at the point $v=v_{cr}$. Its amplitude then grows for large λ exactly in the same manner as for $\mu>0$.

The integral characteristics of both solutions (with $\mu>0$ and $\mu<0$) are different. Thus, the total number of particles in the soliton solution for $\mu>0$,

$$N=\frac{2}{\sqrt{C_1}}\left[\frac{\pi}{2}-\tan^{-1}\left\{\frac{\sqrt{16\lambda^2C_1+\mu^2}-\mu}{4\lambda\sqrt{C_1}}\right\}\right],$$

reaches its maximum value $N_{cr}=\pi/\sqrt{C_1}$ at $\lambda=0$ and decreases smoothly to $N_{cr}/2$ as $\lambda\to\infty$ (see Fig. 1). For negative μ the number of particles N for small λ increases as λ and then asymptotically approaches $N=N_{cr}/2$ from below. It is important that the derivatives $\partial N/\partial\lambda$ has different signs: For solitons with a jump this derivative is negative, while for solitons with $\mu<0$ it is positive.

4. LYAPUNOV STABILITY OF SOLITONS

As noted above, both types of solitons (with $\mu>0$ and $\mu<0$) are stationary points of energy E with a fixed number of particles,

$$\delta(E+\lambda^2N)=\delta H=0, \tag{4.1}$$

where the energy in accordance with Eq. (3.6) is given by the expression

$$E=\int\left[r_x^2+\frac{\mu}{2}r^4-(C+\beta^2)r^6+r^2(\varphi_x+\beta r^2)^2\right]dx. \tag{4.2}$$

As is well known (see, for example, Ref. 20), in the case of Hamiltonian systems a stationary point will be Lyapunov-stable if it yields a minimum or maximum of any integral of motion, for example, the energy.

In the case at hand, if we find conditions for which the energy will have a lower limit with a fixed number of particles (it is obvious that E unbounded from above), the stationary point corresponding to the minimum of E will be stable. Since the solution of the variational problem (4.1) is unique (up to a constant phase factor) for fixed λ^2, which is equivalent to fixing N, the soliton solution (3.4) will by Lyapunov-stable in this case.

We shall now consider scale transformations of the soliton solution

$$\psi_s(x)\to\frac{1}{\sqrt{l}}\psi_s\left(\frac{x}{l}\right),$$

that preserve the number of particles. Under such a transformation the energy E will be a function of the scaling parameter l:

$$E=(I_1-C_1I_2)\frac{1}{l^2}+\frac{\mu I_3}{2l},$$

where

$$I_1=\int r_x^2dx,\quad I_2=\int r^6dx,\quad I_3=\int r^4dx.$$

(We note that the last integral in the expression for the energy (4.2) is identically zero at the soliton solution.) By virtue of Eq. (4.1)

$$\left.\frac{\partial E}{\partial l}\right|_{l=1}=0\quad\text{or}\quad I_1-C_1I_2=-\frac{\mu}{4}I_3.$$

Hence it follows, in the first place, that the soliton energy likewise depends strongly on the constant μ:

$$E=\frac{\mu}{4}I_3\quad(I_3>0).$$

This quantity is positive for solitons with weak repulsion ($\mu>0$) and negative for solitons with weak attraction ($\mu<0$). In the second place, for $\mu>0$ the energy E as a function of the scaling parameter l,

$$E=-\frac{\mu}{4}I_3\left(\frac{1}{l^2}-\frac{2}{l}\right),$$

is unbounded as $l\to 0$, and for weak attraction it possesses a minimum corresponding to the soliton solution.

We shall now show that the energy E for $\mu<0$ has a lower limit for all deformations that leave N unchanged.

Let us consider the integral $\int r^6 dx = I_2$. This integral can be estimated using the Sobolev–Galiardo–Nierenberg inequality in terms of the integral $I_1=\int r_x^2 dx$ and the number of the particles N:

$$\int r^6 dx \leqslant MN^2 \int r_x^2 dx. \tag{4.3}$$

This inequality can be improved by seeking the smallest value of the constant M. For this, following Ref. 5, we consider the functional

$$M[\psi]=\frac{I_2}{I_1 N^2}.$$

Its minimum value determines the best constant. To find it all stationary points need to be considered and the point giving a minimum value of the functional $M[\psi]$ should be found. It is easy to see that this variational problem, $\delta M=0$, is equivalent to finding the soliton solutions for the ψ^6 model:

$$-\psi+\psi_{xx}+3\psi^5=0.$$

This equation has a unique solution $\psi=1/\sqrt{\cosh 2x}$, whence the best constant is simply found as

$$M_{\text{best}}=(2/\pi)^2.$$

As a result, the inequality (4.3) can be rewritten as

$$\int r^6 dx \leqslant \left(\frac{N}{N_1}\right)^2 \int r_x^2 dx, \tag{4.4}$$

where $N_1=\pi/2$.

Next, substituting this inequality into Eq. (4.2) we obtain for the energy E the estimate

$$E\geqslant\left[1-C_1\left(\frac{N}{N_1}\right)^2\right]\int r_x^2 dx + \int r^2(\varphi_x+\beta r^2)^2 dx+\frac{\mu}{2}\int r^4 dx. \tag{4.5}$$

Hence for $\mu>0$ follows that the energy E is bounded by zero if the coefficient in front of the integral I_1 is positive. This gives an upper bound on the number of particles

$$N<\frac{\pi}{2\sqrt{C+\beta^2}}=\frac{N_{\text{cr}}}{2}. \tag{4.6}$$

We recall that $N_{\text{cr}}/2$ is the lower bound for the soliton family (3.4) with $\mu>0$. Therefore for such solitons it is impossible to draw any conclusion about their stability. However, for soliton solutions with $\mu<0$ the inequality (4.6) holds and, as will be seen from the estimates made below, it is sufficient to prove that such solutions are stable.

Thus, let $\mu<0$ in Eq. (4.5). According to Ref. 5, we have $\int r^4 dx$

$$\int r^4 dx \leqslant \frac{1}{\sqrt{3}}\left(\int r_x^2 dx\right)^{1/2} N^{3/2}.$$

Next, substituting this estimate into Eq. (4.5) we obtain

$$E\geqslant\left[1-C_1\left(\frac{N}{N_1}\right)^2\right]I_1-\frac{|\mu|}{2\sqrt{3}}N^{3/2}I_1^{1/2}+\int r^2(\varphi_x+\beta r^2)^2 dx$$

$$\geqslant-\frac{|\mu|^2N^3}{8\sqrt{3}}\left[1-C_1\left(\frac{N}{N_1}\right)^2\right].$$

The latter inequality holds only if the criterion (4.6) is satisfied. This means that the energy E has a lower limit if

$$N<N_{\text{cr}}/2,$$

which is compatible with the entire region of existence of solitons with $\mu<0$. It should be noted that for $\mu=0$ the NLSE (2.23) is, as is said, a critical equation. For this nonlinearity ($\sim|\psi|^6$ in H) collapse is possible if the energy E is negative. If $N<N_{\text{cr}}/2$, dispersion completely smears out the solution. However, a small negative correction to the Hamiltonian fundamentally changes the situation. A relatively weak four-wave interaction against the background of strong attraction ($\sim|\psi|^6$), leading to collapse (see, for example, Ref. 11), is responsible for the existence of stable coupled stationary states — solitons. Weak localization appears.[12]

5. LINEAR STABILITY OF SOLITONS

The preceding analysis has answered the question of stability only for solitons with weak attraction. From this answer it is impossible to draw any conclusion about the stability of solitons with weak repulsion ($\mu>0$). In this section we shall consider this question, investigating the linear problem of stability.

We shall seek a solution of Eq. (2.23) in the form

$$\psi=(r+a)e^{i(\phi+\alpha)}\approx(r+a+ir\alpha)e^{i\phi}, \tag{5.1}$$

where r and ϕ are the soliton solution (3.4) and (3.5), and a and α are small deviations of the amplitude and phase of the soliton, respectively.

Linearizing Eq. (2.3) it is easily found that the dynamics of the perturbations a and α is determined by the Hamiltonian equations

$$2r\frac{\partial a}{\partial t}=\frac{\delta\tilde{H}}{\delta\alpha},\quad 2r\frac{\partial\alpha}{\partial t}=-\frac{\delta\tilde{H}}{\delta a}. \tag{5.2}$$

Here $\tilde{H}=\delta^2 H$ is the second variation of the Hamiltonian (3.4)

$$\tilde{H}=\langle a|L|a\rangle+\int r^2(\alpha_x+2\beta r a)^2 dx, \tag{5.3}$$

where the (Schrödinger) operator L is given by the expression

$$L=-\frac{\partial^2}{\partial x^2}+\lambda^2+2\mu r^2-15C_1r^4. \tag{5.4}$$

If the quadratic form δ^2H is sign-definite, then the soliton solution will be stable. We note that the second term in $\tilde{H}$ (5.3) is positive. Then the positiveness of the entire quadratic form is determined by the average value of the operator L:

$$\langle a|L|a\rangle.$$

In this expression the average is taken not with respect to arbitrary states $|a\rangle$ but only with respect to those states that are orthogonal to $|r\rangle$:

$$\langle r|a\rangle=0. \tag{5.5}$$

This orthogonality condition is a consequence of the conservation of the number of particles and is one of the solvability conditions for the linear system (5.2). In this case, finding the stability criterion for solitons (2.23) is identical to the Vakhitov–Kolokolov derivation[13] (see also Ref. 20) for the NLSE without the β term. To determine the sign it is necessary to find the spectrum of the operator L

$$L\xi=E\xi+Cr. \tag{5.6}$$

Here C is an indefinite Lagrangian multiplier, which is determined from the solvability condition (5.5).

Next, expanding ξ in terms of the eigenfunctions of the operator L ($L\psi_n=E_n\psi_n$) and using the solvability condition (5.5), it is easy to obtain the dispersion relation

$$f(E)\equiv\sum_n{}' \frac{\langle r|\psi_n\rangle\langle\psi_n|r\rangle}{E_n-E}=0. \tag{5.7}$$

Here the prime on the sum means that the term with $E_1=0$ is omitted from the sum because the (shear) eigenmode $\psi_1=r_x$ ($Lr_x=0$) is orthogonal to r. The function r_x possesses a single zero, so that below $E_1=0$ the operator L possesses only one level corresponding to the ground state.

Let us now consider the function $f(E)$ between the ground level $E_0<0$ and the first positive level E_2. In the entire range this function grows monotonically from $-\infty$ at E_0 to $+\infty$ at E_2. Therefore if $f(0)>0$, the spectrum of the operator L has a negative eigenvalue and therefore the average value of the operator L can be negative. For $f(0)<0$ the average of L is always positive.

To find the sign of $f(0)$, it is easy to see that

$$f(0)=\sum_n{}' \frac{\langle r|\psi_n\rangle\langle\psi_n|r\rangle}{E_n}\equiv\langle r|L^{-1}|r\rangle.$$

Next, differentiating Eq. (3.3) with respect to λ^2 we obtain

$$L\frac{\partial r}{\partial\lambda^2}=-r \quad \text{or} \quad \frac{\partial r}{\partial\lambda^2}=-L^{-1}r.$$

Substituting this relation gives[13]

$$f(0)=-\frac{1}{2}\frac{\partial N}{\partial\lambda^2}. \tag{5.8}$$

Therefore if $\partial N/\partial\lambda^2>0$, the quadratic form will be positive-definite. This situation occurs for solitons with weak attraction, and as a result they are stable. This conclusion is in complete agreement with the results of the preceding section.

For solitons with weak repulsion ($\mu>0$) the criterion (5.8) gives a sign-indefiniteness of the quadratic form $\tilde{H}$. This is a necessary condition for instability. This criterion is also sufficiently only in the case $\beta=0$, where the average value of L in Eq. (5.3) can be interpreted as a potential energy, and the integral $\int r^2\alpha_x^2dx$ can be interpreted as the kinetic energy.

Certain arguments can be given in support of the fact that a soliton with weak repulsion is nonetheless unstable for $\beta\neq0$ also. The average value of L can be made negative by taking for a the eigenfunction ξ with $E<0$. For a given value of ξ it is always possible to find a phase α such that the integral

$$\int r^2(\alpha_x+2\beta ra)^2dx$$

vanishes. Thus the Hamiltonian $\tilde{H}$ can be made negative, which can be regarded physically as sufficient for instability. However, strictly speaking, this still requires a definite proof which the present author still cannot provide. An example that refutes this argument is well known. The Hamiltonian $H=-p^2/2-q^2/2$ gives the equation of motion for the usual stable oscillator even though H is negative. However, it can be asserted absolutely that instability will remain for small values of β. Whether or not a threshold with respect to β exists is still unknown, but it is likely. To shed light on this question we turn to the linear equations (5.2), rewriting them in the new variables

$$p=ra \quad \text{and} \quad g_x=\alpha_x+2\beta ra.$$

In these variables the equations of motion have the form

$$\frac{\partial g}{\partial t}+4\beta r^2\frac{\partial g}{\partial x}=-\frac{1}{r}L\frac{1}{r}p, \tag{5.9}$$

$$\frac{\partial p}{\partial t}=-\frac{\partial}{\partial x}r^2\frac{\partial}{\partial x}g. \tag{5.10}$$

For $\beta=0$ this reduces to an equation for the perturbation amplitude a

$$\frac{\partial^2}{\partial t^2}a=-L_0La \tag{5.11}$$

with

$$L_0=-\frac{1}{r}\frac{\partial}{\partial x}r^2\frac{\partial}{\partial x}\left(\frac{1}{r}\right),$$

whence the criterion (5.8) was first derived (compare with Refs. 13 and 20). It is important that the operator L_0 in Eq. (5.11) is nonnegative (since $L_0r=0$ and the function r itself has no zeros). Therefore since the operator L has a negative eigenvalue, the right-hand side of Eq. (5.11) gives for a soliton with weak repulsion on the whole a negative ''eigenvalue,'' making the soliton absolutely unstable.

For $\beta \neq 0$ an additional convective-type term appeared in Eq. (5.9). This term can change the character of the instability itself. Whether or not convection can stabilize the instability is still unclear.

But despite the uncertainty with the linear stability, it should be noted that a soliton with weak repulsion is always unstable with respect to finite disturbances. This follows, specifically, from the fact that under scaling transformations

$$\psi_s \to \frac{1}{\zeta^{1/2}} \psi_s\left(\frac{x}{\zeta}\right),$$

leaving the number of particles N unchanged the energy $E=H-\lambda^2 N$ as a function of the scaling parameter ζ with $N>N_{cr}/2$ has no lower bound as $\zeta \to 0$. The latter, as is well known (see, for example, the reviews Refs. 18 and 22), is a criterion for wave collapse.

6. CONCLUDING REMARKS

In conclusion, we would like to discuss the possibility of the occurrence of the hard regime of soliton excitation in optics. Since in three- and two-dimensional media solitons are unstable near the threshold for modulation instability, the only possibility of realizing this regime is an optical fiber.

As is well known (see, for example, Refs. 23 and 24), the Kerr constant is positive for most media, including glass. For this reason, there are two possibilities for changing the character of the interaction (from attraction to repulsion).

The first possibility is due to a decrease in the matrix element itself as a result of the three-wave interaction. In glasses, the symmetry precludes an intrinsic quadratic nonlinearity and the corresponding nonlinear susceptibility tensor $\chi_{ijk} \equiv 0$.[25] Therefore striction — the interaction of light with sound — remains the only process. The equations first obtained in Ref. 26 can be used to assess the role of this mechanism:

$$i\psi_t = \frac{\delta H}{\delta \psi^*}, \quad n_t = \frac{\delta H}{\delta \Phi}, \quad \Phi_t = -\frac{\delta H}{\delta n}, \tag{6.1}$$

where ψ is the (dimensionless) envelope of the electromagnetic field, n are low-frequency density fluctuations, and Φ is the hydrodynamic potential. Here the Hamiltonian H consists of the Hamiltonian of electromagnetic waves

$$H = \int \left(-i v_{gr} \psi^* \psi_x + \frac{\omega''}{2} |\psi_x|^2 - |\psi|^4 \right) dx, \tag{6.2}$$

the Hamiltonian of acoustic oscillations

$$H = \frac{1}{2} \int (c_s^2 n^2 + \Phi_x^2) dx \tag{6.3}$$

and the sound–light interaction Hamiltonian

$$H = \int \gamma n |\psi|^2 dx. \tag{6.4}$$

In these expressions v_{gr} is the group velocity of high-frequency electromagnetic waves, c_s is the velocity of sound, and γ is the interaction constant of HF and LF waves and is proportional to the density derivative $\partial \varepsilon / \partial n_0$ of the permittivity. The Kerr constant is normalized to 1.

The equations of motion, following from this Hamiltonian, for the HF and LF waves are

$$i(\psi_t + v_{gr}\psi_x) + \frac{1}{2}\omega'' \psi_{xx} + |\psi|^2 \psi = \gamma n \psi,$$

$$n_{tt} - c_s^2 n_{xx} = \gamma |\psi|^2_{xx}.$$

In this system the difference between the group velocity of an electromagnetic wave, which is of the order of the velocity of light, and the sound velocity c_s is important: $v_{gr} \gg c_s$. The latter means that an electromagnetic pulse and the density fluctuations induced by it will move mainly with the group velocity. For this reason, we have for n the local intensity dependence

$$n = \frac{\gamma^2}{v_{gr}^2} |\psi|^2.$$

It is important to underscore that the ponderomotive force leads not to the formation of a dip in the density but rather to an increase in density. As a result, the four-wave matrix element is renormalized as

$$T(\equiv -1) \to -1 + \frac{\gamma^2}{v_{gr}^2},$$

decreasing in value.

Therefore the striction mechanism can weaken the four-wave matrix element, due to the Kerr effect, and in principle change its sign.

The second possibility is a transition from the region of anomalous dispersion into the region of normal dispersion, so that the sign of the dispersion ω'' changes. The second variant is most easily implemented experimentally, but the four-wave matrix element T, as a rule, does not change sign, and it does not have a small constant μ. For this reason, both variants require glasses doped with definite additives that increase the striction constant γ. The glasses chosen must have maximum gains for Mandel'shtam–Brillouin scattering. Such optical fibers would be interesting not only from the standpoint of observing the solitons investigated in the present paper, but also for studying one-dimensional wave collapse.

I thank V. E. Zakharov and F. Dias for helpful discussions and E. I. Kats for a number of valuable remarks. This work was supported by the INTAS program (Grant 96-0413), the Russian Fund for Fundamental Research (Grant 97-01-00093), and NATO Grant OUTR.LG 970583.

*)E-mail: kuznetso@itp.ac.ru

[1)]We recall that by construction these solitons move with a constant velocity. The equation (2.24) itself, however, contains a larger class of localized solutions. However, these solutions are all nonstationary — their phase and group velocities are different.

[2)]Here we do not analyze solutions with $C_1 < 0$ and $\mu < 0$.

[1] M. S. Loguet-Higgins, J. Fluid Mech. **200**, 951 (1989).
[2] G. Iooss and K. Kirchgassner, C. R. Acad. Sci. Paris **311**, I, 265 (1991).
[3] J.-M. Vanden-Broeck and F. Dias, J. Fluid Mech. **240**, 549 (1992); F. Dias and G. Iooss, Physica D **65**, 399 (1993).
[4] M. S. Loguet-Higgins, J. Fluid Mech. **252**, 703 (1993).
[5] V. E. Zakharov and E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **113**, 1892 (1998) [JETP **86**, 1035 (1998)].
[6] L. D. Landau, Dokl. Akad. Nauk **44**, 339 (1944); L. D. Landau and E. M. Lifshitz, *The Theory of Elasticity*, 3rd English edition (Pergamon Press, New York, 1986) [Russian original, Nauka, Moscow, 1953].
[7] F. Dias and G. Iooss, Eur. J. Mech. B/Fluids **15**, 367 (1996).
[8] V. E. Zakharov and E. A. Kuznetsov, Usp. Fiz. Nauk **167**, 1 (1997).
[9] G. Iooss, C. R. Acad. Sci. Paris **324**, 993 (1997).
[10] L. D. Landau and E. M. Lifshitz, *Statistical Physis, Part 1* (Pergamon Press, New York) [Russian original, Nauka, Moscow, 1995, p. 521).
[11] E. A. Kuznetsov and S. K. Turitsyn, Phys. Lett. A **112**, 273 (1985).
[12] E. I. Kats, V. V. Lebedev, and A. R. Muratov, Phys. Rep. **228**, 1 (1993).
[13] N. G. Vakhitov and A. A. Kolokolov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **16**, 1020 (1973).
[14] E. A. Kuznetsov, Phys. Lett. A **101**, 314 (1983).
[15] J. Nycander, Chaos **4**, 253 (1994).
[16] E. M. Gromov and V. I. Talanov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **39**, 735 (1996); Zh. Éksp. Teor. Fiz. **110**, 137 (1996) [JETP **83**, 73 (1996)].
[17] V. E. Zakharov and A. B. Shabat, Zh. Éksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[18] V. E. Zakharov, in *Handbook of Plasma Physics, Vol. 2, Basic Plasma Physics II*, edited by A. Galeev and R. Sudan (North–Holland, Amsterdam, 1984), p. 81.
[19] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[20] E. A. Kuznetsov, A. M. Rubenchik, and V. E. Zakharov, Phys. Rep. **142**, 103 (1986).
[21] E. A. Kuznetsov, J. J. Rasmussen, K. Rypdal, and S. K. Turitsyn, Physica D **87**, 273 (1995).
[22] E. A. Kuznetsov, Chaos **6**, 381 (1996).
[23] N. Bloembergen, *Nonlinear Optics*, (Benjamin, Reading, MA 1977).
[24] G. P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, Boston, 1989) [Russian translation, Mir, Moscow, 1996).
[25] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, New York) [Russian original, Nauka, Moscow, 1982].
[26] V. E. Zakharov and A. M. Rubenchik, Prikl. Mekh. Tekh. Fiz. **13**, 669 (1974).

Translated by M. E. Alferieff

Dynamics of large-amplitude solitons

A. V. Slyunyaev*)

Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia

E. N. Pelinovskiĭ†)

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

(Submitted 4 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 318–335 (July 1999)

The interaction and generation of solitons in nonlinear integrable systems which allow the existence of a soliton of limiting amplitude are considered. The integrable system considered is the Gardner equation, which includes the Korteweg–de Vries equation (for quadratic nonlinearity) and the modified Korteweg–de Vries equation (for cubic nonlinearity) as special cases. A two-soliton solution of the Gardner equation is derived, and a criterion, which distinguishes between different scenarios for the interaction of two solitons, is determined. The evolution of an initial pulsed disturbance is considered. It is shown, in particular, that solitons of opposite polarity appear during such evolution on the crest of a limiting soliton.
 [S1063-7761(99)02307-0]

1. INTRODUCTION

The Korteweg–de Vries equation, which was discovered in 1895 for waves in water, is the most popular model for describing nonlinear waves in a weakly dispersive medium. It essentially served as the first testing grounds for proving the particle-like properties of a nonlinear wave field in the form of stable solitons and the integrability of the evolution problem using a method based on the inverse problem of scattering theory, which provided proof of the exclusive role of solitons in the asymptotic representation of a wave field at large times. The Korteweg–de Vries equation is obtained in first-order perturbation theory for a small wave amplitude and weak high-frequency dispersion; therefore, it is applicable to the description of diverse wave motions in the ocean and the atmosphere, in plasmas and astrophysics, and in nonlinear communication lines. It can be represented in the most general form as

$$\frac{\partial u}{\partial t}+(c+\alpha u)\frac{\partial u}{\partial x}+\beta\frac{\partial^3 u}{\partial x^3}=0, \tag{1}$$

where c, α, and β are constants, which are determined by the specific details of the physical problem. In some cases there is no quadratic nonlinearity, and then the modified Korteweg–de Vries equation is obtained in first-order perturbation theory:

$$\frac{\partial u}{\partial t}+(c+\alpha_1 u^2)\frac{\partial u}{\partial x}+\beta\frac{\partial^3 u}{\partial x^3}=0. \tag{2}$$

In particular, such an equation is obtained for electromagnetic surface waves in an electric field,[1] for waves in quantum-well films,[2] and for internal waves in an ocean with definite stratification.[3] The modified Korteweg–de Vries equation can also be solved exactly by a method based on the inverse problem of scattering theory,[4] and solitons and kinks (shock-like drops) also determine the asymptotic representation of the wave field here. If the quadratic nonlinearity is nonzero, cubic nonlinearity is obtained in the next order of perturbation theory and can be retained in the equation along with the ensuing corrections for dispersion, including nonlinear dispersion (see, for example, Ref. 5). Equations of this type have essentially only now begun to be studied.[6,7] However, it is clear from general arguments that within perturbation theory the effects associated with small higher-order corrections should be small and should not lead to fundamental changes in the form, for example, of solitons (although it is also understood that new features, for example, inelasticity of the interaction of solitons and their slow destruction as a result of radiation effects, can also appear). Nevertheless, there are situations where the coefficient for the quadratic nonlinearity can be so small that the quadratic and cubic nonlinearities are of the same order of magnitude, and, at the same time, there is no need to take into account the next orders with respect to dispersion. The expanded Korteweg–de Vries equation derived here

$$\frac{\partial u}{\partial t}+(c+\alpha u+\alpha_1 u^2)\frac{\partial u}{\partial x}+\beta\frac{\partial^3 u}{\partial x^3}=0 \tag{3}$$

is called the Gardner equation, and in cases where either of the nonlinear coefficients is equal to zero, it gives Eqs. (1) and (2), respectively.

The purpose of this paper is to demonstrate some new physical effects associated with the generation and interaction of large-amplitude solitons, where the quadratic and cubic nonlinearities are of the same order of magnitude.

Apparently, the Gardner equation (3) was first derived rigorously within the asymptotic theory for long internal waves in a two-layer fluid with a density jump at the interface.[8,9] The corresponding expressions for the coeffi-

cients in this equation have the form (in the approximation of a small density jump)

$$c=\sqrt{g\frac{\Delta\rho}{\rho}\frac{h_1h_2}{h_1+h_2}},\quad \beta=\frac{ch_1h_2}{6},\quad \alpha=\frac{3c}{2}\frac{h_1-h_2}{h_1h_2},$$
$$\alpha_1=-\frac{3c}{8h_1^2h_2^2}(h_1^2+h_2^2+6h_1h_2), \tag{4}$$

where u is the vertical displacement of the interface, h_1 and h_2 are the thicknesses of the upper and lower layers, respectively, $\Delta\rho/\rho$ is the relative height of the density jump, and g is the acceleration of gravity. As can be seen from (4), the coefficient of the quadratic nonlinearity can change sign when $h_1\sim h_2$ (and this leads to the interesting features of soliton transformation in the case of a tilted boundary; see Refs. 10–12), while the coefficient of the cubic nonlinearity is negative. The formulas for the coefficients in Eq. (3) were recently generalized to arbitrary density stratification of the fluid,[13] and it was shown that the coefficient of the cubic nonlinearity can be of either sign.[3] If it is assumed that the depths of the layers are roughly equal to one another and much smaller than the wavelength, and that the wave amplitude is small compared with the depth, the quadratic and cubic nonlinearities can be considered small and comparable to one another ($\alpha u/c\sim\alpha_1u^2/c\ll1$), as well as to the dispersion. Since both nonlinearities are then of the same order (in traditional disturbance schemes the cubic nonlinearity is always smaller than the quadratic analog), we can refer to large-amplitude waves in this sense (we note that in the ''everyday'' sense such ocean waves can have large amplitudes amounting to tens of meters). In this paper we examine the case in which the coefficient of the cubic nonlinearity is negative, as in a two-layer fluid ($\alpha_1<0$).

As we know, the single-soliton solution of the Gardner equation can easily be found in an explicit form:

$$u(x,t)=\frac{6\beta\Gamma^2}{\alpha}\frac{1}{1+\sqrt{1+(6\alpha_1\beta\Gamma^2/\alpha^2)\cosh[\Gamma(x-Vt)]}},$$
$$V=c+\beta\Gamma^2, \tag{5}$$

where Γ^{-1} is the effective soliton width. When Γ is small, the solution (5) describes a Korteweg–de Vries soliton with an amplitude $A=3\beta\Gamma^2/\alpha$ (its polarity is specified by the sign of the coefficient of the quadratic nonlinearity), and as $\Gamma\to\alpha/(6|\alpha_1|\beta)^{1/2}$, the soliton acquires a rectangular shape (Fig. 1) with the limiting amplitude

$$A_{\lim}=-\alpha/\alpha_1. \tag{6}$$

It is convenient to normalize the wave field to the limiting amplitude, as well as to vary the distance and time scales:

$$v(y,\tau)=-\frac{\alpha_1}{\alpha}u(x,t),\quad y=\sqrt{-\frac{\alpha^2}{6\beta\alpha_1}}(x-ct),$$
$$\tau=\beta\left(-\frac{\alpha^2}{6\beta\alpha_1}\right)^{3/2}t. \tag{7}$$

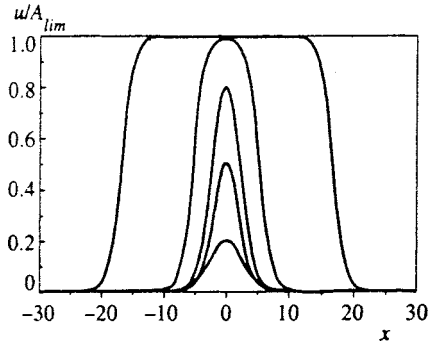

FIG. 1. Forms of solitons with various amplitudes.

Then the Gardner equation can be rewritten in the dimensionless form

$$\frac{\partial v}{\partial\tau}+6v(1-v)\frac{\partial v}{\partial y}+\frac{\partial^3v}{\partial y^3}=0, \tag{8}$$

and the soliton will have positive polarity with an amplitude from zero to unity.

It should first be noted that the generalized Miura transformation[9]

$$Q=-v+v^2+v_y \tag{9}$$

reduces Eq. (8) to the ''classical'' Korteweg–de Vries equation

$$Q_\tau-6QQ_y+Q_{yyy}=0 \tag{10}$$

and thereby proves the integrability of the Gardner equation by a method based on the inverse problem of scattering theory[9,14] and, therefore, the elastic character of soliton collisions. Nevertheless, ''reconstruction'' of the function $v(y,\tau)$ in terms of $Q(y,\tau)$ requires solving the Riccati equation (9) and cannot be expressed in a simple form. On the other hand, the simple replacement of variables

$$v=1/2+w \tag{11}$$

reduces the Gardner equation (8) to the modified Korteweg–de Vries equation relative to $w(y,\tau)$ and also permits the use of the known exact solutions of this equation.

Here we examine the interaction of solitons and their appearance from an initial pulsed disturbance. Soliton interactions and the evolution of the initial perturbation are illustrated by numerical simulation results.

2. SOLITON INTERACTIONS

2.1. Obtaining multisoliton solutions using Darboux transformations

Since the Garner equation can be reduced to the modified Korteweg–de Vries equation using the replacement (11), its N-soliton solution can be sought as the solution of the modified Korteweg–de Vries equation on a pedestal, as was done in Ref. 15 within a method based on the inverse problem of scattering theory. We use a different method, which is based on Miura and Darboux transformations[16] and which is described in Ref. 17, to obtain a two-soliton solution. The method can be described as follows.

As we know, the Korteweg–de Vries equation is the consistency condition for the system of equations (the so-called $L-A$ Lax pair)

$$\hat{L}\Psi=\lambda\Psi,$$
$$\hat{A}\Psi=\Psi_\tau, \tag{12}$$

where

$$\hat{L}=-\frac{\partial^2}{\partial y^2}+Q(y,\tau), \tag{13}$$

$$\hat{A}=-4\frac{\partial^3}{\partial y^3}+6Q\frac{\partial}{\partial y}+3\frac{\partial Q}{\partial y}. \tag{14}$$

The first equation in the system (12) is the stationary Schrödinger equation with the time τ appearing as a parameter. The second equation describes the time dependence of the solution. The consistency condition for the system (12), which is obtained by eliminating Ψ and which transforms the system (12) into the Korteweg–de Vries equation (10), has the form

$$\hat{L}_\tau=[\hat{A},\hat{L}], \tag{15}$$

where $[\hat{A},\hat{L}]=\hat{A}\hat{L}-\hat{L}\hat{A}$ is an operator commutator. It is also known that the $L-A$ pair is covariant with respect to Darboux transformations:

$$\tilde{Q}=Q-2\sigma_y, \quad \sigma=\frac{\Psi_{1y}(y,\lambda_1)}{\Psi_1(y,\lambda_1)},$$
$$\tilde{\Psi}(y,\lambda)=\left(\frac{\partial}{\partial y}-\sigma\right)\Psi(y,\lambda), \tag{16}$$

where $\Psi(y,\lambda)$ and $\tilde{\Psi}(y,\lambda)$ are general solutions of the stationary Schrödinger equation with the potentials $Q(y)$ and $\tilde{Q}(y)$, respectively, and $\Psi_1(y,\lambda_1)$ is the particular solution with the potential $Q(y)$ and the eigenvalue $\lambda=\lambda_1$. The transformed potentials $\tilde{Q}(y,\tau)$ are new nontrivial solutions of Eq. (10). This known procedure makes it possible to obtain multisoliton solutions of the Korteweg–de Vries equation.

It was noted in Ref. 17 that if the function

$$w(y,\tau)=\frac{\Psi_y(y,\tau)}{\Psi(y,\tau)} \tag{17}$$

is included in the treatment and if expressions for Q and Q_y are found from the first equation in the system (12) and substituted into the second equation of the system, the modified Korteweg–de Vries equation relative to $w(y,\tau)$ must hold for consistency of the $L-A$ pair:

$$w_\tau-6(\lambda+w^2)w_y+w_{yyy}=0. \tag{18}$$

This is not difficult to prove. Thus, applying the Darboux transformations, we find new solutions of the modified Korteweg–de Vries equation [and the Gardner equation after the replacement (11)], i.e., the functions $w(y,\tau)$, along with new solutions of the ordinary equation.

The generalized formulas derived by Crum,[18] which permit elimination of the intermediate mathematical operations, are valid for multiple application of a Darboux scheme. For twofold application they have the form

$$\tilde{\tilde{Q}}=Q-2\frac{\partial^2}{\partial y^2}\ln W_2(\Psi_1,\Psi_2),$$
$$\tilde{\tilde{\Psi}}=\frac{W_3(\Psi_1,\Psi_2,\Psi)}{W_2(\Psi_1,\Psi_2)},$$

where

$$W_2=\begin{vmatrix}\Psi_1 & \Psi_2\\ \Psi_1' & \Psi_2'\end{vmatrix}, \quad W_3=\begin{vmatrix}\Psi_1 & \Psi_2 & \Psi\\ \Psi_1' & \Psi_2' & \Psi'\\ \Psi_1'' & \Psi_2'' & \Psi''\end{vmatrix}, \tag{19}$$

and the derivative of the functions is taken with respect to the variables y.

2.2. Two-soliton solution of the Gardner equation

In order to obtain the single-soliton solution of the Gardner equation on the basis of Darboux transformations, we must select the potential

$$Q=1/4, \tag{20}$$

the ''seed'' function

$$\Psi_1=\cosh[(\Gamma_1/2)(y-\delta_1-\Gamma_1^2\tau)], \tag{21}$$

where

$$\delta_1=(1/\Gamma_1)\tanh^{-1}\Gamma_1, \tag{22}$$

and the general solution

$$\Psi=\exp[-(y-\Gamma_1^2\tau)/2] \tag{23}$$

[we selected $\lambda=0$, eliminating the superfluous term in Eq. (18)].

Then, following formulas (16), we have

$$\sigma=\frac{\Gamma_1}{2}\tanh\left[\frac{\Gamma_1}{2}(y-\delta_1-\Gamma_1^2\tau)\right],$$
$$\tilde{Q}=\frac{1}{4}-\frac{\Gamma_1^2}{2\cosh^2[(\Gamma_1/2)(y-\delta_1-\Gamma_1^2\tau)]}, \tag{24}$$

$$w_1(y,\tau)=-\frac{1}{2}+\frac{\Gamma_1^2}{1+\cosh[\Gamma_1(y-\delta_1-\Gamma_1^2\tau)]+\Gamma_1\sinh[\Gamma_1(y-\delta_1-\Gamma_1^2\tau)]}. \tag{25}$$

Next, substituting (25) into (11) and using (22), we obtain the single-soliton expression for the dimensionless Gardner equation (8):

$$v_1(y,\tau)=\frac{\Gamma_1^2}{1+\sqrt{1-\Gamma_1^2}\cosh[\Gamma_1(y-\Gamma_1^2\tau)]}. \tag{26}$$

The soliton (26) attains its limiting amplitude as $\Gamma_1 \to 1$. In this case it is convenient to represent the soliton expression in the form of a combination of a kink and an antikink:

$$v_1(y,\tau)=\frac{1}{2}[\tanh Z_+ - \tanh Z_-],$$

$$Z_\pm=\frac{\Gamma_1}{2}(y-\Gamma_1^2\tau\pm\delta_1), \tag{27}$$

where δ_1 corresponds to the half-width of the limiting soliton.

It is convenient to illustrate the appearance of the potential (20) in our case using a transformation in the Miura form

$$Q=w^2+w_y, \tag{28}$$

which transforms the solutions of Eq. (18) into solutions of the Korteweg–de Vries equation. Substituting (17) into (28), we obtain the first equation of the system (12) with $\lambda=0$, i.e., the consistency condition for the $L-A$ pair in which $w(\Psi)$ and Q are related by (28) is Eq. (18) without the ''velocity'' term. As can easily be seen, the potential (20) is the level at which the Korteweg–de Vries soliton obtained from the solution of the Gardner equation after the transformations (11) and (28) propagates. If we use a Miura transformation in the form (9), we should choose $Q=0$ and $\lambda=-1/4$ in the stationary Schrödinger equation. This gives the same result, but is more convenient for finding the temporal behavior of the solutions from (14).

To obtain the two-soliton solution of the Gardner equation, we must select a second ''seed'' solution:

$$\Psi_2=\sinh[(\Gamma_2/2)(y-\delta_2-\Gamma_2^2\tau)], \tag{29}$$

where

$$\delta_2=(1/\Gamma_2)\tanh^{-1}\Gamma_2. \tag{30}$$

The two-soliton solution thus obtained from formulas (19) and (17) (see Appendix A) can be written as

$$v_2(y,\tau)=\frac{1}{2}(\Gamma_2^2-\Gamma_1^2)\left(\frac{1}{\Gamma_2\coth Z_{2+}-\Gamma_1\tanh Z_{1+}}-\frac{1}{\Gamma_2\coth Z_{2-}-\Gamma_1\tanh Z_{1-}}\right), \tag{31}$$

$$Z_{j\pm}=\frac{\Gamma_j}{2}(y-\Gamma_j^2\tau\pm\delta_j),$$

where $\Gamma_2>\Gamma_1>0$ are arbitrary parameters, which specify the widths of the individual solitons at infinite separation from one another, and δ_j are defined in (22) and (30). The two-soliton expression (31) was obtained using Hirota's method in Ref. 19 and is presented to describe the interaction of a soliton with a kink introduced in Ref. 1. As in the case of a single soliton, the expression (31) can be rewritten in another form, which ''decouples'' the solitons:

$$v_2(y,\tau)=\frac{\Gamma_2^2-\Gamma_1^2}{(\Gamma_2\coth Z_{2+}-\Gamma_1\tanh Z_{1+})(\Gamma_2\coth Z_{2-}-\Gamma_1\tanh Z_{1-})}\times\left(\frac{\Gamma_1^2}{1+\sqrt{1-\Gamma_1^2}\cosh[\Gamma_1(y-\Gamma_1^2\tau)]}-\frac{\Gamma_2^2}{1-\sqrt{1-\Gamma_2^2}\cosh[\Gamma_2(y-\Gamma_2^2\tau)]}\right). \tag{32}$$

Thus, a two-soliton solution has been found analytically. The interaction of two solitons is investigated in the next section.

2.3. Interaction of two solitons

As expected, the interaction of two solitons with amplitudes smaller than the limiting value follows the familiar scenario in the Korteweg–de Vries equation [within (31)]: if the difference in amplitudes is large, the larger soliton, which has a higher velocity, simply overtakes the smaller soliton, forming a single-humped figure at the maximum moment of interaction; if the difference in velocities is small, the trailing (larger) soliton imparts energy to the leading (smaller) soliton and is slowed, while the leading soliton is accelerated (amplitude exchange takes place between the solitons). The case in which one of the solitons (the one which is behind) has an amplitude close to the limiting value and the other soliton (the one which is out in front) has a ''normal'' amplitude is most interesting. Even if the leading soliton is small, the amplitude of the ''combined'' wave formed during the interaction does not exceed the limiting value and a dip characteristic of an exchange interaction appears on it.[20] The interaction process for fairly close amplitude values is shown in Fig. 2 (the amplitude ratio is 0.7): the smaller soliton interacts with the leading edge of the limiting soliton (which moves faster) and flips over on its apex, moves freely along it, and then ''slides'' down the trailing edge of the limiting soliton, recovering its original polarity.

The existence of a soliton of negative polarity on the crest of a limiting soliton can be understood already from the Gardner equation (3): its solution on the pedestal $u=A_{\lim}$ satisfies the same Eq. (3), but with the opposite sign for the quadratic nonlinearity, which leads to the appearance of a soliton of negative polarity. At the same time, as can be seen

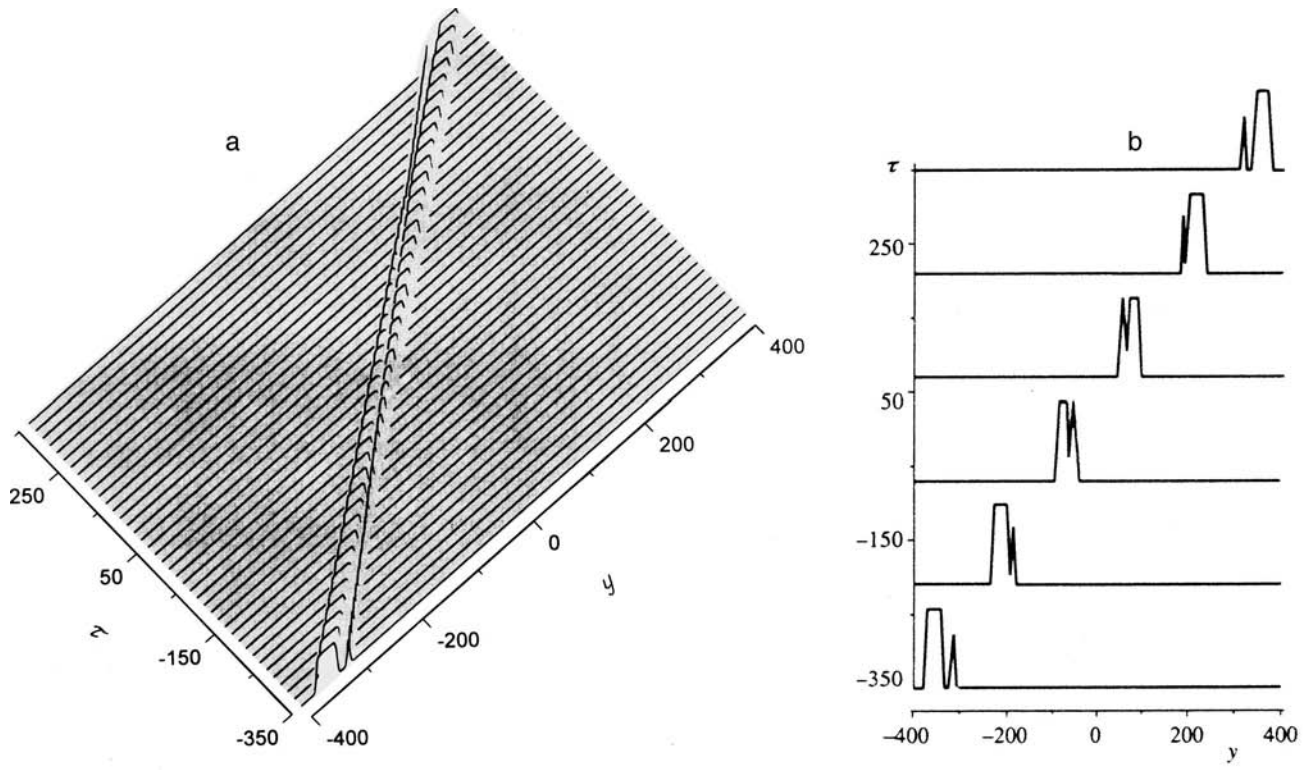

FIG. 2. Interaction of a limiting soliton with a large-amplitude soliton (the amplitude ratio is 0.7): a — dynamics with time; b — at various moments in time.

from Fig. 2b, the small soliton interacts successively with each edge of the limiting soliton, each of which can be interpreted as a kink and an antikink. The interaction of a soliton with a kink (antikink) in its ''pure form'' accompanied by change in polarity has already been analyzed numerically[1] for the modified Korteweg–de Vries equation, to which the Gardner equation can be reduced rigorously by the replacement (11). Successive collisions, first with a kink and then with an antikink, should lead to restoration of the original polarity of the soliton, as follows from (31). Thus, the dip on diagrams of interacting solitons can be interpreted as a soliton of negative polarity.

The exact two-soliton solution (31) permits finding a criterion which determines the type of interaction realized. During an interaction between solitons there is a moment when the two-soliton solution becomes symmetric with respect to the coordinate. As follows from the expression (31), this moment corresponds to the time $\tau=0$. In order to determine the type of soliton interaction, i.e., overtaking or exchange, we must calculate the second derivative of this interaction at the central point ($y=0$). A positive value of the derivative corresponds to an exchange interaction (the presence of a depression), and a negative value corresponds to overtaking. Performing the calculations, we can see (see Appendix B, Sec. B.1) that the following condition must be satisfied for an overtaking interaction:

$$A_2<\frac{3r-1}{3r-1-r^2}: \tag{33}$$

where $r=A_1/A_2$ $(0<A_1<A_2<1)$, and A_1 and A_2 are the soliton amplitudes. For small-amplitude solitons the critical ratio is $r=1/3$, as in the case of the interaction of ''quadratic'' solitons;[21] as the amplitude of the larger soliton increases, the critical ratio decreases, and the limiting soliton always interacts with another soliton according to an exchange scenario (Fig. 3). Thus, the exchange interaction dominates for solitons with larger amplitudes.

As in the case of Korteweg–de Vries solitons, the height of the central point of the two-soliton solution at the ''saddle'' point of the interaction v_2 $(y=0,\ \tau=0)$, at which a symmetric figure forms, is equal to the difference between the soliton amplitudes A_2-A_1 (see Appendix B, Sec. B.2). In the case where one of the solitons tends to the limiting soliton $(\Gamma_2\to 1)$, the expression (31) reduces to the first soliton of opposite polarity moving at the level of the limiting

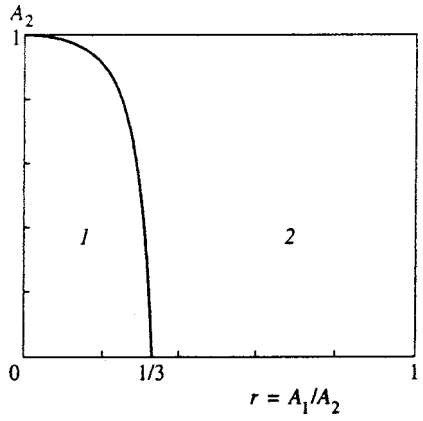

FIG. 3. Overtaking (region *1*) and exchange (region *2*) interaction of two solitons on the parameter plane.

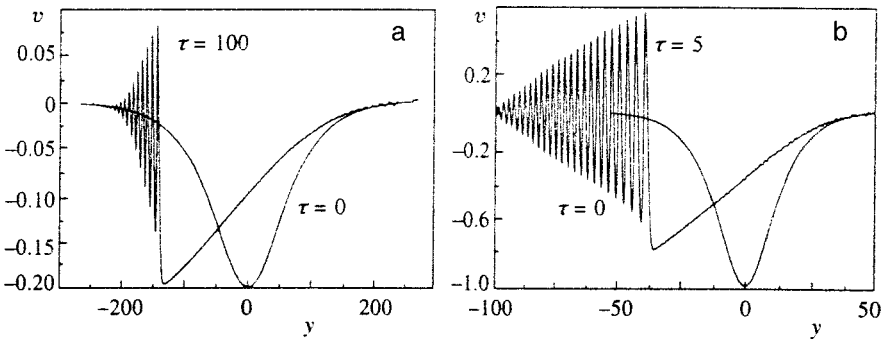

FIG. 4. Destruction of smooth (a) and abrupt (b) negative pulsed disturbances.

amplitude (Appendix B, Sec. B.4). It follows from an analysis of the dependences of the values of the highest point of the symmetric distribution on the amplitude of the larger soliton for various values of r that the solution does not exceed the limiting amplitude and that $r=1/3$ must be chosen to minimize the symmetric wave field.

Solitons acquire a phase shift during an interaction. In the case where one of the solitons is close to the limiting soliton, the appearance of a phase shift clearly follows from the conservation of mass in the Gardner equation: when the smaller soliton reaches the peak of the limiting soliton and its polarity reverses, the deficiency appearing in the mass integral is compensated by a forward shift of the interacting edge of the limiting soliton. As the smaller soliton descends, the mass excess is offset by forward shift of the trailing edge of the limiting soliton. To find its magnitude we must examine the soliton first before the interaction and then after the interaction. The phase shift appearing during a soliton interaction can be found from Eq. (31) (as is seen in Fig. 2a):

$$\Delta y_{1,2}=\pm\frac{2}{\Gamma_{1,2}}\ln\frac{\Gamma_2+\Gamma_1}{\Gamma_2-\Gamma_1} \tag{34}$$

(see Appendix B, Sec. B.3). In Ref. 15 the phase shift (34) was found from the asymptotic solution, and in Ref. 1 the phase shift for the interaction of the small soliton with a shock wave was found from the integrals of the equation.

Thus, the criterion which distinguishes between the types of interaction is the critical ratio between the soliton amplitudes, which is a function of the amplitude of the larger soliton. If this parameter does not exceed the critical value (a large difference between the amplitudes), the interaction between the solitons occurs without the formation of a dip, i.e., overtaking occurs. If, conversely, the parameter is greater than the critical value (similar amplitudes), a dip appears against the background of a large wave at the moment of collision, i.e., only an exchange interaction occurs. Exchange between a small soliton and a large soliton can be interpreted as successive interactions of the small soliton with each edge of the limiting soliton (a kink and an antikink) with a change in its polarity in the intermediate stage. The amplitude of the wave at any moment in time does not exceed the limiting value.

3. DECAY OF THE INITIAL DISTURBANCE

3.1. Destruction of a negative pulse

As we know, the Korteweg–de Vries equation (1) and the modified Korteweg–de Vries equation (2) with negative nonlinearity have self-similar solutions, and the structure of each solution has the form of a certain function which deforms with time according to a definite law. For example, after the replacement

$$w(y,\tau)=(3\tau)^{-1/3}\omega(z),\quad z=(3\tau)^{-1/3}y \tag{35}$$

is made in the modified Korteweg–de Vries equation (18) (we set $\lambda=0$ in it), the function $\omega(z)$ satisfies the equation

$$\omega_{zz}=2\omega^3+z\omega+\alpha, \tag{36}$$

where α is an arbitrary constant.[21] Equation (36) is called Painlevé's second equation (P_{II}), and its solutions are special functions, i.e., Painlevé transcendental functions.

The Korteweg–de Vries equation (10) also has such a self-similar solution.[22] After the replacement

$$Q(y,\tau)=-(3\tau)^{-2/3}q(z),\quad z=(3\tau)^{-1/3}y, \tag{37}$$

the function $q(z)$ satisfies the equation

$$q_{zzz}+6qq_z-(2q+zq_z)=0, \tag{38}$$

whose solutions are uniquely related to the solutions of the Painlevé's second equation:[22]

$$q=-\omega_z-\omega^2,\quad \omega=\frac{q_z+\alpha}{2q-z}. \tag{39}$$

In the case of $\alpha=0$ (where $\omega\to 0$ as $z\to\pm\infty$) the transcendental function $\omega(z)$ acts as an Airy function.[22]

A pulsed disturbance of the modified Korteweg–de Vries equation and a pulse in the Korteweg–de Vries equation with polarity opposite to the sign of the nonlinearity (i.e., in the case where the appearance of a soliton is impossible) evolve, transforming into the self-similar solutions (37) and (35), respectively.

In the case of the Gardner equation, when both (the quadratic and the cubic) nonlinearities are present, a negative pulse (a hyperbolic secant was used in the numerical calculations) will also evolve like the self-similar solutions presented (Fig. 4), although the solution of the Gardner equation cannot be written in a form such as (35) or (37).

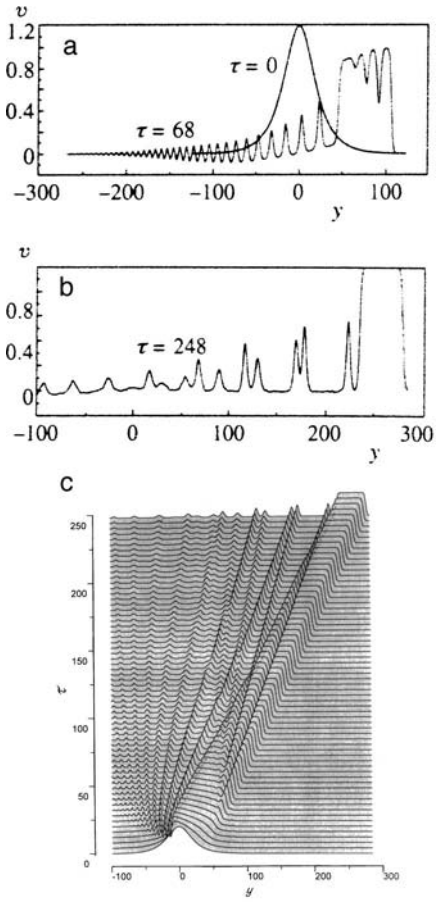

FIG. 5. Evolution of an initial disturbance with an amplitude exceeding the limiting value: a — initial disturbance (thick line) and formation of a group of solitons on the crest of a limiting soliton; b — formation of a limiting soliton and a group of solitons behind it; c — dynamics of the solution with time.

3.2. Evolution of a positive pulse

The decay of a positive initial disturbance with an amplitude smaller than the limiting value follows the scenario of the Korteweg–de Vries equation: the initial disturbance breaks down into a group of solitons (or a single soliton) and an oscillating tail. Interesting features appear as an initial disturbance with an amplitude exceeding the limiting value decays (Fig. 5). In this case two steep drops appear in the first stage and are associated with the cubic nonlinearity [in a pure form relative to the 1/2 level: the solution of the Gardner equation (8) on this pedestal satisfies the modified Korteweg–de Vries equation, for which solutions in the form of drops, i.e., quasishock waves, are valid]. Then the generation of small-scale waves begins on each drop, some of them transform into solitons, and their polarity can be different, depending on the pedestal on which they are generated (Fig. 5a) (the polarity of a soliton in the Gardner equation depends on the pedestal on which it forms: positive solitons form at $0<v_\infty<1/2$, and negative solitons form at $1/2<v_\infty<1$). The negative solitons subsequently interact on the wave crest with an antikink and descend from the limiting soliton, changing their polarity and superimposing themselves on the group of solitons formed by generation of the antikink (Figs. 5b and 5c). One limiting soliton, several small solitons, and a group of small-scale waves form on the asymptote. This process was briefly described in our previous paper.[20] Expansion of an initial disturbance with a ''superlimiting'' amplitude leads to the formation of a broader limiting soliton.

The formation of a limiting soliton was predicted already by the inverse problem of scattering theory in Ref. 24, but the intermediate stage with the formation of solitons of different polarity was not considered.

Thus, the existence of a limiting amplitude for the soliton solution in integrable nonlinear systems leads to new effects in the interaction of solitons and their generation from initial disturbances. The case of the appearance of a limiting soliton associated with the formation and propagation of solitons of opposite polarity on its crest, as well as the case of the interaction of a limiting soliton with another one, are extremely interesting.

This work was carried out with partial support of the INTAS (Grant 95-969), the Russian Fund for Fundamental Research (Grants 96-05-641008, 96-05-64111, 96-15-96592), and the ''Nonlinear Dynamics'' Program of the Ministry of Science and Technology. We thank T. G. Talipova, who took part in writing the numerical program for solving the Gardner equation and in the discussion of the results obtained.

APPENDIX A: DERIVATION OF THE TWO-SOLITON SOLUTION USING DARBOUX TRANSFORMATIONS

To obtain the expression for $v_2(y,\tau)$ which we are seeking, we must plug the functions (29) into the expressions (19) and use the definitions (22) and (30) for $\delta_{1,2}$:

$$W_2=\frac{1}{2}(\Gamma_2\Psi_1\Psi_2'-\Gamma_1\Psi_1'\Psi_2)$$

$$=\frac{1}{4\cosh(\Gamma_2\delta_2)\cosh(\Gamma_1\delta_1)}[\sinh(\Gamma_2\delta_2+\Gamma_1\delta_1)\times\cosh(Z_{2-}-Z_{1-})+\sinh(\Gamma_2\delta_2-\Gamma_1\delta_1)\cosh(Z_{2-}+Z_{1-})], \quad \text{(A1)}$$

$$W_3=\frac{\Psi}{8}[(\Gamma_2^2-\Gamma_1^2)\Psi_1\Psi_2+\Gamma_1(\Gamma_2^2-1)\times\Psi_1'\Psi_2+\Gamma_2(1-\Gamma_1^2)\Psi_1\Psi_2']$$

$$=\frac{\Psi}{16\cosh^2(\Gamma_2\delta_2)\cosh^2(\Gamma_1\delta_1)}\times[\sinh(\Gamma_2\delta_2+\Gamma_1\delta_1)\cosh(Z_{2+}-Z_{1+})+\sinh(\Gamma_2\delta_2-\Gamma_1\delta_1)\cosh(Z_{2+}+Z_{1+})]. \quad \text{(A2)}$$

From formulas (11) and (17) we have

$$v_2(y,\tau)=\frac{1}{2}+\frac{\tilde{\tilde{\Psi}}_y(y,\tau)}{\tilde{\tilde{\Psi}}(y,\tau)}, \quad \text{(A3)}$$

where the function $\tilde{\tilde{\Psi}}(y,\tau)$ is defined in (19). From (A3) we obtain

$$v_2=\frac{\tilde{W}_3'}{\tilde{W}_3}-\frac{W_2'}{W_2},\quad \tilde{W}_3=\frac{W_3}{\Psi}, \tag{A4}$$

$$W_2'=\frac{\sinh(\Gamma_2\delta_2+\Gamma_1\delta_1)\sinh(\Gamma_2\delta_2-\Gamma_1\delta_1)}{8\cosh^2(\Gamma_2\delta_2)\cosh^2(\Gamma_1\delta_1)}\times[\sinh(Z_{2-}-Z_{1-})+\sinh(Z_{2-}+Z_{1-})], \tag{A5}$$

$$\tilde{W}_3'=\frac{\sinh(\Gamma_2\delta_2+\Gamma_1\delta_1)\sinh(\Gamma_2\delta_2-\Gamma_1\delta_1)}{32\cosh^3(\Gamma_2\delta_2)\cosh^3(\Gamma_1\delta_1)}\times[\sinh(Z_{2+}-Z_{1+})+\sinh(Z_{2+}+Z_{1+})]. \tag{A6}$$

Substituting (A1), (A2), (A5), and (A6) into (A4), we have

$$v_2(y,\tau)=\frac{1}{2}\frac{\cosh^2(\Gamma_2\delta_2)-\cosh^2(\Gamma_1\delta_1)}{\cosh^2(\Gamma_2\delta_2)\cosh^2(\Gamma_1\delta_1)}\times\left(\frac{1}{\tanh(\Gamma_2\delta_2)\coth Z_{2+}-\tanh(\Gamma_1\delta_1)\tanh Z_{1+}}-\frac{1}{\tanh(\Gamma_2\delta_2)\coth Z_{2-}-\tanh(\Gamma_1\delta_1)\tanh Z_{1-}}\right). \tag{A7}$$

Using (22) and (30) again, we obtain the two-soliton solution (31); another form of the solution (32) is obtained by converting the terms in (A7) to forms with a common denominator.

APPENDIX B: FEATURES OF THE TWO-SOLITON SOLUTION

B.1. ''Exchange–overtaking'' boundary

In order to determine a criterion which distinguishes between the exchange and overtaking interactions, we must calculate the second derivative of the two-soliton solution with respect to the coordinate at the point ($y=0$, $\tau=0$). It is convenient to rewrite (31) in the form

$$v_2(y,0)=\frac{\Gamma_2^2-\Gamma_1^2}{2}[f(y)+f(-y)], \tag{B1}$$

where

$$f(y)=\left[\Gamma_2\coth\left(\frac{\Gamma_2}{2}(\delta_2+y)\right)-\Gamma_1\tanh\left(\frac{\Gamma_1}{2}(\delta_1+y)\right)\right]^{-1}. \tag{B2}$$

By virtue of (B1)

$$v_2''(y,0)=(\Gamma_2^2-\Gamma_1^2)f''(y), \tag{B3}$$

$$f'(y)=\frac{1}{2}\left[\Gamma_2\coth\left(\frac{\Gamma_2}{2}(\delta_2+y)\right)-\Gamma_1\tanh\left(\frac{\Gamma_1}{2}(\delta_1+y)\right)\right]^{-2}\left\{\Gamma_2^2\left[\coth^2\left(\frac{\Gamma_2}{2}(\delta_2+y)\right)-1\right]+\Gamma_1^2\left[1-\tanh^2\left(\frac{\Gamma_1}{2}(\delta_1+y)\right)\right]\right\}, \tag{B4}$$

$$f''(0)=\frac{1}{2}\left[\Gamma_2\coth\left(\frac{\Gamma_2\delta_2}{2}\right)-\Gamma_1\tanh\left(\frac{\Gamma_1\delta_1}{2}\right)\right]^{-2}\left(\frac{\omega_1^2}{\omega_2}+\omega_3\right), \tag{B5}$$

where

$$\omega_1=\Gamma_2^2\coth^2\left(\frac{\Gamma_2\delta_2}{2}\right)-\Gamma_1^2\tanh^2\left(\frac{\Gamma_1\delta_1}{2}\right)+\Gamma_1^2-\Gamma_2^2,$$

$$\omega_2=\Gamma_2\coth\left(\frac{\Gamma_2\delta_2}{2}\right)-\Gamma_1\tanh\left(\frac{\Gamma_1\delta_1}{2}\right), \tag{B6}$$

$$\omega_3=\Gamma_2^3\coth\left(\frac{\Gamma_2\delta_2}{2}\right)\left[1-\coth^2\left(\frac{\Gamma_2\delta_2}{2}\right)\right]+\Gamma_1^3\tanh\left(\frac{\Gamma_1\delta_1}{2}\right)\left[\tanh^2\left(\frac{\Gamma_1\delta_1}{2}\right)-1\right].$$

Thus, we find that at the moment of closest approach of the solitons (the time $\tau=0$) at the center of the two-soliton equation ($y=0$)

$$v_2''(0,0)=\frac{\omega_1^2/\omega_2+\omega_3}{2[\Gamma_2\coth(\Gamma_2\delta_2/2)-\Gamma_1\tanh(\Gamma_1\delta_1/2)]}. \tag{B7}$$

Going over to the soliton amplitudes

$$A_j=\frac{\Gamma_j^2}{1+\sqrt{1-\Gamma_j^2}},\quad j=1,2,\quad 0\leqslant A_j\leqslant 1 \tag{B8}$$

and using (22) and (30), we have

$$\tanh\frac{\Gamma_j\delta_j}{2}=\frac{A_j}{\Gamma_j},\quad j=1,2. \tag{B9}$$

Substituting (B8) and (B9) into (B6), we can rewrite (B7) as

$$v_2''(0,0)\propto -2A_2[A_2^2(r^3-2r^2-2r+1)+A_2(r^2+8r-3)-6r+2], \tag{B10}$$

where $r=A_1/A_2$.

Thus, in order to determine the sign of the second derivative, we must solve a quadratic equation. The condition for a ''single-humped'' symmetric distribution ($v_2''(0,0)>0$) is

$$A_2<\frac{3r-1}{3r-1-r^2}. \tag{B11}$$

B.2. Height of the central point of the symmetric distribution

The height of the central point is determined from the expression (31) with the coordinates $y=0$ and $\tau=0$:

$$v_2(0,0)=\frac{\Gamma_2^2-\Gamma_1^2}{\Gamma_2\coth(\Gamma_2\delta_2/2)-\Gamma_1\tanh(\Gamma_1\delta_1/2)}. \tag{B12}$$

Using (22), (30), (B8), and (B9), from (B12) we obtain

$$v_2(0,0)=\frac{A_2(2-A_2)-A_1(2-A_1)}{2-A_2-A_1}=A_2-A_1. \tag{B13}$$

B.3. Phase shift

Let us consider the soliton with the index 2 (the one which overtakes the smaller soliton). Its form as $\tau\to-\infty$ is found from (31):

$$v_2(y,\tau)\approx\frac{1}{2}(\Gamma_2^2-\Gamma_1^2)\left(\frac{1}{\Gamma_2\coth Z_{2+}+\Gamma_1}-\frac{1}{\Gamma_2\coth Z_{2-}+\Gamma_1}\right). \tag{B14}$$

Utilizing the identity

$$\Gamma_2\coth\left(\frac{\Gamma_2 z}{2}\right)+\Gamma_1\equiv\frac{\sqrt{\Gamma_2^2-\Gamma_1^2}\cosh[(\Gamma_2 z+\Delta)/2]}{\sinh(\Gamma_2 z/2)}, \tag{B15}$$

where

$$\frac{\Delta}{2}=\tanh^{-1}\frac{\Gamma_1}{\Gamma_2}, \tag{B16}$$

from (B14) we obtain

$$v_2(y,\tau)\approx\frac{\Gamma_2}{2}\left[\tanh\left(Z_{2+}+\frac{\Delta}{2}\right)-\tanh\left(Z_{2-}+\frac{\Delta}{2}\right)\right]. \tag{B17}$$

Long after the interaction ($\tau\to\infty$)

$$v_2(y,\tau)\approx\frac{1}{2}(\Gamma_2^2-\Gamma_1^2)\left(\frac{1}{\Gamma_2\coth Z_{2+}-\Gamma_1}-\frac{1}{\Gamma_2\coth Z_{2-}-\Gamma_1}\right), \tag{B18}$$

and with allowance for (B15) and (B16)

$$v_2(y,\tau)\approx\frac{\Gamma_2}{2}\left[\tanh\left(Z_{2+}-\frac{\Delta}{2}\right)-\tanh\left(Z_{2-}-\frac{\Delta}{2}\right)\right]. \tag{B19}$$

Thus, the total phase shift is

$$\Delta=2\tanh^{-1}\frac{\Gamma_1}{\Gamma_2}. \tag{B20}$$

During the interaction the larger, overtaking, soliton traverses the interaction region, having a velocity greater than its normal velocity (see Fig. 2). After treating the other soliton in a similar manner, we can easily see that it, conversely, slows in the interaction region, acquiring a phase shift $-\Delta$, and instead of (B15) for this soliton we must use the identity

$$\Gamma_2-\Gamma_1\tanh\left(\frac{\Gamma_1 z}{2}\right)\equiv\frac{\sqrt{\Gamma_2^2-\Gamma_1^2}\cosh[(\Gamma_1 z-\Delta)/2]}{\cosh(\Gamma_2 z/2)}. \tag{B21}$$

Representing the phase shift as the coordinate shift

$$\Delta y_j=\frac{2}{\Gamma_j}\Delta, \tag{B22}$$

taking into account the different signs of Δ for different solitons, and switching from a hyperbolic function to a logarithm, we obtain formula (34).

B.4. Soliton at the base of a limiting wave

As the second soliton approaches the limiting soliton, $\Gamma_2\to1$ and the two-soliton expression (31) can be written as

$$v_2(y,\tau)\approx\frac{1-\Gamma_1^2}{2}\left(\frac{1}{1-\Gamma_1\tanh Z_{1+}}-\frac{1}{-1-\Gamma_1\tanh Z_{1-}}\right), \tag{B23}$$

$$Z_{1\pm}=\frac{\Gamma_1}{2}(y-\Gamma_1^2\tau\pm\delta_1).$$

Using (B21) and (B22), we rewrite (B23):

$$v_2(y,\tau)\approx\frac{1}{2}\sqrt{1-\Gamma_1^2}\left(\frac{\cosh Z_{1+}}{\cosh Z_{1-}}+\frac{\cosh Z_{1-}}{\cosh Z_{1+}}\right). \tag{B24}$$

Using (22) again, we obtain a soliton on a pedestal:

$$v_2(y,\tau)\approx1-\frac{\Gamma_1}{2}(\tanh Z_{1+}-\tanh Z_{1-}). \tag{B25}$$

*)E-mail: avs@appl.sci-nnov.ru
†)E-mail: enpeli@appl.sci-nnov.ru

[1] T. L. Perel'man, A. X. Fridman, and M. M. El'yashevich, Zh. Éksp. Teor. Fiz. **66**, 1316 (1974) [Sov. Phys. JETP **39**, 643 (1974)].
[2] E. N. Pelinovskiĭ and V. V. Sokolov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **19**, 536 (1976).
[3] R. Grimshaw, E. Pelinovsky, and T. Talipova, Nonlinear Processes Geophys. **4**, 237 (1997).
[4] 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)].
[5] T. R. Marchant and N. F. Smyth, J. Fluid Mech. **221**, 263 (1990).
[6] A. S. Fokas and Q. M. Liu, Phys. Rev. Lett. **77**, 2347 (1996).
[7] T. R. Marchant and N. F. Smith, J. Appl. Math. **56**, 157 (1996).
[8] T. Kakutani and N. Yamasaki, J. Phys. Soc. Jpn. **45**, 674 (1978).
[9] J. W. Miles, Tellus **31**, 456 (1979).
[10] T. G. Talipova, E. N. Pelinovskiĭ, and R. Grimshaw, JETP Lett. **65**, 120 (1997).
[11] V. Malomed and V. Shrira, Physica D **53**, 1 (1991).
[12] S. J. Knickerbocker and A. C. Newell, Phys. Lett. A **75**, 326 (1980).
[13] K. Lamb and L. Yan, J. Phys. Oceanogr. **26**, 2712 (1996).
[14] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[15] N. N. Romanova, Teor. Mat. Fiz. **39**, 205 (1979).
[16] V. V. Matveev and M. A. Salle, *Darboux Transformations and Solitons*, Springer (1991).
[17] M. Yu. Kulikov and G. M. Fraĭman, Preprint of the Institute of Applied Physics, Russian Academy of Sciences, No. 404, Nizhniĭ Novgorod (1996).
[18] M. Crum, Q. J. Math. **6**, 121 (1955).
[19] Y. Chen and P. L.-F. Liu, Wave Motion **24**, 169 (1996).
[20] E. N. Pelinovskiĭ and A. V. Slyunyaev, JETP Lett. **67**, 655 (1998)].
[21] G. L. Lamb, Jr., *Elements of Soliton Theory*, Wiley, New York (1980) [Russ. transl., Mir, Moscow (1983)].
[22] A. S. Fokas and M. J. Ablowitz, J. Math. Phys. **23**, 2033 (1982).
[23] M. J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering*, Cambridge University Press (1991).
[24] J. W. Miles, Tellus **33**, 397 (1981).

Translated by P. Shelnitz

Space-time dynamics of ultrashort pulses in vacuum

V. A. Mironov*)

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted 23 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 35–46 (July 1999)

An analytical study is made of the evolution of spatially bounded pulses whose length amounts to several periods of the field oscillations. An equation is analyzed that describes unidirectional (reflectionless) propagation of light pulses in vacuum. The method of moments is used to find the variations in length, effective width of the wave field, and other characteristic averaged parameters of a pulse along its propagation path. A broad class of self-similar solutions describing the focusing of the light pulses is found. Finally, by direct integration of the starting equation it is shown that a horseshoe-shaped precursor forms near the leading edge of the pulse. © *1999 American Institute of Physics.* [S1063-7761(99)00407-2]

1. INTRODUCTION

Studies of propagation of electromagnetic radiation with a spectral width of order the carrier frequency have stimulated interest in the prospects for using such pulses. The advantages of broadband signals are well-known from radar and radio communications.[1,2] Probing objects with light signals produces much more information than in the case of radio-frequency pulses.[1–3] In contrast to quasimonochromatic radiation, the use of pulses whose length amounts to only a few oscillation periods of the field makes it possible to form wave structures for which the effects of diffraction smearing are weakened substantially.[4–8] Such structures have become known as electromagnetic projectiles.[5] The advances of optoelectronics in generating microwave radiation with a spectral width of order of several terahertz have found applications in tomography.[3] Possibilities are discussed of using such electromagnetic pulses to accelerate charged particles.[9]

The goal of the present work is a further study of the features of the space-time evolution of light pulses in vacuum, for which the common approximation of slowly varying amplitudes proves inapplicable. In contrast to Refs. 5–8, we will use an approximation corresponding to Fresnel diffraction of the wave field, which makes it possible to describe the evolution of the field in detail. In Sec. 2 we formulate the equations and derive some integral relationships needed in the future investigation. Section 3 is devoted to the method of moments used in analyzing the effective pulse parameters. In Secs. 4 and 5 we discuss the exact solutions of the starting equation. First we will give solutions in self-similar form that describe the transverse focusing of the pulses. Then we will analyze an equation obtained by directly integrating the starting equation for the case of an axisymmetric distribution of the wave field.

2. STATEMENT OF THE PROBLEM AND GENERAL RELATIONSHIPS

Let us examine the propagation of an electromagnetic field in vacuum along the z axis. The wave equation describing this process is

$$\frac{\partial^2\psi}{\partial z^2}+\Delta_\perp\psi-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}=0, \quad \text{where} \quad \Delta_\perp=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}. \tag{1}$$

We seek the solution of this equation in the form $\psi=\psi(z,\tau=ct-z,\mathbf{r}_1)$. Assuming that the shape of the pulse changes little along the path of pulse propagation, we arrive at the equation

$$\frac{\partial^2\psi}{\partial z\partial\tau}=\frac{1}{2}\Delta_\perp\psi. \tag{2}$$

In deriving this equation we used the approximation

$$\frac{\partial^2\psi}{\partial z\partial t}\gg\frac{\partial^2\psi}{\partial z^2},$$

which is a common way of simplifying Eq. (1). Using this approximation is equivalent to allowing for the Fresnel diffraction of a packet of electromagnetic radiation along the transverse coordinate. It is natural, therefore, that in the case of a quasimonochromatic pulse, $\psi=\psi_k\exp\{ik\tau\}$, Eq. (2) yields the well-known parabolic equation for the slowly varying complex-valued amplitude of the wave field, $\psi_k(\tau,z,\mathbf{r}_\perp)$.

Equation (2) resembles an equation obtained as a result of the Brittingham transformation ($\tau=ct-z$, $\eta=ct+z$, and $\mathbf{r}_\perp=\mathbf{r}_\perp$; see Refs. 7 and 11). The difference is that Eq. (2) is an approximate equation. What makes Eq. (2) preferable is that it is a first-order evolutionary equation and the field dynamics is determined by the initial distribution along the characteristic line $z=0$. This equation can be generalized fairly easily to the case where the nonlinearity and dispersion of the medium is taken into account. For example, the linear part of the Khokhlov–Zabolotskiĭ[12] and Kadomtsev–Petviashvili[13] equations has the form of Eq. (2).

Using the Lagrangian of Eq. (2) with the density

$$L=\frac{1}{2}\frac{\partial\psi}{\partial\tau}\frac{\partial\psi}{\partial z}-\frac{1}{4}(\nabla_\perp\psi)^2, \tag{3}$$

and the usual variational procedure,[14] we arrive at the following constants of motion conserved in the evolution of the system:

$$I=\int\left(\frac{\partial\psi}{\partial\tau}\right)^2 d\tau\, d\mathbf{r}_\perp, \tag{4}$$

$$\mathbf{P}=\int\frac{\partial\psi}{\partial\tau}\nabla_\perp\psi\, d\tau\, d\mathbf{r}_\perp, \tag{5}$$

$$H=\int(\nabla_\perp\psi)^2\, d\tau\, d\mathbf{r}_\perp. \tag{6}$$

The fact that such relationships exist is due to the translational symmetry of Eq. (2). In the case of a quasimonochromatic pulse these relationships become the well-known (for the parabolic equation) integral relationships (for the energy, momentum, and Hamiltonian) integrated with respect to τ.

Integrating (3) with respect to τ, we arrive at the equation

$$\frac{\partial\psi}{\partial z}=\frac{1}{2}\Delta_\perp q,\quad q=\int_{-\infty}^{\tau}\psi(t',\mathbf{r}_\perp,z)\,dt', \tag{7}$$

which yields a number of relationships that will prove useful in our further discussions. Multiplying (7) by 2ψ, we obtain one more conservation law,

$$\frac{\partial}{\partial z}\psi^2=\operatorname{div}\psi\nabla_\perp q-\frac{\partial}{\partial\tau}\frac{(\nabla_\perp q)^2}{2}. \tag{8}$$

This implies that the quantity[4]

$$\int\psi^2\, d\mathbf{r}_\perp\, d\tau=I_2 \tag{9}$$

is conserved in the case of a localized distribution of the wave field only if[4]

$$\int_{-\infty}^{\infty}\psi(z,\mathbf{r}_\perp,\tau)d\tau=0. \tag{10}$$

Thus, the area under the curve for the pulse field, Eq. (10), is zero at each point in space. In other words, for all $\mathbf{r}_\perp$ and z the temporal spectrum of the pulse contains no zeroth harmonic. Indeed, the velocity of propagation of a static field is zero, with the result that the field remains near the source of radiation.

When the leading edge of the pulse is steep (e.g., at $\tau=0$), Eq. (7) and condition (10) show that the field distribution at the leading edge,

$$\psi(z,\mathbf{r}_\perp,\tau=0)=\psi_0(\mathbf{r}_\perp),$$

does not change along the path of pulse propagation.[4]

Note that our investigation has yielded two integral relationships expressing conservation laws, Eqs. (4) and (9), which in the case of a quasimonochromatic pulse reduce to a single constant of motion corresponding to the conservation of the number of photons in the parabolic equation. Such degeneracy occurs because in the approximation of slowly varying amplitudes there are no precursors. The presence of two constants of motion reflects a more realistic situation.

3. AVERAGED DESCRIPTION OF THE WAVE FIELD

As in the case of the parabolic equation,[15] the use of the method of moments makes it possible to derive a number of relationships that describe the dynamics of the wave field. The starting point is the equation of continuity,

$$\frac{\partial}{\partial z}\left(\frac{\partial\psi}{\partial\tau}\right)^2=\operatorname{div}_\perp\left(\frac{\partial\psi}{\partial\tau}\nabla_\perp\psi\right)-\frac{1}{2}\frac{\partial}{\partial\tau}(\nabla_\perp\psi)^2. \tag{11}$$

This equation can be derived by multiplying Eq. (2) by $\partial\psi/\partial\tau$ and transforming the right-hand side of the result as follows:

$$\frac{\partial\psi}{\partial\tau}\Delta_\perp\psi=\operatorname{div}_\perp\left(\frac{\partial\psi}{\partial\tau}\nabla_\perp\psi\right)-\nabla_\perp\psi\frac{\partial}{\partial\tau}(\nabla_\perp\psi).$$

Integrating Eq. (11) with respect to $d\tau$ and $d\mathbf{r}$, we readily obtain (4).

We begin with the first moments,

$$\int\tau\left(\frac{\partial\psi}{\partial\tau}\right)^2 d\mathbf{r}\, d\tau=\langle\tau\rangle,\quad \int\mathbf{r}\left(\frac{\partial\psi}{\partial\tau}\right)^2 d\mathbf{r}\, d\tau=\langle\mathbf{r}\rangle,$$

which describe the motion of the center of mass of the wave field $\partial\psi/\partial\tau$. Combining Eq. (11) and the expressions (5) and (6) for constants of motion, we easily arrive at the following relationships for these moments:

$$\frac{\partial\langle\tau\rangle}{\partial z}=\frac{H}{2}=\text{const}, \tag{12}$$

$$\frac{\partial\langle\mathbf{r}\rangle}{\partial z}=-\mathbf{P}=\text{const}. \tag{13}$$

Clearly, the center of mass moves along a straight line determined by the initial conditions (e.g., at $z=0$). For an axisymmetric wave field ($\mathbf{P}=0$), the center of mass moves along the z axis at less than the speed of light. Thus, the diffraction of the wave field ($\nabla_\perp\psi\neq0$) effectively reduces the velocity of propagation of the packet of electromagnetic radiation.

The second moments

$$\int\tau^2\left(\frac{\partial\psi}{\partial\tau}\right)^2 d\mathbf{r}\, d\tau=\langle\tau^2\rangle,\quad \int r_\perp^2\left(\frac{\partial\psi}{\partial\tau}\right)^2 d\mathbf{r}\, d\tau=\langle r_\perp^2\rangle$$

give the effective (longitudinal and transverse) dimensions of the wave field. Their evolution is described by a second-order equation. To find this equation we must know not only the integral relationships (4) and (5) but also the dependence of the rate of variation of the corresponding integral expressions on $(\partial\psi/\partial\tau)\nabla_\perp\psi$ and $(\nabla_\perp\psi)^2$.

First we consider the equation for the pulse length $\langle\tau^2\rangle$. Multiplying Eq. (11) by τ^2 and integrating with respect to $d\tau$ and $d\mathbf{r}$, we arrive at the equation

$$\frac{\partial}{\partial\tau}\langle\tau^2\rangle=\int\tau(\psi_x^2+\psi_y^2)d\tau\, d\mathbf{r}_\perp. \tag{14}$$

The equation for the rate of variation of $\psi_x^2+\psi_y^2$ can easily be derived from (7). Multiplying the latter by τ, integrating, and performing ordinary transformations, we find that

$$\frac{\partial}{\partial z}\int \tau(\psi_x^2+\psi_y^2)d\tau\, d\mathbf{r}_\perp=\frac{1}{2}\int (\Delta_\perp q)^2 d\tau\, d\mathbf{r}_\perp, \tag{15}$$

where q is defined in (7). Thus, the pulse length increases according to the law

$$\frac{d^2\langle(\tau-\langle\tau\rangle)^2\rangle}{dz^2}=\frac{1}{2}\int (\Delta_\perp q)^2 d\tau\, d\mathbf{r}_\perp>0. \tag{14'}$$

More information can be extracted by studying the variation of the transverse dimensions of the wave field. For example, Eq. (11) yields

$$\frac{\partial}{\partial z}\int x^2\psi_x^2 d\tau\, d\mathbf{r}_\perp=-2\int x\left(\frac{\partial\psi}{\partial\tau}\frac{\partial\psi}{\partial x}\right)d\tau\, d\mathbf{r}_\perp. \tag{16}$$

An equation for the variation of a component of the momentum density, $(\partial\psi/\partial\tau)\nabla\psi$, can easily be found from (2) and (7). Multiplying the derived equation by x and performing the necessary transformations, we arrive at the simple equation

$$\frac{\partial}{\partial z}\int x\psi_\tau\psi_x d\tau\, d\mathbf{r}_\perp=-\frac{H}{2}. \tag{17}$$

Similar transformations can be done for the rate of variation of $\langle y^2\rangle$. The result is

$$\frac{\partial^2}{\partial z^2}\int r_\perp^2\psi_\tau^2 d\tau\, d\mathbf{r}_\perp=2H. \tag{18}$$

Thus, as in the case of the parabolic equation,[11,15] the effective transverse size of an axisymmetric wave field increases as follows:

$$\langle r_\perp^2\rangle=\langle r_0^2\rangle+Hz^2, \tag{19}$$

where $\langle r_0^2\rangle$ is the characteristic size at $z=0$.

Comparing (12) and (19), we find that

$$\langle r_\perp^2\rangle-2z\langle\tau\rangle=\langle r_0^2\rangle, \tag{20}$$

which is a formula that links the effective scales $\langle r_\perp^2\rangle$ and $\langle\tau\rangle$. Equation (20) can be interpreted in the following manner. The wavefront of the starting equation (1) is spherical:

$$r^2+z^2-c^2t^2=\text{const}. \tag{21}$$

In terms of the variables used below (r, z, and $\tau=ct-z$), Eq. (21) can be written as

$$r^2-2\tau z+\tau^2=\text{const}. \tag{22}$$

In the approximation described by Eq. (2) we can ignore the term τ^2 in (22). The wavefront becomes parabolic. Averaging (22) under these conditions, we arrive at the same expression (20), which was derived by the method of moments. We see that the variable $\eta=r^2-2\tau z$ becomes a self-similar variable of Eq. (2). Similar relationships can be derived for the moments of the function ψ^2. These, however, prove to be less informative, since their variations are not related to the integral expressions (4)–(6).

4. SELF-SIMILAR STRUCTURES

The above relationships were derived without knowing the exact solution of the equation and describe the behavior of the characteristic parameters (moments) of a spatially localized distribution of the wave field. Let us examine the wave structures of self-similar form,

$$\psi=\varphi(z,\eta=r^2-2z\tau). \tag{23}$$

Inserting (23) in the starting equation (2), we can easily find an equation for the self-similar function:

$$-z\frac{\partial^2\varphi}{\partial z\,\partial\eta}=\eta\frac{\partial^2\varphi}{\partial\eta^2}+\frac{\partial\varphi}{\partial\eta}. \tag{24}$$

This equation has a particular solution

$$\varphi_\beta=z^\beta/\eta^{\beta+1}. \tag{25}$$

We will use the following fact to find the admissible values of β. The starting equation is translationally symmetric in z and τ. The complex generalization of (25), i.e., $z\to z+iz_0$ and $\tau\to\tau+i\tau_0$, has the form

$$\varphi_\beta=\frac{(z+iz_0)^\beta}{[r^2-2(z+iz_0)(\tau+i\tau_0)]^{\beta+1}}. \tag{26}$$

In contrast to (25), this expression has no singularity at $r^2=2\tau z$ and describes a bounded distribution of the wave field if $\tau_0 z_0>0$. The condition (10), which states that the spectrum of the localized solution (26) contains no zeroth harmonic, holds if

$$\beta>0. \tag{27}$$

Here the complex-valued wave field (26) describes the focusing of a spatially localized pulse near $z\approx 0$. The parameter z_0 determines the characteristic size of the focal region, and τ_0 is the length of the incident pulse ($z\gg z_0$). For instance, at $\beta=1$ the distribution of the field is

$$\varphi_1=\left\{4(z+iz_0)\left[\tau-\frac{r^2z}{2(z^2+z_0^2)}+i\left(\tau_0+\frac{r^2z_0}{2(z^2+z_0^2)}\right)\right]^2\right\}^{-1}. \tag{28}$$

We see that the wave field is at its maximum at $\tau^*\approx r^2z/2(z^2+z_0^2)$. The pulse length (the characteristic longitudinal size of the field) $\tau_p\approx\tau_0+r^2z_0/2(z^2+z_0^2)$ increases with the distance from the axis of the system. The field distribution along the axis ($r=0$) is described by the relationships

$$\mathrm{Re}\,\varphi_1=-\frac{2\tau_0\tau}{z_0(\tau^2+\tau_0^2)^2},\quad \mathrm{Im}\,\varphi_1=-\frac{\tau^2-\tau_0^2}{z_0(\tau^2+\tau_0^2)^2}\ \text{ at }\ z=0, \tag{29}$$

$$\mathrm{Re}\,\varphi_1=\frac{\tau^2-\tau_0^2}{z_0(\tau^2+\tau_0^2)^2},\quad \mathrm{Im}\,\varphi_1=-\frac{2\tau_0\tau}{z(\tau^2+\tau_0^2)^2}\ \text{ as }\ z\to\pm\infty. \tag{30}$$

Equation (30) shows that when the pulse passes the focal plane, the amplitude distribution of the field is restored and the phase changes by π. Here the change in the sign of the solution agrees with the invariance of the starting equation under the transformations $z\to -z$ and $\tau\to -\tau$. In the focal

plane ($z \approx 0$ and $x \ll z_0$), the real part of the solution (29) acquires the structure of the imaginary part as $z \to \infty$ and the imaginary part, the structure of the real part. This process corresponds to phase change of $\pi/2$.

By using superposition we can substantially increase the number of self-similar solutions. In particular, for integral values of β, we can use an expansion in a Taylor (Laurent) series in z/η and easily obtain an expression for the solution of the starting equation:

$$\varphi = \frac{1}{z+iz_0} f\left(\frac{r^2}{2(z+iz_0)} - \tau - i\tau_0\right). \tag{31}$$

This can be verified by simply plugging (31) into Eq. (2).

5. SPACE-TIME EVOLUTION OF THE WAVE FIELD

The self-similar structures treated here are similar to Gaussian wave beams in quasioptics (see, e.g., Ref. 12). These structures are ''single-scale'' and do not reflect the specific features of the problem related to the presence of two conserved quantities, (4) and (9). To illustrate these features we will examine the evolution of a pulse with a steep leading edge. In the case of an axisymmetric wave field, the solution of Eq. (2) can be written

$$\psi = \int_0^\infty \int_{-\infty}^\infty \varphi(\omega) R(\chi) \times \exp\left\{ i\frac{\chi^2}{2\omega} z + i\omega\tau \right\} J_0(\chi r)\, d\chi\, d\omega, \tag{32}$$

where

$$\varphi(\omega) = \frac{1}{2\pi} \int \psi(r=0, z=0, \tau) \exp\{-i\omega\tau\}\, d\tau$$

is the spectrum of the pulse along the axis of the system ($r=0$) at $z=0$, and $R(\chi)$ is the spectrum of the transverse distribution of the field at $z=0$. For a pulse that is Gaussian along the transverse coordinate, $\exp\{-r^2/2a^2\}$, we have

$$R(\chi) = a^2 \exp\left\{ -\frac{\chi^2 a^2}{2} \right\}. \tag{33}$$

Integrating (32) with respect to χ, we find that in this case

$$\psi = \int_{-\infty}^\infty \varphi(\omega) \frac{a^2}{a^2 - iz/\omega} \exp\left\{ -\frac{r^2}{a^2 - iz/\omega} + i\omega\tau \right\} d\omega. \tag{34}$$

5.1. Evolution of the pulse on the axis of the system ($r=0$)

At $r=0$ Eq. (34) shows that as the pulse propagates in the system ($z \neq 0$), the spectrum of the field acquires a pole at $\omega = iz/a^2$. The pulse structure is described by the integral

$$\psi(r=0, z, \tau) = \int_{-\infty}^\infty \frac{\varphi(\omega) \exp\{i\omega\tau\}}{\omega - iz/a^2}\, d\omega. \tag{35}$$

To take an example, let us study pulses whose initial shape is

$$\psi(r=0, z=0, \tau) = \psi_n = L_n \exp\{-\tau\}, \tag{36}$$

where L_n is a Laguerre polynomial defined on the interval $0 < \tau < \infty$. For $n \geqslant 1$ this shape meets condition (10). Note that the spectral component of the field at the frequency $\omega = 0$ is zero [see Eq. (35)] for all $z \neq 0$. Hence even for pulses with $\varphi(\omega = 0) \neq 0$ the area under the curve $\psi(r=0, z = +0, \tau)$ proves to be zero. For example for a pulse whose shape is given by (36) at $n=0$ we have

$$\psi(r=0, z=0, \tau) = \psi_0 = \begin{cases} 0, & -\infty < \tau < 0, \\ \exp(-\tau), & 0 \neq \tau < \infty. \end{cases} \tag{37}$$

Integration of (37) gives

$$\psi(r=0, z, \tau) = \frac{\exp(-\tau) - (z/a^2)\exp(-z\tau/a^2)}{1 - z/a^2}. \tag{38}$$

Thus, as a pulse propagates in the system, its shape acquires two scales. The amplitude of the part of the field with the shape of the incident pulse varies according to the law $(1 - z/a^2)^{-1}$. The characteristic time scale of the second term ($\exp\{-z\tau/a^2\}$) is determined by the transverse size of the field at $z=0$ and falls off as $\tau_i \sim a^2/z$ as the pulse propagates in the system. At a distance

$$z_F = a^2 \tag{39}$$

equal to the Fresnel length for a a field with a frequency ν determined by the duration of the incident radiation ($\nu = 1$), τ_p becomes equal to the characteristic time scale of the initial distribution. As a result the shape of the pulse becomes

$$\psi(r=0, z=a^2, \tau \geqslant 0) = L_1 \exp(-\tau), \tag{40}$$

where $L_1 = 1 - \tau$ is the first Laguerre polynomial. Note that a pulse initially ($z=0$) shaped as (40) is transformed in the process of propagation into $\psi(z=a^2) = L_2 \exp\{-\tau\}$, where $L_2 = 1 - 2\tau + \tau^2/2$ is the second Laguerre polynomial. This follows from the fact that for initial distributions of the form (36) the order of the pole in (35) at $z=a^2$ increases by unity, with the result that the shape of the pulse is determined by the polynomial L_{n+1}. Since the functions (36) are not orthogonal, it is impossible to draw a more general conclusion.

In the limit $z \gg z_F = a^2$ the amplitude of a field whose shape is that of the incident pulse [the first term in (38)] decreases as z^{-1}, just as it does in the case of a quasimonochromatic pulse. The characteristic time scale of the second term becomes smaller than the length of the initial pulse, whose amplitude is independent of z. This part of the field is sometimes called the diffraction precursor.[4] Formula (38) describes the formation of such a precursor.

Let us examine the energy characteristic of the field on the axis of the system,

$$W = \int_{-\infty}^\infty \psi^2(r=0, z, \tau)\, d\tau.$$

If we use ψ in the form (38) and integrate, we arrive at the expression

$$W(z) = \frac{1}{2(1 + z/a^2)}. \tag{41}$$

Thus, in the limit $z \gg a^2$ the energy of the pulse on the axis [Eq. (41)] decreases in inverse proportion to z. Obviously, such a gradual decrease is due to the presence of a precursor. The structure of this precursor is determined by the pole at $\omega = iz/a^2$ in the integrand of (35). Clearly, the precursors are the same for all initial distributions of the form (36), with the result that the energy of the field on the axis ($r=0$) of such formations decreases according to the z^{-1} law. This slow decrease of the pulse energy along the path of pulse propagation (slower than z^{-2}) has become known as the electromagnetic-projectile effect[5] and is due to the presence of higher harmonics in the pulse ($\omega \to \infty$), for which the geometrical-optics approximation is valid. The length of the path along which this law is valid is determined by the length of the leading edge of the pulse, τ_p ($\tau_p \ll 1$), and for $z > a^2/\tau_p$ the decrease becomes more rapid ($W \propto z^{-2}$).

5.2. Dynamics of the spatial distribution

To study the spatial structure we need to known the integral (34), which cannot be evaluated in closed form. However, we can get an idea of the dynamics of the spatial distribution by analyzing the coefficients of the Taylor expansion of the wave field in powers of r^2:

$$\psi = \sum_{n=0}^{\infty} \psi_n(z,\tau)\left(\frac{r^2}{2}\right)^n. \tag{42}$$

The simplest way to find the functions ψ_n is to use Eq. (2) directly. Substituting (42) in (2) and equating the coefficients of r^{2n}, we obtain the recurrence relation

$$\psi_{n+1} = \frac{1}{(n+1)^2}\frac{\partial^2 \psi_n}{\partial z\, \partial \tau}. \tag{43}$$

Thus, all the expansion coefficients are found by directly differentiating the wave field on the axis of the system [Eq. (35)], $\psi_0 = \psi(r=0,z,\tau)$. Note that these coefficients meet condition (10): the area under the curve ψ_{n+1} is zero ($\int_{-\infty}^{\infty} \psi_{n+1}\, d\tau = 0$) on the path of pulse propagation along z.

For a pulse of initial shape (37), we substitute (38) in (43) and find the expression for the coefficient of r^2:

$$\psi_1 = -\frac{b[(b-z)+b(1-b)\tau]\exp(-b\tau)+\exp(-\tau)}{a^2(1-b)^2}, \tag{44}$$

where $b = z/a^2$. The coefficients (43) can be calculated in a similar manner.

In the two most interesting cases, $b=1$ and $b \gg 1$, the expressions for the coefficients ψ_n simplify substantially. For instance, at a distance $z = a^2$ corresponding to the Fresnel length ($b=1$), we have

$$\psi_n = (-1)^n \frac{L_{n+1}(\tau)}{n!\, a^{2n}}\exp(-\tau). \tag{45}$$

As a result we find that the space-time structure is described by the expression

$$\psi = \sum_{n=0}^{\infty} (-1)^n \left(\frac{r^2}{2a^2}\right)^n \frac{L_{n+1}(\tau)}{n!}\exp(-\tau). \tag{46}$$

The fairly rapid convergence of the coefficients (45) makes it possible to drop all but a few terms in the series (46) when the problem is analyzed qualitatively.

At the leading edge ($\tau \approx 0$) the series (42) in powers of $r^2/2$ is alternating and describes the field distribution $\exp\{-r^2/2a^2\}$. The roots of the Laguerre polynomials L_n, which determine the behavior of the coefficients in the series (42), are given by the well-known[16] expression

$$\tau_m = \frac{j_m^2}{2(2n+1)}\left[1 + \frac{j_m^2 - 2}{24(2n+1)}\right], \tag{47}$$

where m is the order of the root, and j_m is the root of the zeroth-order Bessel function $J_0(j)$. On the system axis ($r=0$) the field vanishes and changes sign at $\tau = 1$. The study of the roots (47) shows that there are ranges of τ in which two successive expansion coefficients have the same sign. For example, for $1 > \tau > 0.7$, the coefficients ψ_0 and ψ_1 are positive, and allowance for higher-order terms in r^2 facilitates localization of the field distribution in the transverse direction. This means that fairly strong inhomogeneities of the wave field are concentrated near the trailing edge of the pulse ($\tau = 1$).

At distances z much larger than the Fresnel length ($z \gg a^2$ and $b \gg 1$), Eq. (44) yields

$$\psi_1 \approx -\frac{L_1(z\tau/a^2)}{a^2}\exp\left(-\frac{z\tau}{a^2}\right). \tag{48}$$

The dependence of ψ_1 on the self-similar variable $\eta = z\tau/a^2$ simplifies the recurrence relation (43) substantially. More than that, using the well-known functional relationships for the Laguerre polynomials,[17] we can derive expressions similar to Eqs. (45) for the coefficients ψ_n of the Taylor series (42):

$$\psi_n = (-1)^n \frac{L_n(z\tau/a^2)}{n!\, a^{2n}}\exp\left(-\frac{z\tau}{a^2}\right). \tag{49}$$

Note that the coefficient ψ_0, which describes the behavior of the field on the system axis ($r=0$), is not self-similar. In the limit $z \gg a^2$,

$$\psi_0 = -\frac{a^2}{z}\exp(-\tau) + \exp\left(-\frac{z\tau}{a^2}\right). \tag{50}$$

Now let us examine the spatial structure of the part of the field that refers to the diffraction precursor. If in (50) we ignore the first term, we get

$$\psi_{\rm pr} = \sum_{n=0}^{\infty} \frac{(-1)^n L_n}{n!}\frac{z\tau}{a^2}\left(\frac{r^2}{2a^2}\right)^n \exp\left(-\frac{z\tau}{a^2}\right). \tag{51}$$

If we now use the formula for the generating function of the Laguerre polynomials,[17] we arrive at the expression

$$\psi_{\rm pr} = I_0\left(2\sqrt{\frac{z\tau}{a^2}\frac{r^2}{2a^2}}\right)\exp\left(-\frac{r^2}{2a^2} - \frac{z\tau}{a^2}\right), \tag{52}$$

where I_0 is the modified Bessel function.

In view of the linearity of the starting equation, we conclude that (52) is a solution of Eq. (2). The structure of this solution reflects the symmetry of the equation under inter-

change of z and τ. The presence of translational symmetry ($z \to z + iz_0$ and $\tau \to \tau + i\tau_0$) makes it possible, as in Sec. 4, to carry out a complex generalization of the solution (52). In a way similar to (28), this solution describes the transverse focusing of a spatially bounded pulse near $z \approx 0$.

The solution (52) shows that a field localized near the leading edge ($\tau \geqslant 0$), for $r \gg a$ is localized near the parabola $r^2 = 2z\tau$,

$$\psi_{\rm pr}(r \gg a) \approx \frac{a}{(2z\tau r^2)^{1/4}} \exp\left[- \frac{(r - \sqrt{2z\tau})^2}{2a^2} \right], \tag{53}$$

with a characteristic transverse scale of order a and a time (longitudinal) scale of order a^2/z. The amplitude value of the field ($r = \sqrt{2z\tau}$) decreases along the path of pulse propagation according to the law

$$\psi_{\rm pr} \propto \frac{1}{\sqrt{z\tau}}, \tag{54}$$

i.e., much slower than by the z^{-1} law for quasimonochromatic pulses

We have studied some aspects of the evolution of spatially bounded pulses with a spectral width on the order of the carrier frequency in the process of unidirectional (reflectionless) propagation in vacuum. As in the case of quasimonochromatic radiation, it is possible to analyze the variation in the characteristic parameters of the light pulse by the method of moments. We have shown that the center of mass of the wave field moves in a straight line with a group velocity less than the speed of light and that the effective transverse size increases in proportion to z. We find a broad class of self-similar solutions that modify structures having the form of Gaussian wave beams in quasioptics for light pulses.

For initial distributions of a more general form it is shown that a diffusion precursor is formed in the process of pulse propagation. The spatial structure of this part of the field is horseshoe-shaped [see Eq. (53)]. The amplitude of the field decrease slower than it does in the case of a quasimonochromatic pulse.

The author is grateful to the researches at the Institute of Applied Physics of the Russian Academy of Sciences and especially to A. G. Litvak, V. E. Semenov, A. I. Smirnov, and A. A. Balakin for fruitful discussions. The work was supported by Grants from the Russian Fund for Fundamental Research (Grants 98-02-17205 and 99-02-16399).

*)E-mail: mironov@appl.sci-nnov.ru

[1] H. F. Harmuth, *Nonsinusoidal Waves for Radar and Radio Communication*, Academic Press, New York (1981).
[2] L. D. Bakhrakh and A. A. Bliskavitskiĭ, Usp. Fiz. Nauk **162**(12), 160 (1992) [Sov. Phys. Usp. **35**, 1086 (1992)].
[3] D. M. Mittleman, R. H. Jacobsen, and M. C. Nuss, IEEE J. of selected topics in Quan. Elect. **2**, 679 (1996).
[4] É. M. Belenov and A. V. Nazarkin, JETP Lett. **53**, 200 (1991).
[5] T. T. Wu, J. Appl. Phys. **57**, 2370 (1985).
[6] L. G. Sodin, Radiotekh. Elektron. **37**, 1014 (1991); O. A. Tret'yakov and A. N. Dumin, Électromag. Volny Elektron. Sistemy **3**, 12 (1998).
[7] R. W. Ziolkowski, Phys. Rev. A **39**, 2005 (1989).
[8] N. George and S. Radic, Opt. Commun. **139**, 1 (1997).
[9] V. A. Aleshkevich and V. K. Peterson, JETP Lett. **66**, 344 (1997).
[10] B. Rau, T. Tajima, and H. Hojo, Phys. Rev. Lett. **78**, 3310 (1997).
[11] J. N. Brittingham, J. Appl. Phys. **54**, 1179 (1983).
[12] M. B. Vinogradova, O. V. Rudenko, A. P. Sukhorukov, *Theory of Waves* [in Russian], Nauka, Moscow (1990).
[13] M. I. Rabinovich and D. I. Trubetskov, *Oscillations and Waves in Linear and Nonlinear Systems*, Kluwer Academic, Dordrecht (1989).
[14] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford (1975).
[15] S. N. Vlasov, V. A. Petrishchev, and A. I. Talanov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **14**, 1453 (1971); S. N. Vlasov and A. I. Talanov, *Self-focusing of Waves* [in Russian], Institute of Applied Physics, Russian Academy of Sciences, Nizhniĭ Novgorod (1997).
[16] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun (Eds.), National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).
[17] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products*, Academic Press, New York (1980).

Translated by Eugene Yankovsky

Fine structure of splitting of the separatrix of a nonlinear resonance

V. V. Vecheslavov*)

G. I. Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia
(Submitted 22 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 336–346 (July 1999)

This report is a continuation of an analysis, initiated elsewhere {V.V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **114**, 1516 (1998) [JETP **86**, 823 (1998)]}, of the effect of splitting of the separatrix of a nonlinear resonance for the model of standard mapping, based on results of direct measurements of the splitting angle $\alpha(K)$, where K is the system parameter. Measurements were made in the previously used wide range $0.1 \geqslant \alpha \gtrsim 10^{-208}$ ($1 \geqslant K \geqslant 0.0004$), but with significantly higher relative (better than 10^{-50}) and average ($\sim 10^{-55}$) accuracy. This procedure made it possible to substantially refine the effects observed in Ref. 1 and construct qualitatively new empirical dependences providing reliable extrapolation of the data obtained for the angle and the invariant in the intermediate asymptotic limit $K \lesssim 10^{-2}$ beyond the limits of the investigated region. The results obtained by us can be useful for further development of the theory of separatrix splitting and formation of the stochastic layer of a nonlinear resonance. © *1999 American Institute of Physics.* [S1063-7761(99)02407-5]

1. STATE OF THE PROBLEM

The interaction of nonlinear resonances and the chaotic regimes of dynamical Hamiltonian systems arising as a result are among the most important and complicated problems of the contemporary theory of nonlinear oscillations.[2–5] Usually, the initial states of the system are chosen near one of the resonances, which is assumed to be the main or leading resonance, and the others are treated as perturbations. In many cases the problem reduces to an examination of a dynamical model which can be interpreted as a pendulum (the main resonance) subjected to the action of periodic or quasiperiodic forces. In the vicinity of the separatrix of the main resonance chaotic nonlinear oscillations arise as a result of the action of almost any perturbation, no matter how weak.[2–5]

Studies in this field are generally accompanied by wide-scale numerical experiments, in which instead of differential equations in continuous time it is much more effective to use their discrete analogs, namely, mappings.[1–5] A simple but extraordinarily interesting and very popular model of this kind is Chirikov's so-called standard map:[2]

$$\overline{p} = p + K \cdot \sin x, \quad \overline{x} = x + \overline{p}. \tag{1.1}$$

Here p and x are the action–angle variables, and K is the only parameter of the model which characterizes the effect of the perturbation in the period of the mapping $T = 2\pi/\Omega = 1$.

Many papers, including ours, focused on the study of the chaotic dynamics of the standard mapping. Before discussing the new results obtained by us it would be fitting to briefly review the main definitions and the state of the problem.

The leading resonance of system (1.1) can be described by the Hamiltonian of a ''pendulum''

$$H_1(p, x) = \frac{p^2}{2} + K \cos x. \tag{1.2}$$

Recall that in the case of strong nonlinearity such a simple form of the resonance Hamiltonian turns out to be universal.[2] The most important characteristic of a pendulum (1.2) for the problem under discussion is the separatrix

$$p_s = \pm 2\omega_0 \sin\frac{x}{2}, \quad H_1^{(s)} = \omega_0^2 = K \tag{1.3}$$

the singular trajectory, which separates oscillations of the phase (in resonance) from its rotations (out of resonance). In fact, Eqs. (1.3) describe two spatially coincident branches in the time limits $t \to +\infty$ and $t \to -\infty$, respectively. Each branch is an asymptotic trajectory with infinite period of motion, which departs from the position of unstable equilibrium (a saddle point) and returns to it. Almost any perturbation, no matter how weak, splits the separatrix into two intersecting trajectories, which, as before, depart from the saddle point in opposite directions, but never return to it (this effect was described qualitatively by Poincaré in the last century).[6] The two branches of the split separatrix intersect, in particular, at the angle α at $x = \pi$ (the central intersection) and some $p_s(\pi) \approx p_0 = 2\omega_0$ [see Eqs. (1.3)]. The free ends of the split separatrix form an infinite number of loops of unboundedly growing length,[6,7] which, however, fill up a bounded and narrow region along the unperturbed separatrix, thereby forming the so-called stochastic layer—a nucleus and source of the chaos of nonlinear oscillations.[2–5,8,9] The most important characteristic of this layer from the viewpoint of applications is the energy half-width $w_s = H/K - 1$, which, however, can be found only approximately.[2] In this regard, it is important to mention that for the standard map (1.1) the angle α of the central intersection of the branches of the separatrix turns out to be the only exact parameter of chaos that can be calculated with arbitrary accuracy.

Significant progress in the study of the splitting of the separatrix of the standard map was achieved relatively recently (1984). It was connected with the appearance of a series of mathematical papers,[10–12] in which the angle α was determined quite accurately by numerical solution of an auxiliary equation from which an exponential factor was eliminated. Note that the work of physicists of this period and of earlier periods[2–5,8,9] has been mainly oriented toward studying the effect of violation of adiabaticity and finding approximate estimates of the dimensions of the chaotic layer.

For the asymptotic ($\lambda = 2\pi/\sqrt{K} \to \infty$) value of the splitting angle $\alpha_\infty = \alpha(\infty)$ Lazutkin *et al.*[10] obtained the expression

$$\alpha_\infty = \pi \mathscr{L} \frac{e^{-\pi\lambda/2}}{K}, \tag{1.4}$$

in which an important numerical characteristic of the standard mapping, $\mathscr{L}$, is used (in Ref. 1 it was proposed to call it the Lazutkin constant).

It should be noted that a correction factor $f \sim 1$ is used in the theory of the chaotic layer developed by Chirikov[2] (1979). The physical meaning of this factor is to allow for effects of higher-order perturbing resonances: for the standard mapping. He also obtained a first estimate $f \approx 2.15$ numerically. The Lazutkin constant encountered later turned out to be related to this factor by the formula $\mathscr{L} = 16\pi^3 f$ [see formula (1.16) in Ref. 1 and the commentary on it]; therefore, any refinement of the constant $\mathscr{L}$ is simultaneously a refinement of the factor f in the theory of the stochastic layer. The most accurate of the values of this constant published to date, to the best of our knowledge, is [formula (4.14) in Ref. 1]

$$\mathscr{L} = 1118.827705940900778415146393235 66 \pm 3 \times 10^{-27}. \tag{1.5}$$

Lazutkin *et al.*[10] have also estimated the correction to α_∞ in the intermediate asymptotic region $0 < K \ll 1$:

$$c_\alpha(\lambda) = \frac{\alpha(\lambda)}{\alpha_\infty} - 1, \tag{1.6}$$

which was later studied in detail in Ref. 1 and will be considered in this paper (see Secs. 2 and 3).

The next big step in the study of this problem was made by Gelfreich et al.,[11], where the dependence (1.6) was represented in the form of an asymptotic series in appropriate powers of a small parameter. The values of the first four coefficients of this series were determined by numerical solution of auxiliary equations.

This theory introduced the important change of variables $(K, \alpha) \to (h, \sigma)$, where

$$h(K) = \ln\left(1 + \frac{K}{2} + \sqrt{K + \frac{K^2}{4}}\right) \approx \sqrt{K} \tag{1.7}$$

is a positive characteristic index of the tangent (linearized) map (1.1) at the unstable fixed point $x = p = 0$,

$$\sigma(h) = \nu(h) \sin\alpha \tag{1.8}$$

is the simplectic invariant, and $\nu(h)$ is a norm of the tangent vectors.

An approximate solution of the problem was found in Ref. 11. This solution can be written as a correction to the invariant [by analogy with the correction (1.6) to the angle]:

$$c_\sigma(h) = \sigma(h)/\sigma_\infty - 1, \tag{1.9}$$

where $\sigma_\infty = 4\alpha_\infty$ [see Eq. (1.4)].

It is important to emphasize that these ''corrections'' give the most complete description of the intermediate asymptotic behavior and their study enables one to elucidate the ''fine structure'' of splitting of the separatrix and formation of the stochastic layer.

Just recently, a fundamentally different approach to the problem, based on the results of direct measurements of the separatrix-splitting angle of the standard map (1.1) over a very wide range of variation of the parameter K: $1 \geqslant K \geqslant 0.0004$ ($1 \gtrsim h \gtrsim 0.02$): $0.1 \gtrsim \alpha \gtrsim 10^{-208}$ with guaranteed relative accuracy better than 10^{-25} and average accuracy $\sim 10^{-30}$ was realized in Ref. 1. Accordingly, we developed a special technique implementing a software package,[13] which realized the full potential of the standard computer language FORTRAN with arbitrary numerical accuracy (the number of significant digits of the mantissa in the decimal representation of a real number in Ref. 1 reached 300). To compare it with the theory developed in Ref. 11, we also calculated the invariant σ (1.8); here the function $\nu(h)$ (an analytical expression for it is not known) was determined numerically using a special computer program.[14]

In line with the theory developed in Ref. 11, the corrections (1.6) for the angle and (1.9) for the invariant were sought in the form of a finite series in even powers of h

$$\tilde{c}(h) = a(0) + \sum_{m=1}^{M} a(m) h^{2m}. \tag{1.10}$$

In order to facilitate direct comparison of the experimental data with the results of Ref. 11 for the invariant, in addition to the finite series (1.10), we also used a representation in the form of a Taylor series

$$\frac{\sigma}{\sigma_\infty} \mathscr{L} = \mathscr{L} + \sum_{m=1}^{\infty} b(m) h^{2m}, \quad b(m) = a(m) m! \mathscr{L}. \tag{1.11}$$

The least-squares method has been widely used in the search for the coefficients of series (1.10) and the construction of empirical dependences (see, e.g., Ref. 15); a special importance was attributed to monitoring the accuracy of the calculations and to obtaining a reliable estimate of the errors of the results.

The theory developed in Ref. 11 was confirmed in Ref. 1, both qualitatively [series of the form (1.10)] and quantitatively (see Table II in Ref. 1). The construction of simple empirical dependences (4.2) of the expansion coefficients of series (1.10) on the index m can be assumed to be one of the more important results of this study. With them is connected the possibility of extrapolating these series beyond the limits of the region of direct measurements. The unreliability of such an extrapolation, however, was also noted in Ref. 1,

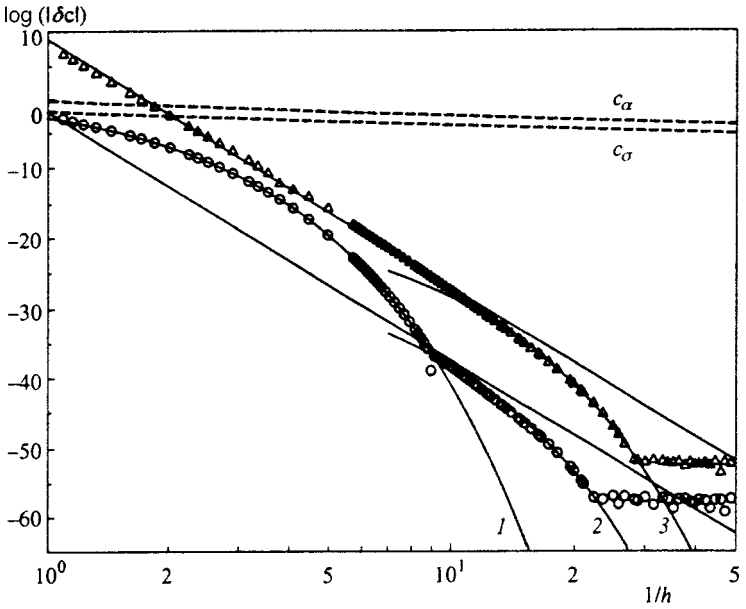

FIG. 1. Results of an interpolation of empirical data on the separatrix splitting angle (triangles) and the invariant (circles): $\delta c(h)$ is the deviation of the correction from the intermediate asymptotic limit (1.10), common logarithm. The downward sloping lines are the first term of the residual term (3.5); curves *1*, *2*, and *3* plot the exponential deviations (see Sec. 3). The upper dashed lines show the total corrections $c(h)$ for the angle (1.6) and the invariant (1.9), respectively.

since the higher coefficients with $m=8,9,10$, both for the angle and for the invariant, clearly deviate from the assumed dependences (see Fig. 3 in Ref. 1 and the commentary). These deviations were called ''anomalies, whose nature remains an open question and requires further study.'' Jumping ahead, let us clarify the reason for the appearance of these ''anomalies,'' which was clarified later. The relative accuracy of measurements of the angle 10^{-25} used in Ref. 1 turns out (as paradoxical as it may seem) to be insufficient to calculate the necessary number of coefficients for reliable extrapolation (see Sec. 2).

Any extrapolation, even an pproximate one, facilitates not only identification but also interpretation of the characteristic properties of the intermediate asymptotic behavior, which can be of substantial assistance in advancing the theory of separatrix splitting and formation of the stochastic layer of a nonlinear resonance. For this reason, it was deemed necessary to extend the analysis of the model of the standard map based on direct measurements of the splitting angle initiated in Ref. 1.

In this paper we investigate the problem of separatrix splitting for the standard mapping within the previous wide range of variation of the splitting angle $0.1\geqslant\alpha\geqslant10^{-208}$ ($1\geqslant K\geqslant0.0004$), but with significantly higher relative accuracy (better than 10^{-50}) and average accuracy ($\sim10^{-55}$). This factor, together with the extremely accurate value of the Lazutkin constant (1.5), has made it possible not only to refine the effects observed earlier in Ref. 1 (Sec. 2), but also to obtain qualitatively new empirical dependences ensuring reliable extrapolation of series (1.10) for the angle and the invariant beyond the limits of the investigated region (Sec. 3).

2. RESULTS OF MEASUREMENTS

In the measurements of the angle we used the scheme described in detail in Ref. 1: the central intersection point of the branches of the separatrix $x_s=\pi$, $p_s(\pi)$ is sought; then two points are calculated, one on each side of it. This makes it possible to fit the branches by second- and fourth-degree polynomials and calculate two values of their angle of intersection: α_2 and α_4. In general, some number $\mathcal{N}_\alpha$ of first significant digits (with allowance for the roundoff) in the values of α_2 and α_4 coincide. We were able to reach values of $\mathcal{N}_\alpha\geqslant50$, which gave a relative accuracy of the angle no worse than 10^{-50}. Actually, the relative accuracy in the intermediate asymptotic region proved to be somewhat higher and stood on average at $\sim10^{-55}$ (see Fig. 1).

Following this scheme, we found the dependence $\alpha(K)$ for 104 values of the perturbation parameter in the interval $1\geqslant K\geqslant0.0004$, which provided the initial empirical material for all subsequent analysis. The value of the angle was recorded to 100 significant digits, which ensured an accuracy of processing of the results much higher than the accuracy of the experimental data (see below).

It was noted in Ref. 1 that the main difficulty of interpolating using a series of the form (1.10) is that different terms of this series differ by many orders of magnitude and the matrix of the normal system of equations of the least-squares method for searching for the coefficients $a(m)$ in (1.10), as a rule, turns out to be degenerate in its computer representation. It became necessary not only to subject the data to a final processing with accuracy $\sim10^{-100}$, but also to introduce a change of scale of the variables of the problem: $(h,\tilde{c})\to(H=Sh,\ C=S\tilde{c})$, where the scale factor $S\gg1$ had to be suitably chosen. However, in Ref. 1 we had to deal with ten coefficients (1.10), whereas in the present work this number grew to seventeen (see Tables I and II) and the simple change of scale ceased to help. The problem was successfully solved by inverting the direct matrix by first partitioning it into blocks (see, e.g., Ref. 16), where the dimensions of the blocks also had to be suitably chosen. Recall that in the least-squares method the diagonal elements of the inverse matrix determine the weights with which the errors of the unknown quantities are summed, and therefore the inversion operation could not be eliminated.[15]

As was noted in Ref. 1, the number of terms of the series

TABLE I. Coefficients $a_\alpha(m)$ of series (1.10) for the angle.

m	$a_\alpha(m)$	Δ	$\langle\delta\rangle$
1	−0.233376428864381610627639715 651844925562242	0.397×10^{-41}	0.384×10^{-42}
2	−0.290818155124688860367364364 03046408113	0.370×10^{-36}	0.334×10^{-37}
3	−0.014824955534894051786788200 5854733	0.212×10^{-32}	0.177×10^{-33}
4	0.043182190148644921649679410 03978	0.836×10^{-29}	0.647×10^{-30}
5	−0.041519239478427464679586098	0.241×10^{-25}	0.171×10^{-26}
6	−0.131373309408107983341741	0.525×10^{-22}	0.342×10^{-23}
7	−0.319169849155133631687	0.885×10^{-19}	0.525×10^{-20}
8	−1.060531457633276423	0.117×10^{-16}	0.630×10^{-17}
9	−4.38156420631767	0.122×10^{-12}	0.593×10^{-14}
10	−21.62868101831	0.101×10^{-09}	0.440×10^{-11}
11	−126.24207274	0.662×10^{-07}	0.255×10^{-08}
12	−861.834118	0.337×10^{-04}	0.115×10^{-05}
13	−6810.600	0.131×10^{-01}	0.395×10^{-03}
14	−61716.8	$0.380\times10^{+01}$	$0.100\times10^{+00}$
15	$-6.3755\times10^{+05}$	$0.771\times10^{+03}$	$0.178\times10^{+02}$
16	$-7.15\times10^{+06}$	$0.981\times10^{+05}$	$0.199\times10^{+04}$
17	$-1.3\times10^{+08}$	$0.591\times10^{+07}$	$0.107\times10^{+06}$

Note. For the number of digits of the coefficient exceeding the width of the column, the remaining digits are written in the same column, one row lower.

M accessible in practice was bounded from above by errors of calculation due mainly to the ''noise'' caused by the finite accuracy of the empirical data $\alpha(h)$. A radical increase in this accuracy (from 10^{-25} to 10^{-50}) and, as a result, a radical reduction of the level of this ''noise'' made it possible to increase the number of coefficients found in the present work to $M=17$ (versus 10 in Ref. 1). With further increase of M, we not only do not obtain new coefficients, but we lose the ones we already have (see Fig. 1 in Ref. 1).

The accuracy of the empirical dependence (1.10) is characterized by the root-mean-square error

TABLE II. Coefficients $b(m)$ of series (1.11) for the invariant.

m	$b_\sigma(m)$	Δ	$\langle\delta\rangle$
1	18.598911958209297358817149041692488164817654712988	0.113×10^{-44}	0.142×10^{-45}
2	−4.34114127056816253677582933 04979013479225149	0.158×10^{-40}	0.183×10^{-41}
3	−4.183263759091894137233272356503100172193 6	0.199×10^{-36}	0.212×10^{-37}
4	−4.934139590730879408563420529301 57	0.226×10^{-32}	0.219×10^{-33}
5	−10.645486442818204235356421215457	0.230×10^{-28}	0.203×10^{-29}
6	−35.860081669350475971059555 3	0.209×10^{-24}	0.166×10^{-25}
7	−177.6036528919052715929381	0.169×10^{-20}	0.121×10^{-21}
8	−1239.435334988728110840	0.121×10^{-16}	0.769×10^{-18}
9	−11806.115613907542005	0.757×10^{-13}	0.427×10^{-14}
10	−149425.48996254799	0.412×10^{-09}	0.204×10^{-10}
11	−2454832.5606990	0.193×10^{-05}	0.835×10^{-07}
12	−51297631.6298	0.762×10^{-02}	0.286×10^{-03}
13	$-1.339922127\times10^{+09}$	$0.248\times10^{+02}$	$0.802\times10^{+00}$
14	$-4.3097121\times10^{+09}$	$0.647\times10^{+05}$	$0.178\times10^{+04}$
15	$-1.6850\times10^{+12}$	$0.126\times10^{+09}$	$0.295\times10^{+07}$
16	$-7.88\times10^{+13}$	$0.166\times10^{+12}$	$0.331\times10^{+10}$
17	$-4.9\times10^{+15}$	$0.110\times10^{+15}$	$0.191\times10^{+13}$

$$\Delta c=\langle(c(h)-\tilde{c}(h))^2\rangle^{1/2}, \tag{2.1}$$

where the angular brackets denote averaging over the interpolation interval. Clearly, we cannot use the entire empirical interval as this interval since the power-law dependence (1.10) in the theory developed in Ref. 11 characterizes only the intermediate asymptotic limit. Therefore, the deviation

$$\delta c(h)=c(h)-\tilde{c}(h) \tag{2.2}$$

contains the most valuable and interesting information about the additional nonadiabatic effects not comprehended in the theory. Thus, as in Ref. 1, it was necessary in a given set of empirical data to also choose the optimal interpolation interval (h_1-h_2), where $h_1=0.02$ was the minimum value in the initial data. The smallest value of Δc (2.1), together with the accuracy of the found coefficients of series (1.10), served as the main criterion here (see Tables I and II).

The quality of the interpolation worsened not only when h_2 was increased, as was noted above, but also when it was decreased due to the small contribution of the higher powers of h for small value of h, and also due to a decrease in the number N_p of points participating in the interpolation. The main results were obtained using the standard interpolation of minimizing the variance $(\Delta c)^2$ [see (2.1)] and are presented in Tables I and II and Fig. 1.

The accuracy of the found coefficients, as in Ref. 1, was estimated by two different methods. First, the standard deviation (root-mean-square error) of the interpolation was calculated[14], $\langle\delta\rangle$ (column 4 in the tables), i.e., the expected error of the coefficients of the random mean-square error of the initial empirical data. To take errors of a different nature into account, namely systematic errors, the values of the coefficients were also determined as means over several interpolations with a different number of data points in each: $N_p=20-28$ for the angle and $N_p=25-35$ for the invariant. In fact, these values are listed in the tables (column 2). A relatively weak dependence of the means on N_p served as the main criterion in our choice of these two groups. As the error we adopted the root-mean-square error of the values of the coefficients in a group, Δ (column 3 in the tables). It can be seen that the error in a group is greatest (and therefore defining) in all cases. The difference in the two errors is a definite indication of substantial systematic errors. The values of the root-mean-square errors in a group probably also determine the number of reliable decimal places of the coefficients.

3. DISCUSSION OF RESULTS AND CONCLUSIONS

Following the scheme laid out in Ref. 1, let us consider first of all the behavior of the coefficients of expansions (1.10), starting with the data in Tables I and II.

Let us begin from the fact observed in Ref. 1 that the ratio of the coefficients of the angle and the invariant has an exponential form (the straight line passing through the squares in Fig. 2):

$$a_\alpha/a_\sigma(m)=R\,e^{\gamma m}, \quad R=0.7471\pm0.0539,$$
$$\gamma=1.3579\pm0.0089. \tag{3.1}$$

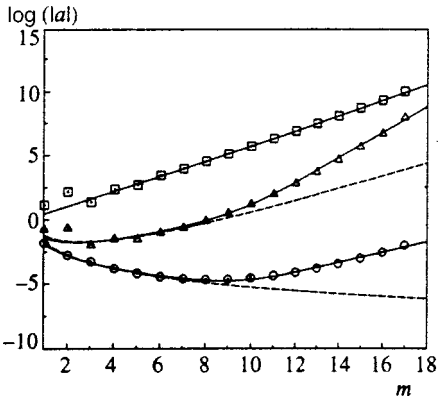

FIG. 2. Variation of the coefficients of the intermediate asymptotic limit (1.10) for the angle a_α (triangles), the invariant (a_σ) (circles), and their ratio a_α/a_σ (squares). The solid curves plot the empirical dependences (3.1) and (3.2) based on the data of our work. The thick points inside the symbols for $m\leqslant 10$ mark the values of the coefficients from Ref. 1, the dashed lines plot the interpolated dependences obtained there.

Taking this into account, and the analysis presented in Ref. 1, we can describe the interpolated dependences by means of the approximate expressions

$$a_\sigma(m)\approx\frac{A_\sigma}{m^p}(1+Qe^{qm}) \tag{3.2a}$$

for the invariant (the curve passing through the circles in Fig. 2) and

$$a_\alpha(m)\approx A_\sigma R\frac{e^{\gamma m}}{m^p}(1+Qe^{qm}) \tag{3.2b}$$

for the angle (the curve passing through the triangles in Fig. 2). Using the values of R and γ from Eqs. (3.1), we replace the coefficients for the angle $a_\alpha(m)$ by their reduced values

$$\tilde{a}_\alpha(m)=\frac{a_\alpha(m)}{R}e^{-\gamma m}=\frac{A_\sigma}{m^p}(1+Qe^{qm})\,. \tag{3.3}$$

Comparison of expressions (3.2a) and (3.3) shows that finding the values of the unknown quantities A_σ, p, Q, and q by the least-squares method can be done by using a combined set $\{a_\sigma(m),\tilde{a}_\alpha(m)\}$ after first removing some of the first terms from it which clearly deviate from the interpolated dependence (see Fig. 2 and Tables I and II). In fact, the interpolation was not done over the coefficients themselves, but over the common logarithms of their absolute values (see Fig. 2). The interpolation involved 29 points and yielded the following results:

$$A_\sigma=-0.01571\pm0.00480,\quad p=3.188\pm0.171,$$
$$Q=2.224\times10^{-5}\pm9.93\times10^{-6},\quad q=1.127\pm0.028. \tag{3.4}$$

The root-mean-square error of this interpolation was ≈0.092, and results of this interpolation are plotted in Fig. 2.

Comparison of expressions (3.2) (they correspond to the solid curves in Fig. 2) with (4.2) from Ref. 1, which correspond to the dashed curves in Fig. 2, shows that processing substantially more accurate and more extensive experimental data obtained by us leads to qualitatively different experimental dependences in regard to the ratio of the coefficients of series (1.10) because of the appearance in expressions (3.2) of the additional factor $[1+Q\exp(qm)]$. For $m<8$, however, the two families of curves are quite similar. It is interesting to note that dependence (3.1) for the ratio of the coefficients of the angle and the invariant did not undergo any change.

It is important to emphasize that since the empirical expressions (3.2) obtained here in the intermediate asymptotic limit $h\lesssim0.1$ do not demonstrate any ''anomalies'' (see Fig. 2), their use is not limited to the number of coefficients actually found. This makes it possible, with the help of (3.2) and (3.4), to estimate quite accurately the residual terms of series (1.10) not included in the interpolation (the solid downward-sloping lines in Fig. 1):

$$R(h,M)=\sum_{m=M+1}^{\infty}a(m)\,h^{2m}\approx a(M+1)h^{2M+2}\,. \tag{3.5}$$

The above analysis of the coefficients makes it possible to more completely represent the global behavior of the investigated dependences. First of all, both series—the series for the angle and the series for the invariant—diverge within the investigated range for some $h>h_{cr}$: for the angle $h_{cr}\approx\exp(-(\gamma+q)/2)\approx0.287$, and for the invariant $h_{cr}\approx\exp(-q/2)\approx 0.569$. However, the character of these dependences for $1\gtrsim h\gtrsim0.1$ (before the intermediate asymptotic region) is qualitatively different.

The variation of the invariant is very well described by a ''cascade of exponentials.'' The first of these exponentials (curve *1* in Fig. 1)

$$|\delta c_1(h)|\approx63e^{-\pi^2/h}, \tag{3.6}$$

which was discovered and received a clear explanation in Ref. 1, significantly exceeds the residual term (3.5). It describes the perturbation of the separatrix by a more distant resonance with frequency $2\Omega=4\pi$. The simple theory developed in Ref. 2 predicts a pre-exponential factor of 8, i.e., almost an order of magnitude less. This difference, however, is completely explainable by a very complicated (in the case under consideration) system of resonances of higher approximations.

The second exponential (curve *2* in Fig. 1, the interpolation is over 34 points, the root-mean-square error ≈0.13)

$$|\delta c_2(h)|=A_2e^{-\beta_2/h},\quad \log(A_2)=-23.45\pm0.11,$$
$$\beta_2=3.512\pm0.017, \tag{3.7}$$

in contrast, is found entirely below the residual term. The mechanism of its appearance for the present state of the theory is completely unclear. The similarity of the value of the factor β_2 in the argument of the exponential to π is noteworthy.

A curious situation arises for the angle: in the region of divergence of the series representing it, no singularities or anomalies are observed in the behavior of the function $c_\alpha(h)$ or its deviation $\delta c_\alpha(h)$ from the interpolation. Moreover, the latter is described completely satisfactorily by the residual term (3.5), which is also represented in Fig. 1 only by its first

term. Here the transition to the intermediate asymptotic region is realized by the exponential (curve *3* in Fig. 1, the interpolation is over 22 points, the root-mean-square error ≈ 0.12):

$$|\delta c_3(h)| = A_3 e^{-\beta_3/h}, \quad \log(A_3) = -15.84 \pm 0.15,$$

$$\beta_3 = 2.906 \pm 0.018. \tag{3.8}$$

In this case, the value of the factor β_3 in the argument of the exponential is also close to π.

All these arguments, taken collectively, show that the behavior of the angle and the invariant at large values of h, $h \sim 1$, differs substantially from their behavior in the intermediate asymptotic region, $h \lesssim 0.1$

In conclusion, we point out that the data obtained by us in many ways confirm and refine the conclusions and results of our earlier study (Ref. 1). But what should probably be considered to be the main result here is the construction of qualitatively new experimental dependences (3.1)–(3.4), which makes it possible to reliably extrapolate series (1.10) beyond the limits of the investigated region of direct measurements. This factor, in our view, can markedly advance further development of the current state of the theory of separatrix splitting and formation of the chaotic layer of a nonlinear resonance.

I am deeply grateful to B. V. Chirikov for discussions and advice. This work was carried out with the partial financial support of the Russian Fund for Fundamental Research (Grant No. 97-01-00865).

*)E-mail: vecheslavov@inp.nsk.su

[1] V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **114**, 1516 (1998) [JETP JETP **86**, 823 (1998)].
[2] B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[3] G. M. Zaslavskiĭ and B. V. Chirikov, Usp. Fiz. Nauk **105**, 3 (1971) [Sov. Phys. Usp. **14**, 549 (1972)].
[4] R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* (Harwood, Chur, 1988).
[5] A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer, 1992).
[6] A. Poincaré, *Les méthodes nouvelles de la mécanique céleste* (Paris, 1892), p. 226.
[7] V. K. Mel'nikov, Trudy Mosk. Mat. Obshchestva **12**, 3 (1963).
[8] N. N. Filonenko, R. Z. Sagdeev, and G. M. Zaslavsky, Nucl. Fusion **7**, 253 (1967).
[9] G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1968) [Sov. Phys. JETP **27**, 851 (1968)].
[10] V. F. Lazutkin, ''Splitting of the separatrix of the standard Chirikov mapping'' [in Russian] Dep. VINITI 6372-84 (1984); V. F. Lazutkin, I. G. Schachmanski, and M. B. Tabanov, Physica D **40**, 235 (1989); V. G. Gelfreich, V. F. Lazutkin, and M. B. Tabanov, Chaos **1**, 137 (1991).
[11] V. G. Gelfreich, V. F. Lazutkin, and N. V. Svanidze, Physica D **71**, 82 (1994).
[12] V. G. Gelfreich, ''A proof of the exponentially small transversality of the separatrices for the standard map.'' Freie Universität Berlin, Preprint 9/98 (1998), p. 56 (to appear in *Commun. Math. Phys.*).
[13] D. H. Bailey, *ACM Transactions on Mathematical Software*, Vol. 19 (1993), p. 288.
[14] V. G. Gelfreich, private communication.
[15] B. M. Shchigolev, *Mathematical Processing of Observations* [in Russian] (Fizmatgiz, Moskva, 1960); D. Hudson, *Statistics* (Geneva, 1964)
[16] C. Lanczos, *Applied Analysis* (Prentice-Hall, Englewood Cliffs, N. J., 1957).

Translated by Paul F. Schippnick

Photon echo and stimulated photon echo study of various collisional-relaxation channels

N. N. Rubtsova, L. S. Vasilenko, and E. B. Khvorostov*)

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090, Novosibirsk, Russia

(Submitted 28 October 1998)
Zh. Éksp. Teor. Fiz. **116**, 47–56 (July 1999)

Collisional relaxation in SF_6 gas and its mixtures with He and Xe is studied by photon echo and stimulated photon echo methods from the standpoint of the possibility of identifying the contributions of different types of collisions. The nonexponential nature of the kinetic curve of the photon echo is clearly observed for pure SF_6, it is weaker in the mixture SF_6+Xe, and it is virtually completely absent for high degrees of dilution of SF_6 with helium. These features can be explained on the basis of estimates, made from experimental data, of the critical delay between the exciting pulses (for which the nonexponential behavior should be most strongly manifested). In pure SF_6 it is possible to distinguish the contribution of the inelastic channel (rotational relaxation) and the contribution of weak collisions. To distinguish successfully the relaxation channels in mixtures with buffer gases a heavier buffer gas and a much better time resolution must be used. It is shown that data obtained on the orientation and alignment relaxation rates by the stimulated photon echo method can serve as an upper limit for the rates of inelastic processes which cannot be measured by the photon echo method. The combined use of photon echo and stimulated photon echo methods made it possible to obtain data on the cross sections for elastic and inelastic scattering of the collisional pairs SF_6-SF_6, SF_6-Xe, and SF_6-He. [S1063-7761(99)00507-7]

1. INTRODUCTION

The shape of an isolated spectral line of a neutral gas is determined by the translational motion of the particles (inhomogeneous line broadening — Doppler effect, line narrowing — Dicke effect), the interaction of particles with external electromagnetic fields or zero-point vibrations of the field (splitting or shift of a spectral line due to the dynamic Stark effect or natural line broadening), as well as by the interaction of particles in collisions (collisional homogeneous broadening). These processes can be closely interrelated, example being the control of the translational motion of particles by laser radiation[1] (localization of atoms in the field of a standing wave), the detection of collisional transitions in a nonresonant radiation field[2] (radiative collisions), and the observation of asymmetry of the Doppler contour due to the statistical dependence of collisions and translational motion of the particles.[3]

The methods of nonlinear laser (so-called Doppler-free) spectroscopy largely eliminate the contribution of translational motion to line broadening, and performing experiments in weak electromagnetic fields decreases the contribution of field-induced defects to the natural line broadening, which for molecular vibrational-rotational IR transitions is small by virtue of the low probabilities of spontaneous transitions. For this reason, the methods of Doppler-free spectroscopy of molecules make it possible to concentrate investigations on homogeneous collisional line broadening.

Collisions change the internal state of the active particle (absorbing or emitting) and/or the state of the translational motion of the particle. A number of parameters of the active particle can change in a single collision, so that distinguishing individual types of collisions becomes problematic. Nonetheless, a description of collision types distinguished by the final result is often used. Collisions with interruption of the phase of the transition dipole moment, elastic scattering by small angles (collisions with a change in the velocity of translational motion), depolarizing collisions (responsible for destroying the polarization moments induced by polarized radiation on degenerate resonance levels; these collisions characterize the asymmetry of the interaction potential of the particles), as well as inelastic scattering (for vibrational-rotational transitions this is mainly rotational relaxation) are distinguished.

Knowing the contributions of different collisional relaxation channels could be helpful for studying the characteristic features of the interparticle interaction potentials and for investigating spectral line shapes. When stationary Doppler-free spectroscopy methods are used, information about collisions is contained in the shape of the homogeneously broadened line, and it could be quite difficult to distinguish the contributions of various channels. Our aim in the present paper is to analyze the possibility of using nonstationary, coherent, Doppler-free spectroscopy to distinguish elastic and inelastic scattering channels of molecules.

This work was performed by photon echo and stimulated photon echo methods. The coherent nature of these phenomena gives these Doppler-free spectroscopy methods a number of advantages over saturated-absorption methods:[4] Under optimal echo excitation conditions coherent spectroscopy meth-

ods give a large increase in the signal/noise ratio and are completely free of field-induced distortions.

2. MANIFESTATION OF VARIOUS INTERMOLECULAR COLLISIONS IN PHOTON-ECHO PHENOMENA

Coherent transient processes, including photon and stimulated photon echo phenomena, have been successfully used now for more than thirty years to investigate relaxation processes in matter. For gases the theoretical description of these phenomena is well developed.[5,6] This makes it possible to use the photon echo and a modification of this method to check theoretical models of collisional interaction of particles experimentally.

We shall present a qualitative picture of the formation of echo responses in a gas of two-level atoms under the action of unidirectional radiation pulses with the same frequency.

Modern photon echo theory is based on the semiclassical approximation. The state of the medium is described by a density matrix, and particle motion as well as relaxational processes are taken into account. The interaction with radiation is studied in the dipole approximation, and the radiation field is assumed to be classical. The dynamics of the interaction of a quantum system with resonant radiation is well known to be determined by the Rabi frequency $\chi = d \cdot \mathscr{E}_i / \hbar$, which characterizes the transition dipole moment d and the intensity $\mathscr{E}_i$ of the electric field of the resonant electromagnetic radiation of the ith pulse. If the duration of a pulse of resonant radiation is sufficiently short, then under the action of the radiation field a two-level system can complete only a part of the period of the Rabi oscillations. This result in the formation of nonequilibrium populations of energy levels and nonzero polarization of the medium. An important parameter of photon echo theory is the area of the exciting pulse

$$\theta_i = \int d \cdot \mathscr{E}_i(t) dt.$$

If echo formation occurs on a narrow spectral line, i.e., if the width of the spectrum of the exciting radiation is greater than the Doppler width, the largest deviation from the equilibrium state is produced by a pulse with area $\theta_1 = \pi/2$. Likewise, in the opposite limiting case of echo formation on a wide spectral line there exists an optimal area of the exciting pulse for the formation of the photon echo signal.

At the moment when the exciting pulse ends, the particles which have interacted with the exciting radiation and have different velocity projections v_z are in-phase. As a result, a macroscopic polarization arises in the system. This polarization engenders coherent spontaneous emission from the gas sample at the frequency of the exciting radiation. However, such a coherence state does not last long, since particles belonging to different velocity groups emit at different frequencies, shifted relative to the line center ω_0 by the amount $\Delta\omega = \omega - \omega_0 = kv_z$ on account of the Doppler effect. In a time of the order of the duration of the exciting pulse after the pulse ends, the difference of the Doppler phases $\phi = kv_z t$ between particles moving different velocities v_z increases linearly with time (Doppler dephasing), as a result of which the coherent emission from the medium decays. This dephasing phenomenon can be detected as a decay of free polarization agains the background formed by the trailing edge of the exciting pulse. Doppler dephasing does not lead to irreversible damping of the excitation — the microscopic polarization of individual particles continues to exist.

The action of a second resonant pulse (if the photon echo is formed on a narrow spectral line, the optimal pulse area is π) can change the sign of the phase for each particle. Further growth of Doppler phases according to the previous law (in proportion to the time t and the velocity projection v_z) creates conditions for restoring the phasing of the radiation of the particles that have interacted with the second exciting pulse at a moment approximately equal to twice the delay time T_{12} between the exciting pulses. At this moment the macroscopic polarization of the medium is restored and spontaneous emission from the particles is detected as coherent emission — the so-called photon-echo signal. The duration of the echo pulse is likewise of the order of the duration of the exciting pulses.

The intensity of a photon echo is sensitive to inelastic collisions, as a result of which the particles leave the energy levels participating in the formation of the photon echo, and to elastic collisions changing the translational velocity (this change affects the Doppler phases). If the change in the longitudinal projection of the velocity is large, which corresponds to a strong-collisions model, the particles which have undergone such collisions cannot then participate in the formation of a coherent response. For small changes in the velocity (weak collisions) the result depends on the delay time between the pulses. Such collisions can have a negligible influence on the amplitude of the echo response for short delays and contribute to damping of the photon echo only for large delays T_{12}. As a result, nonexponential decay kinetics of a photon echo with increasing relaxation rate is observed for long delays.[7] The damping of a photon echo on the initial section is determined by inelastic processes — rotational relaxation, and the total contribution of elastic and inelastic scatterings characterizes the decay rate for long delays T_{12}.

The effect of a second pulse can be not only to change the sign of the Doppler phase, which is equivalent to a phase jump by the amount $\Delta\phi = -2kv_z T_{12}$. The second pulse also converts the nonequilibrium values of the off-diagonal elements of the density matrix to on-diagonal elements, i.e., microscopic polarization into nonequilibrium populations of the energy levels. The Doppler phases $kv_z T_{12}$ accumulated up to this moment are stored in the populations of the energy levels, and they can be stored in the medium for a long time, determined by the irreversible relaxation of the energy levels. A third pulse in the form of a traveling wave arriving with a time delay T_{23} after the second pulse can convert these nonequilibrium populations once again into the off-diagonal components of the density matrix and produce by the phase jump mechanism described earlier the conditions required to observe a coherent echo response at the time $t = T_{23} + 2T_{12}$ — a stimulated photon echo. If an echo is formed on a narrow spectral line, the optimal areas of all three pulses for the amplitude of the stimulated photon echo

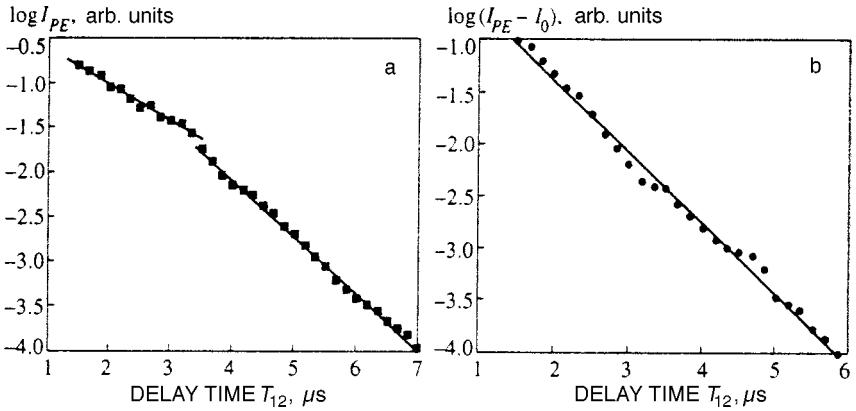

FIG. 1. a — Decay kinetics of the photon echo signal in SF_6 gas at pressure 2.5 mtorr. The gradual change in the slope of the curve reflects the increasing role of elastic scattering by small angles with increasing delay time. The critical delay time is approximately 3.3 μs. The straight lines are least-squares lines drawn through the points before and after a delay of 3.3 μs. b — Decay kinetics of the photon echo signal in the gas mixture 2.5 mtorr SF_6 + 3.5 mtorr He. I_0 — Photon echo intensity at the same SF_6 pressure but without a buffer gas. The curve can be approximated by a straight line, since elastic and inelastic collisions appear in the entire accessible range of delays. The line is a least-squares line drawn through all experimental points.

are $\pi/2$. It is obvious that a stimulated photon echo should be sensitive only to relaxation processes acting on the populations of the resonant energy levels during the delay time T_{23} between the second and third exciting pulses. Weak elastic collisions during this time interval should not affect the amplitude of the stimulated photon echo.

Experiments in a molecular gas ordinarily employ degenerate transitions (typical angular momenta $J = 10 - 100$). This degeneracy must be taken into account when studying the polarization properties of coherent responses. On the other hand, the polarization features of echo generation open up new possibilities for studying the characteristic features of collisional interactions. Thus, in Ref. 8 it was suggested that a stimulated photon echo be produced using light pulses with specially chosen polarizations, making it possible to study the so-called depolarizing collisions, which are simply collisions that destroy the nonequilibrium polarization states (orientation, alignment, and others) formed on the magnetic sublevels of degenerate resonant levels. In in this method, investigation of the decay kinetics of a stimulated photon echo as a function of the delay time T_{23} makes it possible to determine three collisional decay rates for the resonant quantum levels: $\gamma^{(0)}$ — the population relaxation rate, $\gamma^{(1)}$ — the orientation decay rate (the orientation of an energy level by resonant polarized radiation corresponds physically to the production of a macroscopic magnetic-dipole moment in the medium), and $\gamma^{(2)}$ — the alignment decay rate (the alignment of an energy level corresponds to the induction of a macroscopic electric quadrupole moment in the sample). The differences in the values of these relaxation constants make it possible to judge the presence or absence of asymmetry in the interaction potential of the colliding particles.

The above-enumerated features of coherent transient processes give hope that various types of collisions can in principle be studied separately.

3. EXPERIMENTAL RESULTS

The experimental technique, based on the application of radiation from a continuous-wave narrow-band frequency-tunable CO_2 laser, a reference heterodyne laser, a system of electro-optic shutters for forming pulses, as well as laser locking and measurement automating systems, is described in detail in Ref. 4. The photon echo was investigated in the range of delays T_{12} from 1.5 to 10 μs in pur SF_6 gas and in mixtures of SF_6 gas with He and Xe buffer gases on the transition $P(33)A_2^1$ of the vibrational mode ν_3. The exciting pulses were formed, using an electro-optic shutter, from the radiation of a continuous-wave CO_2 laser with of the order of 10 kHz linewidth and with generation frequency near the center of the experimental transition in SF_6.

Figure 1a demonstrates the typical kinetics of the logarithm of the intensity of the photon echo in SF_6 gas with a nonlinear section at the beginning of the curve, attesting to nonexponential decay of the echo signal. As already mentioned in Sec. 2, the nonexponential behavior for short delays between the exciting pulses is due to weak collisions. The additional dephasing of the ensemble of excited particles as a result of weak collisions is determined by the expression

$$\delta\phi = k\,\delta v_z T_{12},$$

where δv_z is the collision-induced change in the projection of the velocity. If $\delta\phi \ll 1$, then the contribution of velocity changing collisions to the decay of the photon echo can be neglected and it can be assumed that the decrease of the photon echo with increasing delay T_{12} is determined only by inelastic relaxation. On the other hand, the condition $k\,\delta v_z T_c \approx \pi$ determines the critical delay T_c between the pulses, so that for $T_{12} > T_c$ the contribution of weak collisions to decay of a photon echo can no longer be neglected and the decay rate reflects the total contribution of inelastic processes and elastic scattering by small angles. Thus, for $T_{12} < T_c$ the decay kinetics of a photon echo is determined by the relaxation rate $\gamma^{(0)}$ of the nonequilibrium populations of the energy levels (rotational relaxation), and the section $T_{12} > T_c$ makes it possible to measure the total relaxation rate

$$\Gamma_{\text{tot}} = \gamma^{(0)} + \Gamma_{\text{vcc}},$$

where Γ_{vcc} is the decay rate of the photon echo as a result of elastic relaxation (velocity changing collisions). In accordance with this interpretation, both relaxation rates $\gamma^{(0)}$ and Γ_{vcc} for pure SF_6 gas can be determined in photon-echo experiments performed for a wide range of delays T_{12}.

The decay kinetics of a photon echo in gas mixtures due to collisions of resonant molecules with buffer-gas particles is somewhat different from decay in a pure gas. For a mixture with xenon it is still possible to distinguish at the start of the kinetic curve a deviation from exponential behavior, whereas in SF_6–He collisions (for delays $T_{12} > 1.5$ μs acces-

TABLE I. Inelastic relaxation rate $\gamma^{(0)}$ and the total rate Γ_{tot} of elastic and inelastic processes measured by the photon echo method and estimates of the characteristic diffraction scattering angle and critical delays T_c.

Collision partners	$\gamma^{(0)}$, 10^6 $s^{-1}\cdot torr^{-1}$	Γ_{tot}, 10^6 $s^{-1}\cdot torr^{-1}$	ρ_W, Å	λ_D, Å	θ, 10^{-3} rad	δv_z, cm/s	T_c, 10^{-6} s
SF_6-SF_6	$\leqslant 36\pm 4.5$	49.8 ± 6.3	8.3	0.03	3.6	145	3.4
SF_6-Xe	not measured	27.6 ± 4.7	6.1	0.029	4.8	199	2.5
SF_6-He	not measured	44.2 ± 6.7	3.8	0.13	3.5	4995	0.1

sible experimentally) a monoexponential decay of the intensity of the photon echo is observed, as Fig. 1b demonstrates.

We shall endeavor to estimate the characteristic scattering angle as the diffraction angle,[9,10] using the ratio $\theta=\lambda_D/\rho_W$, where the de Broglie wavelength $\lambda_D=\hbar/\mu v_{rel}$, μ is the reduced mass of the collision partners, and $v_{rel}=\sqrt{8k_BT/\pi\mu}$ is the relative average velocity. The Weisskopf radius ρ_W can be determined from the experimental data according to the formula $\rho_W\approx 0.365\sqrt{\Gamma_{tot}/n_{buf}v_{rel}}$,[11,12] where n_{buf} is the buffer-gas density. The diffraction-scattering angles calculated in this manner can be used to estimate the average change in the translational velocity in weak collisions according to the relation $\delta v_z\approx\theta v_{rel}$. In turn, the velocity change in elastic scattering by small angles makes it possible to estimate $T_c=\pi/k\delta v_z$ and compare it with the experimentally observed value.

Table I shows the values of $\gamma^{(0)}$, Γ_{tot}, ρ_W, λ_D, θ, δv, and T_c for the collisional pairs SF_6-SF_6, SF_6-He, and SF_6-Xe. Estimates of the critical delay show good agreement with experiment in pure gas, and they explain the weak nonexponential behavior in the mixture with xenon and the absence of such behavior in the mixture with helium. It can be concluded from the results obtained that the investigation of the decay kinetics of a photon echo in a wide range of delays makes it possible to distinguish elastic and inelastic contributions, if both ranges of delays between the exciting pulses are accessible experimentally: $T_{12}<T_c$ and $T_{12}>T_c$. This condition is easier to satisfy by choosing a collision partner with a large mass.

The stimulated photon echo method provides another possibility for distinguishing different types of collisions. A stimulated photon echo was produced in SF_6 gas on the same transition $P(33)A_2^1(0)\rightarrow 1$ of the vibrational mode ν_3 under the action of resonant radiation pulses cut out, using an electrooptic shutter, from continuous-wave linearly polarized radiation from a CO_2 laser operating on the $10P(18)$ line. An additional electro-optic crystal placed after the forming shutter made it possible to rotate by 90° the linear polarization of any of the three excitation pulses separated by time delays T_{12} and T_{23}. The intensity (amplitude) of the stimulated photon echo signal arising at the time $t=2T_{12}+T_{23}$ depends quite strongly on the relative polarization of the exciting pulses — it is highest when the polarization vectors of all three excitation pulses are parallel and lowest when the polarization plane of the first pulse is rotated. Optically heterodyning was used to increase the sensitivity of detection of stimulated photon echo signals corresponding to rotations of the polarization of the first or second exciting pulses.[13] The working laser was tuned to the center of the transition in SF_6, and the laser heterodyne frequency was off set by a fixed amount using a frequency lock system. The strength of the detected signal is proportional to the amplitude of the electromagnetic field of the coherent response of the medium.

The chosen vibrational-rotational transition in SF_6 makes it possible to use the large angular momentum approximation[6,8] (J is the angular momentum of a energy level). Another substantial simplification is due to the fact that the relaxational characteristics of the upper and lower levels of IR transitions are, as a rule, close in magnitude (as is confirmed by the small contribution of phase-interruption collisions to the line broadening[14,15]).

The collisional decay rates of the populations $\gamma^{(0)}$, the orientation $\gamma^{(1)}$, and the alignment $\gamma^{(2)}$ measured by the stimulated photon echo method[13] are presented in Table II.

For gas mixtures $\gamma^{(0)}$ could not be measured by the stimulated photon echo method because of the weakness of the echo signals. However, the measured orientation and alignment relaxation rates can easily serve as an upper limit for the inelastic relaxation rate $\gamma^{(0)}$, since $\gamma^{(\kappa)}=\gamma^{(0)}+\Gamma^{(\kappa)}$,[8] where the positive additive correction $\Gamma^{(\kappa)}$ describes the collisional decay of the multipole moments produced on degenerate resonant levels. Indeed, as one can see by comparing the rates $\gamma^{(\kappa)}$ in Table II for pure SF_6, the population, orientation, and alignment decay rates differ by 10–25%.

We also call attention to the fact that the value obtained

TABLE II. The results of stimulated photon echo measurements of the populations $\gamma^{(0)}$, orientation $\gamma^{(1)}$, and alignment $\gamma^{(2)}$ relaxation rates.

Collision partners	$\gamma^{(0)}$, $s^{-1}\cdot torr^{-1}$	$\gamma^{(1)}$, $s^{-1}\cdot torr^{-1}$	$\gamma^{(2)}$, $s^{-1}\cdot torr^{-1}$
SF_6-SF_6	$(29\pm 3)\cdot 10^6$	$(32\pm 3)\cdot 10^6$	$(38\pm 3)\cdot 10^6$
SF_6-Xe	not measured	$(16\pm 2)\cdot 10^6$	$(16\pm 2)\cdot 10^6$
SF_6-He	not measured	$(36\pm 4)\cdot 10^6$	$(47\pm 4)\cdot 10^6$

TABLE III. Total cross sections σ_{tot} and cross sections for elastic and inelastic scatterings, σ_{vcc} and σ_{rot}, obtained by comparative analysis of the photon echo and stimulated photon echo experimental data. Estimates of the gas kinetic cross section σ_{kin} are presented for comparison.

Collision partners	σ_{vcc}, Å^2	σ_{rot}, Å^2	σ_{tot}, Å^2	σ_{kin}, Å^2
SF_6-SF_6	$150\leqslant\sigma_{vcc}\leqslant 185$	$350\leqslant\sigma_{rot}\leqslant 385$	535	120
SF_6-Xe	$120\leqslant\sigma_{vcc}\leqslant 280$	$\leqslant 160$	280	95
SF_6-He	$20\leqslant\sigma_{vcc}\leqslant 105$	$\leqslant 85$	105	55

for $\gamma^{(0)}$ in polarized stimulated photon echo experiments is less than value obtained for $\gamma^{(0)}$ from the initial section of the kinetic curve of the standard photon echo ($T_{12}<T_c$) and presented in Table I. The reason for the discrepancy is that elastic collisions cannot be absolutely neglected even for times $0<T_{12}<T_c$; such collisions are always present. Since their role increases with T_{12}, it can be concluded that the discrepancy in the values of $\gamma^{(0)}$ would not have been so noticeable if it were possible to advance along the kinetic curve of the photon echo to shorter delays T_{12}, which is technically very difficult to accomplish. In addition, when the delay time decreases to approximately half the pulse duration it becomes necessary to take account of the relaxation during the application of a pulse.[16]

4. DISCUSSION

The data in Tables I and II make it possible to estimate the cross sections for inelastic (rotational) relaxation of SF_6 in collisions with He and Xe as well as the cross sections for scattering of SF_6 by small angles in collisions with these buffers, using the relations $\sigma_{\rm rot}=\gamma^{(0)}/nv_{\rm rel}$ and $\sigma_{\rm vcc}=(\Gamma_{\rm tot}-\gamma^{(0)})/nv_{\rm rel}$. The results of these estimates are presented in Table III. The gas kinetic cross sections obtained for collisional pairs from data on the viscosity and diffusion of gases[17] in the hard-sphere model are also presented for comparison.

It is evident from Table III that the elastic scattering cross section for a heavy active SF_6 particle is greater for collisions with a heavy buffer than for collisions with a light buffer. This assertion is also true for the cross sections of inelastic processes.

Our experimental value of $\sigma_{\rm rot}$ for SF_6-SF_6 collisions was found to be at least three times greater than the value of the gas-kinetic cross section $\sigma_{\rm kin}$ calculated from the viscosity and diffusion coefficients. The relation $\sigma_{\rm rot}\approx 3\sigma_{\rm kin}$ following from Table III means that the relaxation of the nonequilibrium population of a rotational sublevel of SF_6 is on the average three times more frequent than gas-kinetic collisions. This surprising fact for a nonpolar molecule can be explained if it is assumed that there exists an exchange mechanism of rotational relaxation.[14]

In the rotational relaxation model[18] the average number $Z_{\rm rot}$ of gas-kinetic collisions required to establish equilibrium with respect to the rotational states can be determined as

$$Z_{\rm rot}=\frac{3(1+2b)^2}{8b},$$

which holds for ''rough sphere'' molecules,[18] which general physical considerations show SF_6 molecule to be. Here $b=I/\mu a^2$, where I is the moment of inertia of the active particles, μ is the reduced mass, and a is the sum of the molecular radii of the colliding particles. If identical particles collide (the case of a pure SF_6 gas), then $b=1/2$, $Z_{\rm rot}=3$, and therefore $\sigma_{\rm rot}\approx\sigma_{\rm kin}/3$. This estimate differs by an order of magnitude from the values obtained in our experiments.

5. CONCLUSIONS

Photon and stimulated photon echo methods make it possible to study different types of collisions: rotational-inelastic collisions, elastic scattering by small angles, and depolarizing collisions.

Our investigation of the decay kinetics of a photon echo in a wide range of delays between the exciting pulses has demonstrated that it is possible to distinguish the contributions of the elastic and inelastic processes in the decay of a photon echo. In addition, the sections of the kinetic curve of a photon echo with different decay rates are more clearly distinguishable in pure SF_6 gas. For gas mixtures elastic and inelastic processes can be distinguished by using a heavy buffer gas.

For elastic and inelastic scatterings of SF_6 in gas mixtures, the collision cross sections increase with the mass of the buffer particles.

The stimulated photon echo method made it possible to investigate depolarizing collisions by measuring the relaxation rates of the populations, orientation, and alignment. The orientation and alignment relaxation rates measured by the stimulated photon echo method can be used to determine the upper limit on the inelastic relaxation rates, which cannot be measured directly by the photon echo method in mixtures of SF_6 with a buffer gas.

A combined analysis of the data obtained by photon and stimulated photon echo methods gave a complete set of cross sections for elastic and inelastic processes for the collisional pairs SF_6-SF_6, SF_6-Xe, and SF_6-He.

We thank the Russian Fund for Fundamental Research for supporting this work (Grants Nos. 97-02-18496 and 98-02-16390) and we gratefully acknowledge the support provided by the State Science Program ''Laser Physics.''

*)E-mail: eugeny@isp.nsc.ru

[1] V. G. Minogin and V. S. Letokhov, *The Pressure of Laser Radiation on Atoms* (Nauka, Moscow, 1986).
[2] S. I. Yakovlenko, Usp. Fiz. Nauk **136**(4), 593 (1982) [Sov. Phys. Usp. **25**, 216 (1982)].
[3] S. G. Rautian and I. I. Sobel'man, Usp. Fiz. Nauk **90**, 209 (1966) [Sov. Phys. Usp. **9**, 701 (1967)].
[4] L. S. Vasilenko and N. N. Rubtsova, Laser Phys. **7**, 1021 (1997).
[5] L. Allen and D. G. C. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975) [Russian translation, Mir, Moscow, 1978].
[6] I. V. Evseev, I. V. Ermachenko, and V. V. Samartsev, *Depolarizing Collisions in Nonlinear Electrodynamics* (Nauka, Moscow, 1992).
[7] L. S. Vasilenko and N. N. Rubtsova, Opt. Spektrosk. **58**, 697 (1985) [Opt. Spectrosc. **58**, 422 (1985)].
[8] I. V. Evseev, V. M. Ermachenko, and V. A. Reshetov, Zh. Éksp. Teor. Fiz. **78**, 2213 (1980) [Sov. Phys. JETP **51**, 1108 (1990)].
[9] A. P. Kol'chenko, A. A. Pukhov, S. G. Rautian, and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **63**, 1173 (1972) [Sov. Phys. JETP **36**, 619 (1973)].
[10] V. P. Chebotayev and L. S. Vasilenko, Progr. Quant. Electr. **8**(2) (1983).
[11] A. P. Kol'chenko, S. G. Rautian, and A. M. Shalagin, Preprint No. 46, Institute of Nuclear Physics, Siberian Branch of the Soviet Academy of Sciences, Novosibirsk (1972).
[12] V. P. Kochanov, S. G. Rautian, and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **72**, 1358 (1977) [Sov. Phys. JETP **45**, 714 (1977)].
[13] L. S. Vasilenko, N. N. Rubtsova, and E. B. Khvorostov, Zh. Éksp. Teor. Fiz. **113**, 826 (1998) [JETP **86**, 450 (1998)].

[14] N. N. Rubtsova, Doctoral Dissertation in Physical and Mathematical Sciences, Novosibirsk (1997).
[15] L. S. Vasilenko and N. N. Rubtsova, Laser Phys. **7**, 903 (1997).
[16] A. V. Durrant and J. Manners, Opt. Acta **31**, 1167 (1984).
[17] *Reference Data on Physical Quantities*, edited by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).
[18] A. Callir and J. Lambert, in *Comprehensive Chemical Kinetics*, Vol. 3, (edited by C. H. Bamford and C. F. H. Tipper (Elsevier, Amsterdam-London-New York, 1969).

Translated by M. E. Alferieff

Entropy decrease in a thermodynamically isolated system: Light-induced drift

O. A. Chichigina*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 11 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 57–66 (July 1999)

A new method is proposed for describing selective excitation as the addition of information to a thermodynamic system of atoms, decreasing the entropy of the system as a result. This information approach is used to calculate the light-induced drift velocity. The computational results are in good agreement with experimental data. © *1999 American Institute of Physics.* [S1063-7761(99)00607-1]

1. INTRODUCTION

The observance of the second law of thermodynamics is, for all practical purposes, never considered in the study of the interaction of laser radiation with matter. It is assumed that this law is observed automatically because of the strong increase in the entropy of the photons. Since this law is an inequality for strongly nonequilibrium processes, it cannot be used to calculate some parameters on the basis of the other parameters. However, in certain cases of selective excitation and of the collective process which it engenders, an atomic system whose internal energy remains constant and to which no heat is transferred can be studied separately. The action of the laser radiation can be described as the addition of information to the system. Light appears to mark the atoms that possess a prescribed property. In such a system the process is close to equilibrium, and the inequality in the expression for the second law of thermodynamics becomes an equality.

In Ref. 1 a generalized formulation of the second law of thermodynamics is proposed for an isolated system. According to Ref. 1, the entropy increase dH together with the information dI entering the system satisfy the inequality

$$dH + dI \geqslant 0, \tag{1}$$

where

$$I = k_B \sum_i P_i \ln P_i \tag{2}$$

is the information. The term dI is added in connection with the discussion of the influence of the so-called Maxwell's demon.

The second law of thermodynamics without the information term, which, conventionally, can be neglected since k_B is small, is ordinarily used to describe various thermodynamic processes. In the present paper it is shown that when information is recorded on each atom of the system the indicated term plays an important large role and can decrease the entropy.

As an example of such a process we shall consider light-induced drift, which was predicted in Ref. 2 and has now been well studied.[3–6] The crux of the effect is that there appears a directed macroscopic flux of absorbing particles which interact with a traveling light wave and undergo collisions with buffer-gas particles. Most experiments on light-induced drift have been performed with alkali-metal vapors in an inert-gas atmosphere.

Laser radiation with frequency ω' close to the frequency ω_0 of the main transition of an atom excites predominantly atoms having a velocity such that the corresponding Doppler frequency shift $\mathbf{k}\cdot\mathbf{v}$ ($\mathbf{k}$ is the wave vector) compensates the frequency offset $\omega' - \omega_0$. Therefore a dip near the velocity $\mathbf{v}$ appears in the velocity distribution of atoms in the ground state and a corresponding peak appears in the distribution of the atoms in the excited state. The average velocities of atoms of each kind are nonzero and oppositely directed. Since the excited atoms are subject to a greater resistance due to the buffer gas, the total flux of gas interacting with the radiation is likewise nonzero.

The question of the entropy change in such a system of atoms and photons is briefly considered in Ref. 7. It is shown there that the entropy decrease in the system of atoms is much smaller than the entropy increase in the system of photons. The total entropy increases, and the process is irreversible.

However, the atoms can be studied separately. The pressure in experiments with light-induced drift is much less than atmospheric pressure, and the temperature is near 400 K, so that the ideal-gas model is quite applicable to the mixture. The interaction cross section (which changes when a photon is absorbed) does not appear in the equations describing the behavior of the thermodynamic parameters. Hence it can be assumed that this ideal gas is not subject to external perturbations and that the system is thermodynamically isolated. The entropy and internal energy of such a system do not change. But the light-induced drift itself testifies that the entropy decreases, since part of the internal energy is converted into ordered motion, as a result of which the temperature decreases. This contradiction can be resolved by keeping in mind that the light adds to the system information about the direction of motion of the particles. Light appears to mark the atoms moving in a prescribed direction. Then the entropy decrease in the system does not exceed the amount of information entering the system.

2. SECOND LAW OF THERMODYNAMICS FOR AN ADIABATICALLY ISOLATED SYSTEM

The characteristic transition time to the drift state is identical to the free-flight time τ of a metal atom. When this time has elapsed, the velocities of the atoms change, and information is lost and rerecorded by a new flux of photons. At such short times heat-exchange with the thermostat can be neglected and the system can be assumed to be adiabatically closed. In addition, drift with a constant velocity is an equilibrium process. Then the expression (1) becomes an equality

$$dH = -dI. \tag{3}$$

3. THE ENTROPY CHANGE AND THE DRIFT VELOCITY

According to the law of conservation of momentum, the flux of metal atoms (sodium in most experiments) is compensated by an opposite flux of the buffer gas:

$$u m_{\mathrm{Na}} n_{\mathrm{Na}} = u_g m_g n_g, \tag{4}$$

where u and u_g are the velocities and m_{Na} and m_g the masses of the molecules, and n_{Na} and n_g are, respectively, the concentrations of sodium and buffer gas. Since entropy is additive, the entropy change can be written as

$$\Delta H = N_g \Delta H_g + N_{\mathrm{Na}} \Delta H_{\mathrm{Na}},$$

where N_{Na} and N_g are the numbers of metal and buffer-gas atoms, respectively. In the mixtures employed, the metal concentration is more than 10^6 times lower than the buffer-gas concentration. No large changes in the entropy of the metal occur. Therefore the second term can be neglected, and the buffer-gas concentration can be assumed to be the same as the mixture concentration n.

The ideal gas model can be used to determine the entropy change. Then we have per mole

$$\Delta H = \frac{3}{2} R \ln \frac{T_f}{T_i},$$

where T_i and T_f are the initial and final temperatures and R is the universal gas constant.

The appearance of a macroscopic flux with drift velocity u_g means (for constant internal energy of the system) that the temperature decreases by the amount $m u_g^2 / 3k_B$. Then

$$\Delta H = \frac{3}{2} R \ln \left(1 - \frac{2u_g^2}{3\overline{v_g^2}} \right), \tag{5}$$

where $\overline{v_g^2}$ is the mean-square thermal velocity of the buffer-gas atoms.

Substituting the expression (3) into Eq. (5) and then into Eq. (4) gives the drift velocity as a function of the information:

$$u = \frac{m_g n}{m_{\mathrm{Na}} n_{\mathrm{Na}}} \sqrt{\overline{v_g^2} \left[1 - \exp\left(\frac{2\Delta I}{3} \right) \right]}. \tag{6}$$

4. INFORMATION AND THE PROBABILITY OF ABSORPTION OF A PHOTON

The information recorded on each sodium atom is determined by the expression

$$i(P) = k_B [P \ln P + (1-P) \ln (1-P)], \tag{7}$$

where P is the probability that an atom moving in one of two directions will be marked by being transferred into an excited state. In what follows, we shall study a one-dimensional problem and we shall assume the velocity to be its projection on the direction of propagation of the beam.

Since information is recorded on each sodium atom, the amount of information added to the system is νN_A times greater (ν is the fraction of sodium atoms in the mixture and N_A is Avogadro's number).

From Eq. (7) follows

$$\Delta I = \nu R [P \ln P + (1-P) \ln (1-P)]. \tag{8}$$

The maximum amount of information will be added to the system if a wide-band laser radiation marks all metal atoms moving in the same direction and does not mark any atom moving in a different direction, i.e., if $P = 1/2$. This is possible if the spectrum of the laser radiation ''covers'' exactly half the Doppler spectrum, i.e., the laser spectrum decays rapidly in the wings. This same conclusion can be drawn on the basis of a complicated kinetic investigation.[5]

We shall now calculate the probability P of information begin recorded on a single atom about the direction of motion of the atom.

The information recording time is the time interval during which the velocity of an atom does not change much. This time is close to τ, if the buffer-gas atoms are heavier than the metal atoms and one collision is sufficient for information about velocity to vanish. In the general case it is $\tau(1 + m_{\mathrm{Na}}/m_g)$.

The number of photons with frequency ω (wavelength λ) entering the region of interaction with an atom over this time is determined by the expression[8]

$$M = \frac{J}{\hbar \omega} \sigma \tau \left(1 + \frac{m_{\mathrm{Na}}}{m_g} \right),$$

where J is the radiation intensity and σ is the interaction cross section neglecting Doppler broadening, since the latter will be taken into account separately, and collision broadening, since phase interruption does not occur over the free-flight time. Then

$$\sigma = \frac{\lambda^2 \Delta \omega}{4 \Delta \omega_l},$$

where $\Delta\omega$ is the natural linewidth and $\Delta\omega_l$ is the width of the laser spectrum.

Each of these photons interacts with an atom with probability P_D, taking account of the dependence of the transition frequency of an atom on its velocity. The probability P_D is given by Maxwell's distribution (since the frequency and velocity are uniquely related by the relation $\omega - \omega_0 = 2\pi v/\lambda$):

$$P_D = A \exp \frac{(\omega-\omega_0)^2}{\alpha^2},$$

where $\alpha = 2\pi\sqrt{2k_BT/m_{\mathrm{Na}}\lambda^2}$ and A is a normalization constant.

What is the probability of a given direction being marked? The difficulty of this question is due to the fact that the absorbed photon contributes information not only about the direction of motion of the particle but also about the velocity of the particle. To take account of the effect of the velocity v of the marked particles on the drift velocity u it is convenient to represent v in the discrete form $v_i = \Delta v[v/\Delta v]$, where $[x]$ is the integer part of x and $\Delta v = \lambda\Delta\omega/2\pi$ plays the role of the resolving power of the radiation measuring the velocity. We shall consider the case $\Delta v \ll \sqrt{\overline{v^2}}$. Then the motion of a single particle with velocity v_i is equivalent to the motion of i quasiparticles each moving with velocity Δv in the same direction. Thus, we obtain a gas of quasiparticles with velocities Δv and $-\Delta v$ differing only by the direction of motion. The interaction can be assumed to mark one of the i such particles but with a $i = [v/\Delta v]$ times greater probability

$$P_i = A|\omega_i-\omega_0| \exp \frac{(\omega_i-\omega_0)^2}{\alpha^2}. \tag{9}$$

Here we took account of the fact that $v/\Delta v = (\omega-\omega_0)/\Delta\omega$ and we have incorporated all ω_i-independent factors into the normalization constant. To sum the probabilities (9) in order to calculate the normalization constant, it is convenient to use the fact that $\Delta\omega \ll \alpha$ and to switch to integration. We obtain

$$P_i = \frac{\Delta\omega}{\alpha^2}|\omega_i-\omega_0| \exp \frac{(\omega_i-\omega_0)^2}{\alpha^2}.$$

Switching back to continuous frequencies does not introduce any substantial changes:

$$P(\mathrm{int}) = \frac{\Delta\omega}{\alpha^2}|\omega-\omega_0| \exp \frac{(\omega-\omega_0)^2}{\alpha^2} \tag{10}$$

is the probability that one photon, which has entered the region of interaction with an atom, interacts with the atom (int) and records information about the direction of motion of this atom.

It is necessary that over the free-flight time the interaction has occurred an odd number of times, since if the first interaction transfers the atom into an excited state, then the second interaction leads to induced emission, the third interaction once again excites the atom, and so on. Therefore all even-numbered interactions erase information. The average number of interactions is $P(\mathrm{int})M$. The probability of the number of interactions being odd is probability if $P(\mathrm{int})M \sim 1$. Then the probability of three or more interactions can be neglected. We obtain the probability that over the free-flight time information about the direction of motion of an atom will be recorded on the atom and not erased:

$$P = P(\mathrm{int})M[1-P(\mathrm{int})]^{M-1}. \tag{11}$$

Substituting the expression (11) into Eq. (8) and then into Eq. (6) we find the frequency dependence of the drift velocity. This dependence has a characteristic, centrosymmetric relative to the point $(\omega_0,0)$, tilda-shaped form. The expression (6) makes it possible to find analytically the optimal radiation intensity and the optimal pressure p (on which the free-flight time depends as $\tau \sim 1/p$).

5. THREE-LEVEL MODEL OF LIGHT-INDUCED DRIFT

The dependences of the drift velocity on the frequency and intensity of the laser radiation and on the type and pressure of the buffer gas have been obtained in Ref. 6, where experimental results on light-induced drift of sodium in inert gases (zenon, argon, and helium) are presented. To compare the above-described theory with experiment, a more realistic scheme of Na levels, taking account of the hyperfine splitting of the $3^2S_{1/2}$ ground state into two levels — $F1$ with degeneracy $g_1=3$ and $F2$ with degeneracy $g_2=5$ ($\Delta\omega_{hfs} = 2\pi\cdot 1772$ MHz) — must be considered. This splitting is of the same order of magnitude as the Doppler broadening of the spectral lines and introduces an uncertainty in the recording of information about the direction of motion of a sodium atom.

The probability of excitation of an atom is once again given by Eq. (11), but Eq. (10) cannot be used to determine $P(\mathrm{int})$. The fact that an electron can undergo a transition from the level $F1$ or from the level $F2$ with probabilities

$$P_1 = \frac{\Delta\omega g_1}{(g_1+g_2)\sqrt{\pi}\alpha} \exp \frac{(\omega-\Delta\omega_{hfs})^2}{\alpha^2}$$

$$P_2 = \frac{\Delta\omega g_2}{(g_1+g_2)\sqrt{\pi}\alpha} \exp \frac{\omega^2}{\alpha^2}, \tag{12}$$

respectively, must be taken into account. The frequency is measured with respect to the frequency of the transition frequency from the $F2$ level. The probability $P(\mathrm{int})$ is given by the sum of P_1 and P_2:

$$P(\mathrm{int}) = \frac{\Delta\omega}{(g_1+g_2)\sqrt{\pi}\alpha}\left(g_1 \exp \frac{(\omega-\Delta\omega_{hfs})^2}{\alpha^2} + g_2 \exp \frac{\omega^2}{\alpha^2}\right). \tag{13}$$

If the ground state is split, the interaction itself does not uniquely mark the direction of motion of an atom, as happened in the two-level model. Therefore $P(\mathrm{int})$ does not determine the information recorded on an atom, but rather it simply gives the number of atoms participating in the recording of information. The information itself is determined by the conditional probability $P(\mathrm{t\,dir/int})$ of error-free (t) determination of the direction (dir) provided that an interaction occurred at frequency ω. The expression (8) can be rewritten as

$$\Delta I = \nu RP\, i(P[\mathrm{t\,dir/int}]). \tag{14}$$

If an interaction has occurred at a frequency in the interval $\omega \pm \Delta\omega/2$, then there are only two possible variants: The direction of a particle is determined incorrectly (f) or correctly (t):

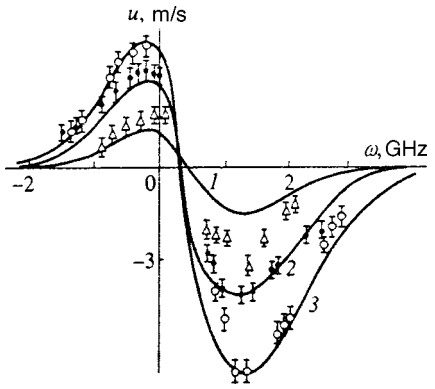

FIG. 1. Drift velocity u versus the offset ω of the laser frequency. Computational results (lines) and experimental data[6] (dots): *1* — helium (Δ), $p=10$ torr, $J=12$ W/cm^2; *2* — argon (●), $p=8.1$ torr, $J=8$ W/cm^2; *3* — xenon (○), $p=2.1$ torr, $J=12$ W/cm^2.

$$P(\text{t dir/int})=1-P(\text{f dir/int}). \tag{15}$$

It is simpler to find $P(\text{f dir/int})$ as

$$P(\text{f dir/int})=\frac{P(\text{f dir})}{P(\text{int})},$$

where $P(\text{f dir})$ is the probability that the interaction occurs at frequency ω and in the process the direction of motion of the atom which has interacted with the light will be indicated incorrectly.

Evidently, the direction of motion of the marked atom depends on the type of ground state: $F1$ or $F2$. Once again there are only two variants: the type of ground state is determined uniquely $P(1)$ or nonuniquely $P(2)$. Hence $P(\text{f dir})$ can be represented as

$$P(\text{f dir})=\sum_{j=1}^{2} P(\text{f dir}/j)P(j).$$

$P(1)$ is the sum of the probabilities that the type of the particles $F1$ and $F2$ is determined uniquely:

$$P(1)=P_1(1-P_2)+P_2(1-P_1),$$

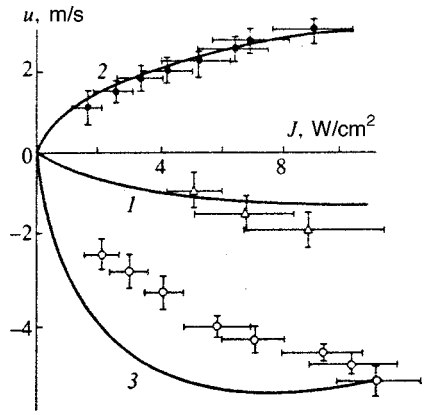

FIG. 2. Drift velocity versus the laser radiation intensity J. Computational results (lines) and experimental data[6] (dots): *1* — helium (Δ), $p=6$ torr, $\omega=1.1$ GHz; *2* — argon (●), $p=8.1$ torr $\omega=-0.3$ GHz; *3* — xenon (○), $p=1.4$ torr, $\omega=1.1$ GHz.

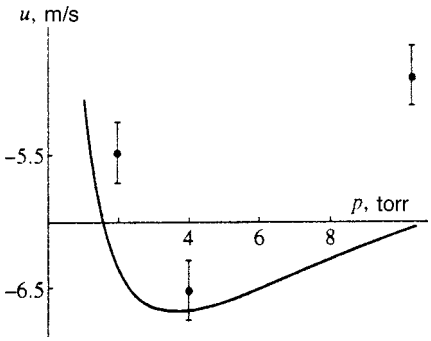

FIG. 3. Drift velocity versus the pressure p for xenon with $J=10$ W/cm^2 and $\omega=1.15$ GHz.

where P_1 and P_2 are given by the expressions (12). Correspondingly, the probability of nonunique determination is

$$P(2)=2P_1P_2$$

and has a maximum between the frequencies of the transitions from the levels $F1$ and $F2$.

If the ground state is determined uniquely, the probability of an error in determining the direction is zero:

$$P(\text{f dir}/1)=0.$$

If the level is not determined uniquely, the probability of an error in determining the direction is the same as the probability of determining any direction. We have similarly to the expression (10)

$$P(\text{f dir}/2)=\frac{\Delta\omega}{(g_1+g_2)\alpha^2}\left|g_1(\omega-\Delta\omega_{hfs})\right.$$
$$\left.\times\exp\frac{(\omega-\Delta\omega_{hfs})^2}{\alpha^2}+g_2\omega\exp\frac{\omega^2}{\alpha^2}\right|. \tag{16}$$

Substituting the expression (15) into Eq. (14) and the expression (14) into Eq. (6) gives the frequency-dependence of the drift velocity.

In Figs. 1–4 the theory is compared with the results of the experiment described in Ref. 6.

To obtain the theoretical curves the ratio ν of the sodium and buffer-gas concentrations was treated as an adjustable parameter, since its value was not given in Ref. 6. However, it should be noted that this parameter alone makes it possible to obtain the correct dependences of the drift velocity on four

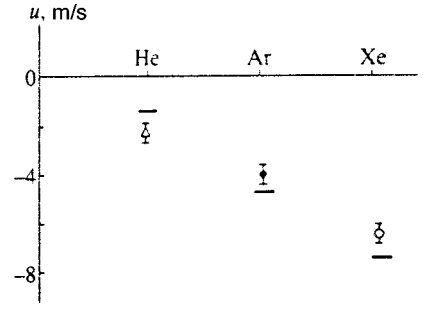

FIG. 4. Drift velocity versus the type of buffer gas, $p=5.4$ Torr, $J=12$ W/cm^2, and $\omega=1.1$ GHz.

different parameters for xenon and argon. For helium the transition time to the drift state is also an adjustable parameter, since the approximations employed are inapplicable if the mass of a metal atom is much greater than the mass of a buffer-gas molecule.

The drift velocity v as a function of the offset of the laser frequency ω, measured from the frequency of the transition $F2$, is shown in Fig. 1 for various buffer gases, radiation intensities J, and pressures p.

It is evident that the theoretical and experimental results for argon and xenon are in good agreement with one another. In contrast to the similar curves in the two-level model, our plot becomes asymmetric.

For helium the theory gives only an approximate description, since if the mass of the buffer-gas atoms is small, it cannot be assumed that information about the direction of motion of a metal atom is erased after one collision. The discrete-time description is not completely applicable for such systems.

Figure 2 shows the drift velocity as a function of the laser radiation intensity J.

The drift velocity increases with the radiation intensity up to a certain value corresponding to saturation. Saturation is possible beacues over the free-flight time an atom can interact twice with the radiation. This leads to induced emission and erasure of information. The existence of a maximum in the drift velocity as a function of pressure can be explained similarly.

Figure 3 shows the drift velocity as a function of pressure p for xenon. Finally, Fig. 4 shows the drift velocity as a function of the type of buffer gas.

An important result is that agreement with the experimental data is achieved neglecting many seemingly important factors, for example, neglecting the change in the scattering cross section of an atom on absorption of a photon or neglecting the real, complicated structure of the energy levels of sodium. This is because the drift velocity is expressed in terms of the thermodynamic parameters and therefore cannot depend on many ''model refining'' parameters, just as it does not depend on the positions and velocities of each particle.

Generalizing this result, it can be stated that the behavior of a complicated (i.e., described by a large number of different parameters) system is similar to that of a random system. The complexity could be due not only to the large number of particles, but it can also be due to the large number of factors influencing the process. Such a complicated system should be described by statistical laws in the language of thermodynamic parameters.

6. CONCLUSIONS

The foregoing analysis has shown that there exist real physical processes whose description requires the second law of thermodynamics in the generalized formulation given by Stratonovich, i.e., taking account of the information entering the system. One such process is selective laser excitation of metal atoms in a buffer gas. Taking account of the information introduced into the gas mixture makes it possible to calculate the light-induced drift velocity and to obtain good agreement with the experimental data on the basis of a comparatively simple model. It has been shown that a thermodynamic system can mean not only a system consisting of a large number of particles but also a system described by a large number of parameters.

I am deeply grateful to Yu. A. Kravtsov for an interesting discussion and a number of valuable remarks and B. A. Grishanin and A. S. Chirkin for reading the manuscript and for helpful suggestions for improving it.

*)E-mail: chichigina@foton.ilc.msu.su

[1] R. L. Stratonovich, *Information Theory* (Sovetskoe radio, Moscow, 1975).
[2] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[3] F. L. Popov, A. M. Shalagin, V. M. Shalaev, and V. V. Yakhnin, Zh. Éksp. Teor. Fiz. **80**, 2175 (1981) [Sov. Phys. JETP **53**, 1134 (1981)].
[4] F. Kh. Gel'mukhanov and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **78**, 1674 (1980) [Sov. Phys. JETP **51**, 839 (1980)].
[5] F. Kh. Gel'mukhanov and A. I. Parkhomenko, Zh. Éksp. Teor. Fiz. **108**, 1589 (1995) [JETP **81**, 871 (1995)].
[6] P. H. G. C. Werij and J. P. Woerdman, Phys. Rep. **169**, 145 (1988).
[7] M. E. Carrera-Patino and R. S. Berry, Phys. Rev. A **34**, 4728 (1986).
[8] D. N. Klyshko, *The Physical Principles of Quantum Electronics* (Nauka, Moscow, 1986).

Translated by M. E. Alferieff

Collapse in the nonlinear Schrödinger equation

Yu. N. Ovchinnikov*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia

I. M. Sigal

University of Toronto, Department of Mathematics, Toronto, Ontario, Canada M5S 3G3
(Submitted 11 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 67–77 (July 1999)

A new class of exact solutions with a singularity at finite time (collapse) is obtained for the nonlinear Schrödinger equation. © *1999 American Institute of Physics.* [S1063-7761(99)00707-6]

The nonlinear Schrödinger equation, where the nonlinearity corresponds to an effective attraction, arises in many physical problems. It appears that such an equation was first obtained in an investigation of strong Langmuir turbulence[1] and in the problem of self-focusing of wave beams.[2] Subsequently, the nonlinear Schrödinger equation was actively investigated irrespective of the physical problem giving rise to it.[3] In dimensionless variables the nonlinear Schrödinger equation (NLS) is ordinarily written in the simple form

$$i\frac{\partial\psi}{\partial t}+\Delta\psi+|\psi|^{2\sigma}\psi=0, \tag{1}$$

where ψ is a scalar function in a d-dimensional space and Δ is the Laplacian operator.

The uniform state $\psi=a\exp(ia^{2\sigma}t)$ is unstable with respect to infinitesimal disturbances for any $\sigma>0$ [the dispersion law is $\omega^2=K^2(K^2-2\sigma a^{2\sigma})$, $\psi=(a+\delta\psi)\exp(ia^2t)$].

For $\sigma d<2$, it appears that states of the form

$$\psi_{(\mathbf{r},t)}=\exp(-i\omega t)\tilde{\psi}(\mathbf{r}),$$

where $\tilde{\psi}(\mathbf{r})$ is a periodic function of the coordinates, are realized. For $\sigma d\geqslant 2$ Eq. (1) possesses solutions with a singularity at a finite time t_0. We shall investigate such solutions below. The equation (1) conserves the total number of ''particles'' and the total energy. We find for the particle flux density j_n the standard equation

$$j_n=i(\psi\nabla\psi^*-\psi^*\nabla\psi),\quad \nabla\equiv\text{grad}. \tag{2}$$

The energy E_V in a volume V is given by the expression

$$E_V=\frac{1}{2}\int_V d^dr\left\{|\nabla\psi|^2-\frac{1}{1+\sigma}|\psi|^{2(1+\sigma)}\right\}. \tag{3}$$

Using Eqs. (1) and (3) we easily find

$$\frac{\partial E_V}{\partial t}=\frac{1}{2}\int_S\left(\nabla\psi^*\frac{\partial\psi}{\partial t}+\frac{\partial\psi^*}{\partial t}\nabla\psi\right)d\mathbf{S}. \tag{4}$$

The equations (2) and (4) will be used below to obtain the singular solutions of Eq. (1).

We shall seek the collapsing solution of Eq. (1) in a d-dimensional space in a spherically symmetric form

$$\psi(\mathbf{r},t)=\hat{\varphi}(\rho,t)\exp(i\chi(\rho,t)), \tag{5}$$

where $\rho=|r|$ and $\hat{\varphi}$ is a real function.

In contrast to Refs. 1 and 3, we shall seek the collapsing solution in the form

$$\hat{\varphi}(\rho,t)=\lambda_{(t)}^{d/2}\varphi(\rho\lambda(t)) \tag{6}$$

without imposing any preliminary constraints on the phase χ.

Let $N(\rho_0,t)$ be the number of particles inside a sphere of radius ρ_0. Then Eqs. (2), (5), and (6) give

$$\frac{\partial N(\rho_0,t)}{\partial t}=\alpha\rho_0\frac{\partial\lambda}{\partial t}(\rho_0\lambda)^{d-1}\varphi^2(\rho_0\lambda)$$

$$=-2\alpha\frac{\partial\chi}{\partial\rho}\lambda^d\rho_0^{d-1}\varphi^2(\rho_0\lambda), \tag{7}$$

where $\alpha\equiv\alpha(d)$ and is determined from the relations

$$dV=\alpha d\rho\rho^{d-1},\quad S_{(\rho)}=\alpha\rho^{d-1},\quad \alpha=\frac{2\pi^{d/2}}{\Gamma(d/2)}. \tag{8}$$

In Eq. (8) $S(\rho)$ is the area of a sphere of radius ρ. From Eq. (7) we find

$$\rho_0\frac{\partial\lambda}{\partial t}=-2\frac{\partial\chi}{\partial\rho}\lambda. \tag{9}$$

The general solution of this equation is

$$\chi(\rho,t)=\chi_0(t)-\frac{\rho^2}{4\lambda}\frac{\partial\lambda}{\partial t}, \tag{10}$$

where the functions $\chi_0(t)$ and $\lambda(t)$ depend only on the time t.

Another equation — expressing the quality of the energy flux through a surface to the change in the energy flux inside the volume — arises from Eqs. (3) and (4):

$$y^{d-1}\left\{\varphi'(y)\left(y\varphi'(y)+\frac{d}{2}\varphi_{(y)}\right)-\frac{y}{2\lambda^2}\varphi_{(y)}^2\left(\frac{\partial\chi_0}{\partial t}\right.\right.$$

$$\left.\left.-\frac{y^2}{4\lambda^2}\frac{\partial}{\partial t}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial t}\right)\right)\right\}=\frac{1}{2}y^d\left[(\varphi')^2+\frac{y^2\varphi^2}{4\lambda^6}\left(\frac{\partial\lambda}{\partial t}\right)^2\right.$$

$$\left.-\frac{\lambda^{d\sigma-2}}{1+\sigma}\varphi^{2(1+\sigma)}\right]+\int_0^y dyy^{d-1}\left\{(\varphi'_{(y)})^2\right.$$

$$+\frac{\varphi^2}{8\lambda}y^2\left[\frac{\partial}{\partial t}\left(\frac{1}{\lambda^4}\left(\frac{\partial\lambda}{\partial t}\right)^2\right)\right]\Bigg/ \frac{\partial\lambda}{\partial t}-\frac{d\sigma\lambda^{d\sigma-2}}{2(1+\sigma)}\varphi^{2(1+\sigma)}\Bigg\}, \tag{11}$$

where $y=\rho_0\lambda$. Let us assume that $\lambda(t)$ is a power-law function of the form

$$\lambda\sim(t_0-t)^{-\nu}. \tag{12}$$

Then from Eq. (11) we find

$$\frac{1}{\lambda^4}\frac{\partial}{\partial t}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial t}\right)\sim(t_0-t)^{4\nu-2},$$

$$\frac{1}{\lambda^6}\left(\frac{\partial\lambda}{\partial t}\right)^2\sim(t_0-t)^{4\nu-2},$$

$$\frac{1}{\lambda}\left[\frac{\partial}{\partial t}\left(\frac{1}{\lambda^4}\left(\frac{\partial\lambda}{\partial t}\right)^2\right)\right]\Bigg/\frac{\partial\lambda}{\partial t}\sim(\nu-1)(t_0-t)^{4\nu-2}. \tag{13}$$

From Eqs. (13) it follows that there exist two values of the parameter ν, $\nu=1$ and $\nu=1/2$, for which Eq. (11) can possess an exact solution in the case of the critical dimension $d\sigma=2$. If $d\sigma\neq2$, then there exists one distinguished value of the parameter ν,

$$\nu=\frac{2}{2+d\sigma}, \tag{14}$$

for which the time-dependence of the ''leading-order'' terms is the same. Setting the sum of these terms to zero in the region $d\sigma>2$ gives an equation for the collapsing function in the leading-order approximation.

For the critical dimension $d\sigma=2$ it follows from Eqs. (13) that the values of the parameter ν fall into two subregions, $\nu<1/2$ and $\nu>1/2$.

The behavior of the function $(t_0-t)^{4\nu-2}$ changes at the boundary ($\nu=1/2$) of these regions.

We shall now consider the most interesting case of the critical dimension $d\sigma=2$. We shall investigate first the particular values $\nu=1$ and 1/2 for which the system of equations (10) and (11) possesses an exact solution.

a. $\nu=1$. In this case [see Eq. (13)]

$$\lambda=\frac{C}{t_0-t}. \tag{15}$$

From Eq. (11) follows an equation for the phase $\chi_0(t)$:

$$\frac{\partial\chi_0}{\partial t}=C_2\lambda^2,\quad \chi_0(t)=\frac{C_2C^2}{t_0-t}+\text{const}. \tag{16}$$

Using Eqs. (15) and (16), Eq. (11) can be written as

$$\left(y^{-d/2+1}\frac{\partial}{\partial y}(y^{d/2}\varphi)\right)\left\{\frac{1}{y^{d-1}}\frac{\partial}{\partial y}\left(y^{d-1}\frac{\partial\varphi}{\partial y}\right)-C_2\varphi+\varphi^{2\sigma+1}\right\}=0. \tag{17}$$

A simple check shows that the function

$$\psi(\rho,t)=\lambda_{(t)}^{d/2}\varphi(\rho\lambda(t))\exp\left\{\frac{iC_2C^2}{t_0-t}-\frac{i\rho^2}{4(t_0-t)}+i\,\text{const}\right\} \tag{18}$$

is an exact collapsing solution of Eqs. (1), where the function $\varphi(\rho)$ is a solution of the ordinary differential equation

$$\frac{1}{\rho^{d-1}}\frac{\partial}{\partial\rho}\left(\rho^{d-1}\frac{\partial\varphi}{\partial\rho}\right)-C_2\varphi+\varphi^{2\sigma+1}=0. \tag{19}$$

It follows from Eq. (19) that a (normalized) solution decreasing rapidly at infinity exists only for $C_2>0$. This solution has been obtained and investigated in Refs. 4–6.

We note that for $C_2>0$ two types of asymptotic behavior are possible at infinity:

$$\varphi\sim(C_2)^{1/2\sigma}+A\sin(\sqrt{2\sigma C_2}\rho+\text{const})/\rho^{(d-1)/2},$$

which arises on a two-parameter family of solutions, and

$$\varphi\sim\frac{\text{const}}{\rho^{(d-1)/2}}\exp(-\sqrt{C_2}\rho),$$

which can be realized only on a one-parameter family of solutions.

We shall now consider the second family of solutions of Eqs. (10) and (11) corresponding to the case $\nu=1/2$.

b. $\nu=1/2$. In this case [see Eq. (13)]

$$\lambda^2=\frac{C}{t_0-t},\quad \frac{\partial\chi_0}{\partial t}=C_2\lambda^2,\quad \chi_0=-C_2C\ln(t_0-t)+\text{const}. \tag{20}$$

Using Eqs. (20), Eq. (11) can be put into the form

$$\left\{y^{-d/2+1}\frac{\partial}{\partial y}(y^{d/2}\varphi)\right\}\left\{\frac{1}{y^{d-1}}\frac{\partial}{\partial y}\left(y^{d-1}\frac{\partial\varphi}{\partial y}\right)-C_2\varphi+\frac{y^2\varphi}{16C^2}+\varphi^{2\sigma+1}\right\}=0. \tag{21}$$

We shall examine the solution φ of Eq. (21) that causes the expression in the second set of braces to vanish:

$$\frac{1}{\rho^{d-1}}\frac{\partial}{\partial\rho}\left(\rho^{d-1}\frac{\partial\varphi}{\partial\rho}\right)-C_2\varphi+\frac{\rho^2\varphi}{16C^2}+\varphi^{2\sigma+1}=0. \tag{22}$$

This equation corresponds to the motion of a particle in a field with the potential U, shown in Fig. 1, to which a term depending on the dimension d of the space must be added:

$$\frac{1}{\rho^2}\frac{(d-1)(d-3)}{4}. \tag{23}$$

In the region $\rho\ll1$ we find from Eq. (21) a solution that is regular at the origin:

$$\varphi=A+\frac{\rho^2A}{2d}(C_2-A^{2\sigma})+\frac{\rho^4A}{4(d+2)}\times\left[\frac{1}{2d}(C_2-A^{2\sigma})(C_2-(2\sigma+1)A^{2\sigma})-\frac{1}{16C^2}\right]+\ldots,$$

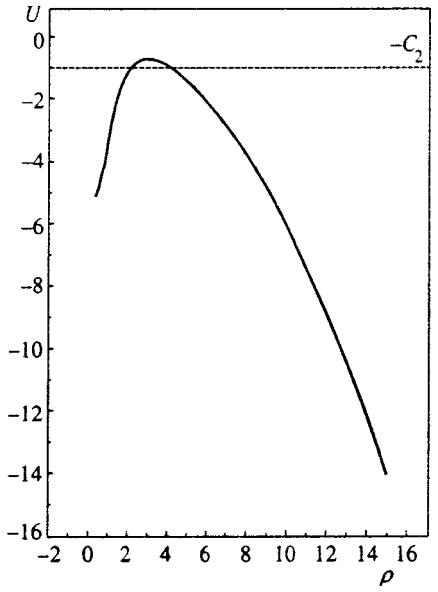

FIG. 1. Form of the potential energy $U=-\varphi^{4/3}-r^2/16C^2$ for Eq. (22), $\{A,C,C_2\}=\{3.4,1,1\}$, $d=3$.

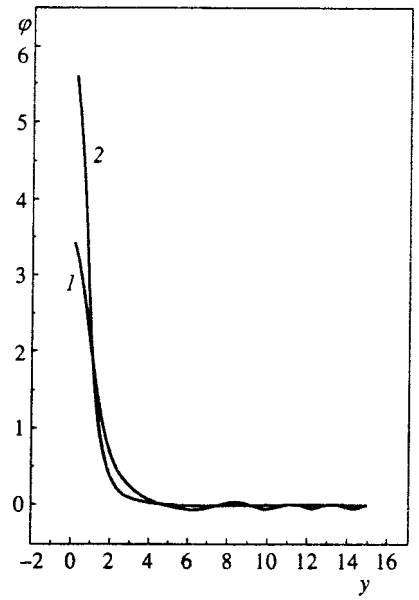

FIG. 2. The function φ for the parameters $\{A,C,C_2\}=\{3.4,1,1\}$ (*1*) and $\{5.5937, 2, 1.5\}$ (*2*).

where $A>0$ is an arbitrary constant. In the region $\rho\gg 1$ the asymptotic behavior of φ is determined by the expression

$$\varphi\sim\frac{B}{\rho^{d/2}}\sin\left\{\frac{\rho^2}{8C}-2CC_2\ln\rho+\text{const}\right\},$$

where B is a constant that depends on $\{A,\ C,\ C_2\}$. It is easy to verify that for all values of the constants $\{A,\ C,\ C_2\}$ the function φ has no singularities at finite values of ρ.

There exists a region of values of the parameters C, C_2, and $\varphi(0)$ where the solution of Eq. (22) outside of the potential barrier is exponentially small. Let

$$\rho^*\equiv\rho^*(C,C_2,\varphi(0)) \tag{24}$$

be the first zero of the function $\varphi(\rho)$. We set

$$\tilde{\varphi}(\rho)=\begin{cases}\varphi(\rho) & \text{for } \rho<\rho^*,\\ 0 & \text{for } \rho>\rho^*.\end{cases} \tag{25}$$

Then the collapsing solution $\psi(\rho,t)$ of Eq. (1) is given by the expression

$$\psi(\rho,t)=\lambda_{(t)}^{d/2}\tilde{\varphi}(\rho\lambda(t))\exp\left\{-iCC_2\ln(t_0-t)\right.$$
$$\left.-\frac{i\rho^2}{8(t_0-t)}+i\ \text{const}\right\}+\psi_1(\rho,t), \tag{26}$$

where $\psi_1(\rho,t)$ is the solution of the linear inhomogeneous equation

$$i\frac{\partial\psi_1}{\partial t}+\Delta\psi_1+|\psi|^{2\sigma}\psi_1+\sigma[|\psi|^{2\sigma}\psi_1+|\psi|^{2(\sigma-1)}\psi^2\psi_1^*]$$
$$-\left(\frac{\partial\varphi_{(\rho)}}{\partial\rho}\right)_{\rho^*}\lambda_{(t)}^{d/2+1}\exp\left\{-iCC_2\ln(t_0-t)\right.$$
$$\left.-\frac{i\rho^2}{8(t_0-t)}+\text{const}\right\}\delta\left(\rho-\frac{\rho^*}{\lambda_{(t)}}\right)=0, \tag{27}$$

$\psi_1(\rho,t^*)=0$, and t^* is the onset time of collapse.

We note that increasing the parameter C increases the width of the subbarrier region and at the same time decreases sharply the probability of formation of an initial fluctuation leading to collapse. Therefore fluctuations for which $C\sim C_2\sim 1$ and C_2 passes near the top of the potential barrier evidently are most likely to lead to collapse.

The values of the function φ in three-dimensional space ($d=3$) are presented in Fig. 2 for two sets of parameters $\{A,\ C,\ C_2\}$ — {3.4, 1, 1} and {5.5937, 2, 1.5}. The values of ρ^* and $\varphi'(\rho^*)$ are, respectively, $\{\rho^*=4.815,\ \varphi'(\rho^*)=-0.082\}$ and $\{\rho^*=13.13,\ \varphi'(\rho^*)=-2.63\cdot 10^{-3}\}$.

c. $\nu>1/2$. In this case

$$\lambda(t)=\frac{C}{(t_0-t)^\nu},\quad \frac{\partial\chi_0}{\partial t}=C_2\lambda^2. \tag{28}$$

For $\nu>1/2$ only the ''leading-order'' terms in Eq. (11), which grow most rapidly as the singular point t_0 is approached, can be set equal to zero. Using Eqs. (28) and (11) we obtain the following equation for the function φ:

$$\left\{y^{-d/2+1}\frac{\partial}{\partial y}(d^{d/2}\varphi)\right\}\left\{-C_2\varphi+\varphi^{2\sigma+1}\right.$$
$$\left.+\frac{1}{y^{d-1}}\frac{\partial}{\partial y}\left(y^{d-1}\frac{\partial\varphi}{\partial y}\right)\right\}=0. \tag{29}$$

The collapsing solution of Eq. (1) for $\nu>1/2$ has the form

$$\psi(\rho,t)=\lambda_{(t)}^{d/2}\varphi(\rho\lambda(t))\exp\left\{\frac{iC_2C^2}{(2\nu-1)(t_0-t)^{2\nu-1}}\right.$$
$$\left.-\frac{i\rho^2\nu}{4(t_0-t)}+i\ \text{const}\right\}+\psi_1(\rho,t), \tag{30}$$

where $\varphi(\rho)$ is the solution of Eq. (19) (for it the expression in the second set of braces in Eq. (29) vanishes), and the function ψ_1 is the solution of the linearized equation (1) with right-hand side arising as a result of the nonvanishing of the small residual terms in Eq. (11). A detailed investigation of the function ψ_1 will be given separately.

d. $\nu<1/2$. Vanishing of the leading terms in Eq. (1) gives the equation

$$\left\{y^{-d/2+1}\frac{\partial}{\partial y}(d^{d/2}\varphi)\right\}\left\{-C_2\varphi+\frac{\nu(1-\nu)}{4}y^2\varphi\right\}=0, \quad (31)$$

where

$$\lambda=\frac{C}{(t_0-t)^{\nu}}, \quad \frac{\partial\chi_0}{\partial t}=\frac{C_2}{C^2}(t_0-t)^{2\nu-2}. \quad (32)$$

Vanishing of the expression in the second set of braces in Eq. (31) does not lead to a nontrivial solution.

e. We shall now consider the case of supercritical dimension $d\sigma>2$. In this case

$$\nu=\frac{2}{2+d\sigma}, \quad \lambda=\frac{C}{(t_0-t)^{\nu}},$$

$$\frac{\partial\chi_0}{\partial t}=\frac{C_2}{\lambda^2(t_0-t)^2}, \quad \frac{1}{\lambda}\frac{\partial\lambda}{\partial t}=\frac{\nu}{t_0-t}. \quad (33)$$

Substituting the expressions (33) into Eq. (11) and equating to zero the terms which are of leading order in the parameter $(t_0-t)^{-1}$, we obtain an equation for the function φ

$$\left\{y^{-d/2+1}\frac{\partial}{\partial y}(y^{d/2}\varphi)\right\}\left\{-C_2\varphi+\frac{\nu(1-\nu)}{4}y^2\varphi + C^{d\sigma+2}\varphi^{2\sigma+1}\right\}=0. \quad (34)$$

From Eq. (34) we find

$$\varphi(y)=\begin{cases}\left[C_2-\dfrac{\nu(1-\nu)}{4}y^2\right]^{1/2\sigma}\Big/\,C^{d+2/\sigma}, & y<2\left(\dfrac{C_2}{\nu(1-\nu)}\right)^{1/2},\\ 0, & y>2\left(\dfrac{C_2}{\nu(1-\nu)}\right)^{1/2}.\end{cases} \quad (35)$$

Thus for $d\sigma>2$ the collapsing solution $\psi(\rho,t)$ is

$$\psi(\rho,t)=\lambda_{(t)}^{d/2}\varphi(\rho\lambda(t))\exp\left\{-\frac{i\nu\rho^2}{4(t_0-t)} + \frac{iC_2}{(1-2\nu)C^2(t_0-t)^{1-2\nu}}+i\,\mathrm{const}\right\}+\psi_1(\rho,t), \quad (36)$$

where the function $\varphi(\rho)$ is determined by Eq. (35). Just as in the case **c** [Eq. (30)], the function $\psi_1(\rho,t)$ is the solution of the linearized equation (1) with right-hand side arising because of the nonvanishing of the residual terms in Eq. (11). A detailed investigation of the function ψ_1 will be given separately.

For $\sigma=1$ a solution of the form (33) and (35) in the three-dimensional case has been obtained in Ref. 7.

f. Weak collapse.

In the supercritical region ($\sigma d>2$) Eq. (1) can have exact collapsing solutions of the form

$$\psi(\rho,t)=\lambda^{\tilde{\nu}}\varphi(\rho\lambda)\exp(i\chi), \quad (37)$$

where $\lambda=\lambda(t)$, $\chi=\chi(\rho,t)$, and $\tilde{\nu}$ is a constant. For $\tilde{\nu}\neq d/2$ Eq. (7) for the phase assumes the form

$$-2\frac{\partial\chi}{\partial\rho}=\frac{\rho}{\lambda}\frac{\partial\lambda}{\partial t}+\frac{2\tilde{\nu}-d}{\varphi^2(y)}\frac{1}{y^{d-1}}\frac{1}{\lambda^2}\frac{\partial\lambda}{\partial t}\int_0^y dy\,y^{d-1}\varphi^2(y). \quad (38)$$

From Eqs. (3) and (37) we find

$$E_V=\frac{\alpha}{2}\int_0^{\rho}d\rho\,\rho^{d-1} \times\left\{\varphi^2\lambda^{2\tilde{\nu}}\left(\frac{\partial\chi}{\partial\rho}\right)^2+\lambda^{2\tilde{\nu}+2}(\varphi')^2-\frac{\lambda^{2\tilde{\nu}(1+\sigma)}}{1+\sigma}\varphi^{2(1+\sigma)}\right\},$$

$$\frac{\partial E_V}{\partial t}=\alpha\lambda^{2\tilde{\nu}}\rho^{d-1}\left\{(\tilde{\nu}\varphi+y\varphi')\varphi'\frac{\partial\lambda}{\partial t}+\frac{\partial\chi}{\partial t}\frac{\partial\chi}{\partial\rho}\varphi^2\right\}. \quad (39)$$

For all terms in Eqs. (38) and (39) to have the same time-dependence the following conditions must be satisfied:

$$\tilde{\nu}\sigma=1, \quad \chi(\rho,t)=\chi_0(t)+\tilde{\chi}(\rho\lambda),$$

$$\frac{1}{\lambda^3}\frac{\partial\lambda}{\partial t}=\mathrm{const}, \quad \frac{\partial\chi_0}{\partial t}\sim\frac{1}{\lambda}\frac{\partial\lambda}{\partial t}. \quad (40)$$

From Eqs. (40) we find

$$\lambda=\frac{C}{\sqrt{t_0-t}}, \quad \chi_0(t)=-\frac{C_1}{2}\ln(t_0-t),$$

$$\frac{1}{\lambda^3}\frac{\partial\lambda}{\partial t}=\frac{1}{2C^2}, \quad \frac{\partial\chi_0}{\partial t}=\frac{C_1\lambda^2}{2C^2}. \quad (41)$$

We now consider the case $d=3$, $\sigma=1$. Using Eqs. (41), Eqs. (38) and (39) for the phase $\tilde{\chi}$ and modulus φ can be put into the form

$$\tilde{\chi}' + \frac{1}{4C^2}\left(y - \frac{1}{y^2\varphi^2}\int_0^y dy\, y^2\varphi^2 \right) = 0,$$

$$\frac{1}{y^2}\frac{\partial}{\partial y}\left(y^2 \frac{\partial \varphi}{\partial y} \right) + \varphi^3 - \frac{\varphi}{2C^2}(C_1 + y\tilde{\chi}') - \varphi(\tilde{\chi}')^2 = 0. \tag{42}$$

We note that any solution of the system of equations (42) is an exact solution of the nonlinear equation (1). Let

$$Z = \int_0^y dy\, y^2\varphi^2. \tag{43}$$

The modulus φ and phase $\tilde{\chi}$ are related simply with the function Z as

$$\varphi = \frac{\sqrt{Z'}}{y}, \quad \tilde{\chi}' = -\frac{yZ' - Z}{4C^2Z'}, \tag{44}$$

and the function Z itself is a solution of the ordinary differential equation[8]

$$Z''' - \frac{(Z'')^2}{2Z'} + \frac{2(Z')^2}{y^2} - \frac{1}{C^2}\left(C_1 Z' - \frac{y}{4C^2}(yZ' - Z) \right) - \frac{(yZ' - Z)^2}{8C^4Z'} = 0. \tag{45}$$

In the region $y \ll 1$ we obtain from Eq. (45)

$$Z(y) = Ay^3 + \frac{y^5}{15}\left(\frac{3AC_1}{2C^2} - 9A^2 \right) + \ldots, \tag{46}$$

where $A > 0$ is an arbitrary constant. In the region $y \gg 1$ the solution of Eq. (45) is

$$Z = By - \frac{2BC^2C_1}{y} + \frac{2BC^4}{y^3}(B - 2C_1^2) + \ldots \tag{47}$$

The equation (45) admits the existence of poles of the form

$$Z \sim -\frac{2y_0^2}{y_0 - y}. \tag{48}$$

However, for $A > 0$ it is impossible to reach such a pole. Since the only possible asymptotic solution of Eq. (45) in the limit $y \to \infty$ is determined by Eq. (47), there exists a three-parameter family of solutions of Eq. (1) of the form (37). These parameters are $\{A, C, C_1\}$. Just as in the case **b**, the region of physical collapse is bounded: The onset time t^* of collapse corresponds to $y < y^*$. For $y > y^*$ we must set $\varphi(y > y^*) = 0$, which will result in the appearance of outgoing waves. For the appropriate choice of the parameters A, C, and C_1 it is possible to make $\varphi(y^*)/\varphi(0) \ll 1$, even if $y^* \lesssim 1$. Weak collapse has been studied in Refs. 7 and 9. We have shown that the problem of weak collapse reduces to solving a single ordinary differential equation for the real function Z, and the solutions form a three-parameter family. Figures 3 and 4 show the functions $\varphi(y)$ and $\tilde{\chi}'(y)$ for two sets of parameters: $(A, C, C_1) = 4,2,2$ and $(A, C, C_1) = 3,1,1$. The ratio $\varphi(0)/\varphi(y^*)$ for the parameters in Fig. 3 is $\varphi(0)/\varphi(y^*) = 71.3$ and $y^* = 2.26$.

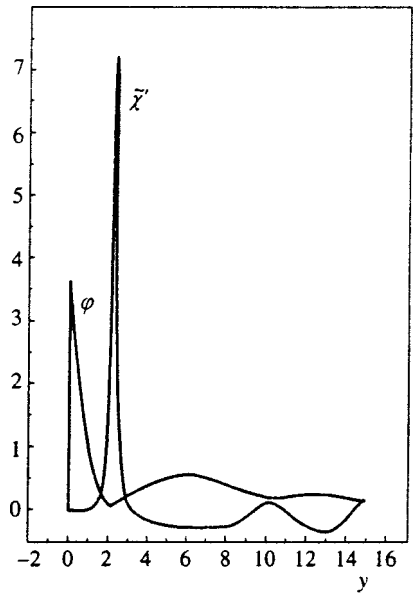

FIG. 3. The functions φ and $\tilde{\chi}'$ for the parameters $\{A, C, C_1\} = \{4,2,2\}$.

g. Weak collapse for arbitrary $\{d, \sigma\}$. Once again [Eq. (43)] we set

$$Z = \int_0^y dy\, y^{d-1}\varphi^2(y), \quad \varphi(y) = \frac{\sqrt{Z'}}{y^{(d-1)/2}}. \tag{49}$$

The equation for the phase (38) assumes the form

$$\tilde{\chi}' = -\frac{yZ' + (2/\sigma - d)Z}{4C^2Z'}. \tag{50}$$

An equation for the modulus $\varphi(y)$ can be easily obtained from Eqs. (1) and (37):

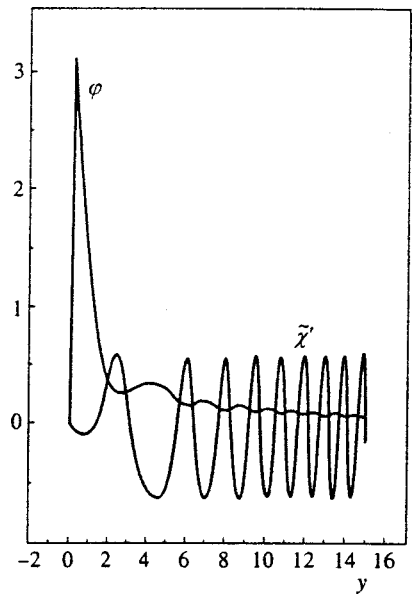

FIG. 4. The functions φ and $\tilde{\chi}'$ for the parameters $\{A, C, C_1\} = \{3,1,1\}$.

$$\frac{1}{y^{d-1}}\frac{\partial}{\partial y}\left(y^{d-1}\frac{\partial\varphi}{\partial y}\right)-\varphi\left[\frac{C_1}{2C^2}+\frac{y}{2C^2}\tilde{\chi}'\right]$$
$$-\varphi(\tilde{\chi}')^2+\varphi^{2\sigma+1}=0. \quad (51)$$

Substituting into Eq. (51) the explicit expressions for the functions φ and $\tilde{\chi}$ [Eqs. (49) and (50)] we obtain a single equation for the real function Z:

$$Z'''-\frac{(Z'')^2}{2Z'}-\frac{(d-1)(d-3)}{2y^2}Z'-\frac{1}{C^2}\left[C_1Z'-\frac{y(yZ'+(2/\sigma-d)Z)}{4C^2}\right]-\frac{(yZ'+(2/\sigma-d)Z)^2}{8C^4Z'}$$
$$+\frac{2(Z')^{\sigma+1}}{y^{(d-1)\sigma}}=0. \quad (52)$$

The equation (52) is an extension of Eq. (45) to arbitrary values of $\{d,\sigma\}$. There exist three values of the parameter $\tilde{\nu}>0$ such that for $y\to 0$ the function Z can be expanded in the series

$$Z(y)=Ay^{\tilde{\nu}}+A_2y^{\tilde{\nu}+2}+\ldots. \quad (53)$$

1. The value $\tilde{\nu}=d-2/\sigma>0$ can occur only if $d>2(1+1/\sigma)$. Then the coefficient A is a single-valued function of $\tilde{\nu}$

$$A\equiv A_{(\tilde{\nu})}=\frac{(d-\tilde{\nu})(\tilde{\nu}+d-4)}{4\tilde{\nu}}. \quad (54)$$

2. The value $\tilde{\nu}=d$ obtaines for arbitrary dimension and arbitrary $A>0$.

3. The value $\tilde{\nu}=4-d$ obtains only for $d<2$. The coefficient $A>0$ is arbitrary. All solutions have the same asymptotic behavior for $y\to\infty$:

$$Z=y^{d-2/\sigma}\left(B-\frac{B_1}{y^2}+\ldots\right). \quad (55)$$

Just as in the case $\{d=3,\ \sigma=1\}$, for arbitrary $\{d,\sigma\}$ φ must be cut off for $y>y^*$. The most appropriate point for this is the position of the deep minimum of the function φ. Evidently, only the values of the parameters $\{A, C, C_1\}$ for which such a minimum exists can be realized.

For the critical dimension $d\sigma=2$ exact solutions of the nonlinear Schrödinger equation which have a singularity at a finite time t_0 were obtained. These solutions correspond to the parameters $\nu=1$ and 1/2. It was shown that collapsing solutions also exist for $\nu>1/2$.

Exact solutions of the NLS which describe weak collapse in the supercritical region $\sigma d>2$ were obtained. These solutions form a three-parameter family $\{A,C,C_1\}$.

A collapsing solution arises in a bounded region of space, and an initial distribution of a special form is required in order for a singularity to arise. In numerical simulation or in real physical objects collapse will arise from fluctuations that lead to collapse and that appear with close to maximum probability.

Yu. N. Ovchinnikov was supported by grant CRDF RP1-194.

*)E-mail: ovc@itp.ac.ru

[1] V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1746 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
[2] V. E. Zakharov and V. S. Synakh, Zh. Éksp. Teor. Fiz. **68**, 940 (1975) [Sov. Phys. JETP **41**, 464 (1976)].
[3] B. J. LeMesurier *et al.*, Physica D **32**, 210 (1988).
[4] V. I. Talanov, JETP Lett. **11**, 199 (1971).
[5] K. Rypdal, I. Rasmussen, and K. Thomsen, Physica D **16**, 339 (1985).
[6] G. M. Fraĭman, Zh. Éksp. Teor. Fiz. **88**, 390 (1985) [Sov. Phys. JETP **61**, 228 (1985)].
[7] V. E. Zakharov and E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **91**, 1310 (1986) [Sov. Phys. JETP **64**, 773 (1986)].
[8] Yu. N. Ovchnikov, JETP Lett. **69**, 418 (1999).
[9] V. E. Zakharov and L. N. Shur, Zh. Éksp. Teor. Fiz. **81**, 2019 (1981) [Sov. Phys. JETP **54**, 1064 (1981)].

Translated by M. E. Alferieff

Simple analytical representation for Delbrück scattering amplitudes at high energies

R. N. Li, A. I. Mil'shteĭn,*) and V. M. Strakhovenko

G. I. Budker Nuclear Physics Institute, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 3 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 78–84 (July 1999)

We use a new representation for the semiclassical Green's function of the Dirac equation in the Coulomb field to find an exact (in the parameter $Z\alpha$) expression for the amplitudes of small-angle Delbrück scattering of high-energy photons. The values of these amplitudes agree with those obtained in earlier calculations, but the structure of the expressions is much simpler than that of previously known representations, which makes numerical calculations much easier.
 [S1063-7761(99)00807-0]

In recent years the process of Delbrück scattering[1] (the coherent photon scattering in the electric field of atoms via virtual electron-positron pairs) has been thoroughly studied by theoretical and experimental methods.[2] From the theoretical viewpoint, the process is interesting because of the important role of the higher-order terms in the perturbation-theory series in the parameter $Z\alpha$($Z|e|$ is the nuclear charge, $\alpha=e^2=1/137$ is the fine-structure constant, e is the electron charge, and $\hbar=c=1$). The range of high photon energies, $\omega\gg m$ (m is the electron mass), is especially important in studies of Delbrück scattering by experimental methods. A significant increase in the accuracy of measuring the Delbrück scattering cross section was achieved in a recent experiment[3] conducted at the G. I. Budker Nuclear Physics Institute with 140–450-MeV photons in the 2.6 to 16.6 mrad scattering-angle range. Exact (in $Z\alpha$) expressions for the amplitudes of Delbrück scattering in the Coulomb field, valid for $\omega\gg m$ and small scattering angles, were obtained by Cheng and Wu,[4–6] who summed the perturbation diagrams for the interaction with the external field in a certain approximation. In Refs. 7 and 8 these amplitudes were found by using the semiclassical Green's function of the Dirac equation in the Coulomb field. In Refs. 9 and 10 the semiclassical Green's function of the Dirac equation was obtained for an arbitrary spherically symmetric decreasing field, making it possible to calculate the Delbrück scattering in a screened potential.

Recently a successful experiment was conducted at the G. I. Budker Nuclear Physics Institute to observe another nonlinear QED effect, the splitting of a photon in the electric field of an atom (the preliminary results of this experiment can be found in Ref. 11). Exact (in $Z\alpha$) expressions for the amplitudes of photon splitting, valid for $\omega\gg m$, can be found in Refs. 12–14, where a new representation for the semiclassical Green's function that substantially simplifies the process of calculating the amplitudes was derived. In the present paper the calculation method developed in Refs. 12–14 is used to find the amplitudes of Delbrück scattering at high energies ($\omega\gg m$) and small photon scattering angles. As expected, the values of the amplitudes obtained by the new formulas coincide with those known from earlier calculations. However, the new representations of amplitudes are much simpler when used in numerical calculations, since they contain integrals of lower multiplicities.

As shown in Ref. 9, it is convenient to write the Delbrück scattering amplitude in a form that contains the Green's function $D(\mathbf{r},\mathbf{r}'|\varepsilon)$ of the squared Dirac equation,

$$D(\mathbf{r}_1,\mathbf{r}_2|\varepsilon)=\left\langle \mathbf{r}_1\left|\frac{1}{\hat{\mathscr{P}}^2-m^2+i0}\right|\mathbf{r}_2\right\rangle,$$

where $\hat{\mathscr{P}}=\gamma^0(\varepsilon+Z\alpha/r)-\gamma\mathbf{p}$, with $\mathbf{p}=-i\boldsymbol{\nabla}$.

In terms of the functions $D(\mathbf{r},\mathbf{r}'|\varepsilon)$, the amplitude of the process has the form[9]

$$\begin{aligned}M=i\alpha\int d\mathbf{r}_1\,d\mathbf{r}_2\exp\{i(\mathbf{k}_1\cdot\mathbf{r}_1-\mathbf{k}_2\cdot\mathbf{r}_2)\}\\ \times\int d\varepsilon\,\mathrm{Tr}[(2\mathbf{e}_2^*\cdot\mathbf{p}_2-\hat{e}_2^*\hat{k}_2)D(\mathbf{r}_2,\mathbf{r}_1|\omega-\varepsilon)]\\ \times[(2\mathbf{e}_1\cdot\mathbf{p}_1+\hat{e}_1\hat{k}_1)D(\mathbf{r}_1,\mathbf{r}_2|-\varepsilon)]+2i\alpha\mathbf{e}_2^*\cdot\mathbf{e}_1\\ \times\int d\mathbf{r}\exp\{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}\}\int d\varepsilon\,\mathrm{Tr}\,D(\mathbf{r},\mathbf{r}|\varepsilon),\end{aligned}\tag{1}$$

where e_1 and k_1 (e_2 and k_2) are the polarization and four-momentum of the initial (final) photon, and $\mathbf{p}_{1,2}=-i\boldsymbol{\nabla}_{1,2}$. In the limit $\omega\gg m$ the main contribution to the cross section of the process is provided by momentum transfers $\Delta\sim m$, which corresponds to small photon scattering angles. In this case the contribution of the last term in (1) to the amplitude can be ignored, since it depends solely on the momentum transfer vector $\boldsymbol{\Delta}=\mathbf{k}_2-\mathbf{k}_1$, and the amplitude for $\omega\gg\Delta$ is proportional to ω (see, e.g., Ref. 2).

The uncertainty relation implies that the lifetime of a virtual electron–positron pair is $\tau\sim|\mathbf{r}_2-\mathbf{r}_1|\sim\omega/(m^2+\Delta^2)$ and that the characteristic impact parameters are $\rho\sim1/\Delta$. Hence for $\omega\gg\Delta\gg m^2/\omega$ the angles between the vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{r}_2$, and $-\mathbf{r}_1$ are small and we can use the corresponding expansion. The characteristic values of the angular momentum are $l\sim\omega\rho\sim\omega/\Delta\gg1$, so that the semiclassical approximation holds. In the case of a screened Coulomb potential, the effect of screening manifests itself only in the limit

$\Delta \sim r_c^{-1} \ll m$, where the screening radius in the Thomas–Fermi model is $r_c \sim (m\alpha)^{-1} Z^{-1/3}$. In the present paper we limit ourselves to momentum transfers $\Delta \gg m^2/\omega, r_c^{-1}$, which provide the principal contribution to the total cross secion of the process. Note that this restriction is not required in calculating Coulomb corrections, and the expressions obtained in the present paper for these corrections are valid for all $\Delta \ll \omega$. A modification of the Born amplitude for $\Delta \sim m^2/\omega$ was studied by Cheng and Wu,[5] and a thorough discussion of the effect of screening can be found in Ref. 9.

In view of momentum conservation, $M(Z=0) \propto \delta(\mathbf{k}_1 - \mathbf{k}_2)$, i.e., for the case at hand $(\Delta \neq 0)$ we have $M(Z=0) = 0$. It is convenient to subtract from the integrand for M in (1) its value at $Z=0$. It is for this difference that the above statement of the smallness of the angles between $\mathbf{r}_2$ and $-\mathbf{r}_1$, which contribute the most to the integral, is valid.

A convenient representation for $D(\mathbf{r}_1, \mathbf{r}_2|\varepsilon)$, the semiclassical Green's function for the squared Dirac equation in the Coulomb field, was obtained in Refs. 12 and 13. For the case of small angles between the vectors $-\mathbf{r}_1$ and $\mathbf{r}_2$ we have (the z axis is directed so that it forms a small angle with the vector $\mathbf{r}_2$)

$$D(\mathbf{r}_1, \mathbf{r}_2|\varepsilon) = \frac{i\kappa}{8\pi^2 r_1 r_2} \exp\{i\kappa(r_1+r_2)\} \int d\mathbf{q} \left[1 + Z\alpha \frac{\boldsymbol{\alpha}\cdot\mathbf{q}}{\kappa q^2}\right] \times \exp\left\{ i\kappa \frac{q^2(r_1+r_2)}{2r_1 r_2} + i\kappa \mathbf{q}\cdot(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) \right\} \times \left(\frac{4r_1 r_2}{q^2}\right)^{iZ\alpha\lambda}, \tag{2}$$

where $\boldsymbol{\alpha} = \gamma^0 \boldsymbol{\gamma}$, $\kappa^2 = \varepsilon^2 - m^2$, $\lambda = \varepsilon/\kappa$, and $\mathbf{q}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ are two-dimensional vectors in the xy plane, with $\boldsymbol{\theta}_1 = \mathbf{r}_{1\perp}/r_1$ and $\boldsymbol{\theta}_2 = \mathbf{r}_{2\perp}/r_2$. Formula (2) contains only elementary functions, and the angles $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ enter only in the factor $\exp\{i\mathbf{q}\cdot(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)\}$. Hence representation (2) of the Green's function is very convenient for calculations.

We direct the z axis parallel to the vector $\mathbf{k}_1$. Note that in the small-angle approximation there is no need to allow for corrections to the transverse part of the polarization vector $\mathbf{e}_2$, and the longitudinal part of $\mathbf{e}_2$ can be expressed in terms of the traverse part by employing the relation $\mathbf{e}_2 \cdot \mathbf{k}_2 = 0$, i.e., $(\mathbf{e}_2)_z = -\mathbf{e}_2 \cdot \boldsymbol{\Delta}/\omega$. Thus, the transverse part of the polarization vector of the final photon with a given helicity can be replaced by the polarization vector of a photon propagating along the z axis and having the same helicity. Below we use the notation $\mathbf{e}$ for this vector in the case of positive helicity; the polarization vector with negative helicity is $\mathbf{e}^*$. Then, to describe Delbrück scattering, we need only find two helical amplitudes, M_{++} and M_{+-}. The other two, M_{--} and M_{-+}, can be found by interchanging $\mathbf{e}$ and $\mathbf{e}^*$.

Substituting (2) in (1), we expand the amplitudes for the case of small angles, when $d\mathbf{r}_1\, d\mathbf{r}_2 \approx r_1^2 r_2^2\, dr_1\, dr_2\, d\boldsymbol{\theta}_1\, d\boldsymbol{\theta}_2$. Calculating the trace of the γ-matrices and evaluating the elementary integral with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, we obtain

$$M = -\frac{i\alpha}{\omega^2} \int_0^\omega d\varepsilon\, \varepsilon\kappa \int_0^\infty \frac{dr_1}{r_1} \int_0^\infty \frac{dr_2}{r_2} \times \int\int \frac{d\mathbf{q}_1\, d\mathbf{q}_2}{(2\pi)^2} \left[\left(\frac{q_1}{q_2}\right)^{2iZ\alpha} - 1 \right] \exp\{i\Phi\}\, T. \tag{3}$$

Here

$$\Phi = \frac{1}{2}\left[\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\mathbf{Q}^2 + \frac{\varepsilon - \kappa}{\omega}\mathbf{Q}\cdot\boldsymbol{\Delta} + \mathbf{q}\cdot\boldsymbol{\Delta} - m^2(r_1+r_2) \right], \tag{4}$$

the function T for different polarizations has the form

$$T_{++} = \frac{2\mathbf{Q}^2}{r_1 r_2} - \frac{\omega^2}{2\varepsilon\kappa}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\left[\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\mathbf{Q}^2 - 2i \right],$$

$$T_{+-} = \frac{4}{r_1 r_2}(\mathbf{e}\cdot\mathbf{Q})^2, \tag{5}$$

and we have introduced the notation $\kappa = \omega - \varepsilon$, $\mathbf{Q} = \mathbf{q}_1 + \mathbf{q}_2$, and $\mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2$. To arrive at (3), we integrated by parts with respect to $\mathbf{q}_1$ and $\mathbf{q}_2$, so that the entire dependence on $Z\alpha$ is contained in the factor $(q_1/q_2)^{2iZ\alpha} - 1$. We also replaced variables according to $r_{1,2} \to (\varepsilon\kappa/\omega) r_{1,2}$. The expression for T_{++} in (5) can be made simpler if in (3) the variables r_1 and r_2 are replaced (temporarily) by the variables $R = r_1 r_2/(r_1 + r_2)$ and $t = r_1/r_2$ and the term in the square brackets in (5) proportional to $\mathbf{Q}^2$ is integrated by parts with respect to R. The result is

$$T_{++} = \frac{2\mathbf{Q}^2}{r_1 r_2} + \frac{\omega^2 m^2}{2\varepsilon\kappa r_1 r_2}(r_1 + r_2)^2. \tag{6}$$

Further transformations amount to the following. We pass from the variables $\mathbf{q}_1$ and $\mathbf{q}_2$ to the variables $\mathbf{q}$ and $\mathbf{Q}$. Then the integral with respect to $\mathbf{q}$ becomes

$$J = \int \frac{d\mathbf{q}}{\mathbf{Q}^2}\left[\left(\frac{|\mathbf{q}+\mathbf{Q}|}{|\mathbf{q}-\mathbf{Q}|}\right)^{2iZ\alpha} - 1 \right] \exp\left\{ -\frac{i}{2}\mathbf{q}\cdot\boldsymbol{\Delta} \right\}. \tag{7}$$

As shown in Ref. 12, this integral can be transformed into

$$J = \int \frac{d\mathbf{q}}{\boldsymbol{\Delta}^2}\left[\left(\frac{|\mathbf{q}+\boldsymbol{\Delta}|}{|\mathbf{q}-\boldsymbol{\Delta}|}\right)^{2iZ\alpha} - 1 \right] \exp\left\{ -\frac{i}{2}\mathbf{q}\cdot\mathbf{Q} \right\}. \tag{8}$$

Using this representation and the parametrization

$$\exp\left\{ i\frac{\mathbf{Q}^2}{2r_1} \right\} = i r_1 \int \frac{d\mathbf{x}}{2\pi} \exp\left\{ -i\frac{r_1 \mathbf{x}^2}{2} - i\mathbf{Q}\cdot\mathbf{x} \right\}, \tag{9}$$

where $\mathbf{x}$ is a two-dimensional vector lying in the same plane as $\mathbf{Q}$, we can easily evaluate the integrals in (3), first with respect to r_1, then with respect to $\mathbf{Q}$, and finally with respect to r_2. As a result we obtain

$$\begin{Bmatrix} M_{++} \\ M_{+-} \end{Bmatrix} = -\frac{i\alpha m^2}{\pi^2 \Delta^2 \omega^2} \int_0^\omega d\varepsilon \int d\mathbf{q} \left[\left(\frac{q_+}{q_-}\right)^{2iZ\alpha} - 1 \right] \times \int \frac{d\mathbf{x}}{(\mathbf{x}^2 + m^2)^2 (\mathbf{v}^2 + m^2)^2} \begin{Bmatrix} m^2(\varepsilon^2 + \kappa^2) + \omega^2 \mathbf{x}\cdot\mathbf{v} \\ 4\varepsilon\kappa(\mathbf{e}\cdot\mathbf{v})^2 \end{Bmatrix}, \tag{10}$$

where $q_\pm = |\mathbf{q}_\pm|$, $\mathbf{q}_\pm = \mathbf{q} \pm \boldsymbol{\Delta}$, and $\mathbf{v} = \mathbf{x} + \mathbf{q}/2 + \boldsymbol{\Delta}(\varepsilon - \kappa)/2\omega$.

For further integration it is convenient to write (10) in a different form. Using the identities

$$\frac{m^2}{(\mathbf{v}^2+m^2)^2}=\frac{1}{2}\boldsymbol{\nabla}_{\mathbf{v}}\frac{\mathbf{v}}{\mathbf{v}^2+m^2}=\boldsymbol{\nabla}_{\mathbf{q}}\frac{\mathbf{v}}{\mathbf{v}^2+m^2},$$

$$\frac{\mathbf{v}}{(\mathbf{v}^2+m^2)^2}=-\frac{1}{2}\boldsymbol{\nabla}_{\mathbf{v}}\frac{1}{\mathbf{v}^2+m^2}=-\boldsymbol{\nabla}_{\mathbf{q}}\frac{1}{\mathbf{v}^2+m^2}$$

and integrating by parts with respect to $\mathbf{q}$, we find that

$$\begin{Bmatrix} M_{++} \\ M_{+-} \end{Bmatrix}=\frac{2\alpha(Z\alpha)m^2}{\pi^2\Delta^2\omega^2}\int_0^{\omega}d\varepsilon\int d\mathbf{q}\left(\frac{q_+}{q_-}\right)^{2iZ\alpha}\times\int\frac{d\mathbf{x}}{(\mathbf{x}^2+m^2)^2(\mathbf{v}^2+m^2)}\left(\frac{\mathbf{q}_+}{q_+^2}-\frac{\mathbf{q}_-}{q_-^2}\right)\times\begin{Bmatrix}\omega^2\mathbf{x}-(\varepsilon^2+\kappa^2)\mathbf{v}\\ 4\varepsilon\kappa(\mathbf{e}\cdot\mathbf{v})\mathbf{e}\end{Bmatrix}. \tag{11}$$

Using Feynman parametrization of the denominators, we evaluate the integral with respect to $\mathbf{x}$ and in the integral with respect to ε introduce the variable $s=2\varepsilon/\omega-1$. The result is

$$\begin{Bmatrix} M_{++} \\ M_{+-} \end{Bmatrix}=\frac{4\alpha(Z\alpha)m^2\omega}{\pi\Delta^2}\int_{-1}^{1}ds\int d\mathbf{q}\left(\frac{q_+}{q_-}\right)^{2iZ\alpha}\times\int_0^1\frac{tdt}{[t(1-t)(\mathbf{q}+s\boldsymbol{\Delta})^2+4m^2]^2}\left(\frac{\mathbf{q}_+}{q_+^2}-\frac{\mathbf{q}_-}{q_-^2}\right)\times\begin{Bmatrix}[t(1-s^2)-2](\mathbf{q}+s\boldsymbol{\Delta})\\ 2t(1-s^2)(\mathbf{e},\mathbf{q}+s\boldsymbol{\Delta})\mathbf{e}\end{Bmatrix}. \tag{12}$$

The integration with respect to $\mathbf{q}$ is done by a trick used in Refs. 13 and 14. We multiply the integrand in (12) by

$$1\equiv\int_{-1}^{1}dy\,\delta\left(y-\frac{2\mathbf{q}\cdot\boldsymbol{\Delta}}{\mathbf{q}^2+\Delta^2}\right)=(\mathbf{q}^2+\Delta^2)\int_{-1}^{1}\frac{dy}{|y|}\,\delta\left(\left(\mathbf{q}-\frac{\boldsymbol{\Delta}}{y}\right)^2-\Delta^2\left(\frac{1}{y^2}-1\right)\right), \tag{13}$$

change the order of integrating with respect to $\mathbf{q}$ and y, and implement the shift $\mathbf{q}\to\mathbf{q}+\boldsymbol{\Delta}/y$. After all this is done, the integral with respect to $\mathbf{q}$ can easily be evaluated. Changing the variable $y=\tanh(\tau-\tau_0)$, where

$$\tau_0=\frac{1}{2}\ln\frac{B+(1+s)^2}{B+(1-s)^2},\quad B=\frac{4m^2}{\Delta^2 t(1-t)}, \tag{14}$$

we find that integrating with respect to τ reduces to calculating two integrals (the same integrals as in Ref. 14) expressed in terms of the derivative of the Legendre function, $P'_{\nu}(x)$:

$$\mathscr{F}_1=a^2\int_0^{\infty}d\tau\frac{\cosh\tau\cos(2Z\alpha\tau)}{(\sinh^2\tau+a^2)^{3/2}}=\frac{2\pi a^2}{\sinh(\pi Z\alpha)}\operatorname{Im}P'_{iZ\alpha}(2a^2-1),$$

$$\mathscr{F}_2=a^2\int_0^{\infty}d\tau\frac{\sinh\tau\sin(2Z\alpha\tau)}{(\sinh^2\tau+a^2)^{3/2}}=-\frac{2\pi a^2}{\sinh(\pi Z\alpha)}\operatorname{Re}P'_{iZ\alpha}(2a^2-1), \tag{15}$$

where $a^2=4B/[(B+(1+s)^2)(B+(1-s)^2)]$. Note that $a^2\leqslant 1$ for all s and $B>0$.

The final expressions for the Delbrück scattering amplitudes are

$$M_{++}=i\frac{\alpha(Z\alpha)\omega}{8m^2}\int_0^1 ds\int_0^1 dt\,a^2t[2-t(1-s^2)]\times[4sB\sin(2Z\alpha\tau_0)\mathscr{F}_1+[B^2-(s^2-1)^2]\cos(2Z\alpha\tau_0)\mathscr{F}_2], \tag{16}$$

$$M_{+-}=i\frac{\alpha(Z\alpha)\omega(\mathbf{e}\cdot\boldsymbol{\Delta})^2}{4m^2\Delta^2}\int_0^1 ds\int_0^1 dt\,a^2t(s^2-1)\times[4sB(1-t)\sin(2Z\alpha\tau_0)\mathscr{F}_1+[B^2(2-3t)+2B(s^2+1)(1-2t)-(s^2-1)^2t]\cos(2Z\alpha\tau_0)\mathscr{F}_2].$$

Let us now discuss the asymptotic behavior of the Delbrück scattering amplitudes for $\Delta\ll m$ and $\Delta\gg m$, which follows from (16). The simplest calculation in this representation is that of the asymptotic expression for $\Delta\ll m$, in the limit $B\sim m^2/\Delta^2\gg 1$, $a^2\approx 4/B\ll 1$, and the functions $\mathscr{F}_{1,2}$ are

$$\mathscr{F}_1\approx 1,\quad \mathscr{F}_2=-2Z\alpha a^2[\ln a+C+\operatorname{Re}\psi(1+iZ\alpha)],$$

where $C=0.577\ldots$ is Euler's constant, and $\psi(x)=d\ln\Gamma(x)/dx$. Also, $\tau_0\sim 1/B\ll 1$. Inserting the asymptotic expressions for $\mathscr{F}_{1,2}$ in (16) yields

$$\begin{Bmatrix} M_{++} \\ M_{+-} \end{Bmatrix}=i\frac{4\alpha(Z\alpha)^2\omega}{m^2}\int_0^1 ds\int_0^1 t\,dt\left[\frac{1}{2}\ln\frac{m^2}{t(1-t)\Delta^2}-C-\operatorname{Re}\psi(1+iZ\alpha)\right]\begin{Bmatrix}[2-t(1-s^2)]\\ 2(1-s^2)(3t-2)(\mathbf{e}\cdot\boldsymbol{\Delta})^2/\Delta^2\end{Bmatrix}. \tag{17}$$

Calculating the elementary integrals, we find that for $m^2/\omega\ll\Delta\ll m$ we have

$$M_{++}=i\frac{28\alpha(Z\alpha)^2\omega}{9m^2}\left[\ln\frac{m}{\Delta}+\frac{41}{42}-C-\operatorname{Re}\psi(1+iZ\alpha)\right],$$

$$M_{+-}=i\frac{4\alpha(Z\alpha)^2\omega(\mathbf{e}\cdot\boldsymbol{\Delta})^2}{9m^2\Delta^2}. \tag{18}$$

To establish the asymptotic behavior of the amplitudes in the limit $\Delta\gg m$, it is convenient to start with (12). The region $1-t\sim m^2/\Delta^2\ll 1$ provides the main contribution to the integral. If in (12) we evaluate the integral with respect to t in this approximation, we get

$$\begin{Bmatrix} M_{++} \\ M_{+-} \end{Bmatrix} = \frac{\alpha(Z\alpha)\omega}{\pi\Delta^2}\int_{-1}^{1} ds \int \frac{d\mathbf{q}}{(\mathbf{q}+s\boldsymbol{\Delta})^2}\left(\frac{q_+}{q_-}\right)^{2iZ\alpha} \times\left(\frac{\mathbf{q}_+}{q_+^2}-\frac{\mathbf{q}_-}{q_-^2}\right)\begin{Bmatrix} -(1+s^2)(\mathbf{q}+s\boldsymbol{\Delta}) \\ 2(1-s^2)(\mathbf{e},\mathbf{q}+s\boldsymbol{\Delta})\mathbf{e} \end{Bmatrix}. \quad (19)$$

We then use the identity (13) to evaluate the elementary integrals with respect to $\mathbf{q}$ and, finally, with respect to s. Then we change variables, $y=\tanh\tau$, and for $\Delta \gg m$ obtain

$$M_{++} = i\frac{4\alpha(Z\alpha)\omega}{3\Delta^2}\int_0^\infty d\tau \sin(2Z\alpha\tau) \times\left(4-3\tanh\frac{\tau}{2}-\tanh^3\frac{\tau}{2}\right) = i\frac{8\alpha\omega}{3\Delta^2}\left\{1-\frac{2\pi Z\alpha}{\sinh(2\pi Z\alpha)}[1-(Z\alpha)^2]\right\},$$

$$M_{+-} = i\frac{16\alpha(Z\alpha)\omega(\mathbf{e}\cdot\boldsymbol{\Delta})^2}{\Delta^4}\int_0^\infty \frac{d\tau\sin(2Z\alpha\tau)}{\sinh^2\tau}(\tau\coth\tau-1) = i\frac{16\alpha(Z\alpha)^2\omega(\mathbf{e}\cdot\boldsymbol{\Delta})^2}{\Delta^4}[1-Z\alpha\,\mathrm{Im}\,\psi'(1-iZ\alpha)]. \quad (20)$$

The asymptotic formulas (18) and (20) coincide with the results obtained in Refs. 6 and 8. We also found that the numerical values of the amplitudes (16) coincide (as expected) with the results obtained in Refs. 6–8 for arbitrary momentum transfers and values of $Z\alpha$.

Each expression for the amplitudes in (16) is a double integral and is simpler than the results of Refs. 6–8, in which the multiplicity of the integrals is higher. Hence the formulas in (16) are very convenient for tabulation.

Thus, in this paper, using Delbrück scattering as an example, we have shown that the method of calculation developed for the photon splitting problem can be successfully used to solve various QED problems in the Coulomb field at high energies.

*)E-mail: A.I.Milshtein@inp.nsk.su

[1] L. Meitner, H. Kösters, and M. Delbrück, Z. Phys. **84**, 137 (1933).
[2] A. I. Milshtein and M. Schumacher, Phys. Rep. **243**, 183 (1994).
[3] Sh. Zh. Akhmadaliev, G. Ya. Kezerashvili, S. G. Klimenko, V. M. Malyshev, A. L. Maslennikov, A. M. Milov, A. I. Milstein, N. Yu. Muchnoi, A. I. Naumenkov, V. S. Panin, S. V. Peleganchuk, V. G. Popov, G. E. Pospelov, I. Ya. Protopopov, L. V. Romanov, A. G. Shamov, D. N. Shatilov, E. A. Simonov, and Yu. A. Tikhonov, Phys. Rev. C **58**, 2844 (1998).
[4] M. Cheng and T. T. Wu, Phys. Rev. **182**, 1873 (1969).
[5] M. Cheng and T. T. Wu, Phys. Rev. D **2**, 2444 (1970).
[6] M. Cheng and T. T. Wu, Phys. Rev. D **5**, 3077 (1972).
[7] A. I. Milstein and V. M. Strakhovenko, Phys. Lett. A **95**, 135 (1983).
[8] A. I. Mil'shteĭn and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **85**, 14 (1983) [Sov. Phys. JETP **58**, 8 (1983)].
[9] R. N. Lee and A. I. Milstein, Phys. Lett. A **198**, 217 (1995).
[10] R. N. Li and A. I. Mil'shteĭn, Zh. Éksp. Teor. Fiz. **107**, 1393 (1995) [JETP **80**, 777 (1995)].
[11] Sh. Zh. Akhmadaliev, G. Ya. Kezerashvili, S. G. Klimenko *et al.*, in *Proc. Int. Conf. PHOTON'97*, Egmond-aan Zee, The Netherlands, A. Buijs and F. C. Erne (Eds.), World Scientific, Singapore (1997), p. 246.
[12] R. N. Li, A. I. Mil'shteĭn, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **112**, 1921 (1997) [JETP **85**, 1049 (1997)].
[13] R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Phys. Rev. A **57**, 2325 (1998).
[14] R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Phys. Rev. A **58**, 1757 (1998).

Translated by Eugene Yankovsky

Large-scale amplification of a magnetic field by turbulent helicity fluctuations

N. A. Silant'ev*)

Instituto National de Astrofisica, Optica y Electronica, Apartado Postal 51 y 216 C. P. 72000 Puebla, Pue. Mexico; Central Astronomical Observatory, Russian Academy of Sciences, 196140 St. Petersburg, Russia

(Submitted 21 October 1998; resubmitted 15 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 85–104 (July 1999)

In this paper the procedure of large-scale averaging of the magnetic-field diffusion equation with the α-term $\mathrm{curl}\alpha(\mathbf{r},t)\mathbf{B}(\mathbf{r},t)$ is used to show that a nonuniform distribution of the turbulent helicity fluctuations (more precisely, the fluctuations of the coefficient α) with a zero average value gives rise to large-scale amplification of the initial magnetic field. A detailed study is carried out of the dependence of the resulting large-scale α effect on the characteristics of the correlator $\langle\langle\alpha(\mathbf{r},t)\alpha(\mathbf{r}',t')\rangle\rangle$ in a rotating medium with a nonuniform distribution of the angular velocity $\omega=\omega(\rho,z)$ (ρ is the distance for the rotation axis z). The effect of helicity fluctuations and the diffusion coefficient on the turbulent diffusion process is also investigated. © *1999 American Institute of Physics.* [S1063-7761(99)00907-5]

1. INTRODUCTION

It is well known that amplification of the initial magnetic field in a turbulent medium requires the presence of helicity, i.e., the helicity correlator $h(1,2)\equiv\langle\mathbf{u}(1)\cdot\mathrm{curl}\,\mathbf{u}(2)\rangle$ must be finite. The case $h>0$ corresponds to right-handed helicity, while $h<0$ corresponds to left-handed helicity. Such a medium may be isotropic, but it certainly must be reflection-asymmetric. The amplification mechanism was developed primarily by Parker[1] and Steenbeck *et al.*[2] and is described in detail in the monographs cited in Refs. 3–5. A straightforward qualitative explanation of the turbulent magnetic dynamo can be found in Ref. 6. It is based on the fact that chaotic helical movements generate an average emf $\langle\mathbf{E}\rangle$ parallel or antiparallel to the average magnetic field: $\langle\mathbf{E}\rangle=\alpha\langle\mathbf{B}\rangle$. According to this common notation, the phenomenon became known as the α effect. Note that $\alpha\propto(-h)$, i.e., left-handed helicity gives rise to induction of a current $\langle\mathbf{j}\rangle$ parallel to the average field $\langle\mathbf{B}\rangle$, while right-handed helicity gives rise to induction of a current opposing $\langle\mathbf{B}\rangle$. The induction of such a current is the reason for the amplification of the average magnetic field $\langle\mathbf{B}\rangle$. The existence of the α effecthas been confirmed by experiments.[7]

The initial equation describing the evolution of the magnetic field is the induction equation

$$\left(\frac{\partial}{\partial t}-D_m\nabla^2\right)\mathbf{B}(\mathbf{r},t)=\mathrm{curl}[\mathbf{u}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t)], \tag{1}$$

where $D_m=c^2/4\pi\sigma$ is the molecular (ohmic) diffusion coefficient, σ is the electrical conductivity of the plasma, and $\mathbf{u}(\mathbf{r},t)$ is the Eulerian velocity of motion of matter. The Navier-Stokes equation, which we will not write here, relates $\mathbf{u}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$, so that actually the velocity $\mathbf{u}(\mathbf{r},t)$ is a functional of $\mathbf{B}(\mathbf{r},t)$, and vice versa. For weak magnetic fields ($B^2/8\pi\ll\rho u^2/2$) one usually ignores the effect of $\mathbf{B}$ on turbulence and considers Eq. (1) in the kinematic representation, where the ensemble of turbulent velocities $\mathbf{u}(\mathbf{r},t)=\mathbf{U}(\mathbf{r},t)+\mathbf{u}'(\mathbf{r},t)$ ($\mathbf{U}=\langle\mathbf{u}\rangle$ and $\langle\mathbf{u}'\rangle=0$) is assumed fixed. The quantities $u_0^2=\langle\mathbf{u}'^2(\mathbf{r},t)\rangle$, τ_0, and R_0 characterize the average amplitude of the turbulent velocity, the average lifetime of correlations of Eulerian velocities, and the average length of such correlations. Sometimes the auxiliary parameter $t_0=R_0/u_0$ is used (it is known as the reversal time of turbulent vortices). The average magnetic field $\langle\mathbf{B}\rangle$ is defined as the average value over the volume $\approx L^3$ and over the averaging time T_0, which are much larger that the characteristic length R_0 and lifetime τ_0 of correlations of turbulent movements (if the reversal time of the turbulent vortices meets the condition $t_0=R_0/u_0\ll\tau_0$ it is sufficient to require that $T\gg t_0$).

By averaging Eq. (1) and assuming that the average field $\langle\mathbf{B}\rangle$ is a smooth function on distances $\approx R_0$ and times $\approx\tau_0$ (or $\approx t_0$ if $t_0\ll\tau_0$) one can obtain (see Refs. 3–5) the equation of diffusion for the average magnetic field:

$$\left(\frac{\partial}{\partial t}-D_m\nabla^2\right)\langle\mathbf{B}\rangle+\mathrm{curl}\,D(\mathbf{r},t)\mathrm{curl}\langle\mathbf{B}\rangle - \mathrm{curl}\,\alpha(\mathbf{r},t)\langle\mathbf{B}\rangle-\mathrm{curl}[\mathbf{U}(\mathbf{r},t)\langle\mathbf{B}\rangle]=\mathbf{0}, \tag{2}$$

where $\mathbf{U}(\mathbf{r},t)$ is the velocity of regular motion of the liquid or gas. Note that the coefficients of turbulent diffusion, $D(\mathbf{r},t)$, and the α effect, $\alpha(\mathbf{r},t)$, in this equation can depend on $\langle\mathbf{B}\rangle$ parametrically, since the turbulent velocity field $\mathbf{u}(\mathbf{r},t)$ depends on $\mathbf{B}$. Kichatinov *et al.*[8,9] used the functional dependence of $D(\mathbf{r},t)$ and $\alpha(\mathbf{r},t)$ on $\mathbf{u}(\mathbf{r},t)$ but, in solving the Navier-Stokes equation, allowed for the fact that the field $\mathbf{u}(\mathbf{r},t)$ depends on $\mathbf{B}$. The derivation of Eq. (2) from Eq. (1) requires only that the average field $\langle\mathbf{B}\rangle$ be smooth, so that we can take it outside the integral sign. Then under the integral sign there remains another, peaked, term (e.g., see Refs. 10 and 11). We also note that here we have assumed, for simplicity, that $D(\mathbf{r},t)$ and $\alpha(\mathbf{r},t)$ are scalar quantities, although in an anisotropic turbulent medium they are pseudotensors of the third and second ranks, respectively (see Ref. 3). If we

neglect D_m in Eq. (1), the equation has an exact solution in Lagrangian variables. The derivation of Eq. (2) from this solution is especially simple,[12,13] but it uses only the smoothness of the average magnetic field $\langle \mathbf{B} \rangle$. Exact formal expressions for the coefficients $D(\mathbf{r},t)$ and $\alpha(\mathbf{r},t)$ in terms of the field $\mathbf{u}(\mathbf{r},t)$ of Eulerian velocities or the field $\mathbf{v}(\mathbf{r},t)$ of Lagrangian velocities can be found in Refs. 10–13. Using these expressions and specifying the statistical properties of the ensemble of initial velocities, we can (at least in principle) fully describe the evolution of the average magnetic field in a turbulent medium by solving Eq. (2) and the Navier-Stokes equation. Lately there has been an increasing number of attempts of computer simulations of the evolution of the magnetic field based directly on Eq. (1) and Navier-Stokes equations. However, in studying the influence of helicity fluctuations on the large-scale α effect we will find it convenient to begin with the diffusion equation (2).

The large-scale averaging procedure

If $\langle \cdots \rangle$ in (2) is assumed to be the average over fairly small-scale turbulent motions, the coefficients $D(\mathbf{r},t)$ and $\alpha(\mathbf{r},t)$ are still random functions with respect to large-scale averaging. We will denote the procedure of large-scale averaging by double angle brackets, $\langle\langle \cdots \rangle\rangle$. Using this notation, we can write the following relationships:

$$
\begin{aligned}
&D(\mathbf{r},t) = D_0 + D'(\mathbf{r},t),\\
&\langle\langle D(\mathbf{r},t)\rangle\rangle \equiv D_0, \quad \langle\langle D'(\mathbf{r},t)\rangle\rangle \equiv 0,\\
&\alpha(\mathbf{r},t) = \alpha_0 + \alpha'(\mathbf{r},t),\\
&\langle\langle \alpha(\mathbf{r},t)\rangle\rangle \equiv \alpha_0, \quad \langle\langle \alpha'(\mathbf{r},t)\rangle\rangle \equiv 0.
\end{aligned}
\tag{3}
$$

Here and below we assume that the large-scale average values D_0 and α_0 are constants. Note that the procedures of small- and large-scale averaging have long been used in statistical physics (see, e.g., Ref. 14).

Substituting (3) in (2), we obtain the master stochastic equation

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t} - (D_m + D_0)\nabla^2\right)\langle \mathbf{B} \rangle - \alpha_0 \operatorname{curl}\langle \mathbf{B} \rangle - \operatorname{curl}[\mathbf{U}\langle \mathbf{B} \rangle]\\
&\quad = \operatorname{curl} \alpha'(\mathbf{r},t)\langle \mathbf{B} \rangle - \operatorname{curl} D'(\mathbf{r},t)\operatorname{curl}\langle \mathbf{B} \rangle.
\end{aligned}
\tag{4}
$$

The right-hand side of Eq. (4) is a stochastic function of coordinates $\mathbf{r}$ and time t. We note once more that the small-scale average field $\langle \mathbf{B} \rangle$ is stochastic with respect to large-scale averaging.

Equation (4) can be written in brief symbolic form:

$$
\hat{L}_{ij}^{(0)}(\mathbf{r},t)\langle B_j(\mathbf{r},t)\rangle = \hat{L}_{ij}(\mathbf{r},t)\langle B_j(\mathbf{r},t)\rangle, \tag{5}
$$

where the nonstochastic operator $\hat{L}^{(0)}$ and the stochastic operator $\hat{L}$ represent the left- and right-hand sides of Eq. (4).

The first to examine this equation for the case of $\mathbf{U}=0$, $D'=0$, and $\alpha_0=0$ and an infinite isotropic medium were Kraichnan[13] and Moffatt.[3] Performing in (4) the procedure of large-scale averaging in the quasilinear approximation, they again arrived at equation (2) for $\langle\langle \mathbf{B}(\mathbf{r},t)\rangle\rangle$ in the diffusion approximation, which, however, contained the renormalized turbulent diffusion coefficient $D_m + D_0 + D$. They found that the term D is negative, i.e., the helicity fluctuations (more precisely, the fluctuations of the coefficient α, which are proportional to helicity) reduce the diffusion of the magnetic field.

Vishniac and Brandenburg[15] used Eq. (4) for computer simulation of the evolution of the average magnetic field in an accretion disk. They also examined the case where there is a chaotic distribution of the helicity fluctuations α' with a zero average $\alpha_0=0$ in the disk. It was found (numerically) that this gives rise to a large-scale distribution of the magnetic field, i.e., α'-fluctuations act as an α effect.

Sokolov[16] found that the presence of helicity fluctuations does indeed cause amplification of the large-scale magnetic field even at zero average helicity. He assumed that there are two averaging ensembles, over the turbulence velocity $\mathbf{u}(\mathbf{r},t)$ proper and over the helicity distribution ensemble, $h(1,2) \equiv \langle \mathbf{u}(1)\cdot\operatorname{curl}\mathbf{u}(2)\rangle$. Physically it is difficult to imagine that these two ensembles can exist separately, since knowing the velocity field $\mathbf{u}(\mathbf{r},t)$ obviously means that we know the helicity correlator $h(1,2)\equiv\langle \mathbf{u}(1)\cdot\operatorname{curl}\mathbf{u}(2)\rangle$ and generally all the correlators. Hence the idea of two averaging ensembles can actually be realized by the use of two averaging procedures, a fairly small-scale and a large-scale one.

The authors of all the cited papers did not do large-scale averaging of Eq. (4) in general form. In this paper we will derive the general formulas for the renormalized coefficients of turbulent diffusion and the α effect, the phenomena caused by helicity correlations in the turbulent medium. Using a medium with differential rotation as an example, we will show both quantitatively and qualitatively the reason for the emergence of a large-scale α effect. We will use the general theory discussed in Refs. 10 and 11, where the procedure of averaging over the ensemble of realizations of fluctuating quantities is carried out by using the stochastic Green's tensor $G_{ij}(\mathbf{r},t;\mathbf{r}',t')$, deriving a renormalized equation for this tensor, and obtaining an hierarchy of nonlinear equations for the average tensor $\langle G_{ij}(\mathbf{r},t;\mathbf{r}',t')\rangle$. The averaging procedure proper is described in Refs. 10 and 11 in general form for an equation of the form (5), which can be applied to our initial stochastic equation (4).

The helicity of a turbulent medium, $h(1,2) \equiv \langle \mathbf{u}(1)\cdot\operatorname{curl}\mathbf{u}(2)\rangle$, is described by a two-point velocity correlator (here and below we use the convenient notation $f(\mathbf{r}_n,t_n)\equiv f(n)$, $d\mathbf{r}_n\,dt_n \equiv dn$, $d\mathbf{r}'\,dt' \equiv d2$, $\mathbf{R}=\mathbf{r}-\mathbf{r}'$, $\tau = t - t'$, etc.). This means that helicity fluctuations are described by velocity correlators of the fourth and higher orders. A direct study of the dynamics of correlations whose orders are so high is extremely difficult. Note that in the classical work of Kazantsev[17] and in a number of later treatments (see Refs. 18 and 19), the dynamics of magnetic-field fluctuations was studied for turbulence without helicity, which requires only knowing the second-order velocity correlators $\langle u_i(1)u_j(2)\rangle$. The advantage of all these treatments is that their authors begin directly from the master induction equation (1) and do not resort to the procedure of the second, large-scale, averaging. The procedure of double averaging used in the present paper makes it possible to study the effect of helicity fluctuations (as fluctuations of the α effect) in a fairly simple manner. At present such an approach is probably the only

one that can be used to study helicity fluctuations.

For practical applications of the theory we will now develop it is important to understand that the helicity fluctuation correlator is a fundamental quantity having the same status for turbulence as the ordinary two-point velocity correlator $\langle u_i(1)u_j(2)\rangle$. More than that, for a given correlator $\langle u_i(1)u_j(2)\rangle$ the helicity fluctuation correlator for non-Gaussian ensembles may be arbitrary (at least in principle). Usually the correlator $\langle u_i(1)u_j(2)\rangle$ is found from experiments or observations or is simply specified in a plausible form for numerical calculations. This is even more true in relation to the helicity fluctuation correlator, which is extremely sensitive to the specific form of the ensemble of initial velocities, the boundary conditions, and the acting forces.

Note that for all practical purposes the second average can be realized either by averaging over large volumes or by averaging over long time intervals.

2. GENERAL FORMULAS FOR THE KINETIC COEFFICIENTS

The general theory, presented in Refs. 10 and 11, makes it possible to obtain from (4) a formally exact integro-differential equation for the large-scale averaged magnetic field $\langle\langle \mathbf{B}\rangle\rangle$. Using the symbolic form of Eq. (4) [see Eq. (5)], for the case of a Gaussian ensemble we obtain

$$\hat{L}_{ij}^{(0)}(1)\langle\langle B_j(1)\rangle\rangle = \text{irreducible part of}$$
$$\times\int d2\,\langle\langle \hat{L}_{in}(1)G_{nm}(1;2)\hat{L}_{mj}(2)\rangle\rangle$$
$$\times\langle\langle B_j(2)\rangle\rangle. \tag{6}$$

If we again use the fact that the field $\langle\langle \mathbf{B}\rangle\rangle$ is smooth, this equation in the diffusion approximation becomes

$$\left(\frac{\partial}{\partial t}-(D_m+D_0)\nabla^2\right)\langle\langle \mathbf{B}\rangle\rangle-\alpha_0\,\mathrm{curl}\langle\langle \mathbf{B}\rangle\rangle$$
$$-\mathrm{curl}[\mathbf{U}\langle\langle \mathbf{B}\rangle\rangle]=\mathrm{curl}\,\mathbf{E}, \tag{7}$$

$$E_i=\alpha_{ij}(\mathbf{r},t)\langle\langle B_j(\mathbf{r},t)\rangle\rangle+D_{ijk}(\mathbf{r},t)\nabla_k\langle\langle B_j(\mathbf{r},t)\rangle\rangle. \tag{8}$$

The left-hand side of Eq. (7) coincides with the left-hand side of Eq. (4). Equations (7) and (8) are the equations of a turbulent magnetic dynamo in the diffusion approximation. The pseudotensors $\alpha_{ij}(\mathbf{r},t)$ and $D_{ijk}(\mathbf{r},t)$ are generalizations of the pseudoscalar coefficient α and the turbulent diffusion coefficient D and describe the effect of the helicity fluctuations $\alpha'(\mathbf{r},t)$ and the fluctuations of the turbulent diffusion coefficient, $D'(\mathbf{r},t)$, on the evolution of the average magnetic field $\langle\langle \mathbf{B}\rangle\rangle$. Their exact expressions are given by the formulas

$$\alpha_{ij}(1)=\text{irreducible part of}$$
$$\times\int d\mathbf{r}'\int_0^t dt'\,[e_{jnm}\langle\langle \alpha'(\mathbf{r},t)G_{in}(\mathbf{r},t;\mathbf{r}',t')$$
$$\times\nabla'_m\alpha'(\mathbf{r}',t')\rangle\rangle-e_{its}e_{rnj}\langle\langle D'(\mathbf{r},t)\nabla_t G_{sr}$$
$$\times(\mathbf{r},t;\mathbf{r}',t')\nabla'_n\alpha'(\mathbf{r}',t')\rangle\rangle]. \tag{9}$$

Here e_{ijk} is the unit antisymmetric pseudotensor ($e_{xyz}=-e_{yxz}$, etc.), $G_{ij}(1;2)$ is the stochastic Green's tensor of the initial equation (4), i.e., the solution of this equation with source $S_{ij}=\delta_{ij}\delta(\mathbf{r}-\mathbf{r}')\delta(t-t')$. Here $G_{ij}^{(0)}(1;2)$ denotes the Green's tensor of Eq. (4) without the right-hand side. This tensor is nonstochastic. Taking the irreducible parts in the expressions means that we must drop the reducible parts of the form

$$\int d2\,\langle\langle F_{is}(1)\rangle\rangle G_{sn}^{(0)}(1;2)\langle\langle K_{nm}(2)\rangle\rangle. \tag{10}$$

The formula for $D_{ijk}(\mathbf{r},t)$ is more cumbersome than formula (9), and we will use the convenient notation mentioned earlier:

$$D_{ijk}(1)=-\text{irreducible part of}$$
$$\times\int d2\,[e_{jrn}\langle\langle \alpha'(1)G_{ir}(1;2)$$
$$\times(\nabla_n^{(2)}\alpha'(2)R_k)\rangle\rangle$$
$$+\langle\langle \alpha'(1)G_{ik}(1;2)\nabla_j^{(2)}D'(2)\rangle\rangle$$
$$-\langle\langle \alpha'(1)G_{ij}(1;2)\nabla_k^{(2)}D'(2)\rangle\rangle$$
$$-e_{its}e_{jrn}\langle\langle D'(1)(\nabla_t^{(1)}G_{sr}(1;2))$$
$$\times(\nabla_n^{(2)}\alpha'(2)R_k)\rangle\rangle-e_{its}\langle\langle D'(1)\nabla_t^{(1)}$$
$$\times G_{sk}(1;2)\nabla_j^{(2)}D'(2)\rangle\rangle+e_{its}\langle\langle D'(1)\nabla_t^{(1)}$$
$$\times G_{sj}(1;2)\nabla_k^{(2)}D'(2)\rangle\rangle]. \tag{11}$$

The function $D'(\mathbf{r},t)$ is even in the helicity h and the function $\alpha'(\mathbf{r},t)$ is odd. This means that the correlator $\langle\langle D'(1)\alpha'(2)\rangle\rangle$ is zero if the average helicity α_0 vanishes. All the formulas become much simpler for this case. Since according to Ref. 20 the turbulent diffusion coefficient is almost independent of helicity for a degree of helicity $a\leqslant 0.5$, the contribution of the cross terms of the form $\langle\langle D'(1)\alpha'(2)\rangle\rangle$ is probably small even for $\alpha_0\neq 0$. Even at 100% helicity the decrease in the diffusion coefficient is insignificant, about 30%. Hence below we will ignore the cross terms.

There are several ways of obtaining an approximation series for the stochastic Green's tensor $G_{ij}(1;2)$ (see Refs. 10 and 11). Inserting this a series into (9) and (11) yields approximation series for the kinetic coefficients $\alpha_{ij}(\mathbf{r},t)$ and $D_{ijk}(\mathbf{r},t)$, respectively.

For the case of the stochastic equation (1) with $D_m\to 0$, plugging the iterations of the ordinary integral equation for $G_{ij}(1;2)$ into the formulas for D_0 and α_0 leads to a series in powers of the parameter $\xi_0=u_0\tau_0/R_0$, which makes it possible to calculate D_0 and α_0 only when $\xi_0\ll 1$. This result is understandable since the expansion is in powers of the Green's function

$$G_m(R,\tau)=\frac{1}{(4\pi D_m\tau)^{3/2}}\exp\left\{-\frac{R^2}{4D_m\tau}\right\},$$

which describes molecular diffusion rather than convective transport by turbulent movements. To obtain D_0 and α_0 in

this case we are forced (see Refs. 10 and 11) to use a renormalized equation for $G_{ij}(1;2)$, with the average Green's function $\langle G_{ij}(1;2)\rangle$, for which there is a hierarchy of nonlinear equations, used for the absolute term. The tensor $\langle G_{ij}(1;2)\rangle$ directly describes the convective nature of the transport of the impurity field, and iterations in powers of $\langle G_{ij}(1;2)\rangle$ lead to an asymptotically convergent series for D_0 and α_0 at all values of ξ_0. A fairly good result is achieved if instead of $G_{ij}(1;2)$ we plug in the average tensor $\langle G_{ij}(1;2)\rangle$ satisfying the first equation in the hierarchy (what is known as the Direct Interaction Approximation, or DIA, equation).

The stochastic equation (4) used in the present paper differs from Eq. (1) in that its left-hand side describes turbulent transport of the magnetic field (to be sure, in the rough, diffusion, approximation) by the coefficients D_0 and α_0. This suggests that we can do without complicated iteration techniques in relation to the renormalized equation for $G_{ij}(1;2)$ and the solution of the nonlinear equation for $\langle G_{ij}(1;2)\rangle$. In this case the iterations of the ordinary integral equation for $G_{ij}(1;2)$,

$$G_{ij}(1;2)=G_{ij}^{(0)}(1;2)+\int d3\, G_{in}^{(0)}(1;3)\hat{L}_{nm}(3)G_{mj}(3;2), \qquad (12)$$

will probably make it possible to obtain fairly accurate values of $\alpha_{ij}(\mathbf{r},t)$ and $D_{ijk}(\mathbf{r},t)$.

For an example that supports these expectations we take the simple problem of funding the field $\mathbf{B}(\mathbf{r},t)$ in an infinite medium that is at rest as a whole and where a helicity fluctuation is of the form $\alpha(\mathbf{r},t)=\alpha_0\exp\{-r^2/r_0^2\}\exp\{-t/\tau_0\}$ for $t>0$. The initial field $\mathbf{B}_0$ is assumed uniform. Solving Eq. (4) with the absolute term

$$G_{ij}^{(0)}(1;2)=\frac{\delta_{ij}}{(4\pi D_0\tau)^{3/2}}\exp\left\{-\frac{R^2}{4D_0\tau}\right\}$$

and the use of the first iteration of Eq. (12), we arrive at the expressions

$$\tilde{\mathbf{B}}^{(0)}(\mathbf{p},t)=\alpha_0V_0i(\mathbf{p}\times\mathbf{B}_0)\times\int_0^t d\tau\,\alpha(t-\tau)\exp\left[-\frac{p^2}{p_0^2}(1+D_0p_0^2\tau)\right],$$

$$\tilde{B}_i^{(1)}=(\delta_{ij}p^2-p_ip_j)B_{0j}\frac{\alpha_0^2V_0^2p_0^3}{8\pi\sqrt{\pi}}\times\int_0^t d\tau\int_0^{t-\tau}d\tau'\frac{\alpha(t-\tau)\alpha(t-\tau-\tau')}{(2+D_0p_0^2\tau')^{5/2}}\times\exp\left[-\frac{p^2}{p_0^2}\left(D_0p_0^2\tau+\frac{1+D_0p_0^2\tau'}{2+D_0p_0^2\tau'}\right)\right], \qquad (13)$$

where we have introduced the notation $V_0=\pi\sqrt{\pi}\,r_0^3$, $p_0=2/r_0$, and $\alpha(t)=\exp\{-t/\tau_0\}$; $\tilde{\mathbf{B}}^{(0)}(\mathbf{p},t)$ and $\mathbf{B}^{(1)}(\mathbf{p},t)$ are the Fourier transforms in the variable $\mathbf{r}$ of the fields $\mathbf{B}^{(0)}(\mathbf{r},t)$ and $\mathbf{B}^{(1)}(\mathbf{r},t)$, respectively. Using these expressions, we will find the total energies of the magnetic fields, $A^{(0)}(t)$ and $A^{(1)}(t)$, generated by the α effect in the entire space in the time τ_0 of action of a helicity fluctuation (it should be recalled that in the final analysis, as $t\to\infty$, turbulent diffusion will destroy these fields). We will examine the limits of small, $\gamma=D_0p_0^2\tau_0\ll1$, and large, $D_0p_0^2\tau_0\gg1$, turbulent diffusion. We have

$$A^{(0)}(\tau_0)\approx\frac{B_0^2}{8\pi}V_0\frac{\alpha_0^2V_0p_0^5\tau_0^2}{72\pi\sqrt{2\pi}},$$

$$A^{(1)}(\tau_0)\approx A^{(0)}(\tau_0)\frac{5\alpha_0^2V_0^2p_0^8\tau_0^2}{64\times72\pi^3\sqrt{2}} \qquad (14)$$

for $\gamma\ll1$ and

$$A^{(0)}(\tau_0)\approx\frac{B_0^2}{8\pi}V_0\frac{\alpha_0^2V_0p_0}{54\pi\sqrt{2\pi}D_0^2},$$

$$A^{(1)}(\tau_0)\approx A^{(0)}(\tau_0)\frac{\alpha_0^2V_0^2p_0^4}{648\times81\pi^3D_0^2} \qquad (15)$$

for $\gamma\gg1$. Let us estimate the ratios $A^{(1)}/A^{(0)}$ for the model of a helicity fluctuation in the form of a separate turbulent region where there are rotational movements of one type of helicity, right-handed or left-handed. Such a model is probably the limit of a helical pattern. According to Ref. 20, the following estimates hold: $D_0\approx u_0^2\tau_0/3$ and $\alpha_0\approx u_0\xi_0/3$ for $\xi_0\ll1$, and $D_0\approx u_0/p_0$ and $\alpha_0\approx u_0$ for $\xi_0\gg1$. The case $\gamma\ll1$ corresponds to $\xi_0^2\ll1$, while the case $\gamma\gg1$ occurs when $\xi_0\gg1$. As a result we obtain

$$\left.\frac{A^{(1)}}{A^{(0)}}\right|_{\gamma\ll1}\approx\frac{\xi_0^4}{45}\ll1,\quad\left.\frac{A^{(1)}}{A^{(0)}}\right|_{\gamma\gg1}\approx0.002, \qquad (16)$$

i.e., the series of iterations in $G_{ij}^{(0)}$ converges well.

We will now do all calculations in the zeroth approximation, where the exact expression for $G_{ij}(1;2)$ is replaced by $G_{ij}^{(0)}(1;2)$. In this approximation, the formulas contain no reducible terms, and so the phrase ''irreducible part of'' can be discarded.

To make our reasoning more graphic, we write the explicit expression for the components of the tensor $\alpha_{ij}(\mathbf{r},t)$:

$$\alpha_{ij}(1)=\int d2\begin{pmatrix} G^{(0)}_{xy}T_z-G^{(0)}_{xz}T_y & G^{(0)}_{xz}T_x-G^{(0)}_{xx}T_z & G^{(0)}_{xx}T_y-G^{(0)}_{xy}T_x \\ G^{(0)}_{yy}T_z-G^{(0)}_{yz}T_y & G^{(0)}_{yz}T_x-G^{(0)}_{yx}T_z & G^{(0)}_{yx}T_y-G^{(0)}_{yy}T_x \\ G^{(0)}_{zy}T_z-G^{(0)}_{zz}T_y & G^{(0)}_{zz}T_x-G^{(0)}_{zx}T_z & G^{(0)}_{zx}T_y-G^{(0)}_{zy}T_x \end{pmatrix}. \tag{17}$$

Here, for brevity, we have used the notation $T_i=\langle\langle\alpha'(1)\nabla_i^{(2)}\alpha'(2)\rangle\rangle$, and $i,j\to x,y,z$.

To simplify matters, in applications one commonly uses the average value of the diagonal elements of the tensor $\alpha_{ij}(\mathbf{r},t)$, which has the meaning of the amplification coefficient of the magnetic field:

$$\alpha(\mathbf{r},t)=\frac{1}{3}\alpha_{ii}(\mathbf{r},t)=e_{inm}\int d2\, G^{(0)}_{in}(1;2)\times\langle\langle\alpha'(1)\nabla_m^{(2)}\alpha'(2)\rangle\rangle. \tag{18}$$

We see that when there is amplification of the field, the Green's tensor must have an antisymmetric part. Similarly, the quantity $D=(1/6)e_{ijk}D_{ijk}$ has the meaning of the average diffusion coefficient. It should be recalled that the total coefficients of diffusion and the α effect are D_m+D_0+D and $\alpha_0+\alpha$, respectively.

If the correlations of the helicity fluctuations are short-lived, $\langle\langle\alpha'(\mathbf{r},t)\alpha'(\mathbf{r}',t')\rangle\rangle\propto\delta(t-t')$, then, allowing for the property $G_{nm}(\mathbf{r},t;\mathbf{r}',t)=\delta_{nm}\delta(\mathbf{r}-\mathbf{r}')$, we find that $\alpha_{ij}=-e_{ijk}\nabla_k\langle\langle\alpha'^2(\mathbf{r},t)\rangle\rangle$ and $\alpha\equiv 0$, i.e., there is no field amplification. This is in full agreement with the conclusion drawn by Sokolov[16] that it is imperative to study the effect of helicity fluctuations, assuming that the time these fluctuations remain correlated is finite.

The isotropic turbulence case

Below we will write the expressions for α and D in the frequently used model of an isotropic turbulent medium with helicity. In this case

$$G^{(0)}_{ij}(\mathbf{R},\tau)=H(\tau)[\,\delta_{ij}g_0(R,\tau)+e_{ijk}\nabla_k g_1(R,\tau)], \tag{19}$$

where $\mathbf{R}=\mathbf{r}-\mathbf{r}'$, $\tau=t-t'$, $H(\tau)=1$ for $\tau>0$, and $H(\tau)=0$ for $\tau<0$. The Fourier transforms of the functions g_0 and g_1 have the following form:

$$\tilde g_0(p,\tau)\equiv\int d\mathbf{R}\, g_0(R,\tau)\exp(-i\mathbf{p}\cdot\mathbf{r})=\cosh(\alpha_0 p\tau)\exp(-D_0p^2\tau), \tag{20}$$

$$\tilde g_1(p,\tau)=-\frac{1}{p}\sinh(\alpha_0 p\tau)\exp(-D_0p^2\tau).$$

In the absence of helicity, $\alpha_0=0$ and $g_1=0$. At this point it is convenient to introduce the fluctuation spectra

$$\langle\langle\alpha'(\mathbf{r},t)\alpha'(\mathbf{r},t+\tau)\rangle\rangle\equiv\int_0^\infty dp\, E_\alpha(p,\tau), \tag{21}$$

$$\langle\langle\alpha'^2(\mathbf{r},t)\rangle\rangle\equiv\alpha_1^2,$$

$$\langle\langle D'(\mathbf{r},t)D'(\mathbf{r},t+\tau)\rangle\rangle\equiv\int_0^\infty dp\, E_D(p,\tau),$$

$$\langle\langle D'^2(\mathbf{r},t)\rangle\rangle\equiv D_1^2.$$

In this notation, the formulas for α and D for the case of an isotropic turbulent medium assume the form

$$\alpha=-\frac{2}{3}\int_0^\infty dp\int_0^t d\tau\, p^2E_\alpha(p,\tau)\tilde g_1(p,\tau), \tag{22}$$

$$D=-\frac{1}{3}\int_0^\infty dp\int_0^t d\tau\left[2p^2E_D(p,\tau)\tilde g_0(p,\tau)+E_\alpha(p,\tau)\tilde g_0(p,\tau)+pE_\alpha(p,\tau)\frac{\partial}{\partial p}\tilde g_0(p,\tau)\right]. \tag{23}$$

Formula (22) shows that helicity fluctuations in an isotropic medium results in amplification of the already existing α effect. To estimate the contribution we take $E_\alpha(p,\tau)=\alpha_1^2\times\delta(p-p_\alpha)\exp(-\tau/\tau_\alpha)$:

$$\alpha=\frac{2}{3}\alpha_0\frac{\alpha_1^2p_\alpha^2\tau_\alpha^2}{(D_0p_\alpha^2\tau_\alpha+1)^2-\alpha_0^2p_\alpha^2\tau_\alpha^2}. \tag{24}$$

The conditions for the diffusion approximation require that the denominator in (24) be positive, i.e., the sign of α must coincide with the sign of the coefficient α_0.

As for the additional contribution of (23) to the turbulent diffusion coefficient, we see that the on the whole the fluctuations $D'(\mathbf{r},t)$ (the first term) and $\alpha'(\mathbf{r},t)$ (the second and third terms) provide a negative contribution, i.e., reduce the initial coefficient D_m+D_0. The fact that the fluctuations $D'(\mathbf{r},t)$ reduce the diffusion coefficient becomes understandable if one recalls that in finding the mean free path of a particle with two scattering mechanisms one should take the sum of the reciprocals: $1/l_{\text{mean}}=1/l_1+1/l_2$. Actually, the turbulent diffusion coefficient is proportional to a certain mixing length l. Hence for an ensemble of two processes with $D=D_0+\Delta D$ and $D=D_0-\Delta D$ the mean value is $D_{\text{mean}}=D_0[1-(\Delta D/D_0)^2]<D_0$. Note that Moffatt,[3] when considering the effect of $\alpha'(\mathbf{r},t)$ on turbulent diffusion, obtained only the second term in (23). This is the result of using a less accurate method in studying the problem. Actually the new term in (23), the third term, provides a positive contribution, whose value may reach two-thirds of the second, negative, term.

Thus, in a uniform and isotropic turbulent medium, helicity fluctuations only enhance the already existing α effect. The situation is different in nonuniform and anisotropic turbulent media. In such media, helicity fluctuations may cause large-scale amplification of the magnetic field even at zero average helicity. This becomes especially evident if we turn to media with differential rotation.

3. THE α EFFECT IN A NONUNIFORMLY ROTATING MEDIUM

The most important examples of such media are the convective shell of the sun and the motion of magma inside the Earth. The accretion disks around stars and interstellar gas also participate in differential rotation.

3.1. Integral equation for the Green's tensor

It is convenient to write the equation for the Green's tensor $G_{ij}^{(0)}(\mathbf{r},t;\mathbf{r}',t')$ satisfying Eq. (4) without the stochastic terms on the right-hand side in the form of an integral equation. For the case considered, $\alpha_0=0$ and $\mathbf{U}(\mathbf{r},t)=\omega(\rho,z)(\mathbf{e}_z\mathbf{r})\equiv\mathbf{e}_\varphi\rho\omega(\rho,z)$, this equation in a cylindrical system of coordinates $(\alpha,\beta=\rho,\varphi,z)$ has the form

$$G_{\alpha\beta}^{(0)}(1;2)=g_{\alpha\beta}^{(0)}(1-2)-\int d3\,\omega(3)g_{\alpha\gamma}^{(0)}(1-3)\frac{\partial}{\partial\varphi_3}G_{\gamma\beta}^{(0)}(3;2)+\int d3\,\rho_3 g_{\alpha\varphi}^{(0)}(1-3)[(\nabla_\rho^{(3)}\omega(3))G_{\rho\beta}^{(0)}(3;2)+(\nabla_z^{(3)}\omega(3))G_{z\beta}^{(0)}(3;2)], \tag{25}$$

where ρ is the distance from the observation point $\mathbf{r}(\rho,\varphi,z)$ to the rotation axis z, the index γ runs through all values (ρ, φ, and z), and the other indices are fixed. The Green's tensor $g_{\alpha\beta}^{(0)}(\mathbf{R},\tau)$ is the ordinary diffusion Green's function $g_{ij}(1-2)=\delta_{ij}g_0(R,\tau)$ $(i,j=x,y,z)$ written in a cylindrical system of coordinates, with

$$g_0(R,\tau)=\frac{1}{(4\pi D_0\tau)^{3/2}}\exp\left\{-\frac{R^2}{4D_0\tau}\right\}. \tag{26}$$

It should be recalled that the transition from Cartesian coordinates to cylindrical is done by using a unitary matrix $U_{\beta k}(\varphi)$ according to the following relationships:

$$A_\beta=U_{\beta k}(\varphi)A_k,$$
$$G_{\alpha\beta}(1;2)=U_{\alpha i}(\varphi_1)G_{ij}(1;2)\tilde{U}_{j\beta}(\varphi_2), \tag{27}$$
$$\alpha_{\gamma\beta}(\mathbf{r},t)=U_{\gamma i}(\varphi)\alpha_{ij}(\mathbf{r},t)\tilde{U}_{j\beta}(\varphi).$$

The tilde indicates the transpose of a matrix: $\tilde{U}_{j\beta}=U_{\beta j}$, and summation is implied over repeated indices. The components of the matrix $U_{\alpha i}(\varphi)$ are

$$U_{\beta\kappa}(\varphi)=\begin{pmatrix}\rho\\ \varphi\\ z\end{pmatrix}\begin{pmatrix}\cos\varphi & \sin\varphi & 0\\ -\sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{pmatrix}. \tag{28}$$

According to (27), the Green's tensor has the form $g_{\alpha\beta}^{(0)}(1-2)=U_{\alpha\beta}(\varphi_1-\varphi_2)g_0(R,\tau)$. One can easily verify that the tensor $G_{ij}^{(0)}(\mathbf{r},t;\mathbf{r}',t')$ depends on the difference of azimuthal angles, $\psi\equiv\varphi_1-\varphi_2$.

3.2. Green's tensor for axisymmetric problems

When Eq. (4) is used to find the magnetic field, it is often assumed that the distribution of field sources or the initial field is axisymmetric. In this case it is enough to know the Green's tensor integrated with respect to the difference of the azimuthal angles, $\psi\equiv\varphi_1-\varphi_2$. We will denote this tensor by $\langle G_{\alpha\beta}^{(0)}\rangle\equiv\langle G_{\alpha\beta}^{(0)}(\rho_1,z_1;\rho_2,z_2;\tau)\rangle$. Integrating Eq. (25) with respect to the angle ψ, we arrive at exact expressions for the components $\langle G_{\alpha\beta}^{(0)}\rangle$:

$$\langle G_{\rho\rho}^{(0)}\rangle=\langle G_{\varphi\varphi}^{(0)}\rangle\equiv G_\perp=\int_0^{2\pi}d\psi\cos\psi\, g_0(R,\tau),$$

$$\langle G_{zz}^{(0)}\rangle\equiv G_\parallel=\int_0^{2\pi}d\psi\, g_0(R,\tau),$$

$$\langle G_{\varphi\rho}^{(0)}\rangle\equiv G_{\varphi\rho}=\int_{-\infty}^{\infty}dz_3\int_0^\infty d\rho_3\,\rho_3^2\times\int_0^\tau d\tau'\,G_\perp(\rho_1,z_1;\rho_3,z_3;\tau-\tau')\times(\nabla_{\rho_3}\omega(\rho_3,z_3))G_\perp(\rho_3,z_3;\rho_2,z_2;\tau'), \tag{29}$$

$$\langle G_{\varphi z}^{(0)}\rangle\equiv G_{\varphi z}=\int_{-\infty}^{\infty}dz_3\int_0^\infty d\rho_3\,\rho_3^2\times\int_0^\tau d\tau'\,G_\perp(\rho_1,z_1;\rho_3,z_3)\tau-\tau')\times(\nabla_{z_3}\omega(\rho_3,z_3))G_\parallel(\rho_3,z_3;\rho_2,z_2;\tau').$$

The other components $\langle G_{\alpha\beta}^{(0)}\rangle$ are equal to zero. Note that in calculating $G_\perp$ and $G_\parallel$, integrals with respect to ψ lead to Bessel functions of imaginary argument, i.e., these components are of an explicit analytic form.

When there is axial symmetry, the scalar fluctuation correlator $A(1;2)\equiv\langle\langle\alpha'(1)\alpha'(2)\rangle\rangle$, obviously, also undergoes no changes under rotations of the system of points 1 and 2 as a whole, i.e., it can be written in the form $A(1;2)=A(z_1,z_2,\rho_1,|\mathbf{r}_\perp^{(1)}-\mathbf{r}_\perp^{(2)}|,\tau)$, where the vector $\mathbf{r}_\perp(\rho,\varphi)$ determines the point of a point in the xy plane. The Green's tensor $\langle G_{\alpha\beta}^{(0)}\rangle$ integrated with respect to the difference ψ of azimuthal angles is the part of $G_{\alpha\beta}^{(0)}$ that is independent of ψ. What is the contribution of this part to the tensor $\alpha_{\gamma\beta}$? If we transform $\langle G_{\alpha\beta}^{(0)}\rangle$ into $\langle G_{ij}^{(0)}\rangle$ by (27) and then calculate first the tensor $\alpha_{ij}(1)$ [see (17)] and then $\alpha_{\gamma\beta}(1)$, we obtain

$$\alpha_{\rho\rho}(\mathbf{r},t)=\alpha_{\rho z}(\mathbf{r},t)=\alpha_{z\rho}(\mathbf{r},t)=\alpha_{zz}(\mathbf{r})=0, \tag{30}$$

$$\alpha_{\rho\varphi}(\rho,z,t)=-\alpha_{\varphi\rho}(\rho,z,t)=-\int d\mathbf{r}'\int_0^t d\tau\times\cos\psi\,G_\perp(z-z',\rho,\rho';\tau)\times\nabla_z'A(z,z',\rho,|\mathbf{r}_\perp-\mathbf{r}_\perp'|,\tau),$$

$$\alpha_{\varphi z}(\rho,z,t)=-\int d\mathbf{r}'\int_0^t d\tau\,G_\perp(z-z',\rho,\rho';\tau)\times\nabla_\rho A(z,z',\rho,|\mathbf{r}_\perp-\mathbf{r}_\perp'|,\tau),$$

$$\alpha_{z\varphi}(\rho,z,t)=\int d\mathbf{r}'\int_0^t d\tau\, G_{\parallel}(z-z',\rho,\rho';\tau)[\cos\psi\,\nabla'_{\rho}A + \sin\psi\,\nabla'_{\varphi}A],$$

$$\alpha_{\varphi\varphi}(\rho,z,t)=\int d\mathbf{r}'\int_0^t d\tau \times\{-\cos\psi\,G_{\varphi\rho}(z,\rho;z',\rho';\tau)\nabla'_z A + \cos\psi\,G_{\varphi z}(z,\rho;z',\rho';\tau)\nabla'_{\rho}A + \sin\psi\,G_{\varphi z}(z,\rho;z',\rho';\tau)\nabla'_{\varphi}A\}.$$

For brevity, in the last two formulas we have omitted the arguments of $A(z,z',\rho,|\mathbf{r}_\perp-\mathbf{r}'_\perp|,\tau)$.

Clearly, when the helicity fluctuation distribution is uniform along the rotation axis ($A(z,z',\rho,|\mathbf{r}_\perp-\mathbf{r}'_\perp|,\tau)=A(|z-z'|,\rho,|\mathbf{r}_\perp-\mathbf{r}'_\perp|,\tau)$), the components $\alpha_{\rho\varphi}=\alpha_{\varphi\rho}$ vanish because they are odd in $z-z'$. In the absence of differential rotation in the z-coordinate ($\omega(\rho,z)=\omega(\rho)$), the component $G_{\varphi z}$ is zero and $G_{\varphi\rho}$ depends on $|z-z'|$, which means that when the distribution of the fluctuations is uniform, the principal component $\alpha_{\varphi\varphi}$, which determines the amplification of the magnetic field, is zero.

Thus, for an α effect to exist ($\alpha_{\varphi\varphi}\neq 0$) to exist, there must be a dependence of the angular velocity ω on the coordinate z or a nonuniform distribution of helicity fluctuations along this coordinate or the two factors must act simultaneously. Here the α effect is highly anisotropic—only the azimuthal component of the magnetic field induces an azimuthal current, which results in amplification of the poloidal component of the average field.

3.3. Qualitative explanation of the α effect

Let us qualitatively examine the generation of the emf $\mathbf{E}_\varphi$ parallel to the average field $\langle\langle\mathbf{B}_\varphi\rangle\rangle$. We will begin with the contribution of the first term in $\alpha_{\varphi\varphi}$ in (30). Suppose that the initial uniform magnetic field is directed parallel to the azimuthal basis vector $\mathbf{e}_\varphi(1)$ at the observation point 1 (see Fig. 1). At points $2'$ and $2''$, which are located at equal distances for the xy plane, this field induces the emf $\mathbf{E}(2')=\alpha'(2')B_0\mathbf{e}_\varphi(1)$ and $\mathbf{E}(2'')=\alpha'(2'')B_0\mathbf{e}_\varphi(1)$. The currents generated by these emf's induce at point 2 in the xy plane two magnetic fields, $\mathbf{B}'$ and $\mathbf{B}''$, which are perpendicular to the initial field but point in opposite directions. The net field $\mathbf{B}(2)=\mathbf{B}'+\mathbf{B}''$ is directed along the rotation axis at an angle $\psi=\varphi_1-\varphi_2$ to the radial basis vector $\mathbf{e}_\rho(2)$, i.e., $B_\rho(2)=-\cos\varphi\,|\mathbf{B}'+\mathbf{B}''|$. Due to radial differential rotation (ω depends on ρ), this radial component transforms, with a probability $\propto G_{\varphi\rho}(1;2)$, into the azimuthal component at point 1 and generates the azimuthal emf $\mathbf{E}_\varphi\propto-\cos\psi G_{\varphi\rho}(1;2)\alpha'(1)\nabla'_z\alpha'(2)B_0\mathbf{e}_\varphi(1)$ at this point. The size of the circles in Fig. 1 corresponds to the relative values of the function $\alpha'(\mathbf{r},t)$. This picture shows again that for this part of the $\alpha_{\varphi\varphi}$ effect to exist the distribution of fluctuations must vary along the z-coordinate or the component $G_{\varphi\rho}$ of the Green's tensor must be nonuniform.

In examining the contributions of the second and third terms it is convenient to place the observation point 1 on the x axis, i.e., $\varphi=0$. Then the sum of these terms can be represented as $G_\varphi(1;2)\nabla'_x A$. As in the case with the first term (Fig. 1), we select the points $2'$ and $2''$ at equal distances from the circle $\rho=\rho_1$ (see Fig. 2). The average magnetic field $B_0\mathbf{e}_\varphi(1)$ in this case is parallel to the axis. The emf's $\mathbf{E}(2')$ and $\mathbf{E}(2'')$ induced by the $\alpha'(\mathbf{r},t)$ field are also parallel to the y axis. The electric currents generated by these emf's induce at the point 2 a finite magnetic field $\mathbf{B}(2)$ directed along the z axis. Due to differential rotation (ω depends on z), this field acquires, with a probability $\propto G_{\varphi z}(1;2)$, an azimuthal component at the observation point 1, where it generates an emf $\mathbf{E}_\varphi\propto G_{\varphi z}(1;2)\nabla'_x A$. For simplicity, in Fig. 2 we depicted the contribution from points $2'$ and $2''$ in the xy plane, but the same is true for other planes parallel to the xy plane.

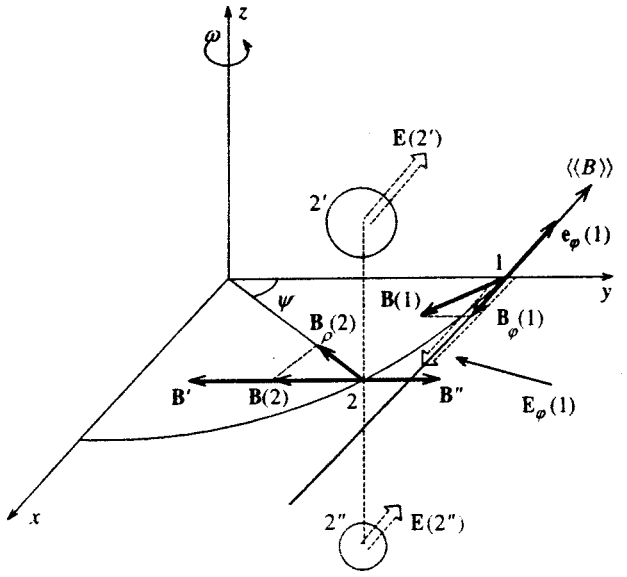

FIG. 1. The mechanism for the occurrence of a transverse α effect (the $\alpha_{\varphi\varphi}$-component) due to the nonuniform distribution of helicity fluctuations along the rotation axis z and the differential rotation in the perpendicular plane with $\omega=\omega(\rho)$.

3.4. An approximate expression for the Green's tensor in the general case

The left-hand side of Eq. (4) (at $\alpha_0=0$) describes the evolution of the average magnetic field due to diffusion ($D_m+D_0\neq 0$) and due to transport of ''frozen-in'' field lines by regular motion. In our integral equation (25) the diffusion process is described by the tensor $g^{(0)}_{\alpha\beta}(1;2)=U_{\alpha\beta}(\varphi_1-\varphi_2)g_0(R,\tau)$, which corresponds to diffusive transport of the field lines in an isotropic medium, i.e., the initial direction of these lines is conserved. It is for this reason that $\gamma^{(0)}_{\alpha\beta}$ contains finite terms $g^{(0)}_{\rho\varphi}=\sin(\varphi_1-\varphi_2)\,g_0=-g^{(0)}_{\varphi\rho}$, according to which, say, the field $\mathbf{B}_\rho$, which was initially radial, acquires a φ-component as a result of diffusive transport to another point, and vice versa. Substitution of $g^{(0)}_{ij}(R,\tau)=\delta_{ij}g_0(R,\tau)$ corresponding to $g^{(0)}_{\alpha\beta}$ into (17) leads to a purely antisymmetric tensor α_{ij}, i.e., the α effect proper (the emf is directed parallel to the average magnetic field) does appear here. Differential rotation with $\omega=\omega(\rho)$ directly creates (due to the frozen-in field lines) an azimuthal component

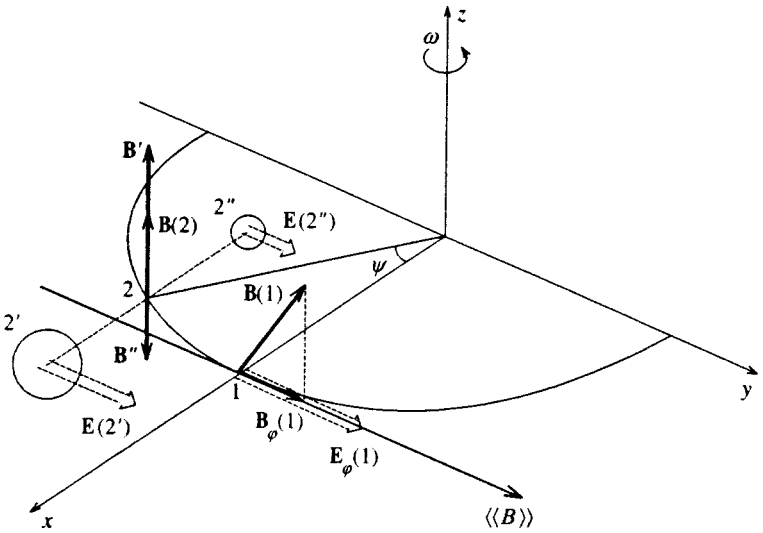

FIG. 2. The mechanism for the occurrence of a transverse α effect (the α_φ-component) due to the nonuniform distribution of helicity fluctuations perpendicular to the rotation axis z and the longitudinal differential rotation with $\omega=\omega(z)$.

field from the radial component and yields an additional contribution to the $G^{(0)}_{\varphi\rho}$-component of the Green's tensor. However, we have seen that diffusion partially transforms this azimuthal component into the radial component, giving rise to an additional contribution to $G^{(0)}_{\rho\rho}$. The inverse sequence, first diffusion and then differential rotation with $\omega=\omega(\rho)$, gives an additional contribution to $G^{(0)}_{\rho\varphi}$. When $\omega=\omega(\rho)$ holds, the z-component of the field cannot transform into the radial or azimuthal component, so that in this case $G^{(0)}_{\varphi z}=G^{(0)}_{\rho z}=0$. In the general case of rotation with $\omega=\omega(\rho,z)$, neither diffusion nor differential rotation can transform the horizontal (in the rotation plane) component of the field into the vertical component. Hence $G^{(0)}_{z\varphi}=G^{(0)}_{z\rho}=0$ always holds. Differential rotation with $\omega=\omega(z)$ generates from the z-component the azimuthal component ($G^{(0)}_{\varphi z}\neq 0$), which through diffusion creates the radial component, leading to $G^{(0)}_{\rho z}\neq 0$.

All these qualitative conclusions are corroborated by the master integral equation (25).

In the axisymmetric case, the second term on the right-hand side of Eq. (25) contributed nothing, which enabled us to find an exact analytic expression for the average Green's function $\langle G^{(0)}_{\alpha\beta}\rangle$. According to (10), to calculate the tensor $\alpha_{ij}(\mathbf{r},t)$ we must know the Green's tensor in general form and not only its axisymmetric part. For this general case we express the Green's tensor $G^{(0)}_{\alpha\beta}(1;2)$ in the form of a sum of the absolute term of Eq. (25), $g^{(0)}_{\alpha\beta}(1-2)$, and the first iteration of the equation. The structure of this expression is of the most general form, i.e., only the components $G^{(0)}_{z\varphi}=G^{(0)}_{z\rho}$ are equal to zero. This expression is probably quite accurate, since it allows for the first orders of the diffusion and differential-rotation processes, while higher-order processes provide small contributions (this follows from general physical considerations). It should alsobe recalled that in magnetic dynamo theory the system of coordinates rotates together with the medium with a constant angular velocity ω_0, so that $\omega(\rho,z)$ stands for the remaining part of the angular velocity, which often (say, for the sun) is much smaller than ω_0. Usually the gradients of the angular velocity, $\nabla_\rho\omega$ and $\nabla_z\omega$, are also smooth functions on the characteristic scales of the problem. All this suggests that the proposed expression for $G^{(0)}_{\alpha\beta}(1;2)$ can be used in practical calculations and estimates. Thus, plugging $g^{(0)}_{\alpha\beta}(1-2)$ into the integral terms of Eq. (18), we arrive at the following expression for $G^{(0)}_{\alpha\beta}(1;2)$:

$$G_{\alpha\beta}(1;2)=U_{\alpha\beta}(\psi)g_0(1-2)-\frac{\partial}{\partial\psi}[U_{\alpha\beta}(\psi)f]+\begin{pmatrix}\rho\\ \varphi\\ z\end{pmatrix}\times\begin{pmatrix}a\sin\psi-c\cos\psi & c\sin\psi-b\cos\psi & h\sin\psi-g\cos\psi\\ a\cos\psi+c\sin\psi & c\cos\psi+b\sin\psi & h\cos\psi+g\sin\psi\\ 0 & 0 & 0\end{pmatrix}. \tag{31}$$

Here we have introduced the notation

$$f(1;2)=\int d3\,\omega(3)g_0(1-3)g_0(3-2),$$

$$a(1;2)=\int d3\,\rho_3(\nabla_\rho\omega(3))\cos^2\psi'\,g_0(1-3)g_0(3-2),$$

$$b(1;2)=\int d3\,\rho_3(\nabla_\rho\omega(3))\sin^2\psi'\,g_0(1-3)g_0(3-2),$$

$$c(1;2)=\int d3\,\rho_3(\nabla_\rho\omega(3))\sin\psi'\cos\psi'\,g_0(1-3)g_0(3-2),$$

$$g(1;2)=\int d3\,\rho_3(\nabla_z\omega(3))\sin\psi'\,g_0(1-3)g_0(3-2),$$

$$h(1;2)=\int d3\,\rho_3(\nabla_z\omega(3))\cos\psi'\,g_0(1-3)g_0(3-2), \tag{32}$$

$\psi=\varphi_1-\varphi_2$, and $\psi'=\varphi_3-\varphi_2$. Interestingly, integrating (31) with respect to ψ leads to the axisymmetric Green's tensor (29), i.e., this approximation incorporates the previous, axisymmetric, approximation.

Using (31), we can easily write the components of the tensor $\alpha_{\beta\gamma}(\mathbf{r},t)$ explicitly. Here we will limit ourselves to

the most important, diagonal, components. For convenience we first write the general formulas for these components:

$$\alpha_{\rho\rho}(\mathbf{r},t)=\int d2\,\{-\sin\psi\,G^{(0)}_{\rho\rho}(1;2)\nabla^{(2)}_z A(1;2)+\cos\psi\,G^{(0)}_{\rho\varphi}(1;2)\nabla^{(2)}_z A(1;2)+[\sin\psi\,\nabla^{(2)}_\rho A(1;2)-\cos\psi\,\nabla^{(2)}_\varphi A(1;2)]G^{(0)}_{\rho z}(1;2)\}, \tag{33}$$

$$\alpha_{\varphi\varphi}(\mathbf{r},t)=\int d2\,\{-\cos\psi\,G^{(0)}_{\varphi\rho}(1;2)\nabla^{(2)}_z A(1;2)-\sin\psi\,G^{(0)}_{\varphi\varphi}(1;2)\nabla^{(2)}_z A(1;2)+[\cos\psi\,\nabla^{(2)}_\rho A(1;2)+\sin\psi\,\nabla^{(2)}_\varphi A(1;2)]G^{(0)}_{\varphi z}(1;2)\}, \tag{34}$$

$$\alpha_{zz}(\mathbf{r},t)=\int d2\,[G^{(0)}_{z\rho}(1;2)\nabla^{(2)}_\varphi A(1;2)-G^{(0)}_{z\varphi}(1;2)\nabla^{(2)}_\rho A(1;2)]. \tag{35}$$

First we see that $\alpha_{zz}(\mathbf{r},t)\equiv 0$, since in a medium with differential rotation with $\omega=\omega(\rho,z)$ the components $G^{(0)}_{z\varphi}=G^{(0)}_{z\rho}$ are equal to zero. We also note that the difference from the corresponding formula in the case of using the axisymmetric Green's tensor [formula (29)] is very large. There $\alpha_{\rho\rho}(\mathbf{r},t)=0$, since $\langle G^{(0)}_{\rho\varphi}\rangle=\langle G^{(0)}_{\rho z}\rangle=0$, and the first term in (33) contributed nothing because it is odd in ψ.

Inserting the components of the tensor (31) into (33) and (34) yields

$$\alpha_{\rho\rho}(\rho,z,t)=\int d2\,\{-f\nabla^{(2)}_z A(1;2)-(a\sin^2\psi+b\cos^2\psi)\nabla^{(2)}_z A(1;2)+h\sin\psi[\sin\psi\,\nabla^{(2)}_\rho A(1;2)-\cos\psi\,\nabla^{(2)}_\varphi A(1;2)]\}, \tag{36}$$

$$\alpha_{\varphi\varphi}(\rho,z,t)=\int d2\,\{-f\nabla^{(2)}_z A(1;2)-(a\cos^2\psi+b\sin^2\psi)\nabla^{(2)}_z A(1;2)+h\cos\psi[\cos\psi\,\nabla^{(2)}_\rho A(1;2)+\sin\psi\,\nabla^{(2)}_\varphi A(1;2)]\}, \tag{37}$$

We see that the components $\alpha_{\rho\rho}$ and $\alpha_{\varphi\varphi}$ are of the same order. This means that in calculating the kinetic coefficients $\alpha_{\beta\gamma}(\mathbf{r},t)$ we must use the Green's tensor in general form. The part of this tensor that is averaged over the azimuthal angle strongly distorts the structure of $\alpha_{\beta\gamma}(\mathbf{r},t)$, although the order of $\alpha_{\varphi\varphi}$ remains unchanged. In a medium with rotation, all the components $\alpha_{\beta\gamma}(\mathbf{r},t)\equiv\alpha_{\beta\gamma}(\rho,z,t)$ are independent of the azimuthal angle, which is the consequence of the fact that the fluctuation correlator is uniform in the angles: $A(1;2)=\langle\langle\alpha'(1)\alpha'(2)\rangle\rangle=A(z_1,z_2,\rho_1,\rho_2,\varphi_1-\varphi_2,\tau)$.

A qualitative explanation of the emergence of the α effect due to helicity fluctuations (see Figs. 1 and 2) given above is fully applicable in relation to the second and third terms in (36) and (37). Attention should be focused on the first terms in these formulas, which appear because of the term with the derivative with respect to the azimuthal angle in the master integral equation (25). We will begin this explanation from fundamental considerations.

The exact solution of Eq. (25) in the absence of differential rotation, i.e., in the case of uniform rotation with a constant angular velocity ω_0, has the following form:[5]

$$G^{(0)}_{\alpha\beta}(1;2)=U_{\alpha\beta}(\psi-\omega_0\tau)g_0(|z-z'|,\rho^2+\rho'^2-2\rho\rho'\cos(\psi-\omega_0\tau),\tau). \tag{38}$$

The physical meaning of this solution is very simple: it describes the ratio of frozen-in magnetic field and, simultaneously, the thermal diffusion of this field in accordance with the varying distance from the moving point 2 to the fixed observation point 1. Susbstituting this solution in (33) and (34) yields

$$\alpha_{\rho\rho}=\alpha_{\varphi\varphi}=-\int d2\,\sin(\omega_0\tau)\,\nabla^{(2)}_z A(1;2)g_0. \tag{39}$$

This formula describes the occurrence of the α effect due to the rotation of the induced magnetic field. Indeed, let the average field $\langle\langle\mathbf{B}\rangle\rangle=\text{const}$ be directed along the y axis (see Fig. 3). In accordance with the values of the function $\alpha'(\mathbf{r},t)$, the emf's generated at points $2'$ and $2''$ are $\mathbf{E}(2')=\alpha'(2')\langle\langle\mathbf{B}\rangle\rangle$ and $\mathbf{E}(2'')=\alpha'(2'')\langle\langle\mathbf{B}\rangle\rangle$. The electric currents corresponding to these emf's induce at point 2 the resulting field $\mathbf{B}'+\mathbf{B}''$ directed parallel to the x axis. Due to its frozen-in nature, the field rotates about the rotation axis, so that a component $\mathbf{B}_\parallel=\sin(\omega_0\tau)\,|\mathbf{B}'+\mathbf{B}''|$ parallel to the average field $\langle\langle\mathbf{B}\rangle\rangle$ but pointing in the opposite direction appears at the observation point 1. This is the component that generates an emf at point 1 in the same direction, i.e., an α effect is present. However, when only uniform rotation is considered no α effect appears, since we implicitly assumed that the average field $\langle\langle\mathbf{B}\rangle\rangle$ is fixed in relation to the fixed observation point 1. Actually $\langle\langle\mathbf{B}\rangle\rangle$ also rotates and no relative rotation through the angle $\omega_0\tau$ occurs. This becomes especially clear if we examine the problem in a rotating system of coordinates, where $\omega_0=0$.

In the presence of differential rotation we have already excluded the constant component of the angular velocity by going over to a rotating system of coordinates. With such a system of coordinates we can assume that the average field is fixed in relation to the point of observation and that the mechanism depicted in Fig. 3 does indeed work. The function $f(1;2)$ corresponds to the solution (38) if we assume that locally the angular velocity $\omega(\rho,z)$ is constant and, more than that, $\omega(\rho,z)\tau$ is much smaller than unity (i.e., $f\to\omega\tau g_0$). Substituting the function $f(1;2)$ for $\sin(\omega_0\tau)\,g_0\approx\omega_0\tau g_0$ we obtain the first terms in (36) and (37). Thus, the term with $f(1;2)$ describes the α effect caused by local rotations of the frozen-in induced magnetic fields.

Sometimes in magnetic dynamo problems the average value $\alpha_\perp(\mathbf{r},t)\equiv(\alpha_{\rho\rho}+\alpha_{\varphi\varphi})/2$. Formulas (36) and (37) yield for this quantity a relatively simple expression:

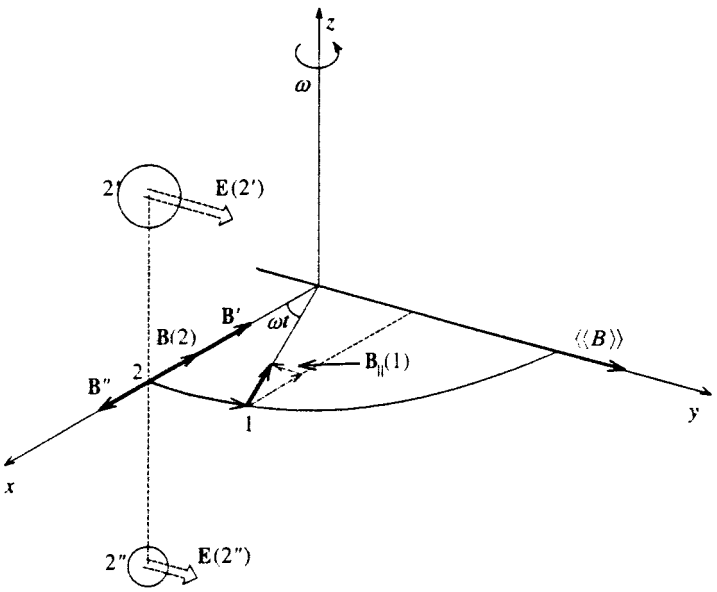

FIG. 3. The mechanism for the occurrence of a transverse α effect due to the nonuniform distribution of helicity fluctuations along the z axis and the rotation of the induced magnetic field.

$$\alpha_\perp(\rho,z,t) = -\int_{-\infty}^{\infty} dz' \int_0^{\infty} d\rho'\, \rho' \int_0^{2\pi} d\psi \left[\left(f(1;2) + \frac{a(1;2)+b(1;2)}{2} \right) \nabla'_z A(1;2) - \frac{h(1;2)}{2} \nabla'_\rho A(1;2) \right]. \quad (40)$$

All the formulas obtained in this paper suggest that a nonuniform distribution of helicity fluctuations causes a large-scale α effect. As for the longitudinal (along the rotation axis) α effect, the general formula (35) requires that finite components $G_{z\varphi}^{(0)}$ and $G_{z\rho}^{(0)}$ are needed for such an effect to exist. The idea that the field lines are frozen-in implies that such an effect is possible if in addition to rotation there are types of motion that eject matter out of the plane of rotation. Primarily this is convective movements of matter accompanied by expansion of the volume occupied by gas.

4. ESTIMATES OF THE α EFFECT AND DISCUSSION

In view of the complexity of the above formulas, estimates of the α effects made by these formulas are qualitative rather than quantitative. First we take the quantities $\omega(\rho,z) \equiv \omega$ and the gradients $\nabla_\rho \omega$ and $\nabla_z \omega$ outside the integral sign, assuming them to be fairly smooth functions. This yields $f(1;2) \approx \omega \tau g_0(1-2)$, $a(1;2) \approx b(1;2) \approx (\nabla_\rho \omega) \tau g_0(1-2) l_{\rm diff}(\tau)$, and $h(1;2) \approx (\nabla_z \omega) \tau g_0(1-2) l_{\rm diff}(\tau)$, where $l_{\rm diff}^2(\tau) \approx 6 D_0 \tau$ determines the diffusion distance in time τ. As noted earlier, the coefficients $\alpha_{\rho\rho}$ and $\alpha_{\varphi\varphi}$ are of the same order, with the result that our estimates refer to the average value of these quantities, $\alpha_\perp$. Using the foregoing estimates of the functions f, a, b, and h and Eq. (40), we arrive at the final estimate:

$$\alpha_\perp \approx \omega \frac{\tau_\alpha^2 \alpha_1^2}{L_z} + \frac{\tau_\alpha^2 \alpha_1^2}{L_\rho} (\nabla_\rho \omega) l_{\rm diff}(\tau_\alpha) + \frac{\tau_\alpha^2 \alpha_1^2}{L_z} (\nabla_z \omega) l_{\rm diff}(\tau_\alpha). \quad (41)$$

Here $\alpha_1^2 \approx \langle\langle \alpha'^2(\mathbf{r},t) \rangle\rangle$ is the level of fluctuations, τ_α is the lifetime of helicity correlations, and L_ρ and L_z are the characteristic lengths of nonuniformity of the fluctuation distribution, which are relatively small, since the nonuniformity of the fluctuations probably manifests itself most strongly near the boundaries of the region (for the sun, say, this is the upper and lower edge of the convective zone).

If we assume that each turbulent vortex transfers, in the process of scale fractionation, its angular moment to smaller vortices, then to make estimates we can assume that $\alpha_1^2 \approx u_0^2$ (which is usually done in estimates), where u_0^2 is the characteristic velocity of turbulent movements. What is important is that helicity fluctuations are inherent in any turbulence and are in no way related to Coriolis forces. As is known, Coriolis forces imply, when there is convection and differential rotation, the existence of an average helicity α_0, which we set to zero in our discussion in ''pure form'' of the novel mechanism for amplification of the average magnetic field by helicity fluctuations. These independent coefficients are relatively weak and can be added, with the quadratic effects being ignored. We also note that u_0^2 can strongly depend on the magnetic field. We must take the turbulent velocities and the distribution of the helicity fluctuations at the values that they had when formed by all the factors: the temperature, pressure, magnetic field, boundary conditions, etc. The characteristic lifetime of a fluctuation correlation is probably longer than the lifetime τ_0 of a turbulent vortex, and so for estimates we assume that $\tau_\alpha \geqslant \tau_0$. The turbulent diffusion coefficient is usually estimated at $D_0 \approx u_0 R_0$, where R_0 is the characteristic size of a turbulent vortex (the

mixing length). Plugging all these quantities into (41), we arrive at a detailed estimate:

$$\alpha_\perp \geqslant \xi_0^2(\omega R_0)\frac{R_0}{L_z} + \xi_0^{5/2}(R_0^2\nabla_\rho\omega)\frac{R_0}{L_\rho} + \xi_0^{5/2}(R_0^2\nabla_z\omega)\frac{R_0}{L_z}, \tag{42}$$

where $\xi_0 = u_0\tau_0/R_0$ is a dimensionless parameter characterizing turbulence (the Strouhal number). It is usually assumed that $\xi_0 \simeq 1$. This parameter can be also be written as the ratio of the vortex lifetime to the time of one rotation, $\xi_0 = \tau_0/t_0$, where $t_0 = R_0/u_0$. If the vortex has time to make many rotations, then $\xi_0 \gg 1$.

In estimating the α effect caused by the Coriolis force it is usually assumed that $\alpha_0 \approx R_0\omega$ (see, e.g., Ref. 21). The estimate (42) shows that the coefficient $\alpha_\perp$, which reflects the nonuniformity in the distribution of the helicity fluctuations, may become equal to α_0 or even larger than α_0. True, in contrast to the mechanism of Coriolis forces, this novel mechanism may be important only near the boundaries of the region considered, where the distribution of fluctuations varies most.

In all the cases considered in this paper (see Figs. 1–3) we have seen that for an emf to be generated parallel to the average magnetic field there must be not only a nonuniform distribution of the helicity fluctuations but also a rotation of the magnetic field induced by the currents, i.e., differential rotation of the plasma is required. Generally speaking, as the general formula (18) implies, for the α effect to manifest itself there must be one vector, the gradient of the correlator of the helicity fluctuations, and one pseudoscalar, $G_i = e_{ijk}G_{jk}^{(0)}$. The pseudovector is not necessarily the angular velocity ω. In problems with complicated geometry, the quantity **G** may be related, say, to reflection asymmetry of the volume considered. Note that in experiments of Shteenbek *et al.*[7] the liquid-sodium conductors were reflection-symmetric.

5. CONCLUSION

Here are the main conclusions that can be drawn from the results of this investigation. The method of large-scale averaging of the equation of diffusion of a magnetic field was used to develop the novel effect first detected by Vishniac and Brandenburg[15] and Sokolov,[16] the large-scale amplification of the magnetic field by turbulent helicity fluctuations. It was found that for turbulence with a zero average helicity the necessary condition for such an effect is the presence of a nonuniform distribution of helicity fluctuations in the conducting turbulent medium. For isotropic, uniform, and reflection-asymmetric turbulence, helicity fluctuations only amplify an already existing α effect. In a medium with differential rotation, a nonuniform distribution of the turbulent helicity fluctuations produces large-scale amplification of the average magnetic field even at zero average helicity. For this case the α effect is highly anisotropic: the average emf not directed parallel to the longitudinal component (along the rotation axis) of the magnetic field, i.e., $\alpha_{zz} \equiv \alpha_\parallel = 0$. All transverse components of the tensor $\alpha_{ij}(\mathbf{r},t)$ are of the same order and can become equal to the coefficients α_0 caused by Coriolis forces. It is also shown that the fluctuations of the diffusion coefficient and helicity reduce the initial turbulent diffusion coefficient.

*)E-mail: silant@inaoep.mx

[1] E. N. Parker, Astrophys. J. **122**, 293 (1955).
[2] M. Steenbeck, F. Krause, and K.-H. Radler, Z. Naturforsch. Teil A **21**, 369 (1966).
[3] H. K. Moffatt, *Magnetic Field Generation in Electrically Conducting Fluids*, Cambridge Univ. Press, Cambridge (1978).
[4] E. N. Parker, *Cosmical Magnetic Fields*, Clarendon Press, Oxford (1980).
[5] F. Krause and K.-H. Radler, *Mean-Field Magnetohydrodynamics and Dynamo Theory*, Pergamon Press, Oxford (1980).
[6] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).
[7] M. Shteenbek, I. M. Kirko, A. Gaĭlitis, A. P. Klyavinya, F. Krause, I. Ya. Laumanis, and O. A. Lielausis, Dokl. Akad. Nauk SSSR **180**, 326 (1968) [Sov. Phys. Dokl. **13**, 43 (1968)].
[8] L. L. Kichatinov, V. V. Pipin, and G. Rudiger, Astron. Nachr. **315**, 157 (1994).
[9] G. Rudiger and L. L. Kichatinov, Astron. Astrophys. **269**, 581 (1993).
[10] A. Z. Dolginov and N. A. Silant'ev, Geophys. Astrophys. Fluid Dyn. **63**, 139 (1992).
[11] N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **101**, 1216 (1992) [Sov. Phys. JETP **74**, 650 (1992)].
[12] H. K. Moffatt, J. Fluid Mech. **65**, 1 (1974).
[13] R. H. Kraichnan, J. Fluid Mech. **75**, 657 (1976).
[14] Yu. L. Klimontovich, *Turbulent Motion and the Structure of Chaos: A New Approach to the Statistical Theory of Open Systems*, Kluwer Academic, Dordrecht (1991).
[15] E. T. Vishniac and A. Brandenburg, Astrophys. J. **475**, 263 (1997).
[16] D. D. Sokolov, Astron. Zh. **74**, 75 (1997) [Astron. Rep. **41**, 68 (1997)].
[17] A. P. Kazantsev, Zh. Éksp. Teor. Fiz. **53**, 1806 (1967) [Sov. Phys. JETP **26**, 1031 (1968)].
[18] V. G. Novikov, A. A. Ruzmaĭkin, and D. D. Sokolov, Zh. Éksp. Teor. Fiz. **85**, 909 (1983) [Sov. Phys. JETP **58**, 527 (1983)].
[19] A. P. Kazantsev, A. A. Ruzmaĭkin, and D. D. Sokolov, Zh. Éksp. Teor. Fiz. **88**, 487 (1985) [Sov. Phys. JETP **61**, 285 (1985)].
[20] N. A. Silant'ev, Zh. Éksp. Teor. Fiz. **112**, 1312 (1997) [JETP **85**, 712 (1997)].
[21] N. O. Weiss, in *Lectures on Solar and Planetary Dynamo*, M. R. E. Proctor and A. D. Gilbert (Eds.), Cambridge Univ. Press, Cambridge (1994).

Translated by Eugene Yankovsky

The quantum poisson–Lie *T*-duality and mirror symmetry

S. E. Parkhomenko*)

Landau Institute for Theoretical Physics, 142432 Chernogolovka, Moscow Region, Russia
(Submitted 17 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 11–25 (July 1999)

Poisson–Lie *T*-duality in quantum $N=2$ superconformal Wess–Zumino–Novikov–Witten models is considered. The Poisson–Lie *T*-duality transformation rules of the super-Kac–Moody algebra currents are found from the conjecture that, as in the classical case, the quantum Poisson–Lie *T*-duality transformation is given by an automorphism which interchanges the isotropic subalgebras of the underlying Manin triple in one of the chirality sectors of the model. It is shown that quantum Poisson–Lie *T*-duality acts on the $N=2$ super-Virasoro algebra generators of the quantum models as a mirror symmetry acts: in one of the chirality sectors it is a trivial transformation while in another chirality sector it changes the sign of the $U(1)$ current and interchanges the spin-3/2 currents. A generalization of Poisson–Lie *T*-duality for the quantum Kazama–Suzuki models is proposed. It is shown that quantum Poisson–Lie *T*-duality acts in these models as a mirror symmetry also. © *1999 American Institute of Physics.* [S1063-7761(99)00207-3]

1. INTRODUCTION

Target-space (*T*) dualities in superstring theory relate backgrounds with different geometries and are symmetries of the underlying conformal field theory.[1,2]

The mirror symmetry[3] discovered in superstring theory is a special type of *T*-duality. At the level of conformal field theory it can be formulated as an isomorphism between two theories, amounting to a change of sign of the $U(1)$ generator and an interchange of the spin-3/2 generators of the left-moving (or rightmoving) $N=2$ superconformal algebra.

Mirror symmetry has mostly been studied in the context of Calabi–Yau superstring compactification. Important progress has been achieved in this direction in the last few years, based on the ideas of toric geometry.[4] In particular, in Ref. 5 toric geometry mirror pair construction was proposed. Though it seems quite certain that pairs of Calabi–Yau manifolds constructed by these methods are mirror, one needs to show that the proposed pairs correspond to isomorphic conformal field theories, to prove that they are indeed mirror. Progress in this direction was made in Ref. 6, but a complete arguments has yet to be carried out. In fact, the only rigorously established example of mirror symmetry, the Greene–Plesser construction,[7] is based on the tensor products of the $N=2$ minimal models.[8] For a review of mirror symmetry and toric geometry methods in Calabi–Yau superstring compactifications see the lectures of Greene.[9]

Recently, Strominger, Yau, Zaslow[10] related mirror symmetry in superstring theory to the quantum Abelian *T*-duality in fibers of toricaly fibrated Calabi–Yau manifolds.

The Poisson–Lie (PL) *T*-duality, recently discovered by Klimcik and Severa in their excellent work,[11] is a generalization of Abelian and non-Abelian *T*-dualities.[12–14] This generalized duality is associated with two groups forming a Drinfeld double,[15] and the duality transformation exchanges their roles. Many aspects of these ideas have been developed in Refs. 16–26. In particular, in Ref. 26 it was shown that PL *T*-duality in the classical $N=2$ superconformal WZNW (SWZNW) and Kazama–Suzuki models is a mirror duality. It is reasonable to expect that PL *T*-duality in the quantum versions of these models will be a mirror duality also. Moreover, it is tempting to conjecture that PL *T*-duality is an adequate geometric structure underlying mirror symmetry in superstring theory. Motivated by this we propose a quantization of PL *T*-duality transformations in the $N=2$ SWZNW and Kazama–Suzuki models.

Quantum equivalence among PL *T*-duality related σ-models was studied perturbatively in Ref. 27 and Ref. 22, and it was shown that PL dualizability is compatible with renormalization at 1 loop. In particular it was shown in Ref. 22 that 1-loop beta functions for the coupling and the parameters in the two simplest examples of PL *T*-duality related models are equivalent. This allows us to suggest that their equivalence extends beyond the classical level with appropriate quantum modification of PL *T*-duality transformations rules.

In the present note the PL *T*-duality transformation rules of the fields in quantum $N=2$ SWZNW models will be found starting from the conjecture that as in the classical case, quantum $N=2$ SWZNW models are PL self-dual and the PL *T*-duality transformation is given by an automorphism of the super-Kac–Moody algebra in the rightmoving sector. Then we obtain PL *T*-duality transformation rules using the Knizhnik–Zamolodchikov equation, Ward identities and a quantum version of the classical formula which relates the generators of rightmoving super-Kac–Moody algebra to its PL *T*-duality transformed. We show that the generators of the $N=2$ super-Virasoro algebras transform under PL *T*-duality like a mirror duality: the $U(1)$ current changes sign and the spin-3/2 currents permute. Thus, the results are in agreement with the conjecture proposed in Ref. 28 that mirror symmetry can be related to a gauge symmetry (auto-

morphism) of the self-dual points of the moduli space of the $N=2$ superconformal field theories (SCFTs) (for the $N=0$ version of this conjecture see Ref. 29). Then we consider quantum PL T-duality in the Kazama–Suzuki models and propose a natural generalization of the quantum PL T-duality transformation. We show that as in the SWZNW models quantum PL T-duality in the Kazama–Suzuki models is a mirror duality also.

The structure of the paper is as follows: In Sec. 2 we briefly review PL T-duality in the classical $N=2$ SWZNW model following.[26] In Sec. 3 we describe Manin triple construction of the quantum $N=2$ SWZNW models on the compact groups and obtain the PL T-duality transformation rules of the quantum fields. We show that PL T-duality transformation is given by an automorphism of the underlying Manin triple which permutes isotropic subalgebras of the triple. Then we obtain transformation rules of the rightmoving $N=2$ super-Virasoro algebra generators. In Sec. 4 we present the Manin triple construction of the Kazama–Suzuki models. We show that they can be described as (*Manin triple*)/(*Manin subtriple*)-cosets. We define quantum PL T-duality transformation in the Kazama–Suzuki models as the subset of the transformations of the numerator triple which stabilizes the denominator subtriple. Then we easily find transformation rules of the rightmoving $N=2$ super-Virasoro algebra generators of the coset. At the end of the section PL T-duality in the $N=2$ minimal models considered briefly as an example.

2. POISSON–LIE *T*-DUALITY AND MIRROR SYMMETRY IN THE CLASSICAL *N*=2 SUPERCONFORMAL WZNW MODELS

In this section we briefly review PL T-duality in the classical $N=2$ SWZNW models, following.[25,26]

We parameterize the super world-sheet by introducing the light cone coordinates $z_\pm$ and Grassman coordinates $\Theta_\pm$ (we use the $N=1$ superfield formalism). The generators of the supersymmetry and covariant derivatives satisfying the standard relations are given by

$$Q_\mp=\frac{\partial}{\partial\Theta_\pm}+i\Theta_\pm\partial_\mp,\quad D_\mp=\frac{\partial}{\partial\Theta_\pm}-i\Theta_\pm\partial_\mp. \tag{1}$$

The superfield of the $N=2$ SWZNW model

$$G=g+i\Theta_-\psi_+ +i\Theta_+\psi_- +i\Theta_-\Theta_+F \tag{2}$$

takes values in a compact Lie group $\mathbf{G}$ so that its Lie algebra $\mathbf{g}$ is endowed with an ad-invariant nondegenerate inner product $\langle,\rangle$. The action of the model is given by

$$S_{\mathrm{SWZ}}=\int d^2xd^2\Theta(\langle G^{-1}D_+G,G^{-1}D_-G\rangle)-\int d^2xd^2\Theta dt \times\left\langle G^{-1}\frac{\partial G}{\partial t},\{G^{-1}D_-G,G^{-1}D_+G\}\right\rangle \tag{3}$$

and possesses manifest $N=1$ superconformal and super-Kac–Moody symmetries:[30]

$$\begin{aligned}&\delta_{a_+}G(z_+,x_-,\Theta_+,\Theta_-)\\&\quad=a_+(z_-,\Theta_+)G(z_+,z_-,\Theta_+,\Theta_-),\\&\delta_{a_-}G(z_+,z_-,\Theta_+,\Theta_-)\\&\quad=-G(z_+,z_-,\Theta_+,\Theta_-)a_-(z_+,\Theta_-),\end{aligned} \tag{4}$$

$$\begin{aligned}&G^{-1}\delta_{\epsilon_+}G=(G^{-1}\epsilon_+(z_-)Q_+G),\\&\delta_{\epsilon_-}GG^{-1}=\epsilon_-(z_+)Q_-GG^{-1},\end{aligned} \tag{5}$$

where $a_\pm$ are $\mathbf{g}$-valued superfields.

An additional ingredient demanded by the $N=2$ superconformal symmetry is a complex structure J on the finite-dimensional Lie algebra of the model which is skew-symmetric with respect to the inner product $\langle,\rangle$.[31–33] That is, we should demand that the following equations be satisfied on $\mathbf{g}$:

$$\begin{aligned}&J^2=-1,\quad \langle Jx,y\rangle+\langle x,Jy\rangle=0,\\&[Jx,Jy]-J[Jx,y]-J[x,Jy]=[x,y]\end{aligned} \tag{6}$$

for any elements x, y in $\mathbf{g}$. It is clear that the corresponding Lie group is a complex manifold with left (or right) invariant complex structure. In the following we shall denote the real Lie group and the real Lie algebra with the complex structure satisfying (6) by the pairs $(\mathbf{G},J)$ and $(\mathbf{g},J)$ respectively.

The complex structure J on the Lie algebra defines the second supersymmetry transformation[31]

$$\begin{aligned}&(G^{-1}\delta_{\eta_+}G)^a=\eta_+(z_-)(J_l)^a_b(G^{-1}D_+G)^b,\\&(\delta_{\eta_-}GG^{-1})^a=\eta_-(z_+)(J_r)^a_b(D_-GG^{-1})^b.\end{aligned} \tag{7}$$

where J_l, J_r are the left invariant and right invariant complex structures on $\mathbf{G}$ which correspond to the complex structure J.

The notion of Manin triple is closely related to a complex structure on a Lie algebra. By definition,[15] a Manin triple $(\mathbf{g},\mathbf{g}_+,\mathbf{g}_-)$ consists of a Lie algebra $\mathbf{g}$ with nondegenerate invariant inner product $\langle,\rangle$ and isotropic Lie subalgebras $\mathbf{g}_\pm$ such that the vector space $\mathbf{g}=\mathbf{g}_+\oplus\mathbf{g}_-$.

With each pair $(\mathbf{g},J)$ one can associate the complex Manin triple $(\mathbf{g}^{\mathbb{C}},\mathbf{g}_+,\mathbf{g}_-)$, where $\mathbf{g}^{\mathbb{C}}$ is the complexification of $\mathbf{g}$ and $\mathbf{g}_\pm$ are $\pm i$ eigenspaces of J. Moreover, it can be proved that there exists a one-to-one correspondence between a complex Manin triple endowed with an anti-linear involution which conjugates isotropic subalgebras τ: $\mathbf{g}_\pm\to\mathbf{g}_\mp$ and a real Lie algebra endowed with an ad-invariant nondegenerate inner product $\langle,\rangle$ and complex structure J which is skew-symmetric with respect to $\langle,\rangle$.[32] The conjugation can be used to extract a real form from a complex Manin triple.

Now we have to consider some geometric properties of the $N=2$ SWZNW models closely related to the existence of complex structures on the groups. We shall follow Ref. 25.

Let us fix some compact Lie group with the left invariant complex structure $(\mathbf{G},J)$ and consider its Lie algebra with the complex structure $(\mathbf{g},J)$. The complexification $\mathbf{g}^{\mathbb{C}}$ of $\mathbf{g}$ has the Manin triple structure $(\mathbf{g}^{\mathbb{C}},\mathbf{g}_+,\mathbf{g}_-)$. The Lie group version of this triple is the double Lie group

$(\mathbf{G}^{\mathbb{C}},\mathbf{G}_+,\mathbf{G}_-)$,[34–36] where the exponential subgroups $\mathbf{G}_\pm$ correspond to the Lie algebras $\mathbf{g}_\pm$. The real Lie group $\mathbf{G}$ is extracted from its complexification with the help of conjugation τ (it will be assumed in the following that τ is the hermitian conjugation)

$$\mathbf{G}=\{g\in\mathbf{G}^{\mathbb{C}}|\,\tau(g)=g^{-1}\}. \tag{8}$$

Each element $g\in\mathbf{G}^{\mathbb{C}}$ from the vicinity $\mathbf{G}_1$ of the unit element from $\mathbf{G}^{\mathbb{C}}$ admits two decompositions:

$$g=g_+g_-^{-1}=\tilde{g}_-\tilde{g}_+^{-1}. \tag{9}$$

Taking into account (8) and (9) we conclude that the element $g(g\in\mathbf{G}_1)$ belongs to $\mathbf{G}$ if

$$\tau(g_\pm)=\tilde{g}_\mp^{-1}. \tag{10}$$

These equations mean that we can parameterize the elements of

$$\mathbf{C}_1\equiv\mathbf{G}_1\cap\mathbf{G} \tag{11}$$

by the elements of the complex group $\mathbf{G}_+$ (or $\mathbf{G}_-$), i.e., we can introduce complex coordinates (they are just matrix elements of g_+ (or g_-)) in the start $\mathbf{C}_1$.

To generate (9) and (10) one has to consider the set W (which we shall assume in the following to be discrete and finite) of classes $\mathbf{G}_+\backslash\mathbf{G}^{\mathbb{C}}/\mathbf{G}_-$ and choose a representative w for each class $[w]\in W$. It gives us the stratification of $\mathbf{G}^{\mathbb{C}}$.[35]

$$\mathbf{G}^{\mathbb{C}}=\cup_{|w|\in W}\mathbf{G}_+w\mathbf{G}_-=\cup_{|w|\in W}\mathbf{G}_w. \tag{12}$$

There is a second stratification:

$$\mathbf{G}^{\mathbb{C}}=\cup_{|w|\in W}\mathbf{G}_-w\mathbf{G}_+=\cup_{|w|\in W}\mathbf{G}^w. \tag{13}$$

We shall assume, in the following, that the representatives w have been chosen to be unitary:

$$\tau(w)=w^{-1}. \tag{14}$$

It allows us to generalize (9) as follows:

$$g=wg_+g_-^{-1}=w\tilde{g}_-\tilde{g}_+^{-1}, \tag{15}$$

where

$$g+\in\mathbf{G}_+^w,\ \ \tilde{g}_-\in\mathbf{G}_-^w \tag{16}$$

and

$$\mathbf{G}_+^w=\mathbf{G}_+\cap w^{-1}\mathbf{G}_+w,\quad \mathbf{G}_-^w=\mathbf{G}_-\cap w^{-1}\mathbf{G}_-w. \tag{17}$$

In order for the element g to belong to the real group $\mathbf{G}$ the elements $g_\pm,\tilde{g}_\pm$ from (15) must satisfy (10). Thus, the formulas (10) and (15) define the mapping

$$\phi_w^+:\mathbf{G}_+^w\to\mathbf{C}_w\equiv\mathbf{G}_w\cap\mathbf{G}. \tag{18}$$

In a similar way one can define the mapping

$$\phi_w^-:\mathbf{G}_-^w\to\mathbf{C}_w\equiv\mathbf{G}_w\cap\mathbf{G}. \tag{19}$$

In Refs. 25 and 26 the following statements were proved.

1) The mappings (18) are holomorphic and define the natural (holomorphic) action of the complex group $\mathbf{G}_+$ on $\mathbf{G}$; the set W parameterizes the $\mathbf{G}_+$-orbits $\mathbf{C}_w$.

2) The $(\mathbf{G},J)$-SWZNW model admits PL symmetry,[11,37] with respect to $\mathbf{G}_+$ action so that we may associate with each external surface $G_+(z_+,z_-,\Theta_+,\Theta_-)\subset\mathbf{G}_+$, of the model a mapping (''Noether charge'') $V_-(z_+,z_-,\Theta_+,\Theta_-)$ from the super world-sheet into the group $\mathbf{G}_-$. The pair $[G_+(z_+,z_-,\Theta_+,\Theta_-),\ V_-(z_+,z_-,\Theta_+,\Theta_-)]$ can be lifted into the the double $\mathbf{G}^{\mathbb{C}}$:

$$\Phi(z_+,z_-,\Theta_+,\Theta_-)=G_+(z_+,z_-,\Theta_+,\Theta_-)\times V_-(z_+,z_-,\Theta_+,\Theta_-). \tag{20}$$

Moreover, the surface (20) can be rewritten in the form

$$\Phi(z_\pm,\Theta_\pm)=G(z_\pm,\Theta_\pm)H_-^{-1}(z_+,\Theta_-). \tag{21}$$

Here $G(z_\pm,\Theta_\pm)\subset\mathbf{G}$ is a solution of the $\mathbf{G}$-SWZNW model and the superfield H_- is given by the solution of the equation

$$H_-^{-1}D_+H_-=2(I_+)^-, \tag{22}$$

where $(I_+)^-$ is $\mathbf{g}_-$-projection of the conservation current $I_+=G^{-1}D_+G$ of the model.

3) With the appropriate modifications the above statements are true also for the mappings (19) and $\mathbf{G}_-$ action on $\mathbf{G}$. Thus, one can represent the surface (20) in the ''dual'' parameterization[11]

$$\Phi(z_\pm,\Theta_\pm)=\check{G}(z_\pm,\Theta_\pm)H_+^{-1}(z_+,\Theta_-), \tag{23}$$

where $\check{G}(z_\pm,\Theta_\pm)$ is the dual solution of the $\mathbf{G}$-SWZNW model and the superfield H_+ is given by the similar equation

$$H_+^{-1}D_+H_+=2(\check{I}_+)^+, \tag{24}$$

where $(\check{I}_+)^+$ is the $\mathbf{g}_+$-projection of the dual conserved current $\check{I}_+\equiv\check{G}^{-1}D_+\check{G}$.

4) Under PL T-duality

$$t:G(z_\pm,\Theta_\pm)\to\check{G}(z_\pm,\Theta_\pm)=G(z_\pm,\Theta_\pm)H(z_+,\Theta_-), \tag{25}$$

where

$$H\equiv H_-^{-1}H_+, \tag{26}$$

the conserved rightmoving current I_+ transforms as

$$t:(I_+)^-\to(\check{I}_+)^+,\ \ (I_+)^+\to(\check{I}_+)^-, \tag{27}$$

while the conserved leftmoving current $I_-\equiv D_-GG^{-1}$ transforms identically:

$$t:(I_-)^\pm\to(I_-)^\pm. \tag{28}$$

Moreover, the classical rightmoving $N=2$ super-Virasoro algebra maps under PL T-duality as follows:[26]

$$t:\Sigma^\pm\to\check{\Sigma}^\mp,\ \ T\pm i\partial K\to\check{T}\mp i\partial\check{K}, \tag{29}$$

where $\Sigma^\pm$ are the spin-3/2 currents, T is the stress-energy tensor, and K is the $U(1)$ current, while the leftmoving $N=2$ super-Virasoro algebra maps identically. Thus, PL T-duality in the classical $N=2$ SWZNW models is a mirror duality.

3. POISSON–LIE *T*-DUALITY AND MIRROR SYMMETRY IN THE QUANTUM *N*=2 SUPERCONFORMAL WZNW MODELS

We start with the Manin triple construction of the $N=2$ Virasoro algebra generators of the quantum SWZNW model on the group $(\mathbf{G},J)$.[32,33,38]

Let us specify an orthonormal basis

$$\{E^a,E_a,a=1,...,d\} \tag{30}$$

in the Manin triple $(\mathbf{g}^{\mathbb{C}},\mathbf{g}_+,\mathbf{g}_-)$, so that $\{E^a\}$ is a basis in g_+, and $\{E_a\}$ is a basis in g_-. The commutation relations and Jacoby identity in this basis take the form

$$\begin{aligned}
&[E^a,E^b]=f_c^{ab}E^c,\\
&[E_a,E_b]=f_{ab}^c E_c,\\
&[E^a,E_b]=f_{bc}^a E^c-f_b^{ac}E_c,\\
&f_d^{ab}f_e^{dc}+f_d^{bc}f_e^{da}+f_d^{ca}f_e^{db}=0,\\
&f_{ab}^d f_{dc}^e+f_{bc}^d f_{da}^e+f_{ca}^d f_{db}^e=0,\\
&f_{mc}^a f_d^{bm}-f_{md}^a f_c^{bm}-f_{mc}^b f_d^{am}+f_{md}^b f_c^{am}=f_{cd}^m f_m^{ab}.
\end{aligned} \tag{31}$$

Let us introduce the matrices

$$B_a^b=f_c f_a^{cd}+f^c f_{ca}^b,\quad A_a^b=f_{ac}^d f_d^{bc}. \tag{33}$$

Let $j^a(z)$, $j_a(z)$ be the generators of the affine Kac–Moody algebra $\hat{g}^{\mathbb{C}}$, corresponding to the fixed basis $\{E^a,E_a\}$, so that the currents j^a generate the subalgebra $\hat{g}_+$ and the currents j_a generate the subalgebra $\hat{g}_-$ (we shall omit in the following the super-world-sheet indices $\pm$, keeping in mind that we are in the rightmoving sector). The singular operator product expansions (OPEs) between these currents are the following:

$$\begin{aligned}
j^a(z)j^b(w)&=-(z-w)^{-2}\frac{1}{2}k(E^a,E^b)\\
&\quad+(z-w)^{-1}f_c^{ab}j^c(w)+\mathrm{reg},\\
j_a(z)j_b(w)&=-(z-w)^{-2}\frac{1}{2}k(E_a,E_b)\\
&\quad+(z-w)^{-1}f_{ab}^c j_c(w)+\mathrm{reg},\\
j^a(z)j_b(w)&=-(z-w)^{-2}\frac{1}{2}(q\delta_b^a+k(E^a,E_b))\\
&\quad+(z-w)^{-1}(f_{bc}^a j^c-f_b^{ac}j_c)(w)+\mathrm{reg},
\end{aligned} \tag{34}$$

where $k(x,y)$ denotes the Killing form for the vectors x, y of $\mathbf{g}^{\mathbb{C}}$. Let $\psi^a(z)$, $\psi_a(z)$ be free fermion currents which have the following singular OPEs:

$$\psi^a(z)\psi_b(w)=-(z-w)^{-1}\delta_b^a+\mathrm{reg}. \tag{35}$$

Then the $N=2$ Virasoro superalgebra currents and the central charge are given by[31–33,38]

$$\Sigma^+=\frac{2}{\sqrt{q}}\left(\psi^a j_a+\frac{1}{2}f_{ab}^c:\psi^a\psi^b\psi_c:\right),$$

$$\Sigma^-=\frac{2}{\sqrt{q}}\left(\psi_a j^a+\frac{1}{2}f_c^{ab}:\psi_a\psi_b\psi^c:\right), \tag{36}$$

$$K=\left(\frac{2B_a^b}{q}-\delta_a^b\right):\psi^a\psi_b:-\frac{2}{q}(f_c j^c-f^c j_c),$$

$$T=-\frac{1}{q}:(j^a j_a+j_a j^a):-\frac{1}{2}:(\partial\psi^a\psi_a-\psi^a\partial\psi_a):,$$

$$c=3\left(d-\frac{2A_a^a}{q}\right). \tag{37}$$

The set of currents (36) can be combined into the superfields

$$\Gamma^{\pm}=\frac{1}{\sqrt{2}}\Sigma^+ +\Theta\left(T\mp\frac{1}{2}\partial K\right), \tag{38}$$

so that the energy-momentum super-tensor is given by the sum

$$\Gamma=\frac{1}{2}(\Gamma^+ +\Gamma^-)=-\frac{1}{q}:\langle DI,I\rangle:+\frac{2}{3q^2}:\langle I,:\{I,I\}:\rangle:. \tag{39}$$

Here I denotes Lie algebra valued super-Kac–Moody currents of the affine superalgebra $\hat{\mathbf{g}}$:

$$\begin{aligned}
&I\equiv I^a E_a+I_a E^a,\\
&I^a=-\sqrt{\frac{q}{2}}\,\psi^a+\Theta\left(j^a+\left(\frac{1}{2}f_{bc}^a:\psi^b\psi^c:+f_c^{ab}:\psi_b\psi^c:\right)\right),\\
&I_a=-\sqrt{\frac{q}{2}}\,\psi_a+\Theta\left(j_a+\left(\frac{1}{2}f_a^{bc}:\psi_b\psi_c:+f_{ab}^c:\psi^b\psi_c:\right)\right).
\end{aligned} \tag{40}$$

We now propose a quantum version of the PL *T*-duality transformation. Perhaps the most comprehensive way to find PL *T*-duality transformation rules for the quantum fields of the model is to quantize canonically the Sfetsos canonical transformations for PL *T*-duality related σ-models[21] and then define and solve the quantum version of the equations (22), (24), and (26). Though developing this approach for the $N=2$ superconformal field theory is an important problem and worth solving, it is beyond our reach at the present moment.

Instead we determine the quantum counterpart of the mapping (25) as an automorphism of the operator algebra of the quantum fields, defined by right multiplication by the rightmoving matrix-valued function $H(Z)$, which implies that $N=2$ SWZNW model is PL self-dual. We propose a very simple way to find the matrix elements of H using super-Kac–Moody Ward identities and the Knizhnik–Zamolodchikov equation.

In the $N=1$ superfield formalism an arbitrary conformal superfield is defined by the following OPEs.[39]

$$\begin{aligned}
&I^a(Z_1)F^\Lambda(Z_2)=Z_{12}^{-1/2}E^a F^\Lambda(Z_2)+\mathrm{reg},\\
&I_a(Z_1)F^\Lambda(Z_2)=Z_{12}^{-1/2}E_a F^\Lambda(Z_2)+\mathrm{reg}.
\end{aligned} \tag{41}$$

Here E^a, E_a denote the generators of the $\mathbf{g}^{\mathbb{C}}$ in the representation with the highest weight Λ,

$$\Gamma(Z_1)F^\Lambda(Z_2)=Z_{12}^{-3/2}\Delta F^\Lambda(Z_2)+Z_{12}^{-1}\frac{1}{2}DF^\Lambda(Z_2)$$
$$+Z_{12}^{-1/2}\partial F^\Lambda(Z_2)+\mathrm{reg}, \tag{42}$$

where the conformal dimension Δ is given by

$$\Delta=C_\Lambda/q,\quad C_\Lambda\equiv-(E^aE_a+E_aE^a). \tag{43}$$

and we have used the standard notations for even and odd world-sheet super-intervals between a pair of points $Z_i=(z_i,\Theta_i)$, $i=1,2$:

$$Z_{12}\equiv z_1-z_2-\Theta_1\Theta_2,\quad \Theta_{12}=Z_{12}^{1/2}\equiv\Theta_1-\Theta_2, \tag{44}$$

so that

$$Z_{12}^{n+1/2}=Z_{12}^n\Theta_{12},\quad n\in\mathbb{Z}. \tag{45}$$

We postulate the quantum version of the formula (25):

$$t{:}F^\Lambda(Z)\to\check F^\Lambda(Z)=F^\Lambda(Z)H(Z), \tag{46}$$

which is the quantum counterpart of (25) (here and in what follows the leftmoving coordinate dependence of the fields will be omitted for simplicity). It follows from the Sugawara formula (39) and the OPEs (41) and (42) that the conformal superfield $F^\Lambda(Z)$ of the model satisfies the Knizhnik–Zamolodchikov equation[39]

$$\frac{q}{2}DF^\Lambda(Z)+{:}F^\Lambda I{:}(Z)=0, \tag{47}$$

which is a quantization of the classical relation $I=G^{-1}DG$. In view of (46): the dual field $\check F^\Lambda$ satisfies the similar equation

$$\frac{q}{2}D\check F^\Lambda(Z)=-{:}\check F^\Lambda\check I{:}(Z)$$
$$=-{:}\check F^\Lambda H^{-1}IH{:}(Z)+\frac{q}{2}\check F^\Lambda H^{-1}DH(Z). \tag{48}$$

Let us go back for a moment to the classical case and consider Eqs. (22), (24), and (26). Using them we can write

$$H^{-1}DH=2(\check I^+-H^{-1}I^-H). \tag{49}$$

As its quantum version we propose

$$\frac{q}{2}\check F^\Lambda H^{-1}DH(Z)=-2{:}\check F^\Lambda(\check I^+-H^{-1}I^-H){:}(Z). \tag{50}$$

The substitution (50) converts (48) into

$${:}\check F^\Lambda(\check I^--\check I^+){:}(Z)={:}\check F^\Lambda(H^{-1}(I^+-I^-)H){:}(Z). \tag{51}$$

Using the left-invariant complex structure J on the group $\mathbf{G}$ one can rewrite it in the form

$${:}\check F^\Lambda(J\mathrm{End}(H)J\check I){:}(Z)={:}\check F^\Lambda I{:}(Z), \tag{52}$$

where we have introduced the notation $\mathrm{End}(H)x=HxH^{-1}$, $x\in\mathbf{g}^{\mathbb{C}}$ and we imply that $\mathrm{End}(H)$ belongs to the group of super-Kac–Moody algebra automorphisms. The equation (52) means that $\mathrm{End}(H)$ interchanges the isotropic subalgebras of the Manin triple because it anticommutes with the complex structure J.

By virtue of (52) Eq. (48) takes the form

$$\frac{q}{2}\check F^\Lambda H^{-1}DH(Z)={:}\check F^\Lambda((\mathrm{End}(H^{-1})J\mathrm{End}(H)J-1)\check I){:}(Z). \tag{53}$$

Using super-Kac–Moody Ward identities[39] it is easy to see that (53) decays into the system of equations

$$H^{-1}DH=0,$$
$$\mathrm{End}(H^{-1})J\mathrm{End}(H)J-1=0. \tag{54}$$

Its solution is given by the constant matrix anticommuting with J:

$$DH=0,\quad J\mathrm{End}(H)+\mathrm{End}(H)J=0. \tag{55}$$

In the orthonormal basis we have chosen, any matrix which anti-commutes with J should have the form

$$\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}h&0\\0&\overline{h}\end{pmatrix}, \tag{56}$$

where h is an arbitrary complex matrix (the bar denotes complex conjugation). Let us denote by $\mathrm{Aut}(\mathbf{g},J)$ the group of automorphisms of $\mathbf{g}$ which commute with J. It is clear that

$$\mathrm{End}(H)=\begin{pmatrix}0&1\\1&0\end{pmatrix}. \tag{57}$$

is a solution of (55). Hence each solution of (55) should have the form:

$$\mathrm{End}(H)=\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}m&0\\0&\overline{m}\end{pmatrix},\quad \begin{pmatrix}m&0\\0&\overline{m}\end{pmatrix}\in\mathrm{Aut}(\mathbf{g},J). \tag{58}$$

In view of (52) $\mathrm{End}(H)$ should be also an automorphism of the algebra $\hat{\mathbf{g}}$. It imposes on the matrix m the relation

$$m^{cb}\overline{m}_{ab}=\delta^c_a. \tag{59}$$

The next condition we should demand is $t^2=1$ (that is, PL T-duality is an involution). It gives the second relation for m:

$$m^{cb}\overline{m}_{ba}=\delta^c_a. \tag{60}$$

Therefore the set of PL T-duality transformations in the $N=2$ superconformal WZNW model on the group manifold $\mathbf{G}$ is given by the set of matrices (58) satisfying (59) and (60). Hence, under the quantum PL T-duality the currents (40) transform as

$$t{:}I^a\to m^{ab}I_b,\quad I_a\to\overline{m}_{ab}I^b, \tag{61}$$

or in components,

$$t{:}\psi^a\to m^{ab}\psi_b,\quad j^a\to m^{ab}j_b,\quad \psi_a\to\overline{m}_{ab}\psi^b,\quad j_a\to\overline{m}_{ab}j^b. \tag{62}$$

Taking into account (36), (59), and (62) we find the PL T-duality transformation of the $N=2$ Virasoro superalgebra currents:

$$t{:}\overset{\pm}{\sum}\to\overset{\mp}{\sum},\quad t{:}K\to-K,\quad T\to T. \tag{63}$$

Notice that, as in the classical case, PL *T*-duality acts in the leftmoving sector as an identity transformation. Therefore we may conclude that quantum PL *T*-duality in the $N=2$ superconformal WZNW models is a mirror duality and has a geometric realization which is given by PL $\mathbf{G}_{\pm}$-holomorphic action on the target space of the model.

Here a remark is in order. In many examples of the $N=2$ SWZNW models on the compact groups $(SU(3), SU(2)\times U(1),\ldots)$ the transformations (61) coincide with Weyl reflections. In these cases mirror symmetry was interpreted by the authors of Ref. 28 as a gauge symmetry. They presented also a contradictory example, $SU(2)\times SU(2)$ SWZNW model, where the Weyl reflections failed to give mirror symmetry. It follows from our formula (61) that in this example mirror symmetry is given by an external automorphism of the Lie algebra $su(2)\times su(2)$. This example illustrates the general picture: PL *T*-duality is given by an automorphism (internal or external) which interchanges the isotropic subalgebras of the underlying Manin triple.

4. POISSON–LIE *T*-DUALITY AND MIRROR SYMMETRY IN QUANTUM KAZAMA—SUZUKI MODELS

In this section we consider PL *T*-duality in Kazama–Suzuki models. Kazama and Suzuki have studied[40] the conditions under which an $N=1$ superconformal coset model can have an extra supersymmetry, giving rise to an $N=2$ superconformal model. Then the $N=2$ superconformal coset theories were classified more accurately in Ref. 41. Their conclusion can be reformulated as follows. Suppose the Manin triple $(\mathbf{g}^{\mathbb{C}}, \mathbf{g}_+, \mathbf{g}_-)$ associated with the pair $(\mathbf{g}, J)$ has a Manin subtriple $(\mathbf{h}, \mathbf{h}_+, \mathbf{h}_-)$, that is, $\mathbf{h}_+ \subset \mathbf{g}_{\pm}$ are subalgebras of $\mathbf{g}_{\pm}$ such that $\mathbf{h} \equiv \mathbf{h}_+ \oplus \mathbf{h}_-$ is a subalgebra of $\mathbf{g}^{\mathbb{C}}$ and $\tau: \mathbf{h}_+ \to \mathbf{h}_-$. Notice that the Manin subtriple specified above defines (with the help of the involution τ) a pair $(\mathbf{k}, J)$ such that $\mathbf{k}^{\mathbb{C}} = \mathbf{h}$ and $\mathbf{k} \subset \mathbf{g}$.

Assume that the basis (30) is chosen so that the subbases

$$\{E^i, i=1,\ldots,d_h\}, \quad \{E_i, i=1,\ldots,d_h\} \tag{64}$$

are bases in the subalgebras $\mathbf{h}_+$ and $\mathbf{h}_-$, respectively. Let us consider a vector subspace

$$\mathbf{a} = \mathbf{g}^{\mathbb{C}}/\mathbf{h} \tag{65}$$

generated (over $\mathbb{C}$) by the vectors

$$\{E^{\alpha}, \alpha = d_h+1,\ldots,d\}, \quad \{E_{\alpha}, \alpha = d_h+1,\ldots,d\}. \tag{66}$$

The Manin triple construction of the Kazama–Suzuki models is given by the following.

Proposition

Suppose the isotropic subspaces

$$\mathbf{a}_{\pm} \equiv \mathbf{a} \cap \mathbf{g}_{\pm} \tag{67}$$

are Lie subalgebras. Then the currents

$$\begin{aligned}
\sum_{cs}^{+} &= \frac{2}{\sqrt{q}}\left(\psi^{\alpha} j_{\alpha} + \frac{1}{2} f^{\gamma}_{\alpha\beta} : \psi^{\alpha}\psi^{\beta}\psi_{\gamma} : \right), \\
\sum_{cs}^{-} &= \frac{2}{\sqrt{q}}\left(\psi_{\alpha} j^{\alpha} + \frac{1}{2} f^{\alpha\beta}_{\gamma} : \psi_{\alpha}\psi_{\beta}\psi^{\gamma} : \right), \\
K_{cs} &= \left(\frac{2C^{\beta}_{\alpha}}{q} - \delta^{\beta}_{\alpha} \right) : \psi^{\alpha}\psi_{\beta} : - \frac{2}{q}(\hat{f}_c j^c - \hat{f}^c j_c), \\
T_{cs} &= -\frac{1}{q} : (j^a j_a + j_a j^a) : - \frac{1}{2} : (\partial\psi^a \psi_a - \psi^a \partial\psi_a) : \\
&\quad + \frac{1}{q} : (u^k u_k + u_k u^k) :,
\end{aligned} \tag{68}$$

where

$$\hat{f}^a = f^{a\gamma}_{\gamma}, \quad \hat{f}_a = f^{\gamma}_{a\gamma}, \quad C^{\alpha}_{\beta} = \hat{f}^a f^{\alpha}_{a\beta} + \hat{f}_a f^{a\alpha}_{\beta}, \tag{69}$$

$$u^k = j^k - f^{k\alpha}_{\beta} : \psi^{\beta}\psi_{\alpha} :, \quad u_k = j_k + f^{\beta}_{k\alpha} : \psi^{\alpha}\psi_{\beta} :, \tag{70}$$

satisfy the OPEs of the $N=2$ super-Virasoro algebra with the central charge

$$c_{cs} = c_g - c_h. \tag{71}$$

This is just the $N=2$ extension[42] of the Goggard–Kern–Olive construction formulated in terms of Manin triples and can be checked by direct calculations.

The Kazama–Suzuki model based on the coset $\mathbf{G}/\mathbf{K}$ can be obtained from the SWZNW model on the group $\mathbf{G}$ by gauging an anomaly-free subgroup $\mathbf{K}$.[43] In view of the Manin triple construction (68) and (71) this implies classically that the currents corresponding to the Manin subtriple $(\mathbf{h}, \mathbf{h}_+, \mathbf{h}_-)$ should vanish:

$$I^i(Z) = I_i(Z) = 0. \tag{72}$$

In quantizing the theory canonically one should impose in some way such constraints on physical states. We impose

$$I^i(Z_1)\Phi(Z_2) = \text{reg}, \quad I_i(Z_1)\Phi(Z_2) = \text{reg}, \tag{73}$$

that is, the physical states of the coset are the highest vectors of the trivial $\hat{\mathbf{h}}$-representation.

Under PL *T*-duality (61) the set of constraints (73) will transform, in general, into another set of constraints giving another coset model. Therefore we should define PL *T*-duality transformations in the Kazama–Suzuki model as the subset of (58)–(60) which stabilizes the set (73), or equivalently, as the subset which stabilizes the Manin subtriple $(\mathbf{h}, \mathbf{h}_+, \mathbf{h}_-)$. Taking into account this condition and using (61) we obtain PL *T*-duality transformation rules for the currents (68) of the $N=2$ super-Virasoro algebra,

$$t: \sum_{cs}^{\pm} \to \sum_{cs}^{\mp}, \quad t: K_{cs} \to -K_{cs}, \; T_{cs} \to T_{cs}, \tag{74}$$

which are similar to (63). It is clear that PL *T*-duality in the leftmoving sector is given by the identity transformation.

Let us consider an example of the Kazama–Suzuki model based on the coset $U(2)/(U(1)\times U(1))$ (the $N=2$ minimal model). The complexification of $u(2)$ is the Lie algebra $gl(2,\mathbb{C})$. In this case the commutation relations (31) in the orthonormal basis (30) are given by

$$[E^0,E^1]=E^1, \quad [E_0,E_1]=E_1, \quad [E^1,E_1]=-E^0+E_0. \tag{75}$$

The isotropic subalgebras $\mathbf{g}_+$ and $\mathbf{g}_-$ of the complex Manin triple are generated by the vectors E^0, E^1 and E_0, E_1 respectively. The currents of the super-Kac–Moody algebra $\hat{gl}(2,\mathbb{C})$ are characterized by the following OPEs

$$\begin{aligned} &I^a(Z_1)I^b(Z_2)=Z_{12}^{-1/2}f_c^{ab}I^c(Z)+\text{reg},\\ &I_a(Z_1)I_b(Z_2)=Z_{12}^{-1/2}f_{ab}^{c}I_c(Z)+\text{reg},\\ &I^a(Z_1)I_b(Z_2)=-Z_{12}^{-1}\frac{q}{2}\delta_b^a+Z_{12}^{-1/2}(f_{bc}^aI^c-f_b^{ac}I_c)+\text{reg}, \end{aligned} \tag{76}$$

where $a, b, c=0, 1$ and the structure constants are given by (75). The Manin subtriple defining our coset model is given by

$$h=h_+\oplus h_-, \quad h_+=\mathbb{C}E^0, \quad h_-=\mathbb{C}E_0. \tag{77}$$

Thus, the Manin subtriple corresponds to the $N=2$ $U(1)^2$-SWZNW model which is described by the pair of scalar complex free superfields $X^0(Z)$, $X_0(Z)$ with obvious OPEs

$$X^0(Z_1)X_0(Z_2)=-2\log Z_{12}. \tag{78}$$

The currents of the super-Kac–Moody algebra $\hat{gl}(2,\mathbb{C})$ can be realized in terms of the fields $X^0(Z)$, $X_0(Z)$ and superparafermions $S^1(Z)$, $S_1(Z)$:[44]

$$\begin{aligned} &I^0=\frac{\sqrt{q}}{2}DX^0, \quad I_0=\frac{\sqrt{q}}{2}DX_0,\\ &I^1=iS^1\exp\left(-\frac{1}{\sqrt{q}}(X_0-X^0)\right),\\ &I_1=iS_1\exp\left(\frac{1}{\sqrt{q}}(X_0-X^0)\right). \end{aligned} \tag{79}$$

The super-parafermion OPEs are deduced from the OPEs (76) and (78) and the null-vector relation in the trivial $\widehat{su}(2)$-representation.

The most general PL T-duality transformation in $U(2)$-SWZNW model is given by

$$\begin{aligned} &I^0\to I_0, \quad I_0\to I^0,\\ &I^1\to\exp(i\phi)I_1, \quad I_1\to\exp(-i\phi)I^1, \end{aligned} \tag{80}$$

where ϕ is an arbitrary real number. We see that the constraints transform into itself. From these formulas we easily find the PL T-duality transformations of the parafermions of the coset

$$S^1\to\exp(i\phi)S_1, \quad S_1\to\exp(-i\phi)S^1. \tag{81}$$

Thus, the PL T-duality transformation acts in the $U(1)^2$-subspace of the $U(2)$-SWZNW model as the usual $R\to 1/R$ T-duality (at the self-dual point), while the PL T-duality transformation (81) corresponds to the axial-vector duality of the coset $SU(2)/U(1)$[29] (to see this is enough to recover the leftmoving constraints).

It is clear that there is a direct generalization of this example to the coset models $\mathbf{G}/U(I)^r$, where r is the dimension of the maximal torus of the group $\mathbf{G}$. the PL T-duality transformation will act on the maximal torus as an Abelian $R\to 1/R$ T-duality (at the self-dual point), while in the $N=2$ Kazama–Suzuki model it will act as an axial-vector duality.[45] In the non-Abelian coset models the PL T-duality transformation rules of the fields are given by the non-Abelian generalization of the axial-vector duality via Ref. 46. In principle they can be found using the non-Abelian generalization of the super-parafermions (79). Some aspects of this construction in the non supersymmetric case can be found in Ref. 47.

Thus, in summary, we conclude that quantum PL T-duality in the Kazama—Suzuki models is a mirror duality also.

5. CONCLUSION

In this work we have considered the PL T-duality transformation in quantum $N=2$ superconformal WZNW and Kazama–Suzuki models. The PL T-duality transformation rules in the quantum $N=2$ SWZNW are found using the Manin triple construction of the $N=2$ SWZNW models, the Knizhnik–Zamolodchikov equation. Ward identities, and the conjecture that, as in the classical case, PL T-duality is given by constant automorphisms of the rightmoving super-Kac–Moody algebras of the models which interchange the isotropic subalgebras of the underlying Manin triples. We have shown that in these models PL T-duality is a mirror duality. We have thus given a geometric realization of the mirror symmetry in these models. Notice also that our results are in agreement with the conjecture proposed in[28] that mirror symmetry can be considered as a gauge symmetry (which is extended in some cases by the external automorphisms) of the self-dual points of the moduli space of the $N=2$ superconformal field theories.

We have given Manin triple construction of the Kazama–Suzuki models, representing them as *(Manin triple)/(Manin subtriple)*-cosets. By means of this representation we defined PL T-duality transformations in the Kazama–Suzuki models as the subset of PL T-duality transformations of the numerator triple which stabilize the denominator triple. It was shown that, thus defined, PL T-duality is a mirror duality also. An interesting open problem is to find the corresponding geometric picture of PL T-duality and mirror symmetry in the classical Kazama–Suzuki models.

Our results are useful in discussing Calabi–Yau superstring compactifications and allow us to conjecture that PL T-duality is an adequate geometric structure underlying mirror symmetry. The extension of our results to the Gepner construction of superstring vacua[48] (see also, Ref. 49) would be a test of the conjecture.

Another interesting problem is to quantize the equations (22) and (24) and determine the quantum version of (21) and (23). Moreover, their solution is important in the context of quantum PL T-duality and mirror symmetry; it may be useful also in discussing T-duality for open strings and D-branes on

curved backgrounds and will be helpful in ''quantization'' of the existing treatments.[17,18]

This work was supported in part by Grants INTAS-95-IN-RU-690, CRDF RPI-277, RFBR 96-02-16507.

*)E-mail: spark@itp.ac.ru

[1] A. Giveon, M. Porrati, and E. Rabinovici, Phys. Rep. **244**, 79 (1994).
[2] E. Alvarez, L. Alvarez-Gaume, and Y. Lozano, preprint CERN-TH. 7486/94, E-print archives hep-th/9410237.
[3] In *Essays on Mirror Manifolds*, ed. by S. T. Yau, International Press, Hong Kong (1991).
[4] D. R. Morrison and M. R. Plesser, Nucl. Phys. B **440**, 279 (1995).
[5] V. Batyrev and L. A. Borisov, E-print archives alg-geom/9412017.
[6] D. R. Morrison and M. R. Plesser, E-print archives hep-th/9508107.
[7] B. R. Greene and M. R. Plesser, Nucl. Phys. B **338**, 15 (1990).
[8] A. Cappelli, C. Itzykson and J.-B. Zuber, Nucl. Phys. B **280**, 445 (1987).
[9] B. Greene, E-print archives hep-th/9702155.
[10] A. Strominger, S-T. Yau, and E. Zaslow, E-print archives hep-th/9606040.
[11] C. Klimcik and P. Severa, Phys. Lett. B **351**, 455 (1995).
[12] X. De la Ossa and F. Quevedo, Nucl. Phys. B **403**, 377 (1993).
[13] A. Giveon and M. Rocek, Nucl. Phys. B **421**, 173 (1994); E. Alvarez, L. Alvarez-Gaume, and Y. Lozano, Phys. Lett. B **336**, 183 (1994); A. Giveon, E. Rabinovici, and G. Veneziano, Nucl. Phys. B **322**, 167 (1989).
[14] B. E. Fridling and A. Jevicki, Phys. Lett. B **134**, 70 (1984); E. S. Fradkin and A. A. Tseytlin, Ann. Phys. **162**, 31 (1985); T. Curtright and C. Zachos, Phys. Rev. D **49**, 5408 (1994); T. Curtright and C. Zachos, Phys. Rev. D **52**, R573 (1995).
[15] V. G. Drinfield, in *Proc. Int. Cong. Math.*, Berkley, Calif. (1986), p. 798.
[16] C. Klimcik and P. Severa, Preprint CERN-TH/95-330, E-print archives hep-th/9512040; C. Klimcik and P. Severa, Preprint CERN-TH/96-142, E-print archives hep-th/9605212; C. Klimcik and P. Severa, Preprint IHES/P/96/70. E-print archives hep-th/9610198; C. Klimcik, Preprint IHES/P/97/58, E-print archives hep-th/9707194.
[17] C. Klimcik and P. Severa, Preprint CERN-TH/96-254, E-print archives hep-th/9609112.
[18] C. Klimcik and P. Severa, Preprint CERN-TH/95-339, E-print archives hep-th/9512124.
[19] E. Tyurin and R. von Unge, E-print archives hep-th/9512025.
[20] K. Sfetsos, Preprint THU-96/38, E-print archives hep-th/9611199.
[21] K. Sfetsos, E-print archives hep-th/9710163.
[22] K. Sfetsos, E-print archives hep-th/9803019.
[23] S. E. Parkhomenko, JETP Lett. **64**, 877 (1996); S. E. Parkhomenko, Preprint Landau TMP/12/96, E-print archives hep-th/9612034.
[24] S. E. Parkhomenko, Preprint Landau TMP/05/97, E-print archives hep-th/9705233.
[25] S. E. Parkhomenko, Nucl. Phys. B **510**, 623 (1998).
[26] S. E. Parkhomenko, Mod. Phys. Lett. A **13**, 1041 (1998).
[27] J. Balog, P. Forgas, M. Mohammedi, L. Palla, and J. Schnittger, E-print archives hep-th/9806068.
[28] A. Giveon and E. Witten, Preprint IASSNS-HEP-94/30, R1-157-93, E-print archives hep-th/9404184.
[29] A. Giveon and E. Kiritsis, Nucl. Phys. B **411**, 487 (1994).
[30] P. DI Veccia, V. G. Knizhnik, J. L. Petersen, and P. Rossi, Nucl. Phys. B **253**, 701 (1985).
[31] Ph. Spindel, A. Sevrin, W. Troost, and A. van Proeyen, Nucl. Phys. B **308**, 662 (1988); Nucl. Phys. B **311**, 465 (1988/89).
[32] S. E. Parkhomenko, Zh. Éksp. Teor. Fiz. **102**, 3 (1992) [Sov. Phys. JETP **85**, 1 (1992)].
[33] S. E. Parkhomenko, Mod. Phys. Lett. A **11**, 445 (1996).
[34] M. A. Semenov-Tian-Shansky, RIMS, Kyoto Univ. **21**, 1237 (1985). Vol. 21.
[35] A. Yu. Alekseev and A. Z. Malkhin, Commun. Math. Phys. **162**, 147 (1994).
[36] J.-H. Lu and A. Weinstein, J. Diff. Geom. **31**, 501 (1990).
[37] A. Yu. Alekseev and I. T. Todorov, Nucl. Phys. B **421**, 413 (1994).
[38] E. Getzler, Preprint MIT(1994).
[39] S. Nam, Phys. Lett. B **187**, 340 (1987); E. B. Kiritsis and G. Siopsis, Phys. Lett. B **184**, 353 (1987).
[40] Y. Kazama and H. Suzuki, Nucl. Phys. B **321**, 232 (1989).
[41] C. Schweigert, Commun. Math. Phys. **149**, 425 (1992).
[42] C. M. Hull and B. Spence, Preprint QMC/PH/89/24 (1989).
[43] K. Gawedski and A. Kupiainen, Nucl. Phys. B **320**, 625 (1989).
[44] N. Ohta and H. Suzuki, Preprint OS-GE 01-89 (1989).
[45] M. Henningson, Nucl. Phys. B **423**, 631 (1994).
[46] E. Kiritsis, E-print archives hep-th/9406082.
[47] M. Yu. Lashkevich, Int. J. Mod. Phys. A **8**, 5673 (1993); Mod. Phys. Lett. A **8**, 851 (1993).
[48] D. Gepner, Nucl. Phys. B **296**, 757 (1988); Phys. Lett. B **199**, 380 (1987).
[49] T. Eguchi, H. Ooguri, A. Taormina and S. K. Yang, Nucl. Phys. B **315**, 193 (1989): A. Taormina, Preprint CERN-TH.5409/89.

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

Strong thermal self-action of a laser beam in gases and liquids

A. N. Kucherov*)

N. E. Zhukovskiĭ Central Aerohydrodynamic Institute, 140160 Zhukovskiĭ, Moscow Region, Russia
(Submitted 28 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 105–129 (July 1999)

Solutions which approximately describe the effect of strong thermal self-action of a laser beam in weakly absorbing media (gases and liquids) have been obtained. This paper considers the regimes of thermal conductivity, transverse flows of gases at subsonic and supersonic velocities, transonic nonlinear regime, and gravitational convection in a horizontal beam. Assuming that the shape of transverse intensity distribution is constant, and that the wave front can be approximated by a second-power polynomial, ordinary differential equations and their solutions for average transverse dimensions of beams have been obtained. These approximate solutions are in satisfactory agreement with exact solutions. © *1999 American Institute of Physics.* [S1063-7761(99)01007-0]

1. INTRODUCTION

Propagation of an intense laser beam through nonlinear absorbing media (liquids and gases) is described by a system of equations that includes the Navier–Stokes aerohydrodynamic equations and the wave-optic equation in the parabolic approximation,[1,2] which is the nonlinear Fresnel or Schrödinger equation. The related aero-optic effects occur on optical paths in the laboratory, in the atmosphere, and in technological laser facilities. The optical configurations involved can vary: horizontal, vertical, focused, parallel, and divergent beams; optical configurations defined by mirrors, lenses, telescopes and other devices. In the hydrodynamic aspect of the problem, the following situations and effects are of great interest: a uniform air flow which is either longitudinal or transverse with respect to the laser beam; effect of thermal conductivity, viscosity, acoustic perturbations (pressure variations); subsonic, transonic, supersonic, and hypersonic regimes; gravitational convection. Theoretical and experimental investigations of the entire variety of different regimes is labor-consuming. In the case of a strong thermal self-action, a numerical solution in the limit of geometrical optics is impossible because of large local gradients.

In the case of small perturbations of the beam due to heating and changes in the refraction index of the medium, a linearized Gebhardt–Smith solution of optical equations[3] (see also Refs. 4 and 5), obtained for a transverse air flowing through a parallel laser beam in the approximation of geometrical optics without allowance for the viscosity, thermal conductivity, and acoustic perturbations in the medium, has played an important role. This solution was used in describing gas-dynamic regimes of thermal self-refraction in a uniform transverse air flow.[6] Generalization of the Talanov transformation[7] of transverse coordinates of focused beams to the case of a variable beam radius defined in vacuum and utilization of similarity transformations in hydrodynamic equations allowed us to derive linearized solutions of aero-optic equations for complex optical configurations in various gas-dynamic conditions.[8]

Under conditions of strong thermal self-action, quasi-stabilization (or saturation) of perturbation parameters (namely, the peak intensity, its shift, and transverse intensity distribution) was detected in experiments[9–11] and numerical calculations.[12–15] With the help of averaged characteristics (transverse dimensions and deviation angles associated with a smoothed wave front) and dimensionality considerations, approximate solutions have been obtained in the case of purely convective air flow and in the case of gravitational convection in a horizontal beam.[16] A polynomial approximation of the wave front phase at large Fresnel numbers allowed us to relate the average characteristics to the coefficients of this polynomial and derive ordinary differential equations for average radii and shifts in the simple, purely convective regime of transverse air flow.[17]

In this paper we present the derivation of equations and solutions for average radii and shifts in the case of a transverse air flowing through a laser beam with velocities ranging between zero and hypersonic values, including the transonic regime, for the case of gravitational convection in a horizontal beam, and also for a stagnant, thermally conductive medium. The assumptions and conditions used in deriving these equations are formulated and generalized.

2. STATEMENT OF THE AERO-OPTIC PROBLEM

Propagation of an intense laser beam in a nonlinear medium is described by a parabolized wave equation of the paraxial (small-angle) optics[2] for the transverse component of a slowly varying electric field amplitude E. In the physical variables ''radiation intensity $I=E^*E$ — deviation angle $\boldsymbol{\vartheta}=\nabla_\perp\Phi$,'' where Φ is the wave front phase of the beam, $\nabla_\perp=\mathbf{e}_x\partial/\partial x+\mathbf{e}_y\partial/\partial y$, we obtain a system of equations for the eikonal and energy conservation

$$\left(\frac{\partial}{\partial z}+\boldsymbol{\vartheta}\nabla_\perp\right)\boldsymbol{\vartheta}=N\nabla_\perp\rho_1+\frac{1}{2F^2}\nabla_\perp\left\{\frac{1}{\sqrt{I}}\nabla_\perp^2\sqrt{I}\right\}, \qquad (1)$$

$$\left(\frac{\partial}{\partial z}+\boldsymbol{\vartheta}\boldsymbol{\nabla}_{\perp}\right)\ln I+\boldsymbol{\nabla}_{\perp}\boldsymbol{\vartheta}=-N_{\alpha}. \tag{2}$$

The coordinate z is measured in units of the optical path length L, the transverse coordinates x and y are measured in units of the characteristic (initial) beam radius r_0, the angle $\boldsymbol{\vartheta}$ is measured in units of r_0/L, and the radiation intensity I is measured in units of $I_0=P_0/\pi r_0^2$, where P_0 is the total initial power of the beam. The parameters of optical similarity are the following: the Fresnel number $F=2\pi r_0^2/\lambda L$, where λ is the radiation wavelength; the radiation attenuation (absorption) number $N_{\alpha}=\alpha L$, where α is the absorption coefficient of the medium; the self-action number $N=(L/z_T)^2$, where $z_T=r_0/\sqrt{Q(T_0/n_0)(\partial n/\partial T)}$ is the length of thermal self-action, $Q=\alpha P_0\tau/(\pi r_0^2\rho_0 C_p T_0)$ is the parameter of temperature and density perturbation, ρ_0, T_0, C_p, and n_0 are the density, temperature, specific heat at a constant pressure, and the refraction index in the unperturbed medium, and τ is the heating time constant. Here we assume a linear relation between the refraction index and changes in the density and temperature of the medium:

$$\Delta n=\frac{\partial n}{\partial T}\Delta T,\quad \frac{\Delta T}{T_0}=QT_1 \tag{3a}$$

in the case of liquid and

$$\Delta n=\frac{\partial n}{\partial \rho}\Delta\rho,\quad \frac{\Delta\rho}{\rho_0}=Q\rho_1,$$

$$n=1+\gamma\frac{\rho}{\rho_s}=n_0+(n_0-1)\frac{\Delta\rho}{\rho_0} \tag{3b}$$

in the case of gas, where γ is the Gladstone–Dayle constant, and ρ_s is the density under standard conditions. For air, in particular, $\rho_s=1.225$ kg/m^3, and for the visible and infrared optical ranges we have $\gamma=2.9\times10^{-4}$. The perturbations of density, ρ_1, and temperature, T_1, are derived from the conditions of mass, momentum, and energy conservation, and from the medium equation of state:

$$\frac{d\rho}{dt}+\rho\,\mathrm{div}\,\mathbf{V}=0, \tag{4}$$

$$\rho\frac{\partial\mathbf{V}}{dt}+\mathrm{Eu}\,\boldsymbol{\nabla}p=-\frac{\mathbf{e}_y}{\mathrm{Fr}}(1+Q\rho_1)+\frac{1}{\mathrm{Re}}\Delta'\mathbf{V}+\cdots, \tag{5}$$

$$\frac{dT_1}{dt}+\frac{\kappa}{(\kappa-1)Q}\frac{dp}{dt}=I(x,y,z,t)+\frac{1}{\mathrm{Pe}}\Delta' T_1+\cdots, \tag{6}$$

$$p=\rho T, \tag{7}$$

where

$$\frac{d}{dt}=\frac{\partial}{\partial t}+\mathbf{V},\boldsymbol{\nabla},\quad \Delta'=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\left(\frac{r_0}{L}\right)^2\frac{\partial^2}{\partial z^2},$$

$$\mathrm{Re}=\frac{r_0^2}{\nu\tau},\quad \mathrm{Pe}=\frac{r_0^2}{\chi\tau},\quad \mathrm{Pr}=\frac{\nu}{\chi},$$

$$\mathrm{Fr}=\frac{V_L^2}{Lg\beta T_0},\quad \mathrm{Eu}=\frac{p_0}{\rho_0 V_L^2},$$

$\mathbf{V}$ is the medium velocity, ν, χ, β, and κ are the coefficients of the kinematic viscosity, thermal diffusivity, thermal expansion [for gases $\beta T_0=-(T_0/\rho_0)(\partial\rho/\partial T)\equiv1$], and the adiabatic constant; and g is gravitational acceleration. The density ρ, temperature T, and pressure p are divided by their initial values ρ_0, T_0, and p_0. Let us consider the following situations: in a stagnant medium the heat flow is due to the quasi-stationary thermal conductivity; under a transverse air flow along the x-axis at a constant velocity u_0, $\tau=r_0/u_0$; in the case of steady gravitational convection along the y-axis in a horizontal beam with a characteristic velocity $V_L=v_g=(\alpha P_0 g/\pi\rho_0 C_p T_0)^{1/3}$, $\tau=r_0/v_g$. In the case of gravitational convection, by equating Q to the Froude number Fr we obtain the scales of velocity v_g and density (temperature) perturbations:

$$Q_g=\left(\frac{\alpha P_0}{\pi r_0^2\rho_0 C_p T_0}\right)^{2/3}\left(\frac{r_0}{g}\right)^{1/3}.$$

The scales of gas-dynamic similarity parameters (whether they are smaller or larger than, or of order of unity) are the Euler number Eu (Eu$=1/\kappa M^2$ in gases, where $M=u_0/c$ is the Mach number, $c=\sqrt{\kappa p_0/\rho_0}$ is the speed of sound in the gas), the Reynolds number Re, and the Péclet number Pe depend on the specific situation and determine the gas-dynamic regime of self-action. The Prandtl number is Pr $=$Pe/Re.

3. BASIC IDEAS OF THE APPROXIMATE METHOD

We have a weak self-action at $N\ll1$ in almost parallel and divergent beams. Strong self-action takes place at $N\geqslant1$ and also in tightly focused beams under conditions of a weak or notable nonlinearity of the medium. The approximate description of the weak self-action is based on the following underlying ideas.

1. Let us introduce average transverse dimensions (radii) a_x and a_y, and displacements x_c and y_c of the center of gravity of the intensity distribution, which are functions of the longitudinal coordinate z alone:

$$a_x(z)=\sqrt{\int_{-\infty}^{\infty}\int(x-x_c)^2 q\,dx\,dy},$$

$$x_c(z)=\int_{-\infty}^{\infty}\int xq\,dx\,dy, \tag{8}$$

$$a_y(z)=\sqrt{\int_{-\infty}^{\infty}\int(y-y_c)^2 q\,dx\,dy},$$

$$y_c(z)=\int_{-\infty}^{\infty}\int yq\,dx\,dy, \tag{9}$$

where

$$q(x,y,z)=I(x,y,z)/W, \quad W(z)=\int_{-\infty}^{\infty}\int Idx\,dy.$$

For simplicity the power W is assumed to be constant along the optical path: $N_\alpha \ll 1$. We consider two-dimensional (plane) and three-dimensional beams, in particular, an axially symmetrical one, where the displacement is zero. The average radius of the axially symmetrical beam is given by the formula

$$a(z)=2\pi\sqrt{\int_0^\infty r^2 q(r,z)r\,dr}, \quad W(z)=2\pi\int_0^\infty Irdr. \tag{10}$$

2. It is convenient to use displaced and compressed (extended) transverse coordinates:

$$\xi=\frac{x-x_c(z)}{a_x(z)}, \quad \eta=\frac{y-y_c(z)}{a_y(z)}. \tag{11}$$

3. In the experiments[3,9–11] and numerical calculations for the convective,[3] supersonic,[12] and other gas-dynamic regimes[13–15] perturbations saturate with increasing N and z: the deviation from the initial direction stabilizes, the peak amplitude and form of the transverse intensity distribution $f(\xi,\eta)$ tend to constant characteristics. An increase in the thermal self-action parameter N is equivalent to an increase in the coordinate z. In a uniform gas flow, these characteristics can be maintained approximately constant within certain intervals over long sections of the optical path comparable to its total length L and the thermal self-action length z_T (Ref. 12). A similar quasi-stabilization of perturbations also takes place in the case of gravitational convection due to heating of the medium in a vertical[10,14,15] or horizontal[11,13] laser beam. We assume that the shape of the transverse intensity distribution $f(\xi,\eta)$ is constant along the optical path, and the power density varies due to changes in the mean radii:

$$I(x,y,z)=\frac{f(\xi,\eta)}{a_x(z)a_y(z)} \text{ or } I(r,z)=\frac{f(\xi)}{a^{1+k}(z)},$$

$$\xi=\frac{r-x_c^{1-k}(z)}{a(z)}, \tag{12}$$

where $k=0$ or $k=1$ in a plane or axially symmetrical beam, respectively.

4. Under the conditions of saturation of the thermal self-action effect we approximate in the general case the wave front phase Φ by a second-power polynomial[17]

$$\Phi(x,y,z)=c_0+[x-x_c(z)]c_{1x}+[y-y_c(z)]c_{1y}+(x-x_c)^2\frac{c_{2x}}{2}+(y-y_c)^2\frac{c_{2y}}{2}+(x-x_c)(y-y_c)c_{xy}+\cdots, \tag{13}$$

where

$$c_{1x}(z)=\frac{dx_c}{dz}, \quad c_{1y}(z)=\frac{dy_c}{dz}, \quad c_{2x}(z)=\frac{1}{a_x(z)}\frac{da_x}{dz},$$

$$c_{2y}(z)=\frac{1}{a_y(z)}\frac{da_y}{dz}, \quad c_{xy}=0.$$

For the case of an axially symmetrical beam, we have

$$\Phi(r,z)=c_0+r^2\frac{c_2}{2}+\ldots, \quad c_1=0,$$

and for a plane beam

$$\Phi(x,z)=c_0+[x-x_c(z)]c_1+(x-x_c)^2\frac{c_2}{2}+\ldots$$

We assume that in the three-dimensional configuration there is a symmetry with respect to one coordinate, x or y; in this case the off-diagonal coefficient is zero, $c_{xy}=0$.

5. Applying coordinate transformation (11) to the equations for the medium (4)–(7) and the similarity transformation to the main perturbations of the gas-dynamic parameters like

$$\rho_1(x,y,z)=R(\xi,\eta)A(z), \quad T_1=T(\xi,\eta)B(z),$$
$$p_1=P(\xi,\eta)C(z), \quad u_1(x,y,z)=U(\xi,\eta)D(z), \tag{14}$$
$$v_1=V(\xi,\eta)E(z), \quad w_1=W(\xi,\eta)F(z),$$

we obtain the relations between functions A, B, C, D, E, and F, on the one hand, and average radii, on the other. For functions R, P, T, U, V, and W we formulate an autonomous problem, which does not include z if equations are written for coordinates ξ and η. For example, for the function ρ_1 we obtain formulas like

$$\rho_1(x,y,z)=\frac{R(\xi,\eta)}{a_x^m a_y^n(z)}. \tag{15}$$

6. Let us substitute the approximation of the angle ϑ in the form of polynomial (13) on the left of the eikonal equation (1) and the function ρ_1 in the form given by Eq. (14) or (15) on the right. Let us integrate Eq. (1) with the weight function $q(x,y,z)=I/W$ across the beam. Taking into account the condition of energy conservation (2) with the intensity given by Eq. (12), we obtain an equation for displacements x_c and y_c. Calculating the divergence of each term in Eq. (1) and repeating the integration with the weight function q, we obtain ordinary differential or integral-differential equations for mean radii a_x and a_y [and, in general, for the function $A(z)$]. In the first stage, we omit the correction for diffraction effects, which is of the order of the inverse Fresnel number squared, to avoid complications. Near the focal point, the diffraction term makes the main contribution and should be taken into account.

Below the approximate method for investigating a laser beam self-action will be tested taking as examples various aero-optical problems.

4. QUASI-STEADY THERMAL CONDUCTIVITY

In a stagnant medium where thermal conductivity is the dominant mechanism transmitting perturbations and $\mathrm{Pe}\ll 1$ (and also in the case of a transverse air flow with a small velocity, $V_L<\chi/r_0$), we can set to zero the left-hand side of Eq. (6) in the quasi-steady limit, $t\gg r_0^2/\chi$. Over the charac-

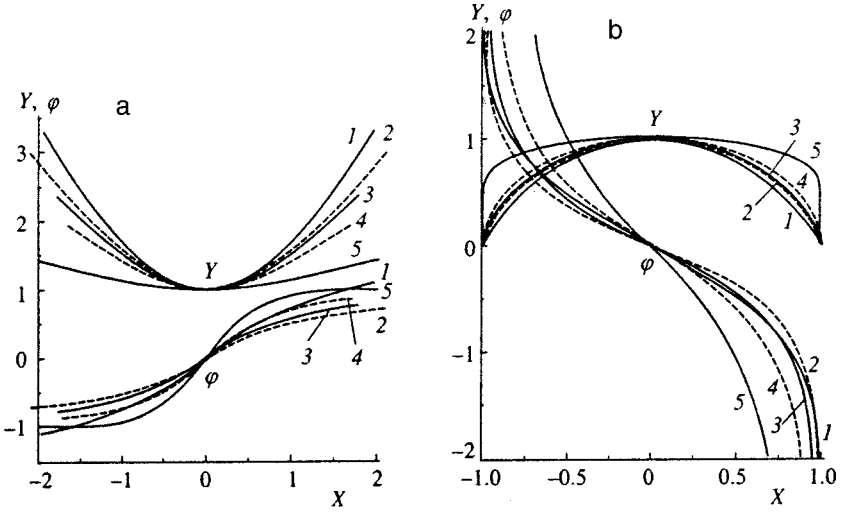

FIG. 1. Normalized universal solution, average radii Y, and expansion angles $\varphi = \text{const}\cdot dY/dX$ as functions of the normalized coordinate X measured along the laser beam in the case of (a) defocusing and (b) self-focusing. The exponent in Eq. (26) is (curves *1*) $m=1$; (*2*) 5/3; (*3*) 2; (*4*) 3; (*5*) 15.

teristic time scale $\tau = r_0^2/\chi$ the Péclet number $\mathrm{Pe}=1$. Since pressure fluctuations are smaller than Q, acoustic perturbations can be ignored, $p \approx \text{const}$. Perturbations of the density and temperature are of the order of Q and, as follows from Eq. (7), have opposite signs: $\rho_1 = -T_1$. The equation (6) of thermal conductivity takes the form

$$\frac{1}{a^2(z)\xi^k}\frac{\partial}{\partial \xi}\xi^k\frac{\partial \rho_1}{\partial \xi} = \frac{f(\xi)}{a^{1+k}(z)}, \quad \rho_1 = a^{1-k}(z)R(\xi), \tag{16}$$

where $k=0$ and 1 for the plane and axially symmetrical configurations, respectively. The procedure described in Sec. 3.6 yields

$$\frac{d^2x_c}{dz^2} = \frac{b_c}{a^k}, \quad x_c(0)=x_{c1}, \quad \frac{dx_c(0)}{dz}=\theta_{c1},$$
$$\frac{d^2a}{dz^2} = \frac{b}{a^k(z)}, \quad a(0)=a_1, \quad \frac{da(0)}{dz}=\theta_1, \tag{17}$$

where

$$b_c = \begin{cases} N\left[\dfrac{dR(+\infty)}{d\xi} - \dfrac{W}{2}\right], & k=0, \\ 0, & k=1, \end{cases}$$
$$b = \frac{N}{W}\int_{\xi_0}^{\infty} f^2(\xi)(\pi\xi)^k d\xi > 0,$$

a_1, x_{c1}, θ_1, and θ_{c1} are the initial values (at $z=z_0=0$) of the average radius, displacement of the center of gravity of the intensity distribution, beam expansion angle, and the rate of the center-of-gravity deviation from the axis, respectively, $\xi_0 = -\infty$, 0 for $k=0$, 1.

In the case of a plane beam ($k=0$) the solution is trivial:

$$a(z) = a_1 + z\theta_1 + \frac{z^2}{2}b, \quad x_c = x_{c1} + z\theta_{c1} + \frac{z^2}{2}b_c. \tag{18}$$

For an axially symmetrical beam ($k=1$), a universal solution of a more general equation

$$\frac{d^2a}{dz^2} = \pm\frac{b}{a(z)}$$

in new independent variables φ, Y, and X,

$$\varphi = \frac{\theta}{\sqrt{2b}}, \quad Y = \frac{a}{a_1}\exp(\pm\varphi_1^2),$$
$$X = \frac{2}{\sqrt{\pi}}\left[\frac{z\exp(\pm\varphi_1^2)}{a_1}\sqrt{\frac{b}{2}} \pm \int_0^{\varphi_1}\exp(\pm t^2)dt\right],$$

can be written in the form

$$Y = \exp(\pm\varphi^2), \quad X = \pm\frac{2}{\sqrt{\pi}}\int_0^{\varphi}\exp(\pm t^2)dt. \tag{19}$$

The normalization conditions for the radius Y and angle φ are selected in such a way that the solution should not include the constants a_1, θ_1, and b, and the focal point be set at $X=\pm 1$. The solution determined by Eq. (19) is shown in Fig. 1 (curves *1*). Note that an absolutely steady regime of thermal conductivity is impossible in both plane and axially symmetrical configurations.

5. AIR FLOW CROSSING A HORIZONTAL LASER BEAM

Let us consider a uniform air flow with a constant velocity $V_L = u_0 \gg \chi/r_0$, ν/r_0 aligned with the x-axis. The viscosity and thermal conductivity can be ignored because Re, $\mathrm{Pe} \gg 1$. The characteristic time is $\tau = r_0/u_0$. Linearization of medium equations (4)–(7) yields the following equation for the principal term in the density perturbation ρ_1:

$$\left[M^2\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)^2 - \nabla_\perp^2\right]\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\rho_1 = \nabla_\perp^2 I. \tag{20}$$

In the steady-state limit $t \gg \tau$ and for a plane beam, this equation and its solution take the form

$$\frac{\partial \rho_1}{\partial \xi} = \frac{f(\xi)}{M^2-1}, \quad \rho_1 = \frac{1}{M^2-1}\int_{-\infty}^{\xi} f(\xi_1)d\xi_1 \equiv R(\xi). \tag{21}$$

For displacement x_c and average radius $a(z)$, we derive the following expressions from the eikonal equation (1):

$$\frac{d^2x_c}{dz^2}=\frac{b_c}{a(z)},\quad b_c=\frac{N}{W(M^2-1)}\int_{-\infty}^{\infty}f^2(\xi)d\xi,$$

$$x_c(0)=x_{c1},\quad \frac{dx_c(0)}{dz}=\theta_{c1},$$

$$\frac{d^2a}{dz^2}=\frac{b}{a(z)},\quad b=0,\quad a(z)=a_1+z\theta_1, \tag{22}$$

$$x_c(z)=x_{c1}+z\theta_{c1}+\frac{b_ca_1}{\theta_1^2}\left\{\left(1+\frac{\theta_1z}{a_1}\right)\times\left[\ln\left(1+\frac{\theta_1z}{a_1}\right)-1\right]+1\right\}.$$

With the initial conditions $x_{c1}=0$ and $\theta_{c1}=0$ in a subsonic flow, the beam shifts downstream with respect to the flow ($b_c<0$), and in a supersonic flow it shifts upstream ($b_c>0$). The heated gas acts as an optical wedge or prism, for it shifts the beam without broadening if it is initially plane-parallel ($\theta_1=0$).

In the case of a three-dimensional beam in the steady-state convection regime, which is realized at the small Mach numbers, $M^2\ll1$, we have the following solutions for the density perturbation:

$$\rho_1=\frac{R(\xi,\eta)}{a_y(z)},\quad R(\xi,\eta)=-\int_{-\infty}^{\xi}f(\xi_1,\eta)d\xi_1,$$

and the corresponding equations for x_c, a_x, and a_y (the variable $y_c=0$):

$$\frac{d^2x_c}{dz^2}=\frac{b_c}{a_xa_y(z)},$$

$$\frac{d^2a_x}{dz^2}=\frac{b_x}{a_xa_y(z)},\quad a_x(z)=a_{x1}+z\theta_{x1}, \tag{23}$$

$$\frac{d^2a_y}{dz^2}=\frac{b_y}{a_y^2(z)},\quad a_y(0)=a_{y1},\quad \frac{da_y(0)}{dz}=\theta_{y1}, \tag{24}$$

where

$$b_c=-\frac{N}{W}\int_{-\infty}^{\infty}\int f^2(\xi,\eta)d\xi\,d\eta,\quad b_x=0,$$

$$b_y=-\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,\eta)\int_{-\infty}^{\xi}\frac{\partial^2f}{\partial\eta^2}d\xi_1d\xi\,d\eta>0.$$

Equation (24) is the free-fall equation (Ref. 18, Sec. 6.188). In the case of zero initial conditions ($x_c=0$, $\theta_c=0$, $\theta_{x1}=0$, and $\theta_{y1}=0$), the beam deviates upstream with respect to the aerodynamic flow and is extended in the transverse direction. A universal solution of Eq. (24) for the average beam thickness a_y in the direction transverse to the flow and for angle θ_y (the case of self-focusing corresponds to the minus sign, $b_y\to\pm|b_y|\equiv\pm b$) is expressed in the new variables $\varphi(X)$ and $Y(X)$:

$$\varphi=\frac{\theta_y}{\sqrt{C_1}},\quad C_1=\left|\theta_{y1}^2\pm\frac{2b}{a_{y1}}\right|,\quad Y=a_y\frac{C_1}{2b},$$

$$X=\frac{2}{\pi}\left[\frac{zC_1^{3/2}}{2b}\pm\frac{\varphi_1}{1\mp\varphi_1^2}+\begin{Bmatrix}\tanh^{-1}\varphi_1\\-\tan^{-1}\varphi_1\end{Bmatrix}\right],\quad \varphi_1=\frac{\theta_{y1}}{\sqrt{C_1}}$$

in the form[16]

$$Y=\frac{1}{1\mp\varphi^2},\quad X=\frac{2}{\pi}\left[\pm\frac{\varphi}{1\mp\varphi^2}+\begin{Bmatrix}\tanh^{-1}\varphi\\-\tan^{-1}\varphi\end{Bmatrix}\right]. \tag{25}$$

In a flow moving with a velocity comparable to the sound velocity, $M>1$ or $M<1$ (the case of a transonic flow will be considered separately), acoustic perturbations of the pressure are essential. A transform like that described by Eq. (14) is possible only under an additional condition $a_y=\text{const}\cdot a_x$, which does not hold, in general, because it is violated as $M^2\to0$ [see Eqs. (23) and (24)]. In specific cases where this condition holds, the solution for $a_x(z)$ is similar to Eq. (25).

6. GENERAL FORM OF THE BASIC EQUATION FOR AVERAGE RADII

The examples discussed above indicate that the average radii (or the maximum one, which is transverse with respect to the flow) are determined by equations like

$$\frac{d^2a}{dz^2}=\pm\frac{b}{a^m(z)},\quad b>0,\quad m>0,$$

$$a(0)=a_1,\quad \theta(0)=\frac{da}{dz}=\theta_1. \tag{26}$$

The exponent m and the form of the similarity transformation of the sought-for functions and independent coordinate measured along the beam that would exclude from the universal solution similarity parameters a_1, θ_1, and b [Eqs. (19) and (25)] depend on the specific gas-dynamic regime. The plus sign corresponds to a defocusing medium ($d\theta/dz>0$, the expansion angle increases), the minus sign corresponds to a focusing medium ($d\theta/dz<0$, the expansion angle decreases). Recall that the initial parameters, namely, the mean radius a_1 and angle θ_1, are divided by the characteristic parameters r_0 and r_0/L. Let $m>1$ (the variants with $m=0$ and 1 were considered in the previous sections). Multiplying Eq. (26) by $2da/dz$, we obtain the first integral

$$\theta^2\equiv\left(\frac{da}{dz}\right)^2=\frac{\pm2b}{(-m+1)a^{m-1}}+C_1,$$

$$C_1=\theta_1^2\mp\frac{2b}{(-m+1)a_1^{m-1}}. \tag{27}$$

From Eqs. (26) and (27) we derive

$$a(z)=\left[\frac{\pm2b}{(m-1)(C_1-\theta^2)}\right]^{1/(m-1)},$$

$$a(z)=\left(\frac{\pm b}{d\theta/dz}\right)^{1/m}. \tag{28}$$

Solutions can be obtained in an explicit form, as will be shown below, only in the specific case of $m=3$ and in the case of large m, $m\gg 1$. Implicit relations can be obtained in the form $z(a)$, $\theta(a)$, but the form $z(\theta)$, $a(\theta)$ is more convenient, in particular, because the sign of derivative $d\theta/dz$ in the first-order equation is unchanged over the entire space $-\infty<z<\infty$ in both focusing and defocusing cases. Using Eq. (28), we derive from Eq. (27) the expression

$$\left[\frac{\pm 2b}{(m-1)(C_1-\theta^2)}\right]^{m/(m-1)} d\theta=\pm b\,dz. \tag{29}$$

In Secs. 4 and 5 we gave solutions of Eq. (26) for $m=0$, 1, and 2 [see Eqs. (18), (19), and (25)]. In the nonlinear optics, the nonlinearity of the third order in the field E (i.e., E^3) is often considered, when the nonlinear correction to the dielectric function of the medium [the function ρ_1 in Eq. (1)] is proportional to the light intensity I (Ref. 19, p. 280). We assume a quasi-steady shape of the transverse intensity distribution and apply the procedure discussed in Sec. 3.6 to the eikonal equation, where we set

$$\rho_1=\text{const}\cdot I(r,z)=\frac{f(\xi)}{a^{1+k}(z)},\quad \xi=\frac{r-x_c^{1-k}(z)}{a(z)}. \tag{30}$$

For a plane beam ($k=0$) the constant $b_c=0$ and the displacement is absent at zero initial conditions ($x_{c1}=0$, $\theta_{c1}=0$). The average radius is determined by Eq. (26) with the exponent $m=2$, and its solution is given by Eq. (25).

In an axially symmetrical beam ($k=1$) the average radius $a(z)$ and angle $\theta(z)$ are described by Eq. (26) at $m=3$. A universal solution in terms of variables φ, Y, and X,

$$\varphi=\frac{\theta}{\sqrt{C_1}},\quad C_1=\left|\theta_1^2\pm\frac{b}{a_1^2}\right|,$$

$$Y=a\sqrt{\frac{C_1}{b}}\quad X=\frac{C_1}{\sqrt{b}}\left[z\pm\frac{a_1\theta_1}{C_1}\right],$$

$$b=-\text{const}\cdot\frac{\pi N}{W}\int_0^\infty \xi[f'(\xi)]^2 d\xi,$$

is derived from Eq. (29) in an explicit form:

$$Y=\frac{1}{\sqrt{1\mp\varphi^2}}=\sqrt{1\pm X^2},\quad \varphi=\frac{\pm X}{\sqrt{1\pm X^2}}. \tag{31}$$

A similar solution was given in Ref. 19, p. 286 only in the beam-axis vicinity and in the case of an approximate description of beams whose wave fronts remained spherical or cylindrical in media with cubic nonlinearity.

In a medium with a fifth-order nonlinearity, we have

$$\rho_1=\text{const}\cdot I^2(r,z)=\frac{f^2(\xi)}{a^{2(1+k)}(z)}. \tag{32}$$

In the case of a plane beam ($k=0$), we have $b_c=0$. There is no displacement at zero initial conditions ($x_{c1}=0$, $\theta_{c1}=0$). For the average radius in Eq. (1), we obtain, following the procedure described in Sec. 3.6, Eq. (26) with the exponent $m=3$, and its solution is given by Eq. (31).

In the case of an axially symmetrical beam ($k=1$), we derive the following expressions from the eikonal equation (1) for the function $a(z)$:

$$\frac{d^2a}{dz^2}=\frac{b}{a^5(z)},\quad b=\text{const}\cdot\frac{\pi N}{W}\int_0^\infty f\frac{d}{d\xi}\left(\xi\frac{d}{d\xi}f^2\right)d\xi. \tag{33}$$

In the case of a focusing medium ($b<0$) the universal solution in variables φ, Y, and X,

$$\varphi=\frac{\theta}{\sqrt{C_1}},\quad C_1=\left|\theta_1^2-\frac{b}{2a_1^4}\right|,\quad Y=a\left(\frac{2C_1}{b}\right)^{1/4},$$

$$X=\frac{\sqrt{2}}{\beta}\left[\left(\frac{2^3C_1^3}{b}\right)^{1/4}z+F\left(\alpha_1,\frac{1}{\sqrt{2}}\right)-2E\left(\alpha_1,\frac{1}{\sqrt{2}}\right)\right],$$

has the form (we assume $b=|b|$)

$$Y=\frac{1}{(1+\varphi^2)^{1/4}},\quad X=\frac{\sqrt{2}}{\beta}\left[F\left(\alpha,\frac{1}{\sqrt{2}}\right)-2E\left(\alpha,\frac{1}{\sqrt{2}}\right)\right], \tag{34}$$

where

$$\alpha=\cos^{-1}\frac{1}{(1+\varphi^2)^{1/4}},\quad \alpha_1=\cos^{-1}\frac{1}{(1+\varphi_1^2)^{1/4}},$$

$$\beta=\sqrt{2}\left[2E\left(\frac{\pi}{2},\frac{1}{\sqrt{2}}\right)-F\left(\frac{\pi}{2},\frac{1}{\sqrt{2}}\right)\right]\simeq 1.198,$$

$F(\alpha,k)$ and $E(\alpha,k)$ are the elliptic integrals of the first and second kind (Ref. 20, Sec. 1.2.79).

7. TRANSONIC STEADY-STATE REGIME

Fundamental differences between the regime, which is nonlinear from the viewpoint of gas dynamics, and the neighboring subsonic and supersonic regimes[21] appear at small values of the transonic similarity parameter $K=(1-M^2)/Q^{2/3}$ such that $|K|\ll 0.29$, as follows from an exact numerical solution.[22] This fact is indirectly confirmed by experiments[23] conducted in a transonic wind tunnel filled with a mixture of air and hexafluoric sulphur SF_6 with an absorption coefficient $\alpha=20\ \text{m}^{-1}$. The beam generated by a CO_2 laser ($\lambda=10.6\ \mu$m) had a power $P_0=500$ W and a diameter $2r_0=1.5$ mm. The parameter $|K|>2$. No shock waves were detected.

The proposed method of average radii allows us to obtain a universal solution in the case of $|K|\ll 1$. Let us consider a three-dimensional beam symmetrical in the coordinate y. In a transonic flow, acoustic perturbations propagating upstream in the flow are driven by the flow at a velocity close to that of sound. Thus perturbations are accumulated in the region of radiation, the region occupied by the perturbations expands in the transverse direction by a factor $Q^{-1/3}$, and the characteristics of perturbations in gas-dynamic parameters, in particular, the density, increase to

$Q^{2/3}$ (Ref. 22). The problem is separated into the internal problem in the region with dimensions $x, y \sim r_0$ and the external one in a wider region $x \sim r_0$, $\tilde{y} \sim yQ^{1/3} \sim 1$. Then their solution are matched. Within the radius r_0 the density perturbations are independent of the transverse coordinate y. The beam is extended and displaced owing to the thermal self-action only along the x-axis directed along the flow. The radiation intensity in the regime of a quasi-steady distribution is described by the equation

$$I(x,y,z)=\frac{f(\xi,y)}{a(z)}, \quad \xi=\frac{x-x_c(z)}{a(z)}.$$

Let us determine the function which describes the main component of the density perturbation $\rho_1=R(\xi)/a^i(z)$, $i=\text{const}$. For the internal problem, we have the following expansions for calculated parameters:

$$\frac{\rho}{\rho_0}=1+Q^{2/3}\rho_1+Q\rho_2+\cdots, \quad \frac{p}{p_0}=1+Q^{2/3}p_1+Qp_2+\cdots,$$

$$\frac{u}{u_0}=1+Q^{2/3}u_1+Qu_2+\cdots, \quad \frac{v}{u_0}=Qv_1+Q^{4/3}v_2+\cdots.$$

From Eqs. (4)–(7) we derive in the first-order approximation of the perturbation theory

$$\begin{gathered}\frac{\partial}{\partial\xi}(\rho_1+u_1)=0, \quad \frac{\partial}{\partial\xi}(\kappa u_1+p_1)=0,\\ \frac{\partial}{\partial\xi}(p_1-\kappa\rho_1)=0, \quad \frac{\partial p_1}{\partial y}=0,\\ \rho_1=-u_1=p_1(\xi,z)/\kappa.\end{gathered} \tag{35}$$

In the second-order approximation the equations have the form

$$\begin{gathered}\frac{\partial}{\partial\xi}(\rho_2+u_2)=-a(z)\frac{\partial v_1}{\partial y}, \quad \frac{\partial}{\partial\xi}\left(u_2+\frac{p_2}{\kappa}\right)=0,\\ \frac{\kappa}{a(z)}\frac{\partial v_1}{\partial\xi}+\frac{\partial p_2}{\partial y}=0, \quad \frac{\partial}{\partial\xi}\left(\frac{p_2}{\kappa}-\rho_2\right)=f(\xi,y),\end{gathered}$$

and their solutions are

$$\begin{gathered}v_1(x,y,z)=\frac{1}{a(z)}\int_0^y f(\xi,y')dy',\\ p_2=-\kappa u_2=-\frac{\kappa}{a(z)}\int_0^y\frac{\partial v_1(\xi,y')}{\partial\xi}dy',\\ \rho_2=\frac{p_2}{\kappa}-\int_{-\infty}^{\xi}f(\xi',y)d\xi'.\end{gathered} \tag{36}$$

The function p_1 is independent of y and determined by the solution of the external problem, like the functions ρ_1 and u_1 related to the former by Eq. (35). The second-order approximation is needed to close the problem in the first-order approximation.

Let us introduce for the external problem a compressed coordinate $\tilde{y}=yQ^{1/3}a^j(z)$ such that $\tilde{y}\sim 1$ for $y\gg 1$, where j is a certain number. The expansions of the sought-for functions in powers of the small parameter up to the third order have the form

$$\begin{gathered}\frac{\rho}{\rho_0}=1+Q^{2/3}\rho_1+Q\rho_2+Q^{4/3}\rho_3\ldots,\\ \frac{p}{p_0}=1+Q^{2/3}p_1+Qp_2+Q^{4/3}p_3\ldots,\\ \frac{u}{u_0}=1+Q^{2/3}u_1+Qu_2+Q^{4/3}u_3\ldots,\\ \frac{v}{u_0}=Q\frac{V_1(\xi,\tilde{y})}{a(z)}+Q^{4/3}v_2+Q^{5/3}v_3\ldots\end{gathered} \tag{37}$$

We have taken into account that the function V_1 for $\tilde{y}\to 0$ should match the solution v_1 of the internal problem as $y\to\infty$. Substitution of expansions (37) in the initial gasdynamic equations (4)–(7) leads to the following system of equations (up to the third order), which are necessary for closing the problem in the lowest-order approximation: in the first-order approximation

$$\rho_1=-u_1=\frac{p_1(\xi,\tilde{y},z)}{\kappa}, \quad \frac{\partial V_1}{\partial\xi}+a^{j+2}(z)\frac{\partial\rho_1}{\partial\tilde{y}}=0, \tag{38}$$

in the second-order approximation

$$\rho_2=-u_2=\frac{p_2(\xi,\tilde{y},z)}{\kappa}, \quad \frac{\partial v_2}{\partial\xi}+a^{j+2}(z)\frac{\partial\rho_2}{\partial\tilde{y}}=0 \tag{39}$$

in the third-order approximation

$$\begin{gathered}\frac{\partial}{\partial\xi}(\rho_3+u_3)=2\rho_1\frac{\partial\rho_1}{\partial\xi}-a^j(z)\frac{\partial V_1}{\partial\tilde{y}},\\ \frac{\partial}{\partial\xi}(\kappa u_3+p_3)=0, \quad \frac{\partial}{\partial\xi}\left(\frac{p_3}{\kappa}-\rho_3\right)=(\kappa-1)\rho_1\frac{\partial\rho_1}{\partial\xi},\\ \kappa\frac{\partial v_3}{\partial\xi}+a^{j+1}(z)\frac{\partial p_3}{\partial\tilde{y}}=0.\end{gathered}$$

Here we have taken into account that the transonic similarity parameter K is zero. From the first three equation we derive

$$(\kappa+1)\rho_1\frac{\partial\rho_1}{\partial\xi}-a^j(z)\frac{\partial V_1}{\partial\tilde{y}}=0. \tag{40}$$

Equations (40) and (38) close the problem for the functions ρ_1 and V_1 in the lowest-order approximation. The similarity transformation

$$\begin{gathered}\rho_1(x,y,z)=\frac{R_1(\xi,\eta)}{(\kappa+1)^{1/3}a^{2/3}},\\ \eta=(\kappa+1)^{1/3}\tilde{y}\equiv[(\kappa+1)Q]^{1/3}\frac{y}{a^{4/3}}, \quad j=-\frac{4}{3}\end{gathered} \tag{41}$$

reduces the problem in the first-order approximation to the universal form which is independent of the similarity numbers and coordinate z along the beam axis:

$$R_1\frac{\partial R_1}{\partial \xi}-\frac{\partial V_1}{\partial \eta}=0,\quad R_1,V_1\big|_{\xi\to-\infty;\,\eta\to+\infty}\to 0,$$

$$\frac{\partial R_1}{\partial \eta}+\frac{\partial V_1}{\partial \xi}=0,\quad V_1(\xi,0)=\int_0^\infty f(\xi,y)\,dy. \tag{42}$$

The gas-dynamic problem is reduced to a nonlinear second-order equation like the Tricomi equation.[24] There are numerical methods for solving this problem.[22]

Thus, we have obtained a result which is important for the optical part of the problem of the laser beam self-action: the density perturbation function within the transverse dimension r_0, according to Eq. (41), has the form

$$\rho_1(x,z)=\frac{R(\xi)}{a^{2/3}(z)},\quad R(\xi)=\frac{R_1(\xi,\eta=0)}{(\kappa+1)^{1/3}}. \tag{43}$$

An initially collimated beam suffers neither displacement nor expansion in the direction perpendicular to the gas flow. From the eikonal equation (1) we derive the following equations for the displacement $x_c(z)$ and average radius $a(z)$ in the direction along the gas flow:

$$\frac{d^2x_c}{dz^2}=\frac{b_c}{a^{5/3}(z)},\quad x_c(0)=x_{c1},\quad \frac{dx_c(0)}{dz}=\theta_{c1},$$

$$\frac{d^2a}{dz^2}=\frac{b}{a^{5/3}(z)},\quad a(0)=a_1,\quad \frac{da(0)}{dz}=\theta_1, \tag{44}$$

where

$$b_c=\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,y)\frac{dR(\xi)}{d\xi}\,d\xi\,dy,$$

$$b=\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,y)\frac{d^2R(\xi)}{d\xi^2}\,d\xi\,dy.$$

The similarity transformation and integration of Eq. (29) at $m=5/3$ in variables φ, Y, X and $b\to\pm|b|\to\pm b$,

$$\varphi=\frac{\theta}{\sqrt{C_1}},\quad C_1=\left|\theta_1^2\pm\frac{3b}{a_1^{2/3}}\right|,\quad Y=a\left(\frac{C_1}{3b}\right)^{3/2},$$

$$X=\frac{zC_1^2}{2(3b)^{3/2}}\pm\frac{\varphi_1(3\mp2\varphi_1^2)}{2(1\mp\varphi_1^2)^{3/2}},\quad \varphi_1=\frac{\theta_1}{\sqrt{C_1}},$$

yield the following universal solution:

$$Y=\frac{1}{(1\mp\varphi^2)^{3/2}},\quad X=\pm\frac{\varphi(3\mp2\varphi^2)}{2(1\mp\varphi^2)^{3/2}}. \tag{45}$$

The displacement x_c can be easily calculated by substituting on the right-hand side of Eq. (44) the expression $(b_c/b)d^2a/dz^2$ and integrating this equation twice with the initial conditions:

$$x_c(z)=\frac{b_c}{b}[a(z)-a_1-\theta_1 z]+x_{c1}+\theta_{c1}z. \tag{46}$$

The differences between the displacement and increase in the average radius are the linear function and the constant factor b_c/b. This result also applies to other regimes with $b_c\neq0$ and $b\neq0$.

In the region around the speed of sound, where the Mach number satisfies the condition $Q^{2/3}\ll|M^2-1|\ll1$, the density perturbation is described by the linearized equation (20), where the gradients in the direction perpendicular to the flow can be ignored, i.e., Eq. (21) and its solution are valid. The constants $b_x=b_y=0$. The average radii follow linear functions:

$$a_x(z)=a_{x1}+z\theta_{x1},\quad a_y(z)=a_{y1}+z\theta_{y1}.$$

The displacement x_c [recall that $y_c=0$ by virtue of the symmetry relation $I(x,y,z)=I(x,-y,z)$] is determined, in general, by the equation

$$\frac{d^2x_c}{dz^2}=\frac{b_c}{(a_{x1}+z\theta_{x1})(a_{y1}+z\theta_{y1})},$$

$$b_c=-\frac{N}{(M^2-1)W}\int_{-\infty}^{\infty}\int f^2(\xi,\eta)\,d\xi\,d\eta,$$

whose solution can be easily obtained in the explicit form. In the specific case of $\theta_{y1}=0$, the displacement x_c is described by the corresponding formula in Eq. (22). In practice, $b_x\neq0$, since the assumption that the form of the intensity distribution is constant is not absolutely accurate. The parameter a_x in this case is given by Eq. (25).

8. DEVELOPED STEADY-STATE GRAVITATIONAL CONVECTION IN A HORIZONTAL LASER BEAM

The characteristic velocity V_L in this case is that of gravitational convection:[13,25,26] $V_L=v_g=(\alpha P_0 g/\pi\rho_0 C_p T_0)^{1/3}$, and the characteristic transient time is $\tau=r_0/v_g$. The viscosity and thermal conductivity are negligible, Pe, Re $\gg1$. We assume that the beam is symmetrical relative to the vertical y-axis: $I(x,y,z)=I(-x,y,z)$. The linearized medium equations (4)–(7) have the form

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0,\quad u,v\big|_{x\to\pm\infty}\to0,\quad v\big|_{y\to-\infty}\to0,$$

$$u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+\frac{\partial p_1}{\partial y}=-\rho_1,\quad p_1,\rho_1\big|_{x\to\pm\infty}\to0, \tag{47}$$

$$u\frac{\partial \rho_1}{\partial x}+v\frac{\partial \rho_1}{\partial y}=-\frac{f(\xi,\eta)}{a_x(z)a_y(z)}.$$

The similarity transformation

$$u=\left(\frac{a_x}{a_y}\right)^{2/3}U(\xi,\eta),\quad v=\left(\frac{a_y}{a_x}\right)^{1/3}V(\xi,\eta),$$

$$\rho_1=\frac{R(\xi,\eta)}{a_x^{2/3}a_y^{1/3}},\quad p_1=\left(\frac{a_y}{a_x}\right)^{2/3}P(\xi,\eta)$$

reduces the system of equations to a form which does not include the coordinate z:

$$\frac{\partial U}{\partial \xi}+\frac{\partial V}{\partial \eta}=0, \quad U\frac{\partial V}{\partial \xi}+V\frac{\partial V}{\partial \eta}+\frac{\partial P}{\partial \eta}=-R,$$

$$U\frac{\partial R}{\partial \xi}+V\frac{\partial R}{\partial \eta}=-f(\xi,\eta).$$

From the optical equations we derive the following system of equations for the average radii and displacements $y_c(z)$:

$$\frac{d^2a_x}{dz^2}=\frac{b_x}{a_x^{5/3}a_y^{1/3}(z)}, \quad a_x(0)=a_{x1}, \quad \frac{da_x(0)}{dz}=\theta_{x1},$$

$$\frac{d^2a_y}{dz^2}=\frac{b_y}{a_x^{2/3}a_y^{4/3}(z)}, \quad a_y(0)=a_{y1}, \quad \frac{da_y(0)}{dz}=\theta_{y1}, \tag{48}$$

$$\frac{d^2y_c}{dz^2}=\frac{b_{cy}}{a_x^{2/3}a_y^{4/3}(z)}, \quad y_c(0)=y_{c1}, \quad \frac{dy_c(0)}{dz}=\theta_{yc1},$$

where

$$b_x=\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,\eta)\frac{\partial^2 R}{\partial \xi^2}d\xi\, d\eta,$$

$$b_y=\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,\eta)\frac{\partial^2 R}{\partial \eta^2}d\xi\, d\eta,$$

$$b_{cy}=\frac{N}{W}\int_{-\infty}^{\infty}\int f(\xi,\eta)\frac{\partial R}{\partial \eta}d\xi\, d\eta.$$

For an estimate, it is usually sufficient to calculate the larger transverse beam radius $a=a_x$. Let us assume that $a_y=\text{const}\cdot a_x$. Then the first line in Eq. (48) reduces to an equation like (29) with the exponent $m=2$. The solution for the average radius is given by Eq. (25), the displacement $y_c(z)$ is determined by Eq. (46) with $a=a_x$ and an unknown numerical factor:

$$y_c(z)=\text{const}\cdot\frac{b_{cy}}{b_y}[a_x(z)-a_{x1}-\theta_{x1}z]+y_{c1}+\theta_{yc1}z. \tag{49}$$

Thus, our technique allows us to estimate perturbations in a laser beam without solving hydrodynamic equations. The constants b_x, b_y, b_{cx}, b_{cy}, etc. proportional to the self-action parameter N can be determined by comparing approximate solutions to experimental data or more accurate numerical calculations obtained previously for typical situations, as was done, following a known algorithm,[13] for model optical configurations in the regime of purely convective transverse gas flow[17] and for the regime of gravitational convection in a horizontal beam.[16] For an annular beam with an outside diameter a factor of 2.86 larger than the inside diameter, the calculations yield $b_y=0.540N$, $b_{cx}=-0.470N$ in the case of a flow along the x-axis; $b_x=0.354N$, $b_y=0.159N$, and $b_{cy}=-0.195N$ in the case of gravitational convection along the vertical y-axis. For a Gaussian focused beam in conditions of convective gas flow we have $b_y=0.354N$ and $b_{cy}=-0.289N$. In other examples with more complicated optical configurations, the numerical factor relating the similarity parameters b_x, b_y, and b_{cy} of the approximate method to the self-action similarity number N was in accurate solutions no greater than unity for the larger average radius and twice as small for the displacement. The system of ordinary differential equations can be easily solved for arbitrary exponents in power functions of average radii a_x and a_y. If there is a universal analytical solution, the estimate of perturbations is, naturally, simpler. A complicated optical configuration can be divided into sections, and perturbations imposed on different analytical solutions can be estimated in each section for different heat-transfer mechanisms.

9. LIMITING CASE OF LARGE m

Equations (28) and (29) in variables φ, Y, and X,

$$\varphi=\frac{\theta}{\sqrt{C_1}}, \quad C_1=\left|\theta_1^2\mp\frac{2b}{(-m+1)a_1^{m-1}}\right|,$$

$$Y=a\left(\frac{m-1}{2b}C_1\right)^{1/(m-1)},$$

$$X=\frac{\pi}{2}\left[z\left(\frac{m-1}{2}\right)^{m/(m-1)}\left(\frac{C_1}{b}\right)^{1/(m-1)}\sqrt{C_1}+\begin{cases}\tanh^{-1}\varphi_1\\ -\tan^{-1}\varphi_1\end{cases}\right],$$

have an obvious asymptotic analytical solution for $m\gg 1$:

$$Y=\frac{1}{(1\mp\varphi^2)^{1/(m-1)}}\approx 1-\frac{1}{m}\ln(1\mp\varphi^2)+O\left(\frac{1}{m^2}\right),$$

$$\varphi=\begin{cases}\tanh(\pi X/2)\\ \tan(-\pi X/2)\end{cases}+O\left(\frac{1}{m}\right), \tag{50}$$

$$X=\frac{2}{\pi}\begin{cases}\tanh^{-1}\varphi\\ -\tan^{-1}\varphi\end{cases}+O\left(\frac{1}{m}\right).$$

Compare solutions (19), (45), (25), (31), (34), and (50) obtained at $m=1$, 5/3, 2, 3, 5, and $m\gg 1$, respectively. Table I lists similarity transformations, equations, and their solutions for all cases discussed above. Figure 1 shows curves for $m=1$, 5/3, 2, 3, and 15 in defocusing and focusing media. The rate of the average radius in defocusing media and over the larger sections of optical paths in focusing media, except sections near the focal points, increases with decreasing exponent m in the second-order equation (26) for the average radius starting with $m=1$. The universal solutions obtained for all m cover the whole range of various situations in the initial statement of the problem: $0\leqslant b<\infty$, $-\infty<\theta_1<\infty$, $0<a_1<\infty$. Setting $z=0$ (and φ_1) and then calculating $\varphi_2=\varphi(z=1)$, we find appropriate sections of universal curves, which describe the beam perturbations plotted in physical variables. This section can contain no waists of the beam or

TABLE I. Universal solutions of Eq. (26) describing focusing and defocusing of a laser beam in a nonlinear medium in the aberration-free approximation of geometrical optics: average transverse dimension (radius) $a=a_{\rm phys}/r_0$ and corresponding expansion angle $\theta=\theta_{\rm phys}/\theta_0\equiv da/dz$ versus the coordinate measured along the beam axis, $z=z_{\rm phys}/L$.

m	Transformations	Equations	Solutions	Heat transfer mechansims; plane beam $D=2$; three-dimensional beam $D=3$
1	$\varphi=\dfrac{\theta}{\sqrt{2b}}$ $Y=\dfrac{a}{a_1}\exp(\pm\varphi_1^2)$ $X=\dfrac{2}{\sqrt{\pi}}\left[\dfrac{z\exp(\pm\varphi_1^2)}{a_1}\times\sqrt{\dfrac{b}{2}}\pm\int_0^{\varphi_1}\exp(\pm t^2)dt\right]$	$\dfrac{dY}{dX}=\sqrt{\pi}\varphi$ $\dfrac{d\varphi}{dX}=\pm\dfrac{\sqrt{\pi}}{2}\exp(\mp\varphi^2)$	$Y=\exp(\pm\varphi^2)$ $X=\pm\dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^{\varphi}\exp(\pm t^2)dt$	thermal conductivity $D=3$ $V_L<\chi/r_0$, $t\gg r_0^2/\chi$; transverse flow, $D=2$ convection $M^2\ll1$; subsonic $M<1$; supersonic $M>1$
5/3	$\varphi=\dfrac{\theta}{\sqrt{C_1}}$ $C_1=\left\vert\theta_1^2\pm\dfrac{3b}{a_1^{2/3}}\right\vert$ $Y=a[C_1/3b]^{3/2}$ $X=\dfrac{zC_1^2}{2(3b)^{3/2}}\pm\dfrac{\varphi_1(3\mp2\varphi_1^2)}{2(1\mp\varphi_1^2)^{3/2}}$	$\dfrac{dY}{dX}=2\varphi$ $\dfrac{d\varphi}{dX}=\pm\dfrac{2}{3}(1\mp\varphi^2)^{5/2}$	$Y=\dfrac{1}{(1\mp\varphi^2)^{3/2}}$ $X=\pm\dfrac{\varphi(3\mp2\varphi^2)}{2(1\mp\varphi^2)^{3/2}}$	transverse transonic flow; $D=3$; $\vert1-M^2\vert\ll Q^{2/3}$
2	$\varphi\dfrac{\theta}{\sqrt{C_1}}$ $C_1=\left\vert\theta_1^2\pm\dfrac{2b}{a_1}\right\vert$ $Y=\dfrac{aC_1}{2b}$ $X=\dfrac{2}{\pi}\left[\dfrac{zC_1^{3/2}}{2b}\pm\dfrac{\varphi_1}{1\mp\varphi_1^2}+\begin{Bmatrix}\tanh^{-1}\varphi_1\\-\tan^{-1}\varphi_1\end{Bmatrix}\right]$	$\dfrac{dY}{dX}=\dfrac{\pi}{2}\varphi$ $\dfrac{d\varphi}{dX}=\pm\dfrac{\pi}{4}(1\mp\varphi^2)^2$	$Y=\dfrac{1}{1\mp\varphi^2}$ $X=\dfrac{2}{\pi}\left[\dfrac{\pm\varphi}{1\mp\varphi^2}+\begin{Bmatrix}\tanh^{-1}\varphi\\-\tan^{-1}\varphi\end{Bmatrix}\right]$	transverse convetion flow, $D=3$ $M^2\ll1$, subsonic supersonic $Q^{2/3}\ll\vert1-M^2\vert\ll1$; gravitational convection in horizontal beam, $D=3$
3	$\varphi=\dfrac{\theta}{\sqrt{C_1}}$ $C_1=\left\vert\theta_1^2\pm\dfrac{b}{a_1^2}\right\vert$ $Y=a\sqrt{\dfrac{C_1}{b}}$ $X=z\dfrac{C_1}{\sqrt{b}}\pm\dfrac{a_1\theta_1}{\sqrt{b}}$	$\dfrac{dY}{dX}=\varphi$ $\dfrac{d\varphi}{dX}=\pm(1\mp\varphi^2)^{3/2}$	$Y=\dfrac{1}{\sqrt{1\mp\varphi^2}}\equiv\sqrt{1\pm X^2}$ $X=\dfrac{\pm\varphi}{\sqrt{1\mp\varphi^2}}$ $\varphi=\dfrac{\pm X}{\sqrt{1\pm X^2}}$	cubic nonlinearity $D=3$; nonlinearity of fifth power, $D=2$

Continuation of Table I.

m	Transformations	Equations	Solutions
		focusing medium (minus sign)	
	$\varphi=\theta/\sqrt{C_1}$ $C_1=b/2a_1^4-\theta_1^2$ $Y=a(2C_1/b)^{1/4}$		$Y=(1+\varphi^2)^{-1/4}$
5	$X=\frac{\sqrt{2}}{\beta}\left[\left(\frac{2^3C_1^3}{b}\right)^{1/4}z+F\left(\alpha_1\frac{1}{\sqrt{2}}\right)\right.$ $\left.-2E\left(\alpha_1,\frac{1}{\sqrt{2}}\right)\right]$	$\frac{dY}{dX}=\frac{\beta}{2}\varphi$ $\frac{d\varphi}{dX}=-\beta(1+\varphi^2)^{5/4}$	$X=\frac{\sqrt{2}}{\beta}\left[F\left(\alpha,\frac{1}{\sqrt{2}}\right)\right.$ $\left.-2E\left(\alpha,\frac{1}{\sqrt{2}}\right)\right]$
	$\beta=1.198$		$\alpha=\cos^{-1}(1=\varphi^2)^{-1/4}$
$m\gg1$	$\varphi=\frac{\theta}{\sqrt{C_1}}$ $Y=a\left(\frac{m-1}{2b}C_1\right)^{1/(m-1)}$	$\frac{dY}{dX}=\frac{\pi}{m-1}\varphi$ $\frac{d\varphi}{dX}=\pm\frac{\pi}{2}(1\mp\varphi^2)^{m/(m-1)}$	$Y=(1\mp\varphi^2)^{-1/(m-1)}$ $\simeq1-\frac{\ln(1\mp\varphi^2)}{m}+O\left(\frac{1}{m^2}\right)$ $\varphi=\left\{\begin{matrix}\tanh(\pi X/2)\\ \tan(-\pi X/2)\end{matrix}\right\}+O\left(\frac{1}{m}\right)$
	$C_1=\left\vert\theta_1^2\mp\frac{2b}{(m-1)a_1^{m-1}}\right\vert$ $X=\frac{\tau}{2}\left[z\left(\frac{m-1}{1}\right)^{m/(m-1)}\left(\frac{C_1}{b}\right)^{1/(m-1)}\sqrt{C_1}+\left\{\begin{matrix}\tanh^{-1}\varphi_1\\ \tan^{-1}\varphi_1\end{matrix}\right\}\right]$		$X=\frac{2}{\pi}\left\{\begin{matrix}\tanh^{-1}\varphi\\ -\tan^{-1}\varphi\end{matrix}\right\}+O\left(\frac{1}{m}\right)$

focal points (in situations with defocusing). A beam can be compressed or extended in both defocusing and focusing media, depending on the initial conditions.

10. COMPARISON BETWEEN APPROXIMATE AND EXACT SOLUTIONS

1. In order to illustrate the performance of the suggested technique, we have obtained an exact numerical solution for the case of a plane collimated Gaussian beam in uniform transverse subsonic and supersonic flows ($m=1$; $D=2$) with the initial conditions

$$I(z=0,x)=\exp(-x^2),\quad \Phi(z=0)=\text{const},\quad \vartheta(z=0)=0.$$

The similarity numbers were selected as follows: the Fresnel number $F\gg1$, absorption number $N_\alpha=0$, self-action number $N=1$; Mach number $M=0-85$; $1.1-2.0$. The numbers of meshpoints are $N_x=256$, 512, and 1024, $N_z=200-800$, the step widths $\Delta x=0.01-0.05$ and $\Delta z=0.0025-0.01$; the dimensions of the studied region are $x_{\text{phys}}=(10-20)r_0$, $z_{\text{phys}}=2z_T$. The results are shown in Figs. 2–4. The profile of a perturbed beam is approximately constant over the section $z/L=1.5-2$. Further calculations are difficult because of large local gradients around the intensity peak. In Fig. 2a the average transverse beam dimension a calculated using the approximate method [see solution (22)] is compared to the exact numerical solution. The error is within 7%. The beam displacement x_c calculated by the approximate method is very close to the exact calculation (Fig. 2b). The intensity profile shape is approximately constant only around the intensity peak, as one can see in Fig. 3. Therefore, the constant

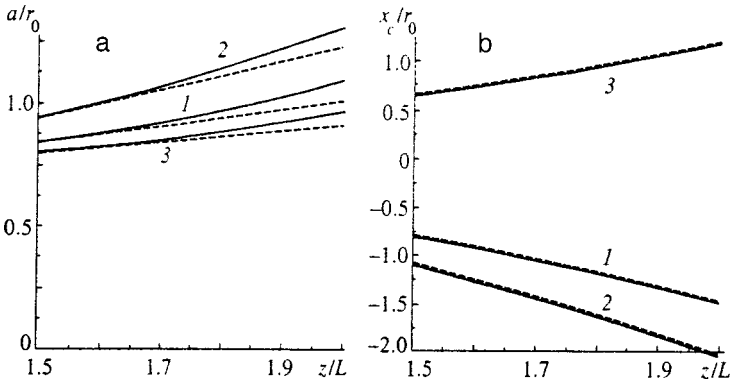

FIG. 2. Comparison between approximate calculations of (a) average beam dimensions a/r_0 and (b) center-of-gravity shifts x_c/r_0 of the intensity distribution (dashed lines) and exact values (solid lines) for the Gaussian beam under a transverse gas flow: (*1*) $M=0.01$; (*2*) 0.5; (*3*) 1.5. The similarity numbers $F\gg1$, $N_\alpha=0$, $N=1$.

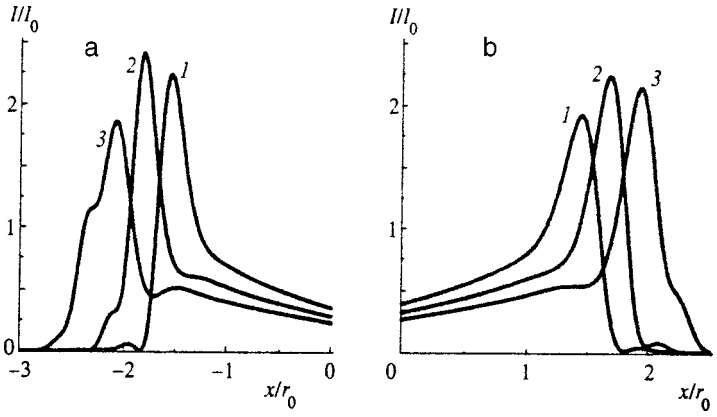

FIG. 3. Profiles of exact numerical solutions for the Gaussian beam propagation under a transverse gas flow: (a) $M=0.1$; (*1*) $z/L=1.5$; (*2*) 1.7; (*3*) 1.9; (b) $M=1.5$; (*1*) $z/L=1/6$; (*2*) 1.8; (*3*) 2.0.

b in solution (22) is in fact nonzero, and in the exact solution the average transverse dimension $a(z)$ approximately follows a linear function. Figure 4a shows the constant b_c versus the Mach number. In the vicinity of the point $M=1$ neither the exact nor approximate solution applies. Figure 4b shows the initial values $a_1(z_0/L=1.5)$, $x_{c1}(z_0/L=1.5)$ as functions of M. The initial slopes of the curves $a(z)$ and $x_c(z)$ are $\theta_1/(r_0/L)=0.337$, 0.344, 0.405, 0.572, 0.956, 1.735, 2.162, 1.123, 0.2175, and 0.0465; $\theta_{c1}/(r_0/L)=-1.135$, -1.149, -1.280, -1.593, -2.370, -3.957, 4.854, 2.711, 0.8813, and 0.3544 at $M=0.01$, 0.1, 0.3, 0.5, 0.7, 0.85, 1.1, 1.2, 1.5, and 2.0, respectively.

2. As a second example, we have calculated an exact numerical solution for an axially symmetrical beam in a quasi-steady-state heat-conducting regime ($m=1$; $D=3$) and compared it with the approximate solution (19).

Figure 5a shows intensity profiles $I(r)/I_0$ at the moment $t=7\tau$ in different cross sections. The intensity profiles are similar. The relative change in the average radius $a(z)$ over the time interval τ is 2.47, 1.42, 0.90, 0.63, and 0.47% at moments $t=3\tau$, 4τ, 5τ, 6τ, and 7τ, respectively. This indicates the existence of a quasi-steady state.

Figure 5b shows the mean radius $a(z)/r_0$, deviation angle $\theta/(r_0/L)$, and ''constant'' b as functions of z along the optical path. Note a slight change in b (as was expected, on the basis of our assumption) and excellent agreement between approximate calculations of the average radius and deviation angle, on the one hand, and accurate calculations in the section $z/L=1.5-3.0$, on the other.

11. SOLUTIONS WITH FOCAL POINTS

In a focusing medium, transverse dimensions of a beam can be reduced by an order of magnitude (Fig. 1b). The minimal dimension at a focal point, a_f, is limited by a diffraction and is of the order of the inverse Fresnel number, $a_f/r_0 \sim F^{-n}$, as follows from the eikonal equation (1). In the previous sections, we have disregarded the wave effects, thus admitting an error of the order of F^{-2}. In most aero-optical problems, the equation for the average transverse beam dimensions, which takes into account the diffraction, can be reduced to a general ordinary differential equation

$$\frac{d^2a}{dz^2}=\frac{1}{a(z)}\left[\frac{\pm b}{a^{m-1}(z)}+\frac{c}{F^2a^2(z)}\right], \quad b,c>0, \quad m>0,$$
$$a(0)=a_1, \quad \theta(0)=\theta_1. \tag{51}$$

This equation is valid at all the Fresnel numbers F, just as the initial equation (1). The first integral of Eq. (51) for $m>1$ is

$$\theta^2\equiv\left(\frac{da}{dz}\right)^2=\frac{\pm 2b}{(-m+1)a^{m-1}}-\frac{c}{F^2a^2}+C_1,$$
$$C_1=\theta_1^2\mp\frac{2b}{(-m+1)a_1^{m-1}}+\frac{c}{F^2a_1^2}. \tag{52}$$

It is obvious that at $m\geqslant 3$ diffraction cannot compensate for self-focusing since the second term on the right side cannot grow faster than the first as $a\to 0$.

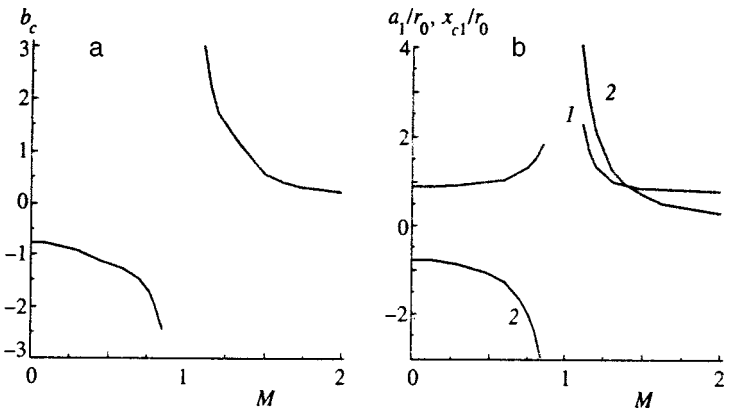

FIG. 4. (a) Parameter b_c and (b, curve *1*) initial average dimensions a_1/r_0 and (b, curve *2*) displacements x_{c1}/r_0 in the cross section at $z_0/L=1.5$ versus the Mach number M.

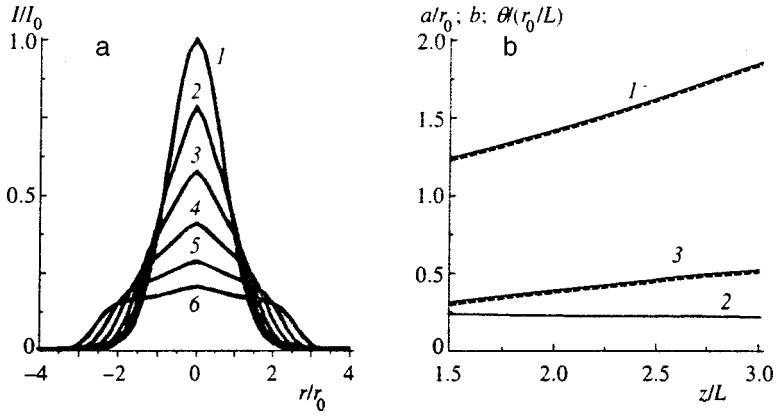

FIG. 5. (a) Transverse distributions in the initially Gaussian beam in the heat-conducting regime of self-action in cross sections (*1*) $z/L=0$; (*2*) 1.0; (*3*) 1.5; (*4*) 2.0; (*5*) 2.5; (*6*) 3.0. (b) Average radius a/r_0 [curves *1*, the exact solution is shown by the solid line, the approximate solution (19) by the dashed line], (*2*) ''constant'' b, and (*3*) angle of expansion $\theta/(r_0/L)$ with respect to initial axis as functions of the optical path length z. The time moment for all curves is $t=7\tau$; similarity numbers $N=1$, $F\gg 1$, $N_\alpha=0$.

1. Let us analyze the solution on an optical path including a focal point, taking as a example an axially symmetrical beam that propagates through a focusing medium in the heat-conducting regime ($m=1$, the sign in front of b in Eq. (51) is minus, and $D=3$). The first integral of Eq. (51) is

$$\theta\equiv\frac{da}{dz}=\pm\sqrt{-2b\ln a-\frac{c}{F^2a^2}+C_1}. \tag{53}$$

Here the plus sign corresponds to the section in front of the focal point, where $\theta=0$, and the minus sign corresponds to the section behind the focal point. In the case of the Gaussian initial intensity distribution and a plane initial wave front, exact numerical calculations performed at $N=-2$ and $F=3$ demonstrated that relative changes in the average radius are within 0.46% at the moments $t>4\tau$, the focal distance is $z_f\approx 2.20L$ at $t=4\tau$, the minimal radius is $a(z_f=2.2)/r_0=0.4969$, and the constants are $b=0.5307$, $c=2.190$, and $C_1=c/F^2$. The intensity profiles are shown in Fig. 6a. Behind the focus, the bell-shaped intensity distribution persists at least on the path section extending to the point $z=3L$. At the focal point, the peak intensity is almost seven times higher, in accordance with the approximate formula $I/I_0\sim(a_f/r_0)^{-2}\sim F^2$.

Figure 6b illustrates the dynamics of formation of a focusing lens. At the moment the quasi-steady state sets in, $t=4\tau$, and the solution of the approximate equation (53) (dashed line) is close to the exact solution (solid curve *5*). There is a satisfactory agreement between these two solutions in front of the focal point, at the focus, and behind the focal point to $z=3L$.

2. A second example of an optical path with a focal point is a focused plane beam ($\theta_1<0$) under transverse subsonic and supersonic gas flows ($m=1$; $D=2$) in the limit of geometrical optics at $F\gg 1$. In this case the constant $b=0$ [Eq. (22)]. The transverse beam dimension decreases following a linear function $a(z)=a_1-z|\theta_1|$. Within the distance $z-z_f\sim F^{-1}$ from the focus at $z_f\approx a_1/|\theta_1|$, we have to take diffraction into account, i.e., obtain a solution in the neighborhood of the focal point in extended coordinates $Z=(z-z_f)F$ and match it to the external solution at $Z\to\pm\infty$ on the scale $z\sim 1$ (Ref. 27). In this specific case, the general solution of Eq. (51), which satisfies the conditions of both the external and internal problems, is the upper branch of solution (31), where b must be replaced by c/F^2. The minimal beam dimension is diffraction limited; it is $a_f=\sqrt{c}/(F|\theta_1|)$.

Using the basic equation (51) at $b\neq 0$ and equating approximately the second term on the right side to the first term, we can estimate the beam compression at the focal point: $a_f\sim 1/F^\alpha$, where $\alpha=2/3$, 1, 3/2, and 2 at $m=0$, 1, 5/3, and 2, respectively. By equating the left-hand and right-hand sides, we estimate the longitudinal dimension of the focal region: $z-z_f\sim 1/F^\beta$, where $\beta=1/3$, 1, 2, and 3 at $m=0$, 1, 5/3, and 2, respectively.

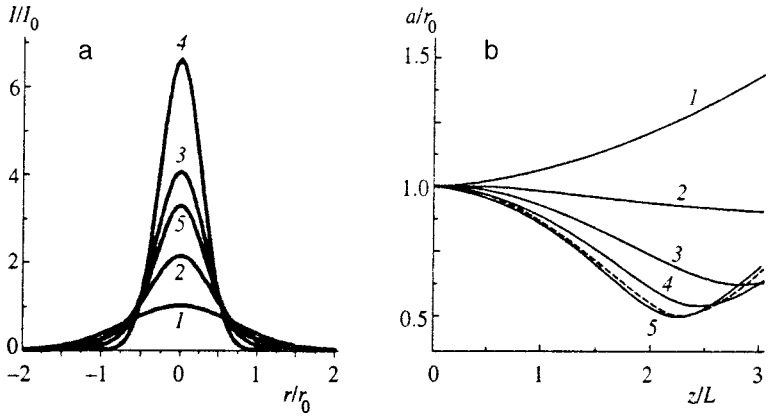

FIG. 6. (a) Intensity profiles for a Gaussian beam in a focusing medium ($N=-2$, $F=3$) in a quasi-steady heat-conducting regime in the following cross sections: (*1*) $z/L=0$; (*2*) 1.0; (*3*) 1.5; (*4*) 2.0; (*5*) 3.0; the time $t=4\tau$. (b) Average radius a/r_0 versus the optical path length on a section including the focal point: the solid curves show exact solutions at moments (*1*) $t=0$; (*2*) 0.0255τ; (*3*) 0.0631τ; (*4*) 1.303τ; (*5*) 4.00τ; the approximate quasi-steady solution is shown by the dashed line. The constants $b=-0.5307$, $c=2.190$, the Fresnel number $F=3$. The initial radius $a(z=0)/r_0=1$, the expansion angle $\theta(z=0)/(r_0/L)=0$.

12. CONCLUSIONS

1. The proposed procedure of approximate description of the laser beam self-action allows us, without solving a complicated system of equations, to draw the following conclusions concerning the self-action effects.

Under a steady-state gas flow, the expansion of a plane collimated beam is negligible (or small), the beam shifts in the downstream direction in a subsonic flow and in the upstream direction in a supersonic flow.

In the convective steady-state regime at $M \ll 1$ of a transverse gas flow, a three-dimensional collimated beam is extended mostly in the direction transverse with respect to the gas flow.

In a transonic gas flow, the predominant effect is expansion (convergence) of the beam in the direction aligned with the gas flow.

2. Universal solutions for all situations under discussion do not contain similarity parameters. In different physical situations, whose only common parameter is the exponent m in Eq. (26), one can use a solution given in Table I.

3. At the focal point, the minimal beam dimension is diffraction-limited in situations when the exponent $m < 3$. It seems that for $m \geqslant 3$ the main assumption of the unchanged shape of intensity distribution fails near the focal point.

4. Comparison with the exact solution at $m = 1$ in the case of transverse flow of a gas with zero viscosity and zero thermal conductivity (a plane laser beam) has demonstrated a difference of several percent. Good agreement between the exact and approximate solutions also takes place in the quasi-steady regime with a finite thermal conductivity at $m = 1$ (axially symmetrical beam) in both focusing and defocusing media.

5. The described procedure can be generalized to the case of $N_\alpha \sim 1$ and is applicable for the laser beam self-action in a longitudinal gas flow, in a vertical laser beam (in the presence of gravitational convection), and in the cases of other mechanisms of optical nonlinearity (Kerr effect, electrostriction, etc.).

The work was financially supported by the Russian Fund for Fundamental Research (Grant No. 99-01-00446) and T&AGI (Central Aerohydrodynamic Institute).

*)E-mail: ank@dept.aerocentr.msk.su

[1] S. A. Akhmanov, A. P. Sukhorukov, and R. V. Khokhlov, Usp. Fiz. Nauk **93**, 19 (1967) [Sov. Phys. Usp. **10**, 609 (1967)].
[2] V. N. Lugovoi and A. M. Prokhorov, Usp. Fiz. Nauk **111**, 203 (1973) [Sov. Phys. Usp. **16**, 658 (1973)].
[3] F. G. Gebhardt and D. C. Smith, IEEE J. Quantum Electron. **QE-7**, 63 (1971).
[4] P. M. Livingston, Appl. Opt. **10**, 426 (1971).
[5] P. V. Avisonis, C. B. Hogge, R. R. Butts, and J. R. Kenemuth, Appl. Opt. **11**, 554 (1972).
[6] M. N. Kogan, and A. N. Kucherov, Dokl. Akad. Nauk SSSR **251**, 575 (1980) [*sic*].
[7] V. I. Talanov, JETP Lett. **11**, 199 (1970).
[8] A. N. Kucherov, M. N. Makashev, and E. V. Ustinov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **36**, 135 (1993).
[9] D. C. Smith and F. G. Gebhardt, Appl. Phys. Lett. **16**, 275 (1970).
[10] R. A. Chodzko and S. C. Lin, AIAA J. **9**, 1105 (1971).
[11] V. A. Petrishchev, N. M. Sheronova, and V. E. Yashin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 963 (1975).
[12] A. N. Kucherov, Dokl. Akad.Nauk SSSR **251**, 309 (1980) [Sov. Phys. Dokl. **25**, 135 (1980)].
[13] A. N. Kucherov, Optika Atmosfery i Okeana **6**, 1519 (1993).
[14] A. N. Kucherov, Optika Atmosfery i Okeana **9**, 1110 (1996).
[15] A. N. Kucherov, Kvantovaya Élektronika **24**, 181 (1997).
[16] A. N. Kucherov, M. N. Makashev, and E. V. Ustinov, Optika Atmosfery i Okeana **6**, 1536 (1993).
[17] A. N. Kucherov, M. N. Makashev, and E. V. Ustinov, Kvantovaya Élektronika **22**, 187 (1995).
[18] E. Kamke, *Handbook on Ordinary Differential Equations* [Russian translation], Nauka, Moscow (1971).
[19] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Theory of Waves* [in Russian], Nauka, Moscow (1979).
[20] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series. Elementary Functions* [in Russian], Nauka, Moscow (1981).
[21] M. N. Kogan and A. N. Kucherov, Izv. Vyssh. Uchebn. Zaved. Fiz. No. 2, 104 (1983).
[22] M. N. Kogan, A. N. Kucherov, V. V. Mikhailov, and A. S. Fonarev, Izv. Akad. Nauk SSSR, Mekhanika Zhidkosti i Gaza, No. 5, 95 (1978).
[23] R. T. Brown and D. C. Smith, Appl. Phys. Lett. **25**, 500 (1974).
[24] L. D. Landau and E. M. Lifshitz, *Hydrodynamics*, Pergamon Press, Oxford–New York (1980).
[25] D. C. Smith, IEEE J. Quantum Electron. **QE-5**, 600 (1969).
[26] B. P. Gerasimov, V. M. Gordienko, and A. P. Sukhorukov, Zh. Tekh. Fiz. **45**, 2485 (1975) [Sov. Phys. Tech. Phys. **20**, 1551 (1975)].
[27] M. Van Dyke, *Perturbation Methods in Fluid Mechanics*, New York (1964).

Translation was provided by the Russian Editorial office.

Thermal destruction of antiferromagnetic Fe–Fe exchange bonds and mechanism of the transition to the spin-glass state in $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ systems with antiferromagnetic competing exchange interactions

N. N. Delyagin,*) A. L. Erzinkyan,†) V. P. Parfenova, and S. I. Reĭman

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

G. M. Gurevich

Institute of Nuclear Research, Russian Academy of Sciences, 117312 Moscow, Russia

J. Dupak

Institute of Scientific Instruments, Academy of Sciences of the Czech Republic, Brno, Czech Republic
(Submitted 24 July 1998)
Zh. Éksp. Teor. Fiz. **116**, 130–140 (July 1999)

The thermal evolution of the competition between the ferro- and antiferromagnetic exchange interactions in $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ alloys, which display different magnetic properties, depending on composition and temperature, is investigated. The distribution functions of the magnetic hyperfine fields $P(B_{hf})$ for ^{57}Fe are determined by Mössbauer spectroscopy in the temperature range 5–300 K for the alloys with $x=0$, 0.024, 0.082, 0.136, 0.195, and 0.252. The temperature dependence of the integrated intensity $I_s(T)$ is analyzed for the low- and high-field portions of $P(B_{hf})$. The features found in the behavior of $I_s(T)$ are interpreted as results of variation of the ratio between the competing exchange interactions of different signs as a result of the thermal destruction of antiferromagnetic Fe–Fe exchange bonds. It is shown that the changes in the spin structure in the low-temperature range are due to the thermal destruction of Fe–Fe exchange bonds. One of the consequences of this destruction is ''reentrance'' (an increase in the hyperfine field with increasing temperature for some of the Fe atoms). The relationship between the thermal destruction of Fe–Fe exchange bonds and the magnetic transitions of the Fe–Ni–Mn system to the spin-glass state is considered. [S1063-7761(99)01107-5]

1. INTRODUCTION

Magnetic systems with competing exchange interactions of different sign display a large variety of types of magnetic ordering (ferromagnetism, antiferromagnetism, spin-glass ordering, and reentrant spin-glass ordering). Concentrated magnetic systems, in which all (or most) of the atoms have a nonzero magnetic moment are most characteristic in this respect. The theoretical analysis of the magnetic behavior of such magnets is very complicated. It is difficult to apply the results of calculations performed for simple model systems to real magnets containing magnetic atoms of different types. Experimental investigations of the relationship between the behavior of competing exchange interactions and the magnetic structure on the microscopic level would be of great interest.

A characteristic example of systems with competing exchange interactions and strong spatial inhomogeneity in the distribution of the exchange fields is provided by disordered Fe–Ni–Mn alloys with an fcc structure. In the ferromagnetic alloy $Fe_{0.65}Ni_{0.35}$ the ferromagnetic (FM) exchange coupling constants J_{FeNi} and J_{NiNi} (which are equal to 450 and 600 K, respectively) are several times greater than the antiferromagnetic (AFM) coupling constant J_{FeFe} (Ref. 1). Nevertheless, in regions with a high local concentration of iron the ferromagnetic spin structure becomes unstable and conditions conducive to the formation of frustrated states and states with antiferromagnetic spin orientations are created (see Refs. 2 and 3 and the references cited there). The Mn–Fe exchange interaction in an fcc lattice is antiferromagnetic and strong (compared with the Fe–Fe exchange interaction); therefore, even a small admixture of Mn increases the competition between the exchange interactions of different sign and causes local disturbances of the ferromagnetic structure of the alloys. The magnetic properties of Fe–Ni–Mn alloys depend strongly on composition and temperature. Investigations of the magnetic phase diagrams of $Fe_{0.65}(Ni_{1-x}Mn_x)_{0.35}$ (Ref. 3), $(Fe_{0.65}Ni_{0.35})_{1-x}(Fe_{0.84}Mn_{0.16})_x$ (Ref. 4), and $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ (Ref. 5) systems have shown that a transition to the spin-glass (SG) state is characteristic of these alloys. Two magnetic transitions (the appearance of a reentrant spin glass), as well as the formation of a mixed (FM+SG) phase, were observed over a broad range of concentrations. Considerable spin polarization was observed for the SG phase of $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ alloys even in very weak magnetic fields (10–60 mT); and significant shifts of

the characteristic temperatures were observed in stronger fields.[5,6]

The unusual behavior of the SG phase in an external magnetic field has been explained within the phenomenological cluster model of Shiga *et al.*,[3] but the question of the mechanism of the formation of the SG phase in the Fe–Ni–Mn system remains open. One necessary condition for the formation of an SG phase is the presence of a sign-changing exchange interaction between the magnetic centers (atoms or magnetic clusters). For ''classical'' (magnetically dilute) spin glasses, the indirect long-range exchange interaction via conduction electrons has been regarded traditionally as such an interaction. In the case of concentrated spin glasses (which include the SG phase of the Fe–Ni–Mn system) this interaction cannot play a decisive role, and the formation of the SG phase should be regarded as a consequence of the competition among the exchange interactions between atoms of different types over short distances. Elucidation of the concrete mechanisms for the formation and decomposition of the SG phase in the Fe–Ni–Mn system is of fundamental importance for developing a theory of concentrated spin glasses.

Mössbauer spectroscopy offers a possibility to observe spin configurations of different types and to classify these states according to the amplitude of the magnetic hyperfine field $B_{\rm hf}$. The possibility of performing measurements over a broad temperature range permits, in particular, obtaining data on the effective exchange fields and the thermal stability of perturbed spin configurations. The use of this method to study disordered Fe–Ni–Mn alloys is based on an analysis of the hyperfine-field distribution functions $P(B_{\rm hf})$. The present work included Mössbauer investigations of $P(B_{\rm hf})$ for ^{57}Fe in $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ alloys in the concentration range $0 \leqslant x \leqslant 0.252$. Attention was focused for the most part on studying the temperature dependence of $P(B_{\rm hf})$ for different ranges of values of $B_{\rm hf}$. The features of this dependence are interpreted as results of the influence of the temperature on the competition between the exchange interactions of different sign. An earlier Mössbauer study of $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ alloys was described in Ref. 7. The dependences of the mean hyperfine field $\langle B_{\rm hf}\rangle$ on composition and temperature were obtained in that work. The temperature dependence of the structure of $P(B_{\rm hf})$ was not examined in Ref. 7.

2. EXPERIMENTAL METHOD

The test samples of $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ were prepared by fusing the metals with a purity no poorer than 99.9%; the ingots were rolled into foils having a thickness of ≈ 10 μm and rapidly cooled. The composition and homogeneity of the alloys were monitored by x-ray fluorescence microanalysis; the deviations of the concentrations of the components from the nominal composition did not exceed 0.1 at.%. The manganese concentration range chosen ($0 \leqslant x \leqslant 0.252$) covers the entire sequence of magnetic phases of this system (FM, FM+SG, SG, and AFM).[5] The Mössbauer absorption spectra were measured in the temperature range 5–300 K. Resonance detectors were employed to increase the magnitude of the resonant absorption effect and the resolving power for detecting Mössbauer radiation with an energy of 14.4 keV.

The distribution functions of the magnetic hyperfine fields $P(B_{\rm hf})$ were calculated using histograms.[8] The resultant Mössbauer spectrum was represented by the convolution of the distribution function $P(B_{\rm hf})$ and an elementary magnetic sextet. The widths of the histogram intervals were chosen so that they would slightly exceed the instrumental width of the components of the magnetic sextets. This allowed us to use the direct method for minimizing the χ^2 functional (using the FUMILI minimization program) and to eliminate the need to employ a smoothing procedure. The variable parameters for minimizing the χ^2 functional were the components of the $P(B_{\rm hf})$ histogram, the relative intensities of the second and fifth components of the magnetic sextets α, the linewidths, the isomer shift, and the correlation coefficients between the hyperfine field and the quadrupole splitting. (This correlation was introduced to take into account the weak asymmetry of the absorption spectra.) For spectra with a resolved hyperfine structure the variable parameter α differed from 2.0 by no more than 10% (which corresponds to random orientations of the spins in the sample). The spectra with a poorly resolved structure were treated with the fixed value $\alpha = 2.0$. Some other features of the method for treating the spectra were described in Ref. 8.

In most cases (the alloys with a high manganese concentration and all the alloys at high temperatures) the satellite structure of $P(B_{\rm hf})$ is complicated and poorly resolved. For this reason, a method based on consideration of the total (integrated) intensities I_s in selected ranges of variation of $B_{\rm hf}$ was employed in analyzing the temperature dependence of the components of the distribution functions. Such a data analysis method enabled us to eliminate the possible errors in the determination of the intensities of the individual components of $P(B_{\rm hf})$ and to represent the temperature dependence of the low- and high-field intensities $I_s(T)$ in a simple and graphic form.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 1 shows the Mössbauer absorption spectra measured at 5 K and the corresponding hyperfine-field distribution functions $P(B_{\rm hf})$ for the alloys with $x = 0$, 0.024, 0.082, 0.136, 0.195, and 0.252. The hyperfine-field distributions $P(B_{\rm hf})$ for the same alloys at various temperatures are shown in Figs. 2 and 3.

3.1. Hyperfine-field distributions $P(B_{\rm hf})$ at 5 K

Let us briefly consider some features of the structure of $P(B_{\rm hf})$ at 5 K. For $Fe_{0.65}Ni_{0.35}$ more than 90% of the intensity is concentrated in the principal maximum, which is centered at $B_{\rm hf} = 34.5$ T. The profile of the principal maximum of $P(B_{\rm hf})$ can be easily interpreted on the assumption of a statistical distribution of Fe and Ni atoms at the lattice points if the magnitude of the magnetic hyperfine field acting on the nucleus of an Fe atom in a given atomic configuration is described by the following familiar formula:

$$B_{\rm hf} = a\mu_{\rm Fe} + b[n_{\rm Fe}\mu_{\rm Fe} + (12 - n_{\rm Fe})\mu_{\rm Ni}], \quad (1)$$

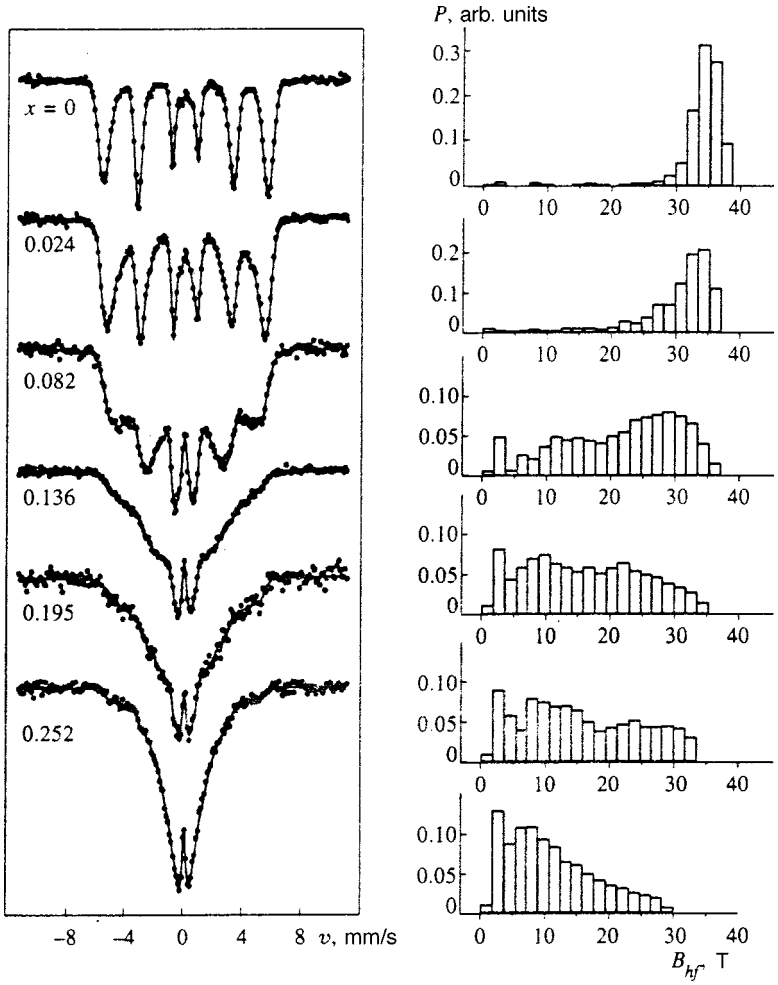

FIG. 1. Mössbauer spectra measured at 5 K (on the left) and corresponding hyperfine-field distribution functions P (on the right) for $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ alloys. The solid lines in the left-hand part were calculated using the method described in the text.

where μ_{Fe} and μ_{Ni} are the magnetic moments of the Fe and Ni atoms, and n_{Fe} is the number of nearest-neighbor Fe atoms in the particular configuration. The first term in this formula is the contribution of the intrinsic magnetic moment of the specific atom to B_{hf}, and the second term is the total contribution of the atoms in the local environment to B_{hf}. If $\mu_{Fe}=2.5\ \mu_B$ and $\mu_{Ni}=0.6\ \mu_B$, a comparison of the observed profile of the principal maximum with a calculation based on formula (1) (with consideration of the binomial distribution of the atoms of different types in the first coordination sphere) yields the following values for the coefficients:

$$a=9.4\ \text{T}/\mu_B, \quad b=0.5\ \text{T}/\mu_B$$

(a doubled number of histogram intervals in comparison to Fig. 1 was used to analyze the profile of the principal maximum.) The values of a and b which we found are typical of ferromagnetic iron alloys. The presence of weak satellite lines (with a total intensity of 6–8%) can be attributed to the formation of perturbed spin configurations in regions with a high local concentration of iron.[2,7]

The replacement of Fe and Ni atoms by Mn atoms quickly destroys the ferromagnetic spin structure of the alloy. The alloy with $x=0.024$ exhibits intense satellites with $B_{hf}\approx 22$ T and $B_{hf}\approx 28$ T to the left of the principal maximum, which can easily be attributed within formula (1) to spin reorientation in some Fe atoms under the influence of the antiferromagnetic Mn–Fe interaction. Raising the concentration of Mn to 8.2% leads to destruction of the principal maximum and the formation of a broad hyperfine-field distribution in the range 9–36 T. When the Mn concentration is increased further, a rapid increase in the intensity of the low-field components and a general shift of the hyperfine-field distribution toward lower values of B_{hf} are observed. The integrated intensity in the range $B_{hf}<4$ T for the alloys with $x=0.082$, 0.136, 0.195, and 0.252 is equal, respectively, to 8, 13, 16, and 33%. Raising the manganese concentration leads to a rapid decrease in the mean value of the hyperfine field $\langle B_{hf}\rangle$, which is equal to 32.6, 28.9, 21.9, 16.1, 14.5, and 11.9 T for the alloys with $x=0$, 0.024, 0.082, 0.136, 0.195, and 0.252, respectively.

The values of B_{hf} exceeding 8–10 T can be explained on the basis of the approximation of Eq. (1) in terms of the spin flip occurring as a result of the influence of antiferromagnetic Fe–Fe and Mn–Fe exchange interactions. The states with $B_{hf}<8$ T cannot be attributed solely to the spin flip. The Fe atoms in such states must have a temperature-dependent mean value of the magnetic moment $\langle\mu\rangle_T$ less than the nominal value of μ_{Fe}. The decrease in $\langle\mu\rangle_T$ means that the exchange fields for some Fe sites are very small and the local magnetization is not in a state of saturation even at

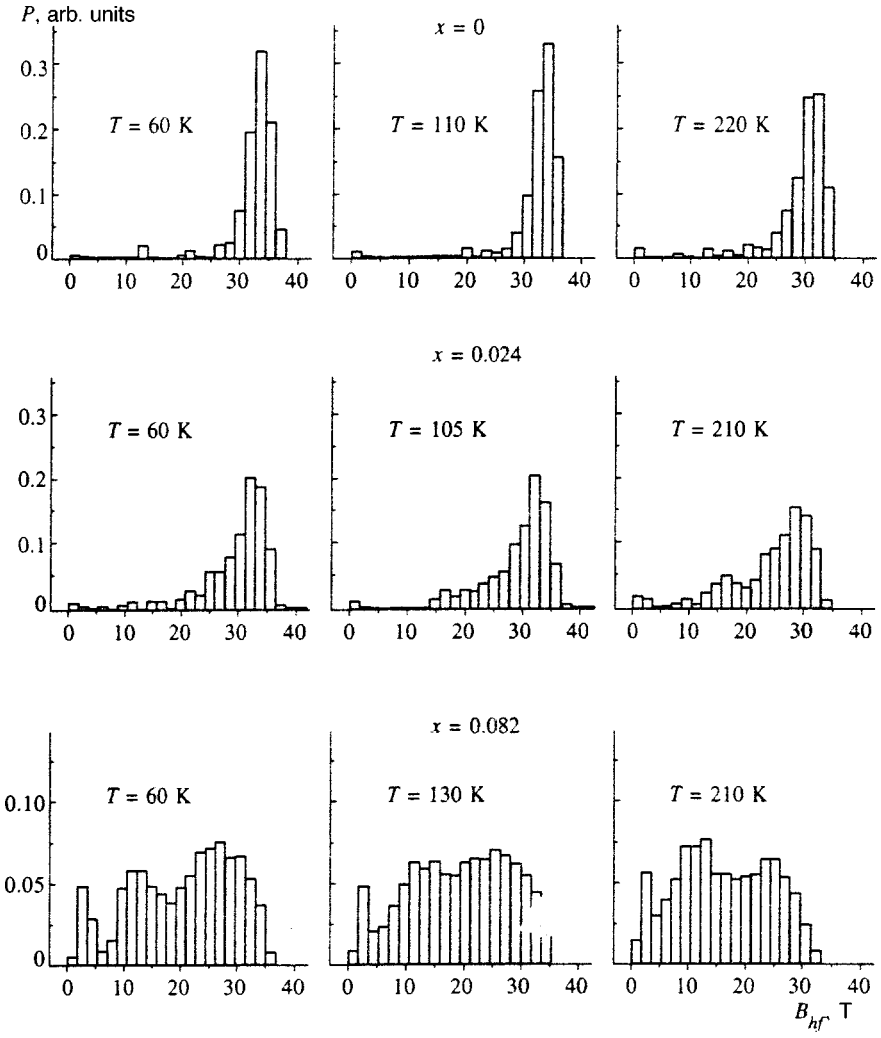

FIG. 2. Hyperfine-field distributions $P(B_{\rm hf})$ for the alloys with $x=0$, 0.024, and 0.082 at various temperatures.

$T=5$ K. The possibility for the formation of such states follows from a theoretical analysis of systems with competing exchange interactions.[9]

3.2. Influence of the thermal destruction of Fe–Fe exchange bonds on the competition between exchange interactions

As can be seen from Figs. 1 and 2, for the alloys with low manganese concentrations the structure of $P(B_{\rm hf})$ is relatively stable over a broad range of temperatures. In particular, a rise in temperature to $\sim$200 K is not accompanied by a significant increase in intensity in the low-field region. This means that at low manganese concentrations most of the perturbed spin configurations are characterized by high thermal stability. It can be concluded that (in agreement with the data in Ref. 10) the exchange coupling constant $J_{\rm MnFe}$ is very large and is probably close to 200 K. For the alloys with $x\geqslant 0.136$ (Figs. 1 and 3) a rise in temperature is accompanied by rapid displacement of the distributions $P(B_{\rm hf})$ toward weaker hyperfine fields and an increase in intensity near $B_{\rm hf}\approx 0$. The increase in intensity near zero values of $B_{\rm hf}$ clearly indicates occurrence of the ''fusing'' of frustrated spins considered in the theory of systems with competing exchange interactions.[9]

A further analysis of the behavior of $P(B_{\rm hf})$ for alloys of various composition was performed using integrated intensities (see Sec. 2). For the alloys with $x\geqslant 0.082$ the temperature dependences of the integrated intensities $I_s(T)$ for the low-field ($B_{\rm hf}<5$ T) and high-field ($B_{\rm hf}>17$ T) ranges, in which the behavioral features of $I_s(T)$ are most clearly displayed, were compared. The low-field range includes only states of Fe atoms for which (as was pointed out above) the exchange fields are known to be small (i.e., frustrated or nearly frustrated states) at 5 K. Conversely, the states of the high-field range with large values of $B_{\rm hf}$ at 5 K are characterized by values of the temperature-dependent mean magnetic moment $\langle\mu\rangle_T$ equal (or close) to the nominal value of $\mu_{\rm Fe}$. For the alloys with $x=0$ and 0.024 the intensity of the low-field satellites is very small; therefore, in these cases only the temperature dependence of $I_s(T)$ for the region of the principal maximum ($B_{\rm hf}=30-38$ T) was considered. The results are shown in Figs. 4 and 5. As is seen from Fig. 4, for alloys with high manganese concentrations the temperature dependences of the high- and low-field integrated intensities have characteristic features, which are most clearly displayed for the alloys with $x=0.136$ and $x=0.195$. A rapid rise in the integrated intensity in the low-

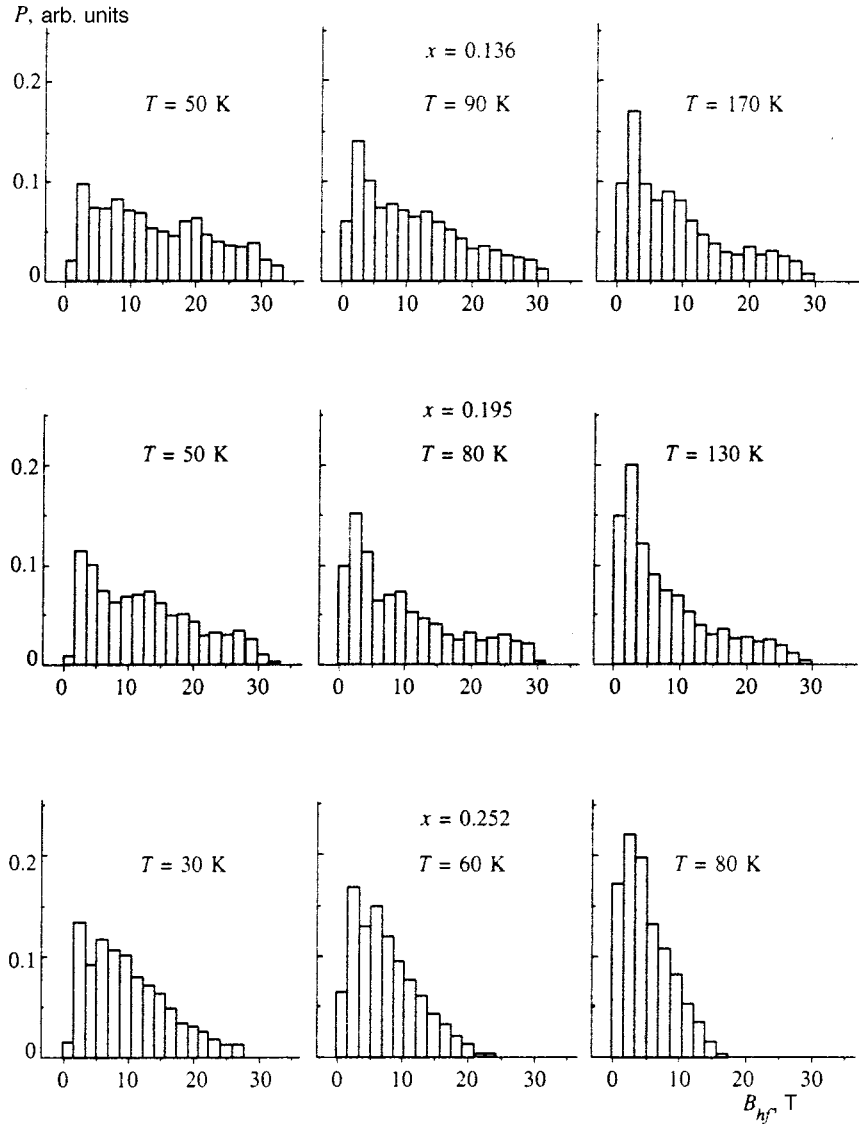

FIG. 3. Hyperfine-field distributions $P(B_{\rm hf})$ for the alloys with $x=0.136$, 0.195, and 0.252 at various temperatures.

field region and a synchronous decrease in the high-field region are observed in the temperature range 40–70 K. When the temperature is raised further, both the increase in intensity at weak fields and the decrease at strong fields cease. In the case of the alloy with $x=0.136$ these intensities remain constant in a broad temperature range. The alloy with $x=0.195$ clearly displays an anomalous increase in the integrated intensity at strong fields, i.e., a return of some of the Fe atoms to states with large values of $B_{\rm hf}$ (''reentrance'') when the temperature is raised in the range $T=110-130$ K. For the alloys with $x=0$ and 0.024 the intensity of the principal maximum decreases rapidly (almost abruptly for the alloy with $x=0$) when the temperature is raised in the range 50–70 K (Fig. 5). In the range 110–130 K the inten-

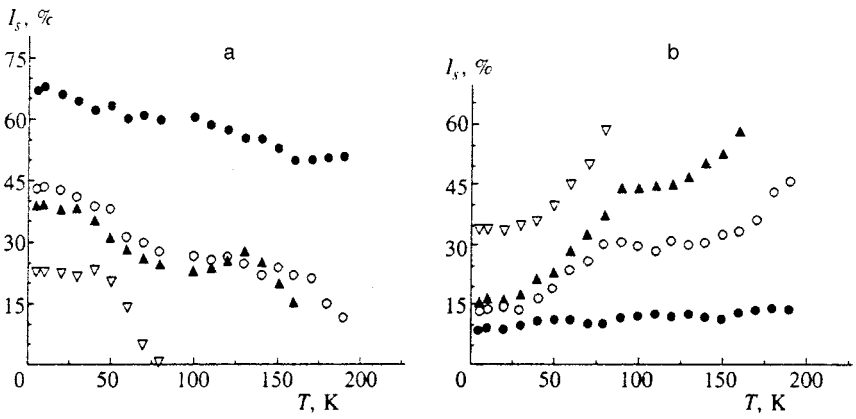

FIG. 4. Temperature dependence of the integrated intensity I_s (T) at strong ($B_{\rm hf}>17$ T) (a) and weak ($B_{\rm hf}<5$ T) (b) fields for the alloys with $x=0.082$ (●), $x=0.136$ (○), $x=0.195$ (▲), and $x=0.252$ (∇). The measurement errors do not exceed the size of the symbols.

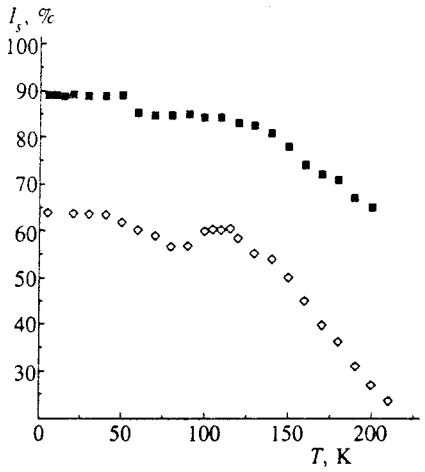

FIG. 5. Temperature dependence of the integrated intensity $I_s(T)$ in the region of the principal maximum ($B_{hf}=30-38$ T) for the alloys with $x=0$ (■) and $x=0.024$ (◇). The measurement errors do not exceed the size of the symbols.

sity remains constant for the alloy with $x=0$, while an increase in the integrated intensity, which corresponds to the ''reentrance'' just described, is clearly seen for the alloy with $x=0.024$.

It is significant that these features of the behavior of $I_s(T)$ are observed in the same temperature ranges for alloys with different manganese concentrations. Clearly, significant reorganization of the spin structure of the alloys takes place at these temperatures due to variation of the ratio between the contributions to the exchange interaction. It is natural to theorize that such variation of the ratio between the exchange interactions of different sign is associated with the antiferromagnetic Fe–Fe exchange interaction, which has a very low characteristic temperature Θ. A rise in temperature should cause the thermal destruction of Fe–Fe exchange bonds and, consequently, a change in the balance between the competing interactions (which is especially significant in regions with a high local concentration of iron). One obvious consequence of the thermal destruction of Fe–Fe exchange bonds is a decrease in the antiferromagnetic contribution to the exchange interaction at temperatures close to Θ. The corresponding change in the balance between the contributions to the exchange field should be manifested in the behavior of the local magnetization and B_{hf}, primarily for the Fe sites at which the energy of exchange interaction is comparable to $k\Theta$. At such sites B_{hf} decreases rapidly with increasing temperature, but at temperatures above ≈ 50 K the thermal destruction of Fe–Fe bonds causes an increase in the resultant exchange field and a corresponding rise in B_{hf}. The increase in B_{hf} with increasing temperature accounts for ''reentrance,'' which is manifested experimentally as a plateau on the plots of the integrated intensity versus temperature at strong fields (or even as a rise in this intensity for the alloys with $x=0.024$ and $x=0.195$). The temperature range in which ''reentrance'' is observed agrees well with the estimate of the exchange coupling constant J_{FeFe} found by low-angle neutron scattering.[1]

The phenomena just described should take place to some extent in all the alloys investigated; however, it is difficult to observe the melting of spins and ''reentrance'' in the alloy with $x=0.082$ because of the low concentration of states with weak exchange fields and the complicated satellite structure of the hyperfine-field distribution. In the alloy with $x=0.252$ ''reentrance'' is not observed because of the low temperature for the transition of this alloy to the paramagnetic state.

The thermal destruction of Fe–Fe exchange bonds accounts for the features of the magnetic phase diagram of the $(Fe_{0.65}Ni_{0.35})_{1-x}Mn_x$ system and the mechanism for formation of the SG phase in these alloys. Let us focus our attention on the fact that the temperature range in which the anomalies of $I_s(T)$ considered above are observed coincides with the temperature range of the transitions to the SG phase (or a mixed FM+SG phase). We also note that the temperatures of these transitions depend weakly on the manganese concentration.[5] This suggests the existence of a direct relationship between the transition to the SG phase and the thermal evolution of the Fe–Fe exchange bonds. The cluster model of the SG phase of the Fe–Ni–Mn system proposed by Shiga *et al.*[3] accounts for the behavior of this phase in external magnetic fields, but does not specify the nature of the sign-changing exchange interaction between the clusters. Iron is the dominant component in the alloys under consideration; therefore, even in the alloys with a considerable manganese concentration the antiferromagnetic Fe–Fe exchange bonds should play a decisive role in the formation of sites with weak sign-changing exchange fields. Magnetic clusters form in regions with large contributions from ferromagnetic Ni–Ni, Ni–Fe, and Ni–Mn exchange bonds, while a relatively weak sign-changing interaction appears in regions with a high local concentration of iron. The temperature range for stability of the SG phase should coincide with the range for stability of the Fe–Fe exchange bonds, as is observed in the alloys considered. The relative compositional independence of the critical temperature for decomposition of the SG phase is a natural consequence of such a mechanism for the formation and decomposition of the sign-changing exchange interaction. The existence of a region with a mixed (FM+SG) phase is attributed to spatial fluctuations of the local concentration of iron and the gradual course of the destruction of the sign-changing exchange interaction as the temperature is raised. The SG phase does not form in the alloys with $x\leqslant 0.04$ because of the low concentration of Fe sites with weak exchange fields.

It is noteworthy that, according to the results of our measurements, in the alloys with $x=0.136$ and 0.195 some of the Fe atoms remain in the magnetically ordered state at temperatures above 50–70 K (i.e., above the nominal temperatures for decomposition of the SG phase found from an analysis of the magnetization curves[5]). The rapid increase in the integrated intensity at weak fields in the range 50–70 K corresponds to decomposition of the SG phase, but complete passage of the system into the paramagnetic state does not occur. As follows from the form of the magnetic phase diagram,[5] this deviation from the data in Ref. 5 cannot be attributed to special features of the sample-preparation process or small variations in the composition of the alloys. It

can be theorized that in the concentration range $x \geqslant 0.14$ there is a mixed-phase region at temperatures above the normal temperature of the transition to the SG phase, which is not manifested in the magnetic measurements because of the small contribution of the magnetic transition to the magnetization and its highly diffuse nature. Our measurements do not permit identification of the type of magnetic ordering, but the existence of a magnetically ordered phase at $T > 70$ K follows unequivocally from the data presented in Figs. 3 and 4. The temperature of the magnetic transition in the alloy with $x = 0.252$ was found to be equal to ~ 90 K, in agreement with the data presented in the magnetic phase diagram.[5] In the control measurements of the Mössbauer spectra which we performed for the alloy with $x = 0.082$ in weak magnetic fields, the temperature of the reentrant transition to the SG phase was found to be ~ 40 K, which is also consistent with the results in Ref. 5.

This work was carried out with financial support from the Russian Fund for Fundamental Research (Project 97-02-16479).

*)E-mail: delyagin@srdlan.npi.msu.su
†)E-mail: erzink@srdlan.npi.msu.su

[1] M. Hatherly, K. Hirakawa, D. Lowder, J. F. Mallett, M. W. Stringfellow, and B. N. Torrie, Phys. Solid State **84**, 55 (1964).
[2] J. B. Müller and J. Hesse, Z. Phys. B **54**, 35 (1987).
[3] M. Shiga, T. Satake, Y. Wada, and Y. Nakamura, J. Magn. Magn. Mater. **51**, 123 (1985).
[4] T. Miyazaki, Y. Ando, and M. Takahashi, Phys. Rev. B **34**, 6334 (1986).
[5] J. Hesse, C. Böttger, A. Wulfes *et al.*, Phys. Status Solidi A **135**, 343 (1993).
[6] C. Böttger, R. Stasch, A. Wulfes, and J. Hesse, J. Magn. Magn. Mater. **99**, 280 (1991); M. Fricke and J. Hesse, Hyperfine Interact. **93**, 1537 (1994).
[7] B. Huck and J. Hesse, J. Magn. Magn. Mater. **78**, 247 (1988).
[8] N. N. Delyagin, G. M. Gurevich, A. L. Erzinkyan, V. P. Parfenova, S. I. Reĭman, S. V. Topalov, and M. Trkhlik, Zh. Éksp. Teor. Fiz. **109**, 1451 (1996) [JETP **82**, 783 (1996)].
[9] W. W. Saslow and G. Parker, Phys. Rev. Lett. **56**, 1074 (1986).
[10] B. Huck, F. Sauerbach, and J. Hesse, Hyperfine Interact. **28**, 479 (1986).

Translated by P. Shelnitz

Nonlinear magneto-optical Kerr effects

A. K. Zvezdin*) and N. F. Kubrakov

General Physics Institute, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted 8 April 1998)
Zh. Éksp. Teor. Fiz. **116**, 141–156 (July 1999)

The polarization state of a second-harmonic wave after reflection from a semi-infinite, optically isotropic magnetic medium is considered for the three characteristic uniform-magnetization directions corresponding to the linear magneto-optical Kerr effects. Expressions for the complex amplitudes of the wave field which specify the nonlinear Kerr effects, viz., the polar, meridional, and equatorial effects, are obtained in a first approximation with respect to the magnetization. The dependences of these effects on the angle of incidence of the inducing wave obtained as a result of a numerical experiment are presented. Analytical formulas are found for them at small angles of incidence. A comparative analysis of the linear and nonlinear Kerr effects is made. © *1999 American Institute of Physics.* [S1063-7761(99)01207-X]

1. INTRODUCTION

Low-dimensional magnetic systems (surfaces, thin films, multilayer structures, quantum dots, and quantum wires) are attracting a great deal of attention. Many unexpected and nontrivial effects associated with the properties of magnetic surfaces and interfaces have been discovered in recent years. These include giant magnetoresistance, significant surface anisotropy, deviations of the magnetic moments on a surface from their bulk values, oscillating exchange interactions between neighboring magnetic layers, and strong biquadratic exchange in multilayer structures. Apart from their unquestionable fundamental significance, these systems are of great practical interest for magnetic memory, sensors, etc.

Some new magneto-optical effects associated with the surfaces of magnetic media, viz., the nonlinear second-harmonic Kerr effects, were recently predicted and detected soon thereafter.[1–6] Although second-harmonic generation is forbidden in materials with an inversion center, and most widely encountered materials (Fe, Co, Ni, FeNi, etc.) are such, space-inversion symmetry is broken at a surface or an interface. Time-reversal symmetry is also broken in magnets. The breaking of these symmetries leads to the appearance of second-harmonic magneto-optical phenomena, which have been found to significantly surpass the corresponding linear effects in magnitude.[4,6] The large value of the polarization plane rotation angle of a second-harmonic wave (relative to the polarization of the inducing wave) ensures high contrast between regions with oppositely directed magnetization. For example, it can exceed 50% in a Co/Cu(100) multilayer structure.[5] In an Fe/Cr multilayer structure and in single-crystal iron whiskers the ratio between the nonlinear and linear Kerr rotation is of the order of 10^3 (Ref. 6). A comparison of the linear and nonlinear equatorial Kerr effects was made in Ref. 7. Nonlinear Kerr effects have been successfully employed for probing buried interfaces in multilayer films[5,8–11] and spin-polarization quantum wells.[12–14] They are also interesting for studying the spin dynamics on magnetic surfaces and in ultrathin layers on the real time scale (in the femtosecond range).[1)]

Thus, nonlinear magneto-optical effects offer a new promising tool for investigating magnetic surfaces and interfaces in magnetic films and multilayer structures with high spatial and temporal resolution, especially surface magnetic anisotropy and interlayer exchange, magnetic domains, quantum wells, noncollinear (canted) surface and interlayer structures, and the relationship between the geometric roughness and the magnetic, tunneling, and transport properties of nanostructures.

However, in order to reliably compare nonlinear and linear magneto-optical effects and to use them to investigate magnetic surfaces and other low-dimensional magnetic structures, tools must be developed to describe nonlinear magneto-optical effects within the familiar approach that is widely used in linear magneto-optics. Within this approach the magneto-optical properties of a material are described by two complex parameters: the refractive index n and the magneto-optical parameter $Q=\varepsilon_{12}/n^2$, where ε_{12} is an off-diagonal element of the dielectric tensor of the medium (see, for example, Ref. 24).

This paper describes a calculation of the complex amplitude of the second harmonic of a wave reflected from a semi-infinite space filled with a ferromagnetic material for arbitrary angles of incidence of the inducing wave in the three geometric configurations which are usually set up in magneto-optical experiments, viz., in the configurations of the polar, meridional, and equatorial Kerr effects. The polarization states of a reflected second-harmonic wave are calculated. Analytical formulas for the nonlinear magneto-optical Kerr effects at small angles of incidence of the inducing wave[2)] and the results of numerical experiments (at arbitrary angles), which characterize these effects as well as the ratios between them and the corresponding linear effects, are presented.

We note that nonlinear magneto-optical effects are described by a T-odd axial fourth-order tensor. On the one

hand, such a tensor corresponds to the great diversity of these effects, and, on the other hand, the large number of components in it creates certain difficulties and uncertainties in the analysis of an experiment. In the present work we utilize a theory of magnetic symmetry and an adequate hierarchy of small parameters, which permit considerable reduction of the number of parameters needed to completely describe nonlinear magneto-optical effects.

2. NONLINEAR POLARIZATION AND MAGNETIZATION OF A MEDIUM

Some theoretical aspects associated with the second harmonic in magnetic media were considered in Refs. 15 and 26–29. According to Refs. 15 and 26–29, nonlinear magneto-optical effects can be described in terms of electrodynamics using the nonlinear electric polarization vector **P**, which includes components that are proportional to the local magnetization vector (or other basis vectors in the case of more complicated magnetic structures). If the polarization **P** is localized on the surface of the medium, the nonlinear magneto-optical effects are caused exclusively by the presence of these surfaces and are determined by the distribution of the magnetization on these surfaces.

The second-order nonlinear surface optical polarization can be written in the form[1,30]

$$P_i = \chi^{(2)}_{ijk}(\mathbf{M}) E_j E_k , \tag{2.1}$$

where the nonlinear surface susceptibility $\chi^{(2)}$ depends on the magnetization **M**, and E_j is a component of the electric field of the light wave. The properties of the third-rank polar tensor $\chi^{(2)}$ and its dependence on magnetization are determined by the time-reversal symmetry and the symmetry of the surface. The time-reversal property (without allowance for dissipation) requires that $\mathrm{Re}\chi^{(2)}$ be an even function of **M** and that $\mathrm{Im}\chi^{(2)}$ be an odd function. It follows from symmetry arguments that the polarization **P** can be represented in the form

$$\mathbf{P} = \mathbf{P}_0 + \mathbf{P}_m , \tag{2.2}$$

where the magnetization-independent contribution is

$$\mathbf{P}_0 = \chi_1 \mathbf{E}(\mathbf{E}\cdot\mathbf{H}) + \chi_2 E^2 \mathbf{N}, \tag{2.2a}$$

and the contribution which is linearly dependent on **M** is

$$\mathbf{P}_m = \chi_3 \mathbf{E}(\mathbf{E}(\mathbf{m}\cdot\mathbf{N})) + \chi_4 E^2 (\mathbf{m}\cdot\mathbf{N}) + \chi_5 (\mathbf{E}\cdot\mathbf{m})(\mathbf{E}\cdot\mathbf{N}) + \chi_6 (\mathbf{E}\cdot\mathbf{N})(\mathbf{E}\cdot\mathbf{m}). \tag{2.2b}$$

Here χ_1 and χ_2 are the nonlinear optical parameters, $\chi_3, \ldots, \chi_6$ are the nonlinear magneto-optical parameters, $\mathbf{m} = \mathbf{M}/M$ is a vector which characterizes the magnetization direction, and **N** is a normal to the surface. Since **P** is a polar vector, only two independent combinations of the polar vectors **N** and **E**, which are second order in **E**, form the polar vector $\mathbf{P}_0$ and only four independent combinations formed from **E** and **N** and the axial vector **m** give $\mathbf{P}_m$.

The relation (2.2) should be regarded as an expansion of $\mathbf{P}(\mathbf{E},\mathbf{N},\mathbf{M})$ in **E**, **N**, and **M**. We restrict the analysis to the terms which are quadratic in **E** and linear in **N** and **M**. The ratio of the light-wave field E to the magnitude of the intra-atomic field E^*, the ratio $\zeta = E_{\mathrm{surf}}/E^*$ of the surface electric field E_{surf} (which breaks the even symmetry at the surface) to E^* (for **N**), and the magnitude of the magneto-optical gyrotropy (for **M**), which is determined by the magneto-optical parameter Q and usually satisfies the condition $Q \ll 1$, are small parameters.

Here expansion in **N** actually means expansion in $\mathbf{N}E_{\mathrm{surf}}$, as can be shown in the following manner. Let us assume that there is a surface layer, in which the properties of the medium vary in such a manner that their variation can be described using the polar vector **A**, which is parallel to **N**. The influence of **A** on the optical properties of the medium will be characterized by E_{surf}, although these quantities, in general, should not be considered identical. The nonlinear optical properties can then be described using the nonlinear susceptibility $\chi^{(3)}$, so that $\mathbf{P} = \chi^{(3)}\mathbf{E}_{\mathrm{surf}}\mathbf{E}\mathbf{E}$ or $\mathbf{P} = \tilde{\chi}^{(3)}\mathbf{A}\mathbf{E}$. In general, the properties of the tensors $\chi^{(3)}$ and $\tilde{\chi}^{(3)}$ can differ somewhat, but this difference does not influence the ensuing arguments. Assuming that the thickness of the surface layer can be less than the wavelength of the light, we can use a δ function to go over to a local description of the surface polarization. The averaged product $\chi^{(3)}\mathbf{E}_{\mathrm{surf}}$ reliably reflects the nonlinear surface susceptibility $\chi^{(2)}$. Consideration of terms of higher order in **N** implies allowance for contributions of a polarization vector of the type $\mathbf{P}^5 = \chi^{(5)}\mathbf{E}_{\mathrm{surf}}\mathbf{E}_{\mathrm{surf}}\mathbf{E}_{\mathrm{surf}}\mathbf{E}\mathbf{E}$, which are clearly at least ζ^2 times smaller than the terms taken into account above. The latter follows from the known relation $\mathbf{P}^{n+1}/\mathbf{P}^n \sim 1/E^*$ (Ref. 30). There may be cases in which $E_{\mathrm{surf}} \sim E^*$; expansion in $\mathbf{N}E_{\mathrm{surf}}$ is then inapplicable, and the theory under consideration is not completely general.

The use of the expansions (2.2a) and (2.2b) reduces the number of parameters needed to describe nonlinear magneto-optical phenomena [in comparison to the general formula (2.1)]. We can show this by comparing (2.1) with (2.2) and (2.2a). Formula (2.2a) can naturally be represented in a matrix form, just as (2.1), where the third-rank tensor has the following form in Voight's notation:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & e_{15} & 0 \\ 0 & 0 & 0 & e_{15} & 0 & 0 \\ e_{31} & e_{31} & e_{33} & 0 & 0 & 0 \end{bmatrix}. \tag{2.3}$$

We recall that it is symmetric relative to interchange of the indices j and k. Such a form for the tensor $\chi^{(2)}_{ijk}$ corresponds to the limiting ∞m symmetry group (the Curie group). A uniform electric field, for example, has this symmetry. It follows from (2.2a) that $e_{33} = e_{31} + 2e_{15}$, i.e., the tensor $\chi^{(2)}_{ijk}$ $(M=0)$ is specified in our case by two independent parameters, rather than three, as required by ∞m symmetry. However, there is no contradiction between formulas (2.2a) and (2.3), since (2.2a) corresponds to the linear approximation with respect to ζ. Taking into account the next term with respect to ζ in the expansion in (2.2a) in, for example, the form $\mathbf{N}(\mathbf{N}\cdot\mathbf{E})^2$, we obtain $e_{33} = e_{31} + 2e_{15} + O(\zeta^2)$.

After reflection from a medium with an inversion center, a second-harmonic wave, in general, includes the influence of not only the surface components, but also the bulk com-

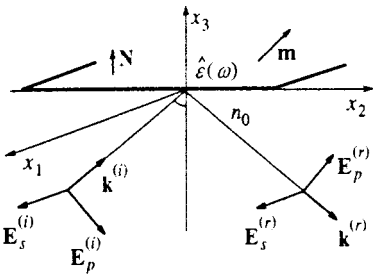

FIG. 1. Determination of the normal modes with a frequency ω for a semi-infinite magnetic medium.

ponents of the polarization **P**. However, the latter are manifested to a considerably weaker extent, especially in the case of metals.[31]

3. POLARIZATION STATE OF REFLECTED WAVES

In order to investigate nonlinear magneto-optical Kerr effects and compare them to the corresponding linear effects, we must determine the polarization state and the intensity of the reflected waves at the frequencies ω and 2ω. This can be done most simply in the case where reflection occurs from a semi-infinite, optically isotropic magnetic medium, in which the uniform-magnetization direction is characterized by the vector **m** (Fig. 1). The crux of the problem is to find the components $E_s^{(r)}$ and $E_p^{(r)}$ of the electric field of the wave at ω and the components $\tilde{E}_s^{(r)}$ and $\tilde{E}_p^{(r)}$ at 2ω. The polarization factor can be introduced as $\chi=-E_p^{(r)}/E_s^{(r)}$ or $\chi=E_s^{(r)}/E_p^{(r)}$, depending on which of the two directions, viz., the direction of *s* or *p* polarization, respectively, is chosen for measuring the polarization plane rotation angle. The polarization factor $\tilde{\chi}$ for the second harmonic (a tilde is used to denote quantities related to the second harmonic) should be introduced by analogy with the one from the definitions of χ. The Kerr rotation (the rotation angle of the major axis of the polarization ellipse) is found from the relation[32]

$$\tan 2\theta=\frac{2\,\mathrm{Re}\,\chi}{1-|\chi|^2}. \tag{3.1}$$

The ellipticity is defined in the following manner:

$$\eta=\frac{1}{2}\arcsin\frac{2\,\mathrm{Im}\,\chi}{1+|\chi|^2}. \tag{3.2}$$

The signs of θ and η correspond to the direction of observation of the wave vector $\mathbf{k}^{(r)}$. If $0<\eta<\pi/4$, the wave has left-elliptic polarization, and if $-\pi/4<\eta<0$, it has right-elliptic polarization. In the case of linear polarization ($\eta=0$), the condition $\mathrm{Im}\chi=0$ should hold. Circular polarization corresponds to $\eta=\pi/4$ (left) and $\eta=-\pi/4$ (right). These definitions clearly also apply to the nonlinear polar and meridional Kerr effects if the quantities appearing in (3.1) and (3.2) are construed as those pertaining to the second harmonic. We note that the signs of θ and η also depend on the representation of the plane wave.

The definition of the equatorial Kerr effect as the relative change $\delta=(I-I_0)/I_0$ in the intensity of a reflected wave in response to passage of the medium from a state with uniform magnetization (I) to a state without magnetization (I_0) is convenient only for a wave with the frequency ω. The analogous definition for the nonlinear case, which was used in Ref. 24, gives an unbounded increase in $\tilde{\delta}$ as the angle of incidence decreases. It is therefore more convenient to use the definition

$$\tilde{\delta}=(I-I_0)/(I+I_0). \tag{3.3}$$

We shall also define the linear equatorial effect by analogy with (3.3).

To examine the nonlinear magneto-optical Kerr effects we must use the results of the solution of the linear problem: the reflection matrix and the representations of the normal modes in a magnetic medium for the orientations of **m** which appear in the definitions of the linear Kerr effects. In this case the dielectric tensor of the magnetic medium is[25]

$$\hat{\varepsilon}(\omega)=n^2\begin{bmatrix}1 & -im_3Q & im_2Q\\ im_3Q & 1 & -im_1Q\\ -im_2Q & im_1Q & 1\end{bmatrix}, \tag{3.4}$$

where n is the complex refractive index ($\mathrm{Im}\,n>0$), and Q is the magneto-optical parameter, which depends linearly on magnetization. The magnetic permeability is assumed to be equal to unity.

An incident wave $\mathbf{E}^{(i)}=\mathbf{E}_0^{(i)}\exp[i(\mathbf{k}^{(i)}\mathbf{x}-\omega t)]$ and a reflected wave $\mathbf{E}^{(r)}=\mathbf{E}_0^{(r)}\times\exp[i(\mathbf{k}^{(r)}\mathbf{x}-\omega t)]$, which propagates in a transparent medium ($x_3<0$) with the refractive index $n_0(\omega)$, can be represented in the form of waves with *s* and *p* polarization, which are related to one another through the reflection matrix

$$\begin{bmatrix}E_s^{(r)}\\ E_p^{(r)}\end{bmatrix}=\begin{bmatrix}r_{ss} & r_{sp}\\ r_{ps} & r_{pp}\end{bmatrix}\begin{bmatrix}E_s^{(i)}\\ E_p^{(i)}\end{bmatrix}, \tag{3.5}$$

whose elements depend on the optical parameters of the two media. In the semi-infinite region of the magnetic medium the two solutions of Fresnel's equation correspond to two normal modes. Knowledge of all three components of the electric field is needed for each of them.

4. REFLECTION MATRIX AND COMPLEX FIELD AMPLITUDES OF A WAVE AT THE FREQUENCY ω

We next present the elements of the reflection matrix and the formulas for the complex field amplitudes appearing in the definition (2.2) of the nonlinear surface polarization[25] for the three directions of **m** (Fig. 1).

4.1. Polar geometry, m=(0,0,1)

The magnetization is orthogonal to the surface of the medium and lies in the plane of incidence. The elements of the reflection matrix are

$$r_{ss}=\frac{X^-}{X^+},\quad r_{pp}=\frac{Y^-}{Y^+}, \tag{4.1}$$

$$r_{sp}=-r_{ps}=\frac{in_0n^2Q}{X^+Y^+}\cos\varphi, \tag{4.2}$$

where $X^{\pm}=n_0\cos\varphi\pm\sqrt{n^2-\alpha^2}$, $Y^{\pm}=n_0\sqrt{n^2-\alpha^2}\pm n^2\cos\varphi$, and $\alpha=n_0\sin\varphi$.

The components of the electric field in the magnetic medium are

$$E_1=\frac{2n_0\cos\varphi}{X^+}\left[E_s^{(i)}+\frac{in^2Q}{2Y^+}E_p^{(i)}\right],$$

$$E_2=\frac{2n_0\cos\varphi}{Y^+}\left[\sqrt{n^2-\alpha^2}E_p^{(i)}-\frac{in^2Q\cos\varphi}{2X^+}E_s^{(i)}\right], \quad (4.3)$$

$$E_3=-\frac{2n_0\alpha\cos\varphi}{Y^+}\left[E_p^{(i)}+\frac{in_0Q}{2X^+}E_s^{(i)}\right].$$

If the incident wave has linear polarization, the reflected wave will have elliptic polarization with the major axis of the polarization ellipse turned through a certain angle, which is determined from (3.1).

4.2. Meridional geometry, m=(0,1,0)

The magnetization is parallel to the surface of the medium and lies in the plane of incidence. This effect, like the polar effect, consists of the appearance of ellipticity and rotation of the polarization plane of the reflected wave if the incident wave is linearly polarized. The elements r_{ss} and r_{pp} are determined from (4.1), and

$$r_{sp}=r_{ps}=\frac{i\alpha n_0n^2Q\cos\varphi}{X^+Y^+\sqrt{n^2-\alpha^2}}. \quad (4.4)$$

The components of the electric field in the magnetic medium are

$$E_1=\frac{2n_0\cos\varphi}{X^+}\left[E_s^{(i)}+\frac{i\alpha n^2Q}{2Y^+\sqrt{n^2-\alpha^2}}E_p^{(i)}\right],$$

$$E_2=\frac{2n_0\cos\varphi}{Y^+}\left[\sqrt{n^2-\alpha^2}E_p^{(i)}+\frac{i\alpha n^2Q\cos\varphi}{2X^+\sqrt{n^2-\alpha^2}}E_s^{(i)}\right], \quad (4.5)$$

$$E_3=-\frac{2n_0\cos\varphi}{Y^+}\left[\alpha E_p^{(i)}-\frac{iQ}{X^+}\left(Y^++\frac{\alpha^2n_0}{2\sqrt{n^2-\alpha^2}}\right)E_s^{(i)}\right].$$

4.3. Equatorial geometry, m=(1,0,0)

The magnetization is parallel to the surface of the medium and orthogonal to the plane of incidence. The equality $r_{sp}=r_{ps}=0$ signifies that the reflected wave has the same polarization as the incident wave. The element r_{ss} is determined from (4.1), and the other diagonal element is

$$r_{pp}=\frac{Y^-}{Y^+}\left[1-\frac{in_0^2n^2Q}{Y^-Y^+}\sin 2\varphi\right]. \quad (4.6)$$

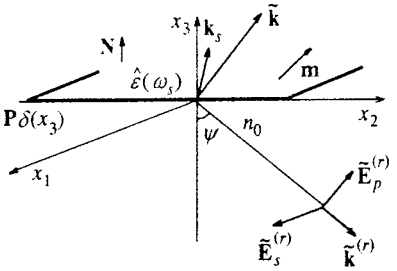

FIG. 2. Waves with the frequency ω_s governed by the nonlinear surface polarization **P**.

The complex amplitudes of the field in the magnetic medium are

$$E_1=\frac{2n_0\cos\varphi}{X^+}E_s^{(i)},$$

$$E_2=\frac{2n_0\cos\varphi}{Y^+}\left[\sqrt{n^2-\alpha^2}-\frac{i\alpha n^2Q}{Y^+}\cos\varphi\right]E_p^{(i)}, \quad (4.7)$$

$$E_3=-\frac{2n_0\cos\varphi}{Y^+}\left[\alpha+\frac{in^2Q}{Y^+}(n_0+\sqrt{n^2-\alpha^2}\cos\varphi)\right]E_p^{(i)}.$$

The equatorial effect clearly takes place only for oblique incidence.

5. NORMAL MODES WITH A FREQUENCY OF 2ω

The surface polarization $\mathbf{P}=(\mathbf{P}_0+\mathbf{P}_m)\exp[i(k_{0s}\alpha x_2-\omega_s t)]$, where $\omega_s=2\omega$ and the wave number $k_{0s}=\omega_s/c$, is a source of plane waves with a frequency ω_s, which are damped in the semi-infinite magnetic medium (Fig. 2). To find the polarization state of these waves, Maxwell's equation must be solved in the two regions.

In the region of the magnetic medium ($x_3>0$)

$$\operatorname{curl}\tilde{\mathbf{H}}=-i\omega_s\varepsilon_0\hat{\varepsilon}(\omega_s)\tilde{\mathbf{E}}, \quad \operatorname{curl}\tilde{\mathbf{E}}=i\omega_s\mu_0\tilde{\mathbf{H}}. \quad (5.1)$$

Here the dielectric tensor

$$\hat{\varepsilon}(\omega_s)=\tilde{n}^2\begin{bmatrix}1 & -im_3\tilde{Q} & im_2\tilde{Q}\\ -im_3\tilde{Q} & 1 & -im_1\tilde{Q}\\ -im_2\tilde{Q} & im_1\tilde{Q} & 1\end{bmatrix}$$

contains the refractive index $\tilde{n}=n(\omega_s)$ and the magneto-optical parameter $\tilde{Q}=Q(\omega_s)$.

In the region of the transparent medium ($x_3<0$) with the refractive index $\tilde{n}_0=n_0(\omega_s)$

$$\operatorname{curl}\tilde{\mathbf{H}}=-i\omega_s\varepsilon_0\tilde{n}_0^2\tilde{\mathbf{E}}, \quad \operatorname{curl}\tilde{\mathbf{E}}=i\omega_s\mu_0\tilde{\mathbf{H}}. \quad (5.2)$$

The boundary conditions at the surface ($x_3=0$) can be obtained as was done in Ref. 33. The following relations are obtained as a result:

$$\tilde{E}_1^+ - \tilde{E}_1^- = 0, \quad \tilde{E}_2^+ - \tilde{E}_2^- = -i\alpha\tilde{n}^{-2}k_{0s}\varepsilon_0^{-1}P_3,$$
$$\tilde{H}_1^+ - \tilde{H}_1^- = \omega_s(m_1\tilde{Q}P_3 - iP_2),$$
$$\tilde{H}_2^+ - \tilde{H}_2^- = \omega_s(m_2\tilde{Q}P_3 + iP_1). \tag{5.3}$$

Equations (5.1), whose solution has the form of a plane wave $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}_0 \exp[i(\mathbf{kx} - \omega_s t)]$ with a wave vector $\tilde{\mathbf{k}} = k_{0s}(0, \alpha, \gamma)$, reduce to the following wave equation in matrix form:

$$\begin{bmatrix} \alpha^2 + \gamma^2 - \tilde{n}^2 & i\tilde{n}^2 m_3\tilde{Q} & -i\tilde{n}^2 m_2\tilde{Q} \\ -i\tilde{n}^2 m_3\tilde{Q} & \gamma^2 - \tilde{n}^2 & -\alpha\gamma + i\tilde{n}^2 m_1\tilde{Q} \\ i\tilde{n}^2 m_2\tilde{Q} & -\alpha\gamma - i\tilde{n}^2 m_1\tilde{Q} & \alpha^2 - \tilde{n}^2 \end{bmatrix} \begin{bmatrix} \tilde{E}_1 \\ \tilde{E}_2 \\ \tilde{E}_3 \end{bmatrix} = 0. \tag{5.4}$$

This equation can be solved only for values of γ which are solutions of Fresnel's equation

$$\gamma^4 - [2(\tilde{n}^2 - \alpha^2) - \tilde{n}^2\tilde{Q}^2(1 - m_3^2)]\gamma^2 - 2\tilde{n}^2\tilde{Q}^2 m_2 m_3 \alpha\gamma + (\tilde{n}^2 - \alpha^2)^2 - \tilde{n}^2\tilde{Q}^2[\tilde{n}^2 - (1 - m_2^2)\alpha^2] = 0. \tag{5.5}$$

Since the medium is assumed to be semi-infinite, of the four roots of Eq. (5.5) only those which correspond to two waves departing from the surface, i.e., normal modes, have physical meaning. Their form depends on the direction of **m**. We next present representations of the normal modes in the first approximation with respect to $\tilde{Q}$ for the orientations of **m** which correspond to the definitions of the three magneto-optical Kerr effects.

agraph *Polar effect, $\mathbf{m} = (0,0,1)$. It follows from (5.5) that

$$\gamma_{1,2} = \sqrt{\tilde{n}^2 - \alpha^2} \mp \frac{1}{2}\tilde{n}\tilde{Q},$$

and Eq. (5.4) gives two modes

$$\tilde{\mathbf{E}}^{(1,2)} = \tilde{\mathbf{E}}_{02}^{(1,2)} \exp(i\tilde{\mathbf{k}}^{(1,2)}\mathbf{x})(\pm\xi, 1, -\alpha(\tilde{n}^2 - \alpha^2)^{-1}\gamma_{1,2}), \tag{5.6}$$

$$\tilde{\mathbf{H}}^{(1,2)} = \frac{\xi k_{0s}\tilde{\mathbf{E}}_{02}^{(1,2)}}{\omega_s\mu_0} \exp(i\tilde{\mathbf{k}}^{(1,2)}\mathbf{x})(\xi\gamma_{1,2}, \pm\gamma_{1,2}, \mp\alpha),$$

where $\xi = i\tilde{n}\sqrt{\tilde{n}^2 - \alpha^2}$, $\tilde{\mathbf{k}}^{(j)} = k_{0s}(0, \alpha, \gamma_j)$, the upper sign corresponds to the first mode, and the lower sign corresponds to the second mode.

agraph *Meridional effect, $\mathbf{m} = (0,1,0)$. To find the normal modes from (5.4) we need the exact values of

$$\gamma_{1,2} = \sqrt{\tilde{n}^2 - \alpha^2 - \frac{1}{2}\tilde{n}^2\tilde{Q}^2 \mp \tilde{n}\tilde{Q}\sqrt{\alpha^2 + \frac{1}{4}\tilde{n}^2\tilde{Q}^2}}.$$

Then

$$\tilde{\mathbf{E}}^{(1,2)} = \tilde{\mathbf{E}}_{02}^{(1,2)} \exp(i\tilde{\mathbf{k}}^{(1,2)}\mathbf{x}) \times \left(\frac{i\tilde{n}^2\tilde{Q}(\tilde{n}^2 - \gamma_{1,2}^2)}{\alpha\gamma_{1,2}(\tilde{n}^2 - \alpha^2 - \gamma_{1,2}^2)}, 1, \frac{\tilde{n}^2 - \gamma_{1,2}^2}{\alpha\gamma_{1,2}} \right), \tag{5.7}$$

$$\tilde{\mathbf{H}}^{(1,2)} = \frac{k_{0s}\tilde{\mathbf{E}}_{02}^{(1,2)}}{\omega_s\mu_0} \times \exp(i\tilde{\mathbf{k}}^{(1,2)}\mathbf{x}) \left(-\frac{\tilde{n}^2}{\gamma_{1,2}}, \frac{i\tilde{n}^2\tilde{Q}(\tilde{n}^2 - \gamma_{1,2}^2)}{\alpha(\tilde{n}^2 - \alpha^2 - \gamma_{1,2}^2)}, -\frac{i\tilde{n}^2\tilde{Q}(\tilde{n}^2 - \gamma_{1,2}^2)}{\gamma_{1,2}(\tilde{n}^2 - \alpha^2 - \gamma_{1,2}^2)} \right).$$

Equatorial effect, $\mathbf{m} = (1,0,0)$. It follows from (5.5) that

$$\gamma_{1,2} \approx \gamma = \sqrt{\tilde{n}^2 - \alpha^2}$$

and the normal modes are

$$\tilde{\mathbf{E}}^{(1)} = \tilde{\mathbf{E}}_{01}^{(1)} \exp(i\tilde{\mathbf{k}}\mathbf{x})(1,0,0),$$
$$\tilde{\mathbf{H}}^{(1)} = \frac{k_{0s}\tilde{\mathbf{E}}_{01}^{(1)}}{\omega_s\mu_0} \exp(i\tilde{\mathbf{k}}\mathbf{x})(0, \gamma, -\alpha),$$
$$\tilde{\mathbf{E}}^{(2)} = \tilde{\mathbf{E}}_{02}^{(2)} \exp(i\tilde{\mathbf{k}}\mathbf{x}) \left(0, 1, -\frac{\alpha\gamma + i\tilde{n}^2\tilde{Q}}{\tilde{n}^2 - \alpha^2} \right), \tag{5.8}$$
$$\tilde{\mathbf{H}}^{(2)} = \frac{k_{0s}\tilde{\mathbf{E}}_{02}^{(2)}}{\omega_s\mu_0} \exp(i\tilde{\mathbf{k}}\mathbf{x}) \left(\frac{\tilde{n}^2(\gamma + i\alpha\tilde{Q})}{\tilde{n}^2 - \alpha^2}, 0, 0 \right).$$

Here the first mode has *s* polarization, and the vector $\tilde{\mathbf{E}}^{(2)}$ lies in the plane of incidence, but is not orthogonal to the wave vector.

It follows from (5.2) that only one plane wave $\tilde{\mathbf{E}}^{(r)} = \tilde{\mathbf{E}}_0^{(r)} \exp[i(\tilde{\mathbf{k}}\mathbf{x} - \omega_s t)]$ with the wave vector $\tilde{k}^{(r)} = k_{0s}\tilde{n}_0(0, \sin\psi, \cos\psi)$ propagates in the transparent medium ($x_3 < 0$) (Fig. 2). In contrast with the normal modes in the magnetic medium, this wave can be represented in the form of a superposition of waves with *s* and *p* polarization. In particular, at the surface

$$\tilde{\mathbf{E}}^{(r)} = (\tilde{\mathbf{E}}_s^{(r)}, \tilde{\mathbf{E}}_p^{(r)} \cos\psi, \tilde{\mathbf{E}}_p^{(r)} \sin\psi),$$
$$\tilde{\mathbf{H}}^{(r)} = \frac{k_{0s}\tilde{n}_0}{\omega_s\mu_0}(\tilde{\mathbf{E}}_p^{(r)}, -\tilde{\mathbf{E}}_s^{(r)} \cos\psi, -\tilde{\mathbf{E}}_s^{(r)} \sin\psi). \tag{5.9}$$

The boundary conditions (5.3) are satisfied if the relation

$$\sin\psi = \frac{n_0}{\tilde{n}_0} \sin\varphi, \tag{5.10}$$

which can be called the reflection law for a wave with the frequency ω_s, holds. If the refractive index of the transparent medium does not depend on frequency, we have $\psi = \varphi$.

6. NONLINEAR POLAR KERR EFFECT

In the case $\mathbf{m} = (0,0,1)$ the expressions (5.6) for the normal modes and the boundary conditions (5.3) give the following representations of the *s* and *p* components of the reflected wave with the frequency ω_s in terms of the components of the nonlinear surface polarization:

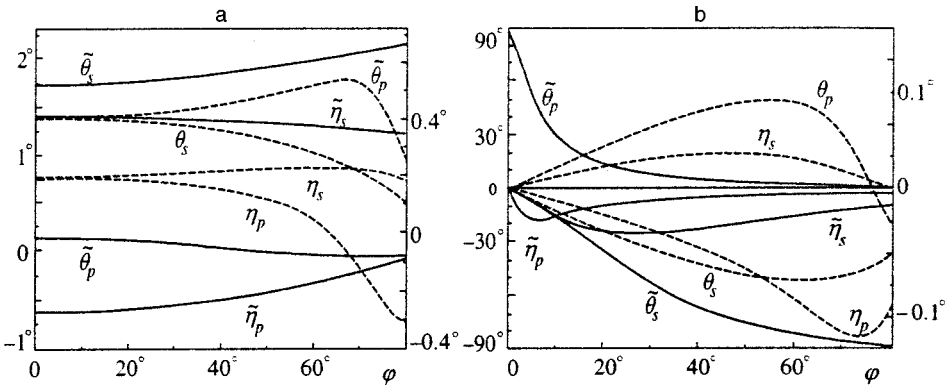

FIG. 3. Polarization plane rotation angle and ellipticity for the nonlinear (solid lines) and linear (dashed lines, to which the right-hand scale refers) polar (a) and meridional (b) Kerr effects for s and p polarization of the incident wave.

$$\tilde{E}_s^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{X}^+}\left[P_1+\frac{i\tilde{Q}}{2\tilde{Y}^+}(P_2\tilde{n}^2\cos\psi-\alpha\tilde{n}_0P_3)\right],$$
$$\tilde{E}_p^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{Y}^+}\left[\sqrt{\tilde{n}^2-\alpha^2}P_2+\alpha P_3-\frac{i\tilde{n}^2\tilde{Q}}{2\tilde{X}^+}P_1\right], \tag{6.1}$$

where $\tilde{X}^+=\tilde{n}_0\cos\psi+\sqrt{\tilde{n}^2-\alpha^2}$, $\tilde{Y}^+=\tilde{n}_0\sqrt{\tilde{n}^2-\alpha^2}+\tilde{n}^2\cos\psi$, and $\alpha=n_0\sin\varphi$. According to (2.2), in the configuration of the polar Kerr effect the components of **P** are

$$P_1=\chi_1E_1E_3+(\chi_5+\chi_6)E_2E_3,$$
$$P_2=\chi_1E_2E_3-(\chi_5+\chi_6)E_1E_3,$$
$$P_3=\chi_2(E_1^2+E_2^2)+(\chi_1+\chi_2)E_3^2,$$

where the complex amplitudes E_1, E_2, and E_3 are known from (4.3). Thus, the s and p components of the reflected wave with the frequency ω_s are now defined in terms of the s and p components of the incident wave $\mathbf{E}^{(i)}$ and the optical parameters of the medium. In particular, if the incident wave has s polarization (Fig. 1), then

$$\tilde{E}_s^{(r)}=\frac{\alpha k_{0s}}{2\varepsilon_0\tilde{X}^+}\left(\frac{2n_0\cos\varphi}{X^+}E_s^{(i)}\right)^2\left[\chi_1\frac{n_0Q}{Y^+}+\chi_2\frac{\tilde{n}_0\tilde{Q}}{\tilde{Y}^+}\right],$$
$$\tilde{E}_p^{(r)}=\frac{i\alpha k_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{X^+}E_s^{(i)}\right)^2\chi_2, \tag{6.2}$$

and if it has p polarization, then

$$\tilde{E}_s^{(r)}=\frac{\alpha k_{0s}}{\varepsilon_0\tilde{X}^+}\left(\frac{2n_0\cos\varphi}{Y^+}E_p^{(i)}\right)^2\times\left\{\frac{1}{2}\chi_1\left[\frac{n^2Q}{X^+}+\frac{\tilde{Q}}{\tilde{Y}^+}(\sqrt{n^2-\alpha^2}\tilde{n}^2\cos\psi+\alpha^2\tilde{n}_0)\right]+\chi_2\frac{\tilde{n}_0n^2\tilde{Q}}{2\tilde{Y}^+}-i(\chi_5+\chi_6)\sqrt{n^2-\alpha^2}\right\},$$
$$\tilde{E}_p^{(r)}=-\frac{i\alpha k_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{Y^+}E_p^{(i)}\right)^2\times[\chi_1(\sqrt{(\tilde{n}^2-\alpha^2)(n^2-\alpha^2)}-\alpha^2)-\chi_2n^2]. \tag{6.3}$$

In each case the reflected wave with the frequency ω_s is elliptically polarized. Substitution of the parameter $\tilde{\chi}$ found from (6.2) or (6.3) into (3.1) and (3.2) gives the results needed: the polarization plane rotation angle and the ellipticity for the second harmonic. In a first approximation with respect to the angle of incidence φ the corresponding formulas for the complex angle $\tilde{\theta}+i\tilde{\eta}$ are fairly compact. In fact, if the incident wave has s polarization (Fig. 2) and $\tilde{\theta}_s$ is measured from the direction for p polarization ($\tilde{\chi}=\tilde{E}_s^{(r)}/\tilde{E}_p^{(r)}$), then

$$\tilde{\theta}_s+i\tilde{\eta}_s=\frac{1}{2i}\left[\frac{n_0\tilde{n}Q}{n(n_0+n)}\frac{\chi_1}{\chi_2}+\frac{\tilde{n}_0\tilde{Q}}{\tilde{n}_0+\tilde{n}}\right]. \tag{6.4}$$

The Kerr angle $\tilde{\theta}_p+i\tilde{\eta}_p$ (p polarization of the incident wave and $\tilde{\chi}=\tilde{E}_s^{(r)}/\tilde{E}_p^{(r)}$) in this approximation likewise does not depend on φ:

$$\tilde{\theta}_p+i\tilde{\eta}_p=i\tilde{n}\left\{\left[\frac{1}{2}\left(\frac{nQ}{n_0+n}+\tilde{Q}\right)-i\frac{\chi_5+\chi_6}{\chi_1}\right]\times\left(\tilde{n}-n\frac{\chi_2}{\chi_1}\right)^{-1}-\frac{\tilde{n}_0\tilde{Q}}{2\tilde{n}(\tilde{n}_0+\tilde{n})}\right\}. \tag{6.5}$$

As an example we present some characteristic plots of the polarization plane rotation angle and the ellipticity versus the angle of incidence φ (Fig. 3a) for the linear (θ,η) and nonlinear ($\tilde{\theta},\tilde{\eta}$) Kerr effects when $\tilde{n}_0=n_0=1$, $\tilde{n}=n=2.36\pm3.48i$, $\tilde{Q}=Q=-0.034+0.003i$ (n and Q correspond to iron), $\chi_2/\chi_1=0.1$, $\chi_j/\chi_1=0.01i$, $\mathrm{Im}\chi_1=0$, and $j=3,\ldots,6$ (the values of these ratios are close to the values given in Refs. 34 and 35). For s polarization of the incident wave, the direction for p polarization was chosen as the direction for measuring the angle $\tilde{\theta}_s$. As can be seen from the curves, the polarizations of the reflected waves with the frequencies ω and ω_s are nearly orthogonal in this case. It should be noted that in the case of normal incidence the nonlinear polarization **P** has only one nonzero component P_3, and, as follows from (6.1), (6.2), and (6.3), the reflected wave is totally absent since $\alpha=0$. However, the polarization factor $\tilde{\chi}$ is nonzero because of the linear dependence of the field components on α. This means that as $\varphi\to0$, the polarization plane rotation angle and the ellipticity can have fairly

large values (as can be seen from Fig. 3), while the intensity of the reflected wave becomes infinitesimally small.

7. NONLINEAR MERIDIONAL KERR EFFECT

If the incident wave is linearly polarized and $\mathbf{m}=(0,1,0)$, the reflected waves with frequencies equal to ω and ω_s will be elliptically polarized with the major axes of the polarization ellipsoids turned through certain angles (Kerr rotation) relative to the s polarization of these waves. In order to find the Kerr angle and the ellipticity for the second harmonic, we must substitute the expressions (5.7) for normal modes into the boundary conditions (5.3) and use (5.9). As a result, the s and p components of the reflected wave with the frequency ω_s will be as follows:

$$\tilde{E}_s^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{X}^+}\left[P_1+\frac{i\alpha\tilde{n}^2\tilde{Q}\cos\psi}{2\tilde{Y}^+\sqrt{\tilde{n}^2-\alpha^2}}P_2 - iQ\left(1+\frac{\alpha^2\tilde{n}_0}{2\tilde{Y}^+\sqrt{\tilde{n}^2-\alpha^2}}\right)P_3\right], \tag{7.1}$$

$$\tilde{E}_p^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{Y}^+}\left[\alpha P_3+\sqrt{\tilde{n}^2-\alpha^2}P_2+\frac{i\alpha\tilde{n}^2\tilde{Q}}{2\tilde{X}^+\sqrt{\tilde{n}^2-\alpha^2}}P_1\right],$$

where, according to (2.2), the components of the nonlinear surface polarization are

$$P_1=\chi_1E_1E_3+(\chi_3+\chi_4)E_1^2+(\chi_4+\chi_6)E_2^2+(\chi_4-\chi_5)E_3^2,$$

$$P_2=\chi_1E_2E_3+(\chi_3-\chi_6)E_1E_2,$$

$$P_3=\chi_2(E_1^2+E_2^2)+(\chi_1+\chi_2)E_3^2+(\chi_3+\chi_5)E_1E_3,$$

and the complex amplitudes E_1, E_2, and E_3 can be found from (4.5).

If the incident wave has s polarization, then

$$\tilde{E}_s^{(r)}=-\frac{k_{0s}}{\varepsilon_0\tilde{X}^+}\left(\frac{2n_0\cos\varphi}{X^+}E_s^{(i)}\right)^2\times\left[\chi_1Q\left(1+\frac{\alpha^2n_0}{2Y^+\sqrt{n^2-\alpha^2}}\right)-\chi_2\tilde{Q}\left(1+\frac{\alpha^2\tilde{n}_0}{2\tilde{Y}^+\sqrt{\tilde{n}^2-\alpha^2}}\right)-i(\chi_3+\chi_4)\right], \tag{7.2}$$

$$\tilde{E}_p^{(r)}=\frac{i\alpha k_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{X^+}E_s^{(i)}\right)^2\chi_2,$$

and if it has p polarization, then

$$\tilde{E}_s^{(r)}=\frac{k_{0s}}{\varepsilon_0\tilde{X}^+}\left(\frac{2n_0\cos\varphi}{Y^+}E_p^{(i)}\right)^2\times\left\{\chi_1\left[\alpha^2\tilde{Q}\left(1+\frac{\alpha^2\tilde{n}_0+\tilde{n}^2\sqrt{n^2-\alpha^2}\cos\psi}{2\tilde{Y}^+\sqrt{\tilde{n}^2-\alpha^2}}\right)+\frac{\alpha^2n^2Q}{2X^+\sqrt{n^2-\alpha^2}}\right]+\chi_2n^2\tilde{Q}\left(1+\frac{\alpha^2\tilde{n}_0}{2\tilde{Y}^+\sqrt{\tilde{n}^2-\alpha^2}}\right)+i\chi_4n^2-i\chi_5\alpha^2+i\chi_6(n^2-\alpha^2)\right\}, \tag{7.3}$$

$$\tilde{E}_p^{(r)}=\frac{i\alpha k_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{Y^+}E_p^{(i)}\right)^2\times[\chi_1(\alpha^2-\sqrt{(\tilde{n}^2-\alpha^2)(n^2-\alpha^2)})+\chi_2n^2].$$

Determining the polarization factor $\tilde{\chi}$ from (7.1) or (7.2) and (7.3) and substituting it into (3.1) and (3.2), we can easily obtain the polarization plane rotation angle $\tilde{\theta}$ and the ellipticity $\tilde{\eta}$ of the reflected wave. For small values of φ and s polarization of the incident wave we have

$$\tilde{\chi}=-\frac{\tilde{E}_p^{(r)}}{\tilde{E}_s^{(r)}}=\frac{in_0\varphi}{\tilde{n}}\frac{\chi_2}{\chi_1Q-\chi_2\tilde{Q}-i(\chi_3+\chi_4)}, \tag{7.4}$$

and for p polarization we have

$$\tilde{\chi}=\frac{\tilde{E}_s^{(r)}}{\tilde{E}_p^{(r)}}=\frac{n\tilde{n}}{i\tilde{n}_0\varphi}\frac{\chi_2\tilde{Q}+i(\chi_4+\chi_6)}{\chi_2n-\chi_1\tilde{n}}. \tag{7.5}$$

In Fig. 3b both $\tilde{\theta}$ and $\tilde{\eta}$ depend on the polarization of the incident wave and the angle of incidence in the case of the linear and nonlinear meridional Kerr effects. We used the same values of the required parameters as those used to construct the plots in Fig. 3a.

8. NONLINEAR EQUATORIAL KERR EFFECT

To calculate the intensities appearing in the definition (3.3) of this effect we need the complex amplitudes $\tilde{E}_s^{(r)}$ and $\tilde{E}_p^{(r)}$, which are found after plugging the normal modes (5.8) into the boundary conditions (5.3) for $\mathbf{m}=(1,0,0)$:

$$\tilde{E}_p^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{Y}^+}\left\{\sqrt{\tilde{n}^2-\alpha^2}P_2+\alpha P_3-\frac{i\tilde{n}^2\tilde{Q}}{\tilde{Y}^+}\times[\alpha\cos\psi P_2-(\tilde{n}_0+\sqrt{\tilde{n}^2-\alpha^2}\cos\psi)P_3]\right\}, \tag{8.1}$$

$$\tilde{E}_s^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{X}^+}P_1,$$

where, as follows from (2.2), the components of the nonlinear polarization

$$P_1=\chi_1E_1E_3-(\chi_3-\chi_6)E_1E_2,$$

$$P_2=\chi_1E_2E_3-(\chi_4+\chi_6)E_1^2-(\chi_3+\chi_4)E_2^2-(\chi_4-\chi_5)E_3^2,$$

$$P_3=\chi_2(E_1^2+E_2^2)+(\chi_1+\chi_2)E_3^2-(\chi_3+\chi_5)E_2E_3$$

are specified in terms of the complex amplitudes E_1, E_2, and E_3, in accordance with (4.7). Regardless of whether the incident wave has s or p polarization, the relation $\tilde{E}_s^{(r)}=0$ holds; i.e., the reflected wave with the frequency ω_s will have only a p component (see Fig. 2). This property is also observed in the case of films, as has been confirmed experimentally.[7] If the incident wave has s polarization, then

$$\tilde{E}_p^{(r)}=\frac{ik_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{X^+}E_s^{(i)}\right)^2\left\{\chi_2\left[\alpha+\frac{i\tilde{n}^2\tilde{Q}}{\tilde{Y}^+}\times(\tilde{n}+\sqrt{\tilde{n}^2-\alpha^2}\cos\psi)\right]-(\chi_4+\chi_6)\sqrt{\tilde{n}^2-\alpha^2}\right\}, \quad (8.2)$$

and for p polarization we have

$$\begin{aligned}\tilde{E}_p^{(r)}=&\frac{ik_{0s}}{\varepsilon_0\tilde{Y}^+}\left(\frac{2n_0\cos\varphi}{Y^+}E_p^{(i)}\right)^2\\&\times\left\{\chi_1\left[-in^2Q\left(\sqrt{\tilde{n}^2-\alpha^2}\left(1-\frac{2\alpha^2\cos\varphi}{Y^+}\right)-\frac{2\alpha^2}{Y^+}(n_0+\sqrt{n^2-\alpha^2}\cos\varphi)\right)\right]\right.\\&+\alpha[\alpha^2-\sqrt{(\tilde{n}^2-\alpha^2)(n^2-\alpha^2)}]\\&+\frac{i\alpha^2\tilde{n}^2\tilde{Q}}{\tilde{Y}^+}[\tilde{n}_0+(\sqrt{\tilde{n}^2-\alpha^2}+\sqrt{n^2-\alpha^2})\cos\psi]\\&+\chi_2n^2\left[\alpha+\frac{2i\alpha^2n_0Q}{Y^+}+\frac{i\tilde{n}^2\tilde{Q}}{\tilde{Y}^+}\times(\tilde{n}_0+\sqrt{\tilde{n}^2-\alpha^2}\cos\psi)\right]+\chi_3\sqrt{\tilde{n}^2-\alpha^2}[\alpha^2\\&-\sqrt{(\tilde{n}^2-\alpha^2)(n^2-\alpha^2)}]-\chi_4n^2\sqrt{\tilde{n}^2-\alpha^2}\\&\left.+\chi_5\alpha^2(\sqrt{\tilde{n}^2-\alpha^2}+\sqrt{n^2-\alpha^2})\right\}. \quad (8.3)\end{aligned}$$

In order to determine the intensity of the reflected wave I_0, all the magnetization-dependent parameters $(Q,\tilde{Q},\chi_3,\ldots,\chi_6)$ must be set equal to zero.

After calculating the intensity from (8.1) or (8.2) and (8.3) and substituting it into (3.3), we obtain the characteristic $\tilde{\delta}$ of the nonlinear equatorial Kerr effect. We note that the asymptotic expansion $\tilde{\delta}=1+O(\varphi^2)$ holds as $\varphi\to0$, while the relative change in intensity $(I-I_0)/I_0\propto\varphi^{-2}$.

Figure 4 shows plots of $\tilde{\delta}$ versus φ for s and p polarizations of the incident wave. The characteristic δ of the linear equatorial effect was determined in a similar manner for comparison. All the parameters correspond to those for which the plots in Fig. 3 were constructed.

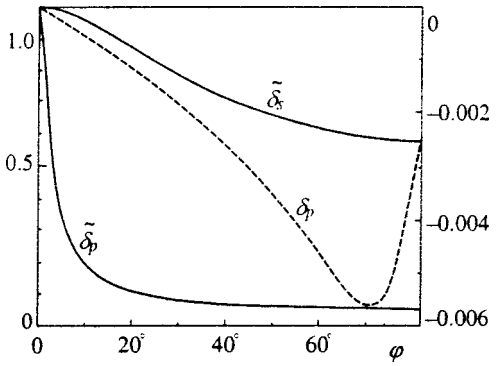

FIG. 4. Dependence of the nonlinear (solid lines) and linear (dashed line, right-hand scale) equatorial Kerr effects on the angle of incidence and polarization of the incident wave.

9. CONCLUSIONS

The principal results of this work include the analytical expressions found for a semi-infinite magnetic medium, which characterize the s and p components of the reflected second-harmonic wave in terms of optical and magneto-optical parameters of the media provided the uniform-magnetization directions correspond to the three directions which were adopted in the classification of the linear magneto-optical Kerr effects. All the results are given in the linear approximation with respect to the magnetization-dependent parameters. Equations which relate the s and p components of the reflected second-harmonic wave to the nonlinear surface electric polarization have been obtained. This made it possible to determine the polarization plane rotation angle (the Kerr rotation) and ellipticity of that wave for arbitrary polarization of the incident wave.

Expressions which characterize the nonlinear magneto-optical Kerr effects, viz., the polar, meridional, and equatorial effects, have been found. Plots of these effects as functions of the angle of incidence for incident waves with s or p polarization have been presented for known values of the parameters. It has been shown that they significantly surpass the corresponding linear magneto-optical Kerr effects. Although application of the method described here to layered media would be of greater interest, the results which we obtained can be useful for illustrating the features of the nonlinear magneto-optical Kerr effects.

*)E-mail: zvezdin@magnof.phys.msu.su

[1)]A great deal of attention has also been attracted by bulk nonlinear magneto-optical effects, especially in so-called magnetoelectric materials, in which the second harmonic appears because of the odd symmetry of the magnetic structure with respect to space inversion.[15,16] Sharp intensification of second-harmonic generation due to the appearance of magnetic order has been observed in $BiFeO_3$ below the transition point to the antiferromagnetic state.[17] The nonlinear magneto-optical effects in the magnetoelectric material Cr_2O_3 were thoroughly studied in Ref. 18. They are also clearly expressed in films of magnetic garnets[19–22] and in the Heussler alloy.[23]

[2)]The nonlinear equatorial Kerr effect was also considered for small angles of incidence in Refs. 24 and 25.

[1] Ru-Pin Pan, H. D. Wei, and Y. R. Shen, Phys. Rev. B **39**, 1229 (1989).
[2] J. Reif, J. C. Zink, C. M. Schneider, and J. Kirschner, Phys. Rev. Lett. **67**, 2878 (1991).
[3] G. Spierings, V. Koutsos, H. A. Wierenga, M. W. J. Prins, D. Abraham, and T. Rasing, Surf. Sci. **287**, 747 (1993).
[4] U. Pustogowa, W. Hübner, and K. H. Bennemann, Phys. Rev. B **49**, 10 031 (1994).
[5] H. A. Wierenga, W. de Jong, M. W. J. Prins, T. Rasing, R. Vollmer, A. Kirilyuk, H. Schwabe, and J. Kirschner, Phys. Rev. Lett. **74**, 1462 (1995).
[6] B. Koopmans, M. Groot Koerkamp, T. Rasing, and H. van den Berg, Phys. Rev. Lett. **74**, 3692 (1995).
[7] R. Vollmer, A. Kirilyuk, H. Schwabe, J. Kirschner, H. A. Wierenga, W. de Jong, and T. Rasing, J. Magn. Magn. Mater. **148**, 295 (1995).
[8] G. Spierings, V. Koutsos, H. A. Wierenga, M. W. J. Prins, D. Abraham, and T. Rasing, J. Magn. Magn. Mater. **121**, 109 (1993).
[9] H. A. Wierenga, M. W. J. Prins, and T. Rasing, Physica B **204**, 281 (1995).
[10] T. M. Crawford, C. T. Rogers, T. J. Silva, and Y. K. Kim, J. Appl. Phys. **81**, 4354 (1997).
[11] T. M. Crawford, C. T. Rogers, T. J. Silva, and Y. K. Kim, IEEE Trans. Magn. **MAG-38**, 3598 (1997).
[12] H. A. Wierenga, W. de Jong, M. W. J. Prins, T. Rasing, R. Vollmer, A. Kirilyuk, H. Schwabe, and J. Kirschner, Phys. Rev. Lett. **74**, 1462 (1995).
[13] T. A. Luce, W. Hübner, and K. H. Bennemann, Phys. Rev. Lett. **77**, 2810 (1996).
[14] A. Kirilyuk, T. Rasing, R. Megy, and P. Beauvillain, Phys. Rev. Lett. **77**, 4608 (1996).
[15] N. N. Akhmediev and A. K. Zvezdin, JETP Lett. **38**, 196 (1983).
[16] N. N. Akhmediev, S. B. Borisov, A. K. Zvezdin, I. L. Lyubchanskiĭ, and Yu. V. Melikhov, Fiz. Tverd. Tela (Leningrad) **27**, 1075 (1985) [Sov. Phys. Solid State **27**, 650 (1985)].
[17] A. M. Agal'tsov, V. S. Gorelik, A. K. Zvezdin, V. A. Murashov, and D. N. Rakov, Kratk. Soobshch. Fiz. **5**, 37 (1989).
[18] M. Fiebig, D. Fröhlich, B. B. Krichevtsov, and R. V. Pisarev, Phys. Rev. Lett. **73**, 2127 (1994).
[19] A. Aktsipetrov, O. V. Braginskaya, and D. A. Esikov, Kvantovaya Élektron. **17**, 359 (1990) [Sov. J. Quantum Electron. **20**, 296 (1990)].
[20] R. V. Pisarev, B. B. Krichevtsov, V. N. Gridnev, V. P. Klin, D. Fröhlich, and C. Pahlke-Lerch, J. Phys. C **5**, 8621 (1993).
[21] G. Petrocelli, S. Martelucci, and M. Richetta, Appl. Phys. Lett. **71**, 1931 (1993).
[22] V. V. Pavlov, R. V. Pisarev, A. Kirilyuk, and T. Rasing, Phys. Rev. Lett. **78**, 2004 (1997).
[23] J. Reif, C. Rau, and E. Matthias, Phys. Rev. Lett. **71**, 1931 (1993).
[24] A. K. Zvezdin, Physica A **241**, 444 (1997).
[25] A. K. Zvezdin and V. A. Kotov, *Modern Magneto-Optics and Magneto-Optical Materials*, Institute of Physics, Bristol, England (1997).
[26] P. S. Pershan, Phys. Rev. **130**, 919 (1963).
[27] S. Kielich and R. Zavodny, Opt. Acta **20**, 867 (1973).
[28] E. B. Graham and R. E. Raab, Philos. Mag. B **66**, 269 (1992).
[29] S. S. Girgel' and T. V. Demidova, Opt. Spektrosk. **62**, 63 (1987) [*sic*].
[30] Y. R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York (1984) [Russ. transl., Nauka, Moscow (1989)].
[31] W. Hübner and K. H. Bennemann, Phys. Rev. B **40**, 5973 (1989).
[32] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam (1977).
[33] B. Jérôme and Y. R. Shen, Phys. Rev. E **48**, 4556 (1993).
[34] U. Pustogowa, W. Hübner, and K. H. Bennemann, Phys. Rev. B **48**, 8607 (1993).
[35] U. Pustogowa, W. Hübner, and K. H. Bennemann, Surf. Sci. **307–309**, 1129 (1994).

Translated by P. Shelnitz

Flattening of vacancy force fields on a kinematic interface between solids

V. V. Meshcheryakov,*) G. D. Kuznetsov, and A. A. Yuldashev

Moscow State Institute of Steel and Alloys, 117936 Moscow, Russia
(Submitted 17 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 157–167 (July 1999)

It is shown that the diffusive formation of the boundary of a crystal moving uniformly over the surface of another crystal should be accompanied by flattening of the displacement fields of the crystal lattice in the vicinity of vacancies. As the relative velocity of the crystals rises, the flattening of vacancies leads to lowering of their dipole moments and an increase in the number of contact atoms on the interface between the crystals. This phenomenon should be manifested most strongly for high rates of relative motion of the bodies and for small contact areas in the nanoscopic range. It is noted that the decrease in the dipole moment of a vacancy into which a contact atom diffuses can be the reason for the passage of the kinematic interface between the contacting crystals into a quasimolten state. It is concluded that friction in a polyatomic contact should differ qualitatively from friction in the monatomic contacts created in atomic-force microscopy. © *1999 American Institute of Physics.* [S1063-7761(99)01307-4]

1. INTRODUCTION

The reviews of experimental and theoretical studies of the features of the friction between solid materials in Refs. 1 and 2 show that the question of the mechanism of friction is still of great interest. A large portion of these studies focused on the properties of the monatomic contacts used in atomic-force microscopy.[3–7] One of their features is generalization of the results of experiments with monatomic contacts to the case of polyatomic contacts, which is based on an implicit assumption that the mechanisms of friction are identical in both systems. Nevertheless, there has been no experimental verification or even theoretical substantiation of this assumption.

Another portion of these studies focused directly on polyatomic contacts. Special mention should be made here of Refs. 8–10, in which the concept of ''stick-slip'' motion was adopted as a basis for developing models of the mechanism of kinematic friction in ''solid-on-solid'' systems. The essence of this mechanism is that one crystal moving over the surface of another crystal, which should be considered rigid and fixed, experiences periodic locking, which is overcome with resultant passage of the original crystal into a sliding state. In this case some physical characteristics, being averages over a lattice period, turn out to differ in absolute value from the characteristics associated with crystals in a static state. These ideas underlie the work in Refs. 8 and 9, which have a theoretical character and in which the microscopic side of the problem of kinematic friction was considered. A periodic lattice potential was proposed as the factor providing for the locked state, and the frictional force was proposed as the quantity which is averaged over the sliding states. The same ideas underlie the experimental study in Ref. 10, in which the macroscopic side of the problem was considered. However, microscopic surface irregularities were proposed as the factor providing for the locked state, and the resistance of the contact to an electron current was proposed as the quantity averaged over the sliding states.

Disregarding the question of the structural level at which slip occurs, we can see something in common in these studies: the original models were constructed on the basis of an interface with a static potential well. It is difficult to imagine how stick-slip motion can be realized without the generation of vacancies by the crystal boundaries (see, for example, Ref. 11) or, in other words, without the diffusion of atoms among vacant lattice sites under the action of external stresses. In fact, if the characteristic value of the force sufficient for detaching one atom from the surface of a crystal is assumed to be of the order of 10^{-4} dyne,[12] then under loads of the order of, for example, 10^5 dyne, the number of contact atoms diffusing into vacancies at a given moment during stick-slip motion is of the order of 10^9. The role of these 10^9 atoms in shaping the frictional forces or electrical conductance of contacts can be taken into account phenomenologically, but a more detailed approach based on evaluation of the variation of the number of diffusing atoms as a function of the external conditions for carrying the experiment is also possible.

In this paper we shall describe the dynamics of the formation of a moving force contact and give an estimate of the variation of the number of contact atoms as a function of the relative velocity of the bodies in the contact under the action of an external force. The development of Tomlinson's old conception[13] of a relationship between the number of contact atoms on an interface with an external load, which was adopted in Ref. 14, is based on a solution of the problem of the motion of a point force dipole in a crystal. A moving dipole simulates an elementary act of diffusion of an atom from a lattice point into a vacancy. As a whole, diffusion should govern two processes. At first a static equilibrium is established on the interface between the deformable crystal and the absolutely rigid substrate in the absence of relative motion between them along the contacting surfaces. This is followed by the establishment of a dynamic equilibrium, un-

der which the deformable crystal moves over the surface of the rigid substrate, causing the dynamic transport of atoms and vacancies, while maintaining a static equilibrium in the direction perpendicular to the interface. In this case the diffusive motion of an atom can be regarded as an elementary act of passage of the moving crystal from a locked state to a sliding state, and the motion formed as result of many diffusive acts of the interface can be regarded as an instantaneous picture of stick-slip motion.

A description of such a detailed mechanism of friction using the ordinary electron-phonon model of crystals would hardly be possible at present. In this paper we therefore restrict the discussion to the model of a crystal in the form of an infinite continuous medium containing moving point force sources. Nevertheless, just such an approach allows us to see in the stage of formulating the problem that the force fields of moving atoms and vacancies, which account for the relative motion of the crystals, must be flattened. The ensuing analysis shows that a consequence of the flattening of the dipole force fields should be an increase in the number of contact atoms in the interacting crystals. This result leads to conclusions which can find application in the analysis of grain-boundary slip in polycrystals and the relative motion of nanocrystals in nanocrystalline materials.

2. FORMATION OF A KINEMATIC INTERFACE

We assume that an interface between crystals or individual crystallites in polycrystals is formed only by the diffusive motion of atoms among vacant sites in the crystal lattice. In this case the movement of atoms into vacancies, whose initial velocities are equal to zero, specifies a static interface. A moving, or kinematic, interface formed by the uniform motion of one crystal relative to the surface of the other is specified by the movement of atoms into vacancies moving with an initial velocity $\mathbf{v}_0$.

The number n of contact atoms at a static interface is determined by the external load and the parameters of the interactions between atoms.[13,14] At a kinematic interface the value of n should also depend on the velocity $\mathbf{v}_0$. The basis for this can be as follows.

The movement of atoms into vacancies should be accompanied by considerable dynamic lattice strains, which can be described using point force dipoles.[14,15] Assuming that a vacancy moving in the direction of an atom fixed by an interatomic bond creates a volumetric strain in the crystal, we consider the equation

$$D_{ik}(\mathbf{r},t)s_k(\mathbf{r},t)=F_i(\mathbf{r},t) \tag{1}$$

for the dynamic displacements of lattice atoms from their equilibrium positions $s_k(\mathbf{r},t)$. The differential operator $D_{ik}(\mathbf{r},t)$ is defined in the form of a matrix:

$$D_{ik}(\mathbf{r},t)=\rho_0\delta_{ik}\frac{\partial^2}{\partial t^2}-H_{ik,mn}\frac{\partial^2}{\partial x_m\partial x_n},$$

where ρ_0 is the density of the medium, and $H_{ik,mn}$ is the Huang tensor,[15] which describes the elastic and crystallographic properties of the crystal. The uniformly moving dipole characterizing the force field of a moving vacancy can be represented by the force density

$$F_i(\mathbf{r},t)=-P_{in}^{(0)}\frac{\partial}{\partial x_n}\delta(\mathbf{r}-\mathbf{v}_0t), \tag{2}$$

where $P_{in}^{(0)}$ is the static force-dipole tensor.[15]

The solution of Eq. (1) with an inhomogeneity of the type (2) is known from problems concerning moving electrodynamic multipoles.[16] The principal result for a multipole moving with a velocity not exceeding the rate of propagation of electromagnetic waves in the medium is flattening of the level lines of its force field. The degree of flattening depends on the velocity of the multipole.

The analogous effect in a deformation problem should be expressed in the form of the flattening of the displacement field $s_k(\mathbf{r},t)$ in the vicinity of a vacancy and, therefore, in the form of the change in its force state, which can be described by a force-dipole tensor. The change in the dipole tensor of a vacancy, in turn, should lead to a change in the force state of the fixed atom which diffuses into that vacancy. Under the conditions of constancy of the external load and static equilibrium of the contact pair (in the direction perpendicular to the interface) a change in the force state of a contact atom should lead to a change in the total number n of contact atoms on the interface between the crystals. Finally, since the degree of flattening of a vacancy should be determined by the dipole velocity $\mathbf{v}_0$, we conclude that a function $n(\mathbf{v}_0)$ exists.

3. FLATTENING OF THE ELASTIC FIELD OF A VACANCY

We restrict the analysis to the solution of Eq. (1) for the case of an infinite crystal, which implies disregard of the possible differences in the bulk structure of the strain fields of vacancies located on the surface of the crystal and in its bulk. The tools of elasticity theory together with the conception of a point force dipole give satisfactory results for a discrete medium,[15] but the problems which can be solved within such an approach have been formulated mainly for an infinite crystal. This is due not only to the technical difficulties in taking into account nontrivial boundary conditions, but also to the futility of quantitatively refining the solutions, which, at best, should have a qualitative character. We shall therefore restrict the discussion below to qualitative results and quantitative estimates with no more than order-of-magnitude accuracy and we shall disregard the known differences in the forces of interatomic bonds and, therefore, the dipole moments of vacancies that move along the surface of a crystal and in its bulk.

We seek the solution of Eq. (1) for one component of the displacements of a point in a continuous isotropic medium described by two elastic constants. Setting

$$H_{ik,mn}=c_{44}\delta_{ik}\delta_{mn}+(c_{11}-c_{44})\delta_{ikmn},$$

where c_{11} and c_{44} are the longitudinal and transverse elastic constants, we obtain the equation

$$\rho_0 \frac{\partial^2 s_x(\mathbf{r},t)}{\partial t^2} - \left[c_{11} \frac{\partial^2}{\partial x^2} + c_{44} \left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] s_x(\mathbf{r},t)$$
$$= - \left[P_{xx}^{(0)} \frac{\partial}{\partial x} + P_{xy}^{(0)} \frac{\partial}{\partial y} + P_{xz}^{(0)} \frac{\partial}{\partial z} \right] \delta(\mathbf{r} - \mathbf{v}_0 t).$$

After we go over to the new variables $\tilde{x} = \xi x$, $\tilde{y} = y$, and $\tilde{z} = z$, where $\xi = \sqrt{c_{44}/c_{11}}$, it takes the following form:

$$\left[\nabla_{\tilde{\mathbf{r}}}^2 - \frac{1}{c_t^2} \frac{\partial^2}{\partial t^2} \right] s_x(\tilde{\mathbf{r}},t)$$
$$= \frac{1}{\sqrt{c_{11}/c_{44}}} \left[\xi P_{xx}^{(0)} \frac{\partial}{\partial \tilde{x}} + P_{xy}^{(0)} \frac{\partial}{\partial \tilde{y}} + P_{xz}^{(0)} \frac{\partial}{\partial \tilde{z}} \right]$$
$$\times \delta(\tilde{x} - \xi v_{0x} t)\, \delta(\tilde{y} - v_{0y} t)\, \delta(\tilde{z} - v_{0z} t), \qquad (3)$$

where $c_t = \sqrt{c_{44}/\rho_0}$ is the velocity of transverse strains of the medium.

The solution of Eq. (3) can be found using an expansion of the displacements in monochromatic plane waves. Using standard transformations, we obtain the Fourier component of the displacement field

$$s_x(\mathbf{k},\omega) = \frac{i[\xi P_{xx}^{(0)} k_x + P_{xy}^{(0)} k_y + P_{xz}^{(0)} k_z] \delta(\xi k_x v_{0x} + k_y v_{0y} + k_z v_{0z} - \omega)}{(2\pi)^3 \sqrt{c_{11} c_{44}} (k^2 - \omega^2/c_t^2)}.$$

Inverse Fourier transformation of the function $s_k(\mathbf{r},t)$ gives the strain field of the medium in the vicinity of a uniformly moving vacancy:

$$s_x(\mathbf{r},t) = \frac{\gamma_{xl} \gamma_{yt} \gamma_{zt} (\mathbf{p}^* \cdot \mathbf{r}^*)}{4\pi r^{*3} \sqrt{c_{11} c_{44}}}. \qquad (4)$$

Here the coefficients $\gamma_{\alpha k} = 1/\sqrt{1 - v_{0\alpha}^2/c_k^2}$, where $\alpha = x,y,z$ correspond to strains propagating either with the longitudinal velocity $c_l = \sqrt{c_{11}/\rho_0}$ when $k = l$ or with the transverse velocity c_t when $k = t$. The vectors $\mathbf{p}^*$ and $\mathbf{r}^*$ have the following components:

$$\mathbf{p}^* = (\gamma_{xl} \xi P_{xx}^{(0)}, \gamma_{yt} P_{xy}^{(0)}, \gamma_{zt} P_{xz}^{(0)}),$$

$$\mathbf{r}^* = [\gamma_{xl} \xi (x - v_{0x} t), \gamma_{yt} (y - v_{0y} t), \gamma_{zt} (z - v_{0z} t)].$$

At $t = 0$ the displacements of a vacancy moving along the z axis with the velocity $\mathbf{v}_0 = (0,0,v_{0z})$

$$s_x(\mathbf{r}, v_{0z}) = \frac{\xi^2 P_{xx}^{(0)} x + P_{xy}^{(0)} y + \dfrac{P_{xz}^{(0)} z}{1 - v_{0z}^2/c_t^2}}{4\pi \sqrt{c_{11} c_{44} \left(1 - \dfrac{v_{0z}^2}{c_t^2} \right) \left(\xi^2 x^2 + y^2 + \dfrac{z^2}{1 - v_{0z}^2/c_t^2} \right)^3}} \qquad (5)$$

have oblate level lines along the direction of motion of the vacancy.

Let us test the validity of formula (5) for the limiting value of the velocity $v_{0z} = 0$ corresponding to the case of an interface between crystals. If we assume that the static dipole tensor is formed by a vector dyad, i.e.,

$$\mathbf{P}^{(0)} = \begin{vmatrix} f_{0x} x_0 & f_{0x} y_0 & f_{0x} z_0 \\ f_{0y} x_0 & f_{0y} y_0 & f_{0y} z_0 \\ f_{0z} x_0 & f_{0z} y_0 & f_{0z} z_0 \end{vmatrix},$$

where $\mathbf{f}_0(f_{0x}, f_{0y}, f_{0z})$ and $\mathbf{r}_0(x_0, y_0, z_0)$ are, respectively, the vectors of the dipole force and its moment arm, for the simplest case of an isotropic medium with one elastic constant $\kappa = c_{11} = c_{44}$ and for a velocity $v_{0z} = 0$ formula (5) yields a relation which describes the displacement field of an elastic medium in the vicinity of a static vacancy,

$$\mathbf{s}(\mathbf{r}) = \frac{1}{4\pi\kappa r^3} \mathbf{f}_0(\mathbf{r}_0 \cdot \mathbf{r}),$$

which obeys the familiar r^{-2} decay law.[17]

The estimates of $\mathbf{s}$ for typical values of the elastic constants show that the displacements (5) amount to $10^{-2} - 10^{-1}$ of the interatomic distance in a region with a characteristic length measured in nanometers. Therefore, the flattening of the displacement fields on a kinematic interface can alter the distance between the moving bodies and make a contribution to the variation of the parameters of the interatomic interactions. This contribution will clearly be competitive toward the change in the forces of the interatomic bonds when a vacancy passes from the bulk to the surface. Nevertheless, we also disregard this contribution.

The static displacements $\mathbf{s}(\mathbf{r})$ are not needed below. We shall also ignore the estimates of the dynamic displacements $\mathbf{s}(\mathbf{r}, v_{0z})$ and the differences between the latter and the static displacements, since, as will be seen from the following, the decisive physical characteristic of a point force dipole moving in a crystal is its dipole moment, rather than the structure of the force field, which is specified by assigning of the dipole tensor. Such an approach is consistent with the generally accepted opinion regarding the role of the dipole tensor in describing the properties of point defects.[15]

4. DIPOLE MOMENT OF A UNIFORMLY MOVING VACANCY

The components of the dipole tensor $\mathbf{P}^{(0)}$ are not independent parameters of the model. Since the medium is assigned by the elastic constants and the atomic volume, the displacement of an atomic volume to infinity or the formation of a vacancy, which amounts to the same thing, leads to the appearance of an extended force field with the parameters $\mathbf{f}_0$ and $\mathbf{r}_0$, which should be functions of the model parameters. For the model of a medium with a moving source, the independent parameters are supplemented by the source velocity $\mathbf{v}_0$. Thus, on the basis of the deduction of flattening of the force field of a moving source, we can pose the question of finding the dependence of $\mathbf{f}_0$ and $\mathbf{r}_0$ or a combination thereof on $\mathbf{v}_0$.

We utilize the fact that the displacement function (5), which is applied to the description of displacements of points in an atomic lattice, should be normalized to the dipole for-

mation volume Ω. Since a dipole describes the force field of a vacancy in the problem under consideration, we set Ω equal to the atomic volume. Then $\Omega = m/\rho_0$, where m is the atomic mass. In the absence of experimental data on the number n of contact atoms, it is reasonable to make estimates with no more than order-of-magnitude accuracy. Therefore, the known possible deviations of the vacancy formation volume from Ω can be ignored.

We assume that the original force field in the vicinity of a vacancy is spherically symmetric. In this case the components of the dipole force $f_{0x} = f_{0y} = f_{0z}$, the components of the moment arm of the dipole force $x_0 = y_0 = z_0$, and the components of the force-dipole tensor $P_{ik}^{(0)} = f_{0z} z_0$. We calculate the normalization integral

$$\int \mathbf{s}(\mathbf{r}, v_{0z})\, d\mathbf{S} = \Omega \tag{6}$$

in the approximation $\xi = 1$, for which the displacement component is

$$s_\alpha = \frac{\gamma f_{0z} z_0 (x + y + \gamma^2 z)}{4\pi\kappa \sqrt{(x^2 + y^2 + \gamma^2 z^2)^3}},$$

where $\gamma = 1/\sqrt{1-\beta^2}$, $\beta = v_{0z}/c_\kappa$, and $c_\kappa = \sqrt{\kappa/\rho_0}$. In this case the integral (6) takes the form

$$A \int \frac{[2xy + (\gamma^2 + 1)(x + y)z]\, dS}{r\sqrt{(x^2 + y^2 + \gamma^2 z^2)^3}} + A \int \frac{dS}{r\sqrt{x^2 + y^2 + \gamma^2 z^2}}, \tag{7}$$

where we have introduced the notation $A = \gamma f_{0z} z_0 / 4\pi\kappa$. The transition to spherical coordinates reveals that the first integral in (7) is equal to zero, and the second integral takes the form

$$A \int \frac{\sin\theta\, d\theta}{\sqrt{\sin^2\theta + \gamma^2 \cos^2\theta}}. \tag{8}$$

A calculation of the integral (8),

$$\frac{2\pi A}{\sqrt{\gamma^2 - 1}} \int_{-1}^{1} \frac{du}{\sqrt{u^2 + 1/(\gamma^2 - 1)}} = \frac{2\pi A}{\sqrt{\gamma^2 - 1}} \ln \left| \frac{\sqrt{1 + 1/(\gamma^2 - 1)} + 1}{\sqrt{1 + 1/(\gamma^2 - 1)} - 1} \right|,$$

and conversion to the parameter β give the value of the integral

$$\int \mathbf{s}(\mathbf{r}, v_{0z})\, d\mathbf{S} = \frac{f_{0z} z_0}{2\kappa\beta} \ln \left| \frac{1 + \beta}{1 - \beta} \right| = \frac{f_{0z} z_0 \tanh^{-1}\beta}{\kappa\beta}. \tag{9}$$

Substituting (9) into the original normalization integral (6) and returning to the vector notation, we obtain the scalar product

$$\mathbf{f}_0 \cdot \mathbf{r}_0 = \frac{3\beta}{\tanh^{-1}\beta} \kappa\Omega, \tag{10}$$

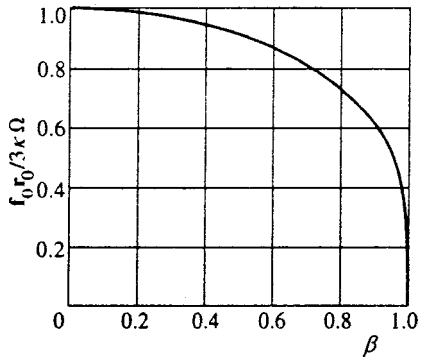

FIG. 1. Variation of the relative dipole moment of a uniformly moving vacancy.

which characterizes the variation of the components of the dipole tensor of a uniformly moving force dipole as a function of its velocity.

Figure 1 presents an illustration of formula (10). It follows from the plot of $\beta/\tanh^{-1}\beta$ that an increase in the velocity of a dipole moving in an elastic medium is accompanied by a decrease in its dipole moment. The physical basis of the decrease in the dipole moment is the condition of conservation of the formation volume of the dipole as its force field is flattened. This condition requires compensation for the increase in the magnitude of the displacements of the deformable medium with increasing velocity of the dipole due to the decrease in the dipole force and its moment arm. In other words, if the formation volume remains constant, the vacancy becomes less rigid as its velocity increases. When $\beta = 1$, i.e., when the dipole moves with the velocity of sound, the force-dipole moment vanishes. It can be stated that a linear continuous medium ''does not see'' its own dipole force source, which moves with the strain propagation velocity in the medium. When $\beta = 0$, formula (10) leads to the familiar relation $\mathbf{f}_0 \cdot \mathbf{r}_0 = 3\kappa\Omega$ (Ref. 14), which specifies the components of the dipole tensor of a static force dipole.

The decrease in the dipole moment of a vacancy into which an atom located on the interface between the media diffuses leads to a decrease in the energy of formation E of the vacancy, since $E \propto P^{(0)}(\beta)$, and to a decrease in the self-diffusion energy. This can correspond to passage of the interface into a quasimolten state similar to the one observed in particles of small dimensions.[18] The quasimelting of the interface can, in turn, be the cause of both the decrease in the kinematic frictional force in comparison to the static frictional force and the increase in the electrical resistance of a moving contact in comparison to a stationary contact. A detailed analysis of these factors requires the use of a more complicated interaction Hamiltonian than the one which leads to Eq. (1). This applies, in particular, to the problem of the electrical conductance of a kinematic contact, whose solution is associated with the need to ascertain the particular features of the percolation of electrons through an interface in a system of ions having translational degrees of freedom.

The relation (10) also allows us to state that the decrease in the dipole moment of a vacancy into which a contact atom diffuses should alter the number of atom-vacancy pairs pro-

viding for the kinematic contact between the solids.

5. NUMBER OF CONTACT ATOMS ON A KINEMATIC INTERFACE

The number of contact atoms can be found by utilizing the condition of static equilibrium between the crystals along the x axis, which is perpendicular to the direction of their relative motion. For an external force F_0 perpendicular to the interface between the crystals, the number of contact atoms should be

$$n(\beta)=\frac{F_0}{f_{0x}(\beta)}. \tag{11}$$

If $f_{0x}<F_0/n$, the external force per interfacial atom exceeds the force of an interatomic bond (without allowance for the difference between surface and bulk bonds), which leads to diffusive redistribution of the atoms at the interface between the crystals under the action of the force F_0 and to an increase in the number of atoms n.

The value of f_{0x} is unknown, but it appears in the scalar product (10). Additional relations which permit the determination of f_{0x} can be established from the solution of the dynamic problem of a moving atom in a crystal. This problem is based on Eq. (1), in which the force nonuniformity,

$$F_i(\mathbf{r},t)=f_i(t)\,\delta(\mathbf{r})-P_{in}(t)\,\frac{\partial}{\partial x_n}\,\delta(\mathbf{r}), \tag{12}$$

is specified by the dynamic dipole tensor

$$\mathbf{P}(t)=\begin{vmatrix} f_{0x}x_0 & f_{0x}y_0 & f_{0x}[z_0+z(t)] \\ f_{0y}x_0 & f_{0y}y_0 & f_{0y}[z_0+z(t)] \\ [f_{0z}+f_z(t)]x_0 & [f_{0z}+f_z(t)]y_0 & [f_{0z}+f_z(t)][z_0+z(t)] \end{vmatrix},$$

where $\mathbf{f}(f_x(t),f_y(t),f_z(t))$ is the vector of the force acting on an atom with a mass m, and $\mathbf{r}(x(t),y(t),z(t))$ is the displacement vector of the atom.

The solution of Eq. (1) with the force density (11) was given in Ref. 14 in the wave-band approximation. The results include the frequency of the field of limiting strains,

$$\omega=(16\kappa c_\kappa/m)^{1/3}, \tag{13}$$

at which the absorption of energy occurs in the system, and the relation between the vectors of the dipole force and its moment arm,

$$\mathbf{f}_0=-m\omega^2\mathbf{r}_0, \tag{14}$$

which ensures that the moments of the forces appearing upon deformation of the crystal are equal to zero.

A combined solution of Eqs. (10), (13), and (14) gives the component of the dipole force vector

$$f_{0x}(\beta)=f_{0x}(0)\sqrt{\frac{\beta}{\tanh^{-1}\beta}}, \tag{15}$$

where $f_{0x}(0)=16^{1/3}\kappa\Omega^{2/3}$, and the component of the vector of the moment arm of the dipole force

$$x_0(\beta)=(\Omega/16)^{1/3}\sqrt{\beta/\tanh^{-1}\beta},$$

which depend on the rate of uniform motion of the crystal relative to the rigid substrate.

Assuming that the magnitude of the force F_0 is assigned, from Eqs. (11) and (15) we find that the number of contact atoms on the kinematic interface in a crystal/absolutely-rigid-substrate system is

$$n(\beta)=n(0)\sqrt{\frac{\tanh^{-1}\beta}{\beta}}, \tag{16}$$

where $n(0)=F_0/f_{0x}(0)$ is the number of contact atoms at the stationary boundary.

An analysis of Eqs. (10) and (16) shows that the decrease in the vacancy dipole moment caused by the flattening of its force field should lead to an increase in the number of contact atoms with increasing velocity of the crystal that moves relative to the substrate. Since the number of contact atoms at the stationary interface varies in the range from the value $n(0)=1$, which is realized in atomic-force microscope, to the highest theoretically possible value, which is restricted by the number of atoms on the surface of the crystal $n(0)\approx N^{2/3}$, where N is the total number of atoms in the sample, the intermediate and most typical value $n(0)\approx 10^9$, which corresponds to a load $F_0=10^5$ dyne, can be used for estimates.

For example, assuming that $c_\kappa\sim 10^5$ cm/s within an order of magnitude, for a crystal or a grain boundary in a polycrystal moving with a velocity $v_0=10^2$ cm/s we obtain an increment of the number of contact atoms

$$\Delta n=n(\beta)-n(0),$$

equal to $\Delta n(10^{-3})\approx 10^2$. Another pair of estimates for other values of β gives $\Delta n(10^{-2})\approx 10^4$ and $\Delta n(10^{-1})\approx 10^6$. These results are illustrated in Fig. 2. The function $\Delta n(\beta)$ has a logarithmic character because the vacancy dipole moment vanishes as $\beta\to 1$. At fairly high rates of relative motion of the crystals the change in the number of contact atoms can be very significant. It thus follows that the flattening of the vacancy force field, which can be ignored in the case of a single diffusion event on a fairly extensive boundary, can make an appreciable contribution to the changes in properties which depend on the number of contact atoms on a kinematic interface. Thus, the phenomenon of the flattening of vacancies and alteration of the number of contact atoms

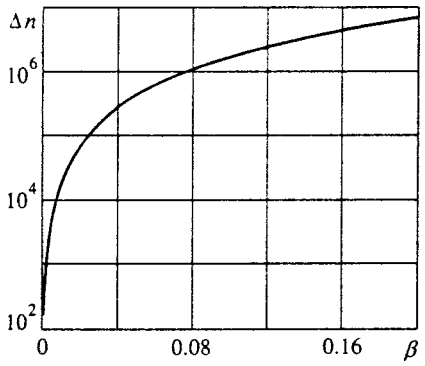

FIG. 2. Dependence of the increment of the number of contact atoms on the relative velocity of crystals having a number of contact atoms $n(0)\approx 10^9$ in the static state.

can be especially significant in two extreme cases: the case of high rates of motion of the bodies, which is realized, for example, during relative motions of grain boundaries,[19] and the case of small dimensions of the contact areas in the nanoscopic range, which is realized in nanocrystalline materials. In the latter case the approach of the number of contact atoms $n(\beta)\approx N^{2/3}$ to the number of atoms on the surface of a moving nanocrystal or an individual crystallite of the nanocrystalline material causes sharp enhancement of the influence of the individual strain fields on the properties not only of the interfaces, but also on their bulk properties.[20] Thus, the softening of the phonon spectrum in the region of the long-wavelength boundary can lead to the appearance of features in the flow of current through a contact pair.

6. CONCLUSIONS

The solution of the problem of the dipole moment of a force source moving uniformly in a crystal relative to a static boundary shows that the changes in the properties of a kinematic interface between crystals in comparison to a static boundary is associated with a need to take into account the strain fields of the crystal created by the dynamics of the atom-vacancy pair. The displacement of atoms in the crystal from static equilibrium positions into moving positions leads to flattening of the vacancy force fields and a change in the number of contact atoms. Hence it follows that friction in a polyatomic contact should differ qualitatively from friction in a monatomic contact. This conclusion imposes a restriction on the possible generalization of the results of experiments with a monatomic atomic-force microscope tip to the case of systems with polyatomic contacts.

The qualitative solution of the problem of the flattening of the vacancy force field does not require knowledge of the features of the interatomic interactions (which is associated with allowance for the character of the conductivity of the materials) or consideration of the difference between the character of the bulk diffusion and that of the surface diffusion. At the same time, the inclusion of these factors would make it possible to give more exact quantitative estimates and, therefore, to examine a concrete experimental situation. On the other hand, at present there is no possibility for observing the dynamics of the formation of a force contact *in situ*, although the alternative of measuring the number of contact atoms as a function of the relative velocity of the bodies has not been ruled out. Such an experiment can be based on measurements of the current flowing through a contact pair. In this case, however, the approach described in this paper will not be adequate for explaining the variation of the current. The increase in the number of contact atoms as the crystal moves should increase the electrical conductance of the entire contact, which is a parallel group of point contacts, but the decrease in the vacancy dipole moment can diminish the conductance of an individual point contact to a considerable extent. Allowance for the latter calls for significant complication of the formulation of the problem of kinematic friction.

We wish to thank A. G. Lyapin and O. A. Kazakov for a discussion of the results of this study and for offering some critical remarks.

*)E-mail: valery@meshcheryakov.misa.ac.ru

[1] I. L. Singer, J. Vac. Sci. Technol. A **12**, 2605 (1994).
[2] U. D. Schwarz, O. Zworner, P. Koster, and R. Wiesendanger, Phys. Rev. B **56**, 6987, 6997 (1997).
[3] W. Zhong and D. Tomanek, Phys. Rev. Lett. **64**, 3054 (1990).
[4] Den Nijs Marcel, Phys. Rev. Lett. **64**, 435 (1990).
[5] Motohisa Hirano and Kasumasa Shinjo, Phys. Rev. B **41**, 11 837 (1990).
[6] M. A. Lantz, S. J. O' Shea, M. E. Welland, and K. L. Johnson, Phys. Rev. B **55**, 10 776 (1997).
[7] R. J. A. Van den Oetelaar and C. F. J. Flipse, Surf. Sci. **384**, L828 (1997).
[8] J. B. Sokolov, Thin Solid Films **206**, 208 (1991).
[9] A. Buldum and S. Ciraci, Phys. Rev. B **55**, 2606 (1997).
[10] M. E. Levinshteĭn and S. L. Rumyantsev, Phys. Solid State (St. Petersburg) **35**, 953 (1993) [Phys. Solid State **35**, 490 (1993)].
[11] V. V. Gorbunov and B. M. Darinskiĭ, Fiz. Tverd. Tela (Leningrad) **34**, 1059 (1992) [Sov. Phys. Solid State **34**, 565 (1992)].
[12] G. Binnig, C. F. Quate, and C. Gerber, Phys. Rev. Lett. **56**, 930 (1986).
[13] I. Tomlinson, Philos. Mag. **7**, 905 (1929).
[14] V. V. Meshcheryakov, Fiz. Tverd. Tela (St. Petersburg) **37**, 43 (1995) [Phys. Solid State **37**, 20 (1995)].
[15] G. Leibfried and N. Breuer, *Point Defects in Metals, Vol. 1: Introduction to the Theory*, Springer-Verlag, Berlin–New York (1978) [Russ. transl., Mir, Moscow (1981), p. 188].
[16] V. V. Batygin and I. N. Toptygin, *Problems in Electrodynamics*, Academic Press, London (1964) [newer edition of Russ. original, Nauka, Moscow (1970), p. 157].
[17] M. A. Krivoglaz, *Theory of X-Ray and Thermal Neutron Scattering by Real Crystals*, Plenum Press, New York (1969) [Russ. original, Nauka, Moscow (1967), p. 69].
[18] É. L. Nagaev, Usp. Fiz. Nauk **162**, 49 (1992) [Sov. Phys. Usp. **35**, 747 (1992)].
[19] S. G. Psakh'e and K. P. Zol'nikov, Pis'ma Zh. Tekh. Fiz. **23**(14), 44 (1997) [Tech. Phys. Lett. **23**(7), 555 (1997)].
[20] V. V. Meshcheryakov, Zh. Éksp. Teor. Fiz. **111**, 1845 (1997) [JETP **84**, 1010 (1997)].

Translated by P. Shelnitz

Strongly correlated bistable sublattice and temperature hysteresis of elastic and thermal crystal properties

A. P. Saĭko*) and V. E. Gusakov

Institute of Solid-State Physics and Semiconductors, National Academy of Sciences, 220072 Minsk, Belarus
(Submitted 30 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 168–193 (July 1999)

It is shown that in crystal lattices with a basis the cooperative behavior of a certain type of atoms performing optical long-wavelength vibrations in a double-well potential of the field of the matrix lattice may lead to the formation of a bistable sublattice. As a result of the interaction of the metastable states of such a sublattice with the vibrational states of the matrix lattice, the elastic and thermal properties of the crystal acquire anomalous, hysteresis-like, temperature curves. The concepts developed in the paper make it possible to obtain a qualitative interpretation, which agrees with the experimental data, of the hysteresis-like temperature dependence of the speed and absorption of ultrasonic waves, the specific heat, and the thermal conductivity in superconducting yttrium and bismuth cuprates. © *1999 American Institute of Physics.*
[S1063-7761(99)01407-9]

1. INTRODUCTION

Precision experiments in which the propagation of ultrasound in high-T_c superconductors[1–15] and ferroelectric conductors (see, e.g., Refs. 16 and 17) was studied detected temperature hysteresis of the speed of ultrasound (and Pal'-Val' *et al.*,[4] Kim *et al.*,[8] and Borisov *et al.*[17] detected temperature hysteresis of the absorption coefficient of ultrasound) that was found to encompass a temperature range from ten to hundred kelvins. More than that, in the same temperature range superconductors revealed hysteresis behavior of specific heat[18–20] and thermal conductivity.[21–25] What is remarkable is the large interval of temperature hysteresis and at the same time the absence of relaxation in the measured parameters in the hysteresis region (some samples were kept at a fixed temperature for several hours). Although there is still no universal opinion concerning the nature of the observed anomalies, it is obvious that they are related in one way or another to the metastable states of the crystal lattice. The absence of relaxation processes may indicate, for instance, that the metastable states form in conditions of a strong correlation of the lattice degrees of freedom, since otherwise local energy fluctuation would rapidly destroy the metastable states.

Below we show that this anomalous, hysteresis, behavior of the elastic and thermal characteristics of such compounds may be due to the presence in them of an anharmonically unstable, strongly correlated sublattice that executes optical long-wavelength vibrations in the field of the matrix lattice.

2. A MODEL OF A STRONGLY CORRELATED BISTABLE SUBLATTICE

2.1. Independent and strongly correlated particles in a double-well crystal potential; a bistable sublattice

In a crystal lattice with a multiatomic basis we examine a sublattice formed by ions of a single species. Suppose that each ion vibrates in a double-well asymmetric potential (Fig. 1) oriented, say, along one of the crystallographic axes. In the absence of ordering long-range forces the ions may be thought of as being almost entirely independent of one another. In this case, in addition to the intrawell vibrations in the limit $\Theta \ll U$ (where $\Theta = k_B T$ is the temperature expressed in energy units and U is the height of the potential barrier), the ions are capable, due to thermal fluctuations, of performing slower movements, say, hop across the potential barrier from one stable position to another with a probability $\propto \exp\{-U/\Theta\}$. At high temperatures ($\Theta \geqslant U/2$) the ions are passing, i.e., they oscillate above the barrier.[26,27]

Under realistic conditions there are always correlations between the displacements of the ions in a crystal. The correlation between the relative displacement of the ions in the sublattice may be so strong that a cooperative effect could arise in which the displacement of one atom would generate similar displacements of the neighboring ions, i.e., a coherent ensemble acting as an integral whole is formed. Such a situation has a large probability of occurring in highly polarized systems. In this case a change in an external parameter, e.g., the temperature, gives rise to a coordinated shift of the atoms of the sublattice considered. The cooperative behavior of the ions of the correlated sublattice makes the sublattice unresponsive to fluctuations since, being ''bombarded'' by the quanta of the reservoir (e.g., by the phonon of the matrix lattice), the sublattice perceives a perturbation as an integral whole. Such unresponsiveness, or rigidity, of the sublattice, which prevents the separate ions from behaving independently, extends over distances of order the coherence length and means that the transition of a separate ion from intrawell dynamics to above-barrier dynamics can occur only when the entire coherent volume undergoes such a transition, since the probability of the entire correlated ensemble consisting of n particles surmounting the barrier simultaneously is proportional to $\exp\{-nU/\Theta\}$ (here U has the meaning of the

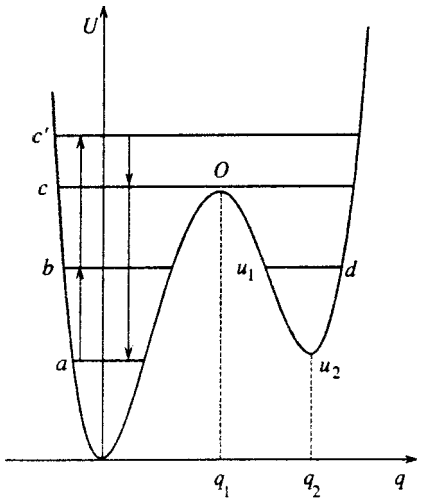

FIG. 1. The schematic of the potential and the dynamics of transitions for a strongly correlated bistable sublattice.

potential of the entire sublattice per particle); for this reason, even at $n\sim 10$ the activation transitions of the correlated ensemble (and hence of each particle comprising the ensemble) across the barrier are unlikely to occur, even at temperatures $\Theta\sim U/2$.

Thus, when being heated, a strongly correlated sublattice will evolve from vibrations in the global minimum to above-barrier vibrations, with the slower component of motion, hops from the global minimum to the local and back, being almost entirely excluded. Of course, due to renormalization, in a strongly correlated sublattice the double-well ion potential differs from the ''bare'' potential inherent in independent particles, so that it would be more natural to speak of a double-well potential for the entire coherent sublattice per particle, and all fluctuation transitions across the barrier for such a sublattice can be ignored, as we have just seen.

2.2. Model Hamiltonian and the derivation of the main relationships

To thoroughly study the dynamics of such a sublattice consisting of N particles we need to write down the sublattice Hamiltonian H_l. Disregarding insignificant details, we may assume that the coherence length extends over the entire sublattice, i.e., the lattice is a single coherent ensemble. Then

$$H_l = NH_{\rm anh}, \tag{1}$$

where $H_{\rm anh}$ is the reduced (to a single ion) Hamiltonian of the strongly correlated lattice. We write the latter Hamiltonian in the form of an anharmonic oscillator in a double-well potential with asymmetric wells formed by the field of the matrix lattice:

$$H_{\rm anh}=\frac{p^2}{2m}+\frac{\alpha}{2}q^2-\frac{\beta}{3}q^3+\frac{\gamma}{4}q^4, \tag{2}$$

where m is the ion mass and q and p are the coordinate and canonically conjugate momentum of the ion along a specified direction fixed, say, by one of the crystallographic axes.

We examine the thermal behavior of the sublattice in the approximation of self-consistent phonons.[28] To this end we introduce the statistical-mean displacement $\langle q\rangle$, the dynamic displacement $\delta q(t)=q(t)-\langle q\rangle$, and the variance $\sigma\equiv\langle(q-\langle q\rangle)^2\rangle$. Then, using the relationships $\langle(\delta q)^{2n}\rangle=(2n-1)\times\ldots\times 3\sigma^n$ and $\langle(\delta q)^{2n+1}\rangle=0$ (n is an integer), valid in the adopted approximation, we find that

$$\langle H_{\rm anh}\rangle=\frac{\alpha}{2}\langle q\rangle^2-\frac{\beta}{3}\langle q\rangle^3+\frac{\gamma}{4}\langle q\rangle^4+m\Omega^2\sigma-\frac{3}{4}\gamma\sigma^2, \tag{3}$$

where

$$\Omega^2\equiv\frac{[\alpha-2\beta\langle q\rangle+3\gamma(\sigma+\langle q\rangle^2)]}{m} \tag{4}$$

is the effective frequency characterizing the sublattice. The variance σ can be found by the fluctuation-dissipation theorem:

$$\sigma=\frac{1}{2m\Omega}\coth\frac{\Omega}{2\Theta} \tag{5}$$

(we set $\hbar=1$ throughout the paper), the relation between σ and $\langle q\rangle$ can be established from the condition for stability of the sublattice, $\langle\partial H_{\rm anh}/\partial(\delta q)\rangle=0$:

$$(\beta-3\gamma\langle q\rangle)\sigma=\alpha\langle q\rangle-\beta\langle q\rangle^2+\gamma\langle q\rangle^3. \tag{6}$$

The free energy F, which we must know in order to give a complete statistical-thermodynamic description of the system, is established by the Bogolyubov variational principle:

$$F\leqslant F_0-\langle H_{\rm anh}-H_0\rangle_0=\frac{\alpha}{2}\langle q\rangle^2-\frac{\beta}{3}\langle q\rangle^3+\frac{\gamma}{4}\langle q\rangle^4+\Theta\ln\left(2\sinh\frac{\Omega}{2\Theta}\right)-\frac{3}{4}\gamma\sigma^2, \tag{7}$$

where F_0 is the free energy corresponding to the Hamiltonian of the pseudoharmonic approximation,

$$H_0=\frac{\alpha}{2}\langle q\rangle^2-\frac{\beta}{3}\langle q\rangle^3+\frac{\gamma}{4}\langle q\rangle^4+\frac{p^2}{2m}+\frac{m}{2}\Omega^2(\delta q)^2. \tag{8}$$

The closed system of equations (3)–(8) makes it possible, at least qualitatively, to describe the thermal behavior of the bistable sublattice model considered here.

2.3. Temperature dependence of the dynamic and statistical characteristics of a bistable sublattice

The results of numerical calculations by formulas (4)–(7) are illustrated by Figs. 2 and 3.

Figure 2(a) depicts the temperature dependence of the mean displacement $\langle q\rangle$ of the sublattice. The solution represented by curve *1*, which lies below the asymptote (the dashed line) and has the shape of a hysteresis curve, describes the transition of the sublattice from the global minimum (see Fig. 1) to passing trajectories (above-barrier oscillations) when the system temperature is raised from absolute zero; in the high-temperature limit, $\langle q\rangle$ approaches its asymptotic value $\langle q\rangle_{\rm as}=\beta/3\gamma$. Above the asymptote there is the solution represented by curve *2*, which is related to the possibility of the sublattice being in the second, local, minimum at low temperatures. However, the probability that the

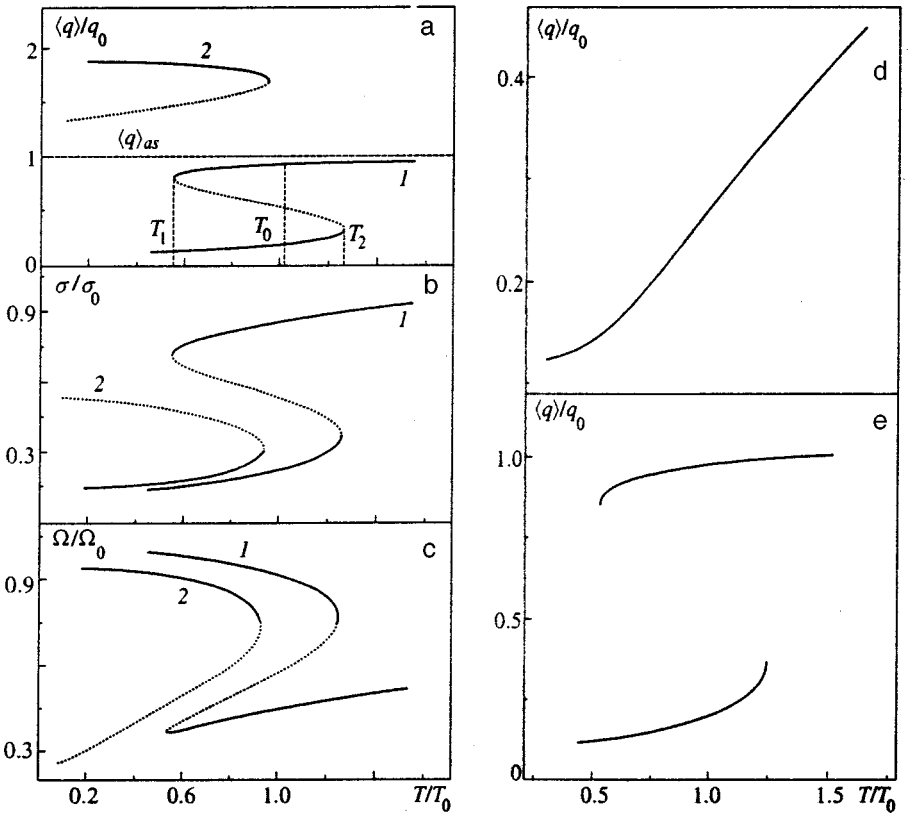

FIG. 2. Temperature dependence of the statistical-mean displacement $\langle q \rangle$ (a), the variance σ (b), and the effective frequency Ω (c) of a bistable sublattice. (d) The statistical-mean displacement $\langle q \rangle$ for an anharmonic oscillator calculated by the molecular-dynamics method with allowance for kinetic-energy fluctuations; (e) the same for the case where the kinetic-energy fluctuations are weaker by a factor of 100; $q_0 = \langle q \rangle_{as}$, $\sigma_0 = \sigma(q_0)$, and $\Omega_0 = \Omega(q_0, \sigma_0)$. The parameters of the bistable potential are $u_1 = 0.03$ eV, $q_1 = 0.073$ Å, $q_2 = 0.14$ Å, and $T_0 = 173$ K. The curves *1* and *2* describe the motion of the lattice in the global and local minima of the potential, respectively. The dotted curves represent unstable solutions.

local minimum will be occupied is extremely low, since the thermal fluctuations that would take the system from the global minimum to the local one are suppressed by the strongly correlated movements of the atoms in the sublattice (see Sec. 2.1). Hence the contribution of this solution [see also Figs. 2(b) and 2(c)], which would be effective for the case of independent particles, will be ignored throughout the paper (nor will it be depicted in the figures, with the exception of Figs. 3 and 5).

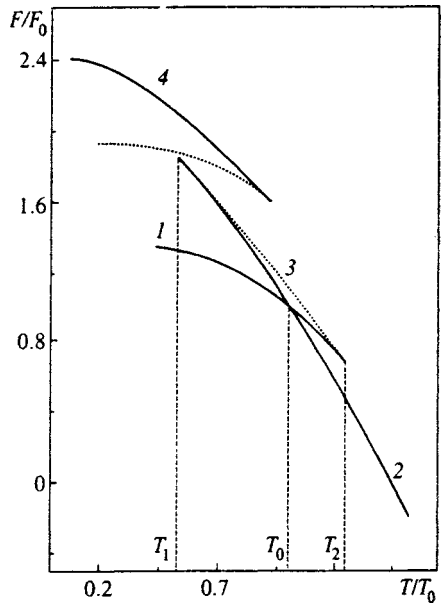

FIG. 3. Temperature dependence of the free energy of a bistable sublattice: curves *1* and *4* correspond to states in the global and local minima, curve *2* corresponds to above-barrier states, and curve *3* corresponds to an unstable solution; $u_1 = 0.03$ eV, $q_1 = 0.073$ Å, $q_2 = 0.14$ Å, and $T_0 = 173$ K.

The hysteresis behavior of the statistical-mean displacement $\langle q \rangle$, the displacement variance σ, and the effective frequency Ω of the sublattice (Fig. 2) is explained by the nature of the temperature dependence of the free energy F per particle (Fig. 3). The ''low-temperature'' branches of the hysteresis curves in Fig. 2 correspond to the free energy of the sublattice in the left, global, minimum (curve *1* in Fig. 3), while the ''high-temperature'' branches correspond to the free energy of above-barrier vibrations (curve *2* in Fig. 2); curve *3* in Fig. 3 describes unstable states; and curve *4* corresponds to solutions that refer to the positions of the sublattice in the right, local, minimum.

At the point T_0 the free energies become equal, and under the condition of total equilibrium at this point there would have been a transition of the sublattice from intrawell vibrations to above-barrier vibrations or back, depending on whether the system is heated or cooled. However, Fig. 3 shows (and so does Fig. 2) that ''overheated'' [in the interval (T_0, T_2)] or ''supercooled'' [in the interval (T_0, T_1)] metastable states may set in (see the discussion below). When heated, the sublattice, reaching the boundary of the metastable region at point T_2, suddenly changes its dynamics: it undergoes a first-order transition from intrawell vibrations to above-barrier vibrations, with the frequency decreasing approximately twofold (see Figs. 2 and 3); when cooled, the sublattice, entering the region of metastable states and approaching the region boundary at point T_1, discontinues its above-barrier motion and ''falls'' into the deeper potential

well, retaining its previous frequency of vibrations in the process. The hysteresis region $\Delta T = T_2 - T_1$ depends primarily on the depth of the local minimum: as the depth decreases, ΔT becomes smaller, and at the transition point T_c vanishes completely; at this point the first and second temperature derivatives of $\langle q \rangle$, σ, and Ω become infinite, which corresponds to a second-order transition.

Thus, at the points T_1 and T_2, where the boundaries of the metastable regions are crossed, at the point T_0 where the free energies are equal, and at the transition point T_c the states of the correlated, ordered, sublattice change, i.e., order-order transitions of the first- or second orders occur.

What is interesting is that the results of molecular-dynamics modeling agree completely with our results obtained by the self-consistent phonon approximation. Molecular-dynamics calculations (Fig. 2) show that when the fluctuation hops across the barrier are effective (the case of independent particles), even at fairly moderate temperatures the second, local, minimum may become occupied (in accordance with the Boltzmann factor $\exp\{-U/\Theta\}$), and we will have a monotonic temperature dependence [Fig. 2(d)] of the displacement and hence of the other parameters, the displacement variance and the effective frequency of the system's vibrations. As the probability of fluctuation transitions drops, due to the realization of a correlated state in the sublattice, the temperature dependence of the displacement and the other parameters acquires the shape of a hysteresis curve.

2.4. Qualitative interpretation of the hysteresis behavior of a bistable lattice with strongly correlated particles

The hysteresis behavior of a strongly correlated sublattice in an asymmetric double-well potential can easily be understood from qualitative physical considerations based on classical statistical physics.

In view of the broken symmetry, at low temperatures the sublattice is in the left, global, minimum. As a result of heating, the oscillation trajectories (the behavior of the system is examined using the tools of classical statistical physics) of the ions of the correlated sublattice gradually rise to the vertex O (Fig. 1) of the potential barrier, because the fluctuations in a coherent ensemble are suppressed (this aspect was discussed earlier). At first glance, when on path b, the sublattice (and hence each ion) could, as its temperature grows, either go over to the closest passing orbit c by reducing its velocity (since the distance between the stopping points increases) or find itself in the right well. However, both cases are impossible. Finding itself, for instance, on the path c, which is directly above the barrier, and hence reducing its kinetic energy, the sublattice would be at a lower temperature (since in the classical limit the temperature is simply the average kinetic energy), and this would violate the isothermal condition (the sublattice is in contact with the thermostat). Hence eventually the sublattice will find itself on a higher path, say c', starting from which it will rise higher and higher as the temperature increases. Neither can the sublattice go from path b in the left well to path d in the right, since it would have to lower its average kinetic energy, i.e., its temperature. We now consider the reverse course that the system takes when the temperature is lowered. As the temperature of the thermostat decreases, the sublattice gradually goes over to lower paths down to path c on which the sublattice average kinetic energy K_c, i.e., the temperature T_c, is much lower than the one ($K_{c'}$ or $T_{c'}$) that the sublattice had when it emerged from the left well and went over to passing trajectories. A further decrease in temperature will force the sublattice to ''fall'' onto one of the low-lying paths in the left well (due to the broken symmetry), say onto path a, without changing its average kinetic energy, i.e., K_a will be equal to K_c, or $T_a = T_c$. Thus, the size of the hysteresis region is $K_{c'} - K_c = K_b - K_a$ or, what is the same thing, $T_{c'} - T_c = T_b - T_a$.

3. INTERACTION BETWEEN THE BISTABLE SUBLATTICE AND THE MATRIX LATTICE

3.1. The total lattice Hamiltonian

The interaction of the anharmonic vibrations of a bistable lattice and the phonon excitations of the matrix lattice may lead to experimentally observable effects. For instance, the scattering of a traveling acoustic mode of the matrix lattice by perturbations caused by the vibrational motion of the bistable lattice in a double-well potential gives rise to singularities in the real and imaginary parts of these modes, which must be observable in experiments, in particular, in the anomalous behavior of the elastic and thermal characteristics of the crystal, such as the speed (the elastic modulus) and decay of ultrasound and the thermal conductivity. Thus, to study these questions we must focus on the interaction between the vibrational degrees of freedom of the matrix and the bistable sublattice.

We write the total lattice Hamiltonian H normalized to the number of atoms in the bistable sublattice as

$$H = H_h + H_{\text{anh}} + H_{\text{int}}. \tag{9}$$

The first term on the right-hand side of Eq. (9), H_h, models the phonon Hamiltonian of the lattice and is taken in the form of the Hamiltonian of a set of harmonic oscillators whose parameters are normalized to the empirical values of the lattice constants of the crystal:

$$H_h = \sum_k \left(\frac{p_k^2}{2\mu_k} + \frac{\mu_k \omega_k^2}{2} x_k^2 \right), \tag{10}$$

where x_k, p_k, μ_k, and ω_k are the displacement, momentum, mass, and frequency of the kth oscillator (mode). The term H_{anh} is the Hamiltonian describing the bistable sublattice; it is defined in (2). The last term in (9), H_{int}, allows for the coupling of the lattice oscillators and the bistable sublattice; it is chosen in the form of a sum of the cubic and quartet interaction terms:

$$H_{\text{int}} = H_{\text{int}}^{(3)} + H_{\text{int}}^{(4)}, \tag{11}$$

$$H_{\text{int}}^{(3)} = q^2 \sum_k \lambda_k x_k, \quad H_{\text{int}}^{(4)} = q^2 \sum_{k,k'} \lambda_{kk'} x_k x_{k'}, \tag{12}$$

where λ_k and $\lambda_{kk'}$ are the coupling coefficients.

3.2. Deriving the Dyson equation for the phonon Green's function

To find the renormalized frequencies of the matrix lattice and the corresponding decay coefficients, we use the method of equations of motion for two-time retarded Green's functions:[29]

$$\langle\langle x_k(t);x_{k'}(t')\rangle\rangle = -i\theta(t-t')\langle[x_k(t),x_{k'}(t')]\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \exp\{-i\omega(t-t')\} \times\langle\langle x_k|x_{k'}\rangle\rangle_\omega, \quad (13)$$

where $\langle\cdots\rangle$ denotes the operation of quantum-statistical averaging, $[\cdots,\cdots]$ stands for a commutator, and $\theta(\tau)$ is the Heaviside step function. Differentiating the Green's function (13) first with respect to t and then with respect to t' as described in Ref. 28, we arrive at three coupled equations for the Fourier transforms of the Green's functions:

$$\sum_{k''}[\mu_k(\omega^2-\omega_k^2)\delta_{kk''}-2(\sigma+\langle q\rangle^2)\lambda_{kk''}]\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega = \delta_{kk'}+\sum_{k''}2\lambda_{kk''}\cdot{}^{\rm ir}\langle\langle Qx_{k''}|x_{k'}\rangle\rangle_\omega+\lambda_k\langle\langle Q|x_{k'}\rangle\rangle_\omega, \quad (14)$$

$$\sum_{k_1}[\mu_{k'}(\omega^2-\omega_{k'}^2)\delta_{k'k_1}-2(\sigma+\langle q\rangle^2)\lambda_{k'k_1}]\langle\langle Q|x_{k_1}\rangle\rangle_\omega = \sum_{k_1}2\lambda_{k'k_1}\langle\langle Q|Qx_{k_1}\rangle\rangle_\omega^{\rm ir}+\lambda_{k'}\langle\langle Q|Q\rangle\rangle_\omega, \quad (15)$$

$$\sum_{k_1}[\mu_{k'}(\omega^2-\omega_{k'}^2)\delta_{k'k_1}-2(\sigma+\langle q\rangle^2)\lambda_{k'k_1}]\cdot{}^{\rm ir}\langle\langle Qx_{k''}|x_{k_1}\rangle\rangle_\omega = \sum_{k_1}2\lambda_{k'k_1}\cdot{}^{\rm ir}\langle\langle Qx_{k''}|Qx_{k_1}\rangle\rangle_\omega^{\rm ir}+\lambda_{k'}\cdot{}^{\rm ir}\langle\langle Qx_{k''}|Q\rangle\rangle_\omega, \quad (16)$$

where

$$Q\equiv 2\langle q\rangle\delta q+(\delta q)^2. \quad (17)$$

The superscript ''ir'' indicates that the corresponding Green's function is irreducible, i.e., it cannot be reduced to lower-order functions by decoupling the product of single-time operators.[28]

We define the Green's function in the lowest-order approximation as

$$\sum_{k''}[\mu_{k'}(\omega^2-\omega_k^2)\delta_{kk''}-2\lambda_{kk''}\times(\sigma+\langle q\rangle^2)]\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega^{(0)}=\delta_{kk'}. \quad (18)$$

Then Eqs. (14)–(16) become

$$\langle\langle x_k|x_{k'}\rangle\rangle_\omega=\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}+\sum_{k'',k_1}2\lambda_{k''k_1}\langle\langle x_k|x_{k''}\rangle\rangle_\omega^{(0)}\times{}^{\rm ir}\langle\langle Qx_{k_1}|x_{k'}\rangle\rangle_\omega+\sum_{k''}\lambda_{k''}\langle\langle x_k|x_{k''}\rangle\rangle_\omega^{(0)}\times\langle\langle Q|x_{k'}\rangle\rangle_\omega, \quad (19)$$

$$\langle\langle Q|x_{k'}\rangle\rangle=\sum_{k''}\lambda_{k''}\langle\langle Q|Q\rangle\rangle_\omega\cdot\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega^{(0)}+\sum_{k'',k_1}2\lambda_{k_1k''}\langle\langle Q|Qx_{k_1}\rangle\rangle_\omega^{\rm ir}\cdot\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega^{(0)}, \quad (20)$$

$${}^{\rm ir}\langle\langle Qx_{k_1}|x_{k'}\rangle\rangle_\omega=\sum_{k''}\lambda_{k''}\cdot{}^{\rm ir}\langle\langle Qx_{k_1}|Q\rangle\rangle_\omega\cdot\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega^{(0)}+\sum_{k'',k_0}2\lambda_{k_0k''}\cdot{}^{\rm ir}\langle\langle Qx_{k_1}|Qx_{k_0}\rangle\rangle_\omega^{\rm ir}\times\langle\langle x_{k''}|x_{k'}\rangle\rangle_\omega^{(0)}, \quad (21)$$

thus yielding

$$\begin{aligned}\langle\langle x_k|x_k\rangle\rangle_\omega=&\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)}+\sum_{k',k''}\lambda_{k'k''}\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}\times\langle\langle Q|Q\rangle\rangle_\omega\cdot\langle\langle x_{k''}|x_k\rangle\rangle_\omega^{(0)}\\&+\sum_{k',k'',k_0,k_1}4\lambda_{k',k''}\lambda_{k_0,k_1}\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}\times{}^{\rm ir}\langle\langle Qx_{k''}|Qx_{k_0}\rangle\rangle_\omega^{\rm ir}\cdot\langle\langle x_{k_1}|x_k\rangle\rangle_\omega^{(0)}\\&+\sum_{k',k'',k_1}2\lambda_{k'k''}\lambda_{k_1}\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}\times{}^{\rm ir}\langle\langle Qx_{k''}|Q\rangle\rangle_\omega\cdot\langle\langle x_{k_1}|x_k\rangle\rangle_\omega^{(0)}\\&+\sum_{k',k'',k_1}2\lambda_{k_1k''}\lambda_{k'}\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}\times\langle\langle Q|Qx_{k_1}\rangle\rangle_\omega^{\rm ir}\cdot\langle\langle x_{k''}|x_k\rangle\rangle_\omega^{(0)}. \quad (22)\end{aligned}$$

The two last ''interference'' terms in (22), containing the Green's functions ${}^{\rm ir}\langle\langle Qx_k|Q\rangle\rangle$ and $\langle\langle Q|Qx_k\rangle\rangle^{\rm ir}$, have an order of smallness in the interaction $H_{\rm int}$ higher than the second and can be discarded. We also ignore the contribution of off-diagonal components of the Green's functions in the lowest-order approximation, assuming that $\langle\langle x_k|x_{k'}\rangle\rangle_\omega^{(0)}=\delta_{kk'}\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)}$, and write Eq. (22) in the form of the Dyson equation:

$$\langle\langle x_k|x_k\rangle\rangle_\omega^{-1}=(\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)})^{-1}-M_k(\omega), \quad (23)$$

where the self-energy part has the form

$$M_k(\omega)=\lambda_k^2\langle\langle Q|Q\rangle\rangle_\omega+\sum_{k',k''}4\lambda_{kk'}\lambda_{k''k}\times{}^{\rm ir}\langle\langle Qx_{k'}|Qx_{k''}\rangle\rangle_\omega^{\rm ir}, \quad (24)$$

with

$$\langle\langle Q|Q\rangle\rangle_\omega=4\langle q\rangle^2\langle\langle\delta q|\delta q\rangle\rangle_\omega+\langle\langle(\delta q)^2|(\delta q)^2\rangle\rangle_\omega+2\langle q\rangle(\langle\langle\delta q|(\delta q)^2\rangle\rangle_\omega+\langle\langle(\delta q)^2|\delta q\rangle\rangle_\omega), \quad (25)$$

$$
\begin{aligned}
{}^{\mathrm{ir}}\langle\langle Qx_k|Qx_{k'}\rangle\rangle_\omega^{\mathrm{ir}} &= 4\langle q\rangle^2\langle\langle \delta q x_k|\delta q x_{k'}\rangle\rangle_\omega \\
&\quad + {}^{\mathrm{ir}}\langle\langle (\delta q)^2 x_k|(\delta q)^2 x_{k'}\rangle\rangle_\omega^{\mathrm{ir}} + 2\langle q\rangle \\
&\quad \times(\langle\langle \delta q x_k|(\delta q)^2 x_{k'}\rangle\rangle_\omega^{\mathrm{ir}} \\
&\quad + {}^{\mathrm{ir}}\langle\langle (\delta q)^2 x_k|\delta q x_{k'}\rangle\rangle_\omega). \qquad (26)
\end{aligned}
$$

[it should be recalled that the operator Q is defined in Eq. (17)].

3.3. An approximate calculation of the self-energy part

It is convenient to express the higher-order Green's functions that enter into $M_k(\omega)$ [see Eqs. (25) and (26)] in terms of correlation functions via the spectral theorem.[29] For instance, for one of these Green's functions we have

$$
\begin{aligned}
&{}^{\mathrm{ir}}\langle\langle (\delta q)^2 x_k|(\delta q)^2 x_{k'}\rangle\rangle_\omega^{\mathrm{ir}} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega'}{\omega-\omega'}\left(\exp\frac{\omega'}{\Theta}+1\right)\int_{-\infty}^{\infty}dt \\
&\times\exp\{-i\omega' t\}\langle{}^{\mathrm{ir}}[(\delta q(t))^2 x_k(t)][(\delta q(t'))^2 x_{k'}(t')]^{\mathrm{ir}}\rangle. \qquad (27)
\end{aligned}
$$

The correlation function in (27) can be decoupled by forming pairwise two-time averages (single-time averages, according to the definition of the ''ir'' operation, are equal to zero):

$$
\begin{aligned}
&\langle{}^{\mathrm{ir}}[(\delta q(t))^2 x_k(t)][(\delta q(t'))^2 x_{k'}(t')]^{\mathrm{ir}}\rangle \\
&\quad \approx 2\delta_{kk'}\langle\delta q(t)\,\delta q(t')\rangle^2\cdot\langle x_k(t)x_k(t')\rangle, \qquad (28)
\end{aligned}
$$

where the factor 2 reflects the two possible ways of pairing the operators δq taken at times t and t'. By analogy with the diagrammatic technique, we can assume that adopting the approximation (28) is equivalent to ignoring the vertex corrections in the processes of interaction between phonons and the vibrations of the bistable sublattice. The spectral theorem can be used to express the one-particle correlators in (28) in terms of the corresponding Green's functions:

$$
\begin{aligned}
&{}^{\mathrm{ir}}\langle\langle (\delta q)^2 x_k|(\delta q)^2 x_{k'}\rangle\rangle_\omega^{\mathrm{ir}} \\
&\approx 2\delta_{kk'}\int_{-\infty}^{\infty}\!\!\int\!\!\int\frac{d\omega_1 d\omega_2 d\omega_3}{\omega-(\omega_1+\omega_2+\omega_3)} \\
&\times\frac{\exp\{(\omega_1+\omega_2+\omega_3)/\Theta\}-1}{[\exp\{\omega_1/\Theta\}-1][\exp\{\omega_2/\Theta\}-1][\exp\{\omega_3/\Theta\}-1]} \\
&\times\left[-\frac{1}{\pi}\mathrm{Im}\langle\langle\delta q|\delta q\rangle\rangle_{\omega_1+i\varepsilon}\right]\left[-\frac{1}{\pi}\mathrm{Im}\langle\langle\delta q|\delta q\rangle\rangle_{\omega_2+i\varepsilon}\right] \\
&\times\left[-\frac{1}{\pi}\mathrm{Im}\langle\langle x_k|x_k\rangle\rangle_{\omega_3+i\varepsilon}\right]. \qquad (29)
\end{aligned}
$$

In the same way one should deal with the remaining Green's functions in the expression for $M_k(\omega)$; some of these, namely those in the parentheses in Eqs. (25) and (26), vanish in view of approximations of the form (28). Next, in calculating (29) we can ignore the self-energy parts of the one-particle Green's functions by writing them in the lowest-order approximation:

$$
\langle\langle\delta q|\delta q\rangle\rangle_\omega\to\langle\langle\delta q|\delta q\rangle\rangle_\omega^{(0)}=[m(\omega^2-\Omega^2)]^{-1}, \qquad (30)
$$

with $\langle\langle\delta q|\delta q\rangle\rangle_\omega^{(0)}$ the Green's function corresponding to the Hamiltonian H_0 [Eq. (8)], and

$$
\begin{aligned}
&\langle\langle x_k|x_k\rangle\rangle_\omega\to\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)} \\
&\quad =[\mu_k(\omega^2-\omega_k^2)-2\lambda_{kk}(\sigma+\langle q\rangle^2)]^{-1}. \qquad (31)
\end{aligned}
$$

This enables us to explicitly calculate the exact Green's function $\langle\langle x_k|x_k\rangle\rangle_\omega$ on the basis of (23).

3.4. Determining the shift and decay of the lattice mode frequencies

The renormalized frequencies of the lattice modes, ε_k, and the decay coefficients Γ_k can be found by solving the equation

$$
(\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)})^{-1}-\mathrm{Re}\,M_k(\tilde\omega_k+i\varepsilon)+i\,\mathrm{Im}\,M_k(\tilde\omega_k+i\varepsilon)=0, \qquad (32)
$$

where

$$
\tilde\omega_k\approx\omega_k+\frac{\lambda_{kk}}{\mu_k\omega_k}(\sigma+\langle q\rangle^2) \qquad (33)
$$

is the pole of the Green's function $\langle\langle x_k|x_k\rangle\rangle_\omega^{(0)}$. From (32) it follows that

$$
\varepsilon_k\approx\tilde\omega_k+\frac{1}{2\mu_k\tilde\omega_k}\mathrm{Re}\,M_k(\tilde\omega_k+i\varepsilon), \qquad (34)
$$

$$
\Gamma_k\approx-\frac{1}{2\mu_k\tilde\omega_k}\mathrm{Im}\,M_k(\tilde\omega_k+i\varepsilon). \qquad (35)
$$

For our further investigations it is enough to determine the contributions to the renormalized frequencies ε_k and decay coefficients Γ_k of the cubic, $H_{\mathrm{int}}^{(3)}$, and quartet, $H_{\mathrm{int}}^{(4)}$, interactions in the first nonvanishing orders: the first in $H_{\mathrm{int}}^{(4)}$ and the second in $H_{\mathrm{int}}^{(3)}$ for ε_k and the second in $H_{\mathrm{int}}^{(4)}$ for Γ_k.

Dealing with the first term in the expression (24) for the self-energy part in the same way as we did in (29), where the exact one-particle Green's functions are replaced by their lowest-order approximations (30) and (31), we arrive at an expression for the contribution to $M_k(\omega)$ of the cubic interaction in second order:

$$
M_k^{(3)}(\tilde\omega_k)=4\lambda_k^2\left[\frac{\langle q\rangle^2}{m(\tilde\omega_k^2-\Omega^2)}+\frac{\sigma}{m(\tilde\omega_k^2-4\Omega^2)}\right], \qquad (36)
$$

$$
\varepsilon_k\equiv\omega_k+\Delta_k=\tilde\omega_k+\frac{1}{2\mu_k\tilde\omega_k}M_k^{(3)}(\tilde\omega_k). \qquad (37)
$$

Reasoning in a similar manner, we find an expression for the second term in (24) resulting from the quartet interaction:

$$M_k^{(4)}(\tilde{\omega}_k)$$
$$= \frac{2}{m\Omega\mu_k\omega_k\tilde{\omega}_k}\sum_{k'}\lambda_{kk'}^2\omega_{k'}$$
$$\times\left\{4\langle q\rangle^2\left[\frac{(\Omega-\tilde{\omega}_{k'})[n(\tilde{\omega}_{k'})-n(\Omega)]}{\tilde{\omega}_k^2-(\Omega-\tilde{\omega}_{k'})^2}\right.\right.$$
$$\left.+\frac{(\Omega+\tilde{\omega}_{k'})[1+n(\tilde{\omega}_{k'})+n(\Omega)]}{\tilde{\omega}_k^2-(\Omega+\tilde{\omega}_{k'})^2}\right]$$
$$+\frac{2\Omega-\tilde{\omega}_{k'}}{m\Omega}\frac{[n(\Omega)+1]^2n(\tilde{\omega}_{k'})-n^2(\Omega)[n(\tilde{\omega}_{k'})+1]}{\tilde{\omega}_k^2-(2\Omega-\tilde{\omega}_{k'})^2}$$
$$+\frac{2\Omega+\tilde{\omega}_{k'}}{m\Omega}\frac{[n(\Omega)+1]^2[n(\tilde{\omega}_{k'})+1]-n^2(\Omega)n(\tilde{\omega}_{k'})}{\tilde{\omega}_k^2-(2\Omega+\tilde{\omega}_{k'})^2}$$
$$\left.+\frac{2\tilde{\omega}_{k'}}{m\Omega}\frac{n(\Omega)[n(\Omega)+1]}{\tilde{\omega}^2-\tilde{\omega}_{k'}^2}\right\}, \tag{38}$$

where $n(x)=[\exp\{x/\Theta\}-1]^{-1}$.

The decay coefficients Γ_k, which can be expressed in terms of $\mathrm{Im}\,M_k^{(4)}(\tilde{\omega}_k+i\varepsilon)$, are due to processes of creation (annihilation) of one or two vibrational quanta of the bistable sublattice accompanied by processes of absorption (emission) of two quanta of the matrix lattice, and also to processes of elastic scattering of the quanta of the matrix lattice that do not change the vibrational state of the bistable sublattice. On the basis of (35) and (38) we can write

$$\Gamma_k=\frac{\pi}{2\mu_k\tilde{\omega}_k m\Omega}\sum_{k'}\frac{\lambda_{kk'}^2}{\mu_{k'}\tilde{\omega}_{k'}}\left\{4\langle q\rangle^2([n(\tilde{\omega}_{k'})-n(\Omega)]\right.$$
$$\times[\delta(\tilde{\omega}_{k'}+\tilde{\omega}_k-\Omega)-\delta(\tilde{\omega}_{k'}-\tilde{\omega}_k-\Omega)]+[1$$
$$+n(\tilde{\omega}_{k'})+n(\Omega)][\delta(\tilde{\omega}_{k'}-\tilde{\omega}_k+\Omega)-\delta(\tilde{\omega}_{k'}+\tilde{\omega}_k$$
$$+\Omega)])+\frac{1}{m\Omega}[(n(\Omega)+1)^2n(\tilde{\omega}_{k'})-n^2(\Omega)$$
$$\times(n(\tilde{\omega}_{k'})+1)][\delta(\tilde{\omega}_{k'}+\tilde{\omega}_k-2\Omega)-\delta(\tilde{\omega}_{k'}-\tilde{\omega}_k$$
$$-2\Omega)]+\frac{2}{m\Omega}n(\Omega)(n(\Omega)+1)[\delta(\tilde{\omega}_{k'}-\tilde{\omega}_k)$$
$$\left.-\delta(\tilde{\omega}_{k'}+\tilde{\omega}_k)]\right\}. \tag{39}$$

To do some estimates, it is enough to examine the one-dimensional model of a lattice in the Debye approximation. Plugging the expression for the coupling coefficient in the form

$$\lambda_{kk'}=\frac{\lambda}{M}\sum_{j=1}^{M}\exp\{i(k'+k)r_j\}\sqrt{\mu_k\mu_{k'}}\omega_k\omega_{k'} \tag{40}$$

($\mathbf{r}_j$ is the radius vector of the jth atom in the matrix lattice, and λ is a constant whose dimensions are cm^{-2}) into (39), we find that for $k<k_D/2$ and $\omega_D\approx\Omega$ (k_D and ω_D are the Debye wave vector and frequency) only elastic processes provide nonvanishing contributions to decay:

$$\Gamma_k\approx 4\pi\lambda^2\left(\frac{\hbar}{2m\Omega}\right)^2 n(\Omega)[n(\Omega+1)]\omega_k\frac{k}{k_D}\frac{1}{\delta^3}, \tag{41}$$

where $\delta=\tilde{\omega}_k/\omega_k$. The above equation shows that the quartet interaction $H_{\mathrm{int}}^{(4)}$ has almost no damping effect on vibrations in the ultrasonic frequency range. Indeed, for reasonable values of the parameters ($\lambda=8\ \text{Å}^{-2}$, $\Omega=10^{13}\,\mathrm{s}^{-1}$, $T=273\,\mathrm{K}$, and $\delta=1$) we have the estimate

$$\frac{\Gamma_k}{\omega_k}\sim 10^{-2}\frac{k}{k_D}, \tag{42}$$

i.e., the decay of acoustic vibrations becomes significant only at maximum frequencies ($k\sim k_D$).

3.5. Temperature hysteresis of the shift and decay of the lattice mode frequencies as a result of interaction with the bistable lattice

The temperature dependence of the frequency shift $\Delta_k(T)$ is determined by the dependence of this shift on the characteristics $\langle q\rangle$, σ, and Ω, which experience temperature hysteresis, and by the competition between negative (for $\Omega>\tilde{\omega}_k$) cubic and positive (for $\lambda_{kk}>0$) quartet contributions. Indeed, for instance, when the sample is cooled, the stable states of the correlated sublattice become metastable, and at the temperature T_1 the sublattice suddenly goes over to another stable branch (see Figs. 2 and 3), as a result of which related abrupt changes are experienced by the frequency shifts Δ_k of the acoustic modes of the matrix [see Eqs. (33), (36), and (37)] interacting with the sublattice (Fig. 4). Now, when heated, the sublattice is on this new branch up to a temperature T_2, after which it again suddenly returns to its old, high-temperature, stable branch, thus bringing about a sudden (discontinuous) change in Δ_k (Fig. 4). Figure 4 depicts the Δ_k vs. T curves calculated by formulas (33), (36) and (37) for different ratios of the competing interactions of the third (λ_k) and fourth (λ_{kk}) orders.

The temperature dependence of the decay of an acoustic mode (Γ_k) for $k\sim k_D$ is depicted in Fig. 5 (note that here we allow for neither the nonlinearity of the matrix lattice proper, a nonlinearity that provides a nonhysteresis contribution to decay, nor for other decay mechanisms). The decay that is the largest (the high-temperature branch of curve 1) for the scattering of lattice modes by perturbations caused by above-barrier vibrations of the bistable sublattice suddenly becomes smaller (the low-temperature branch of the curve 1) when the sublattice abruptly reduces the amplitude of its vibrations after it has been ''captured'' by the global minimum as a result of cooling, The stable branch of curve *2* represents the contribution to scattering of the local minimum of the bistable potential. Actually this contribution will be smaller, since it must be multiplied by a quantity proportional to $\exp\{-\Delta F/k_BT\}$, where ΔF is the difference of free energies of the sublattice in the local and global minima; hence, the system is heated from absolute zero, the increase in decay follows almost exactly the low-temperature branch of curve *1*, and then the decay suddenly increases, going over to the high-temperature branch and thus completing the hysteresis cycle.

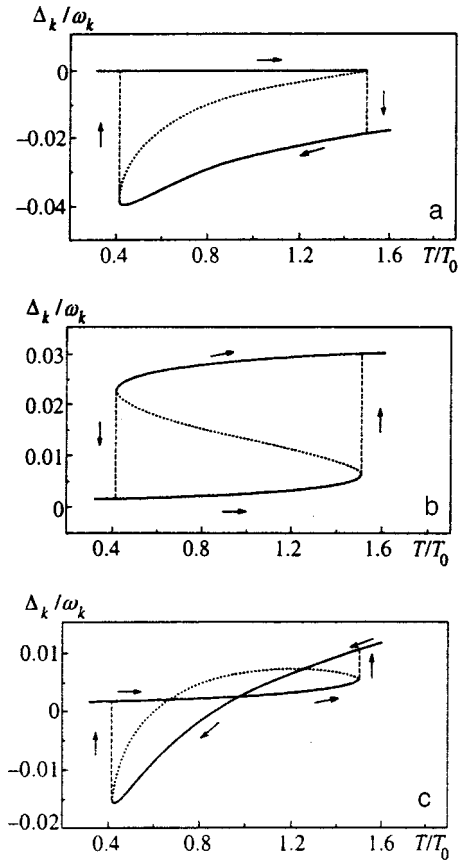

FIG. 4. Temperature dependence of the relative shift of the acoustic modes of the matrix with allowance for cubic interaction (a), quartet interaction (b), and competition of quartet and cubic interactions (c); $u_1=0.03$ eV, $q_1=0.073$ Å, $q_2=0.14$ Å, $T_0=173$ K, $\lambda_{kk}/\mu_k\omega_k^2=8.3$ Å^{-2}, and $\lambda_k=3.74\times10^{-21}$ eV·Å^{-3}.

4. EXPERIMENTAL PREREQUISITES FOR THE EXISTENCE OF A BISTABLE SUBLATTICE IN SUPERCONDUCTING OXIDE CUPRATES

At present there are many papers that point to the important role that the apical oxygen atom plays in the formation of the superconducting properties of $YBa_2Cu_3O_{7-\delta}$ or compounds with a structure containing apical atoms. In the $YBa_2Cu_3O_{7-\delta}$ compound, each apical atom O(4) interacts along the crystallographic axis **c** with the two nearest neighbors, the atoms Cu(1) and Cu(2) ($l_{Cu(1)}=1.80-1.86$ Å and $l_{Cu(2)}=2.30-2.45$ Å; see Refs. 30 and 31), whose coordination in oxygen is not the same. The nature of the bond of the apical atom varies substantially: from covalent for the superconducting compound $YBa_2Cu_3O_{7-\delta}$ to ionic for a nonsuperconducting compound.[32] Participating in the transfer of holes from the basal planes to the CuO_2 planes, an apical atom manifests a number of features in the temperature dependence of the vibrational states. For instance, *x*-ray studies,[33–35] ion-channeling experiments,[36,37] Raman spectroscopy,[38–40] and neutron scattering measurements[41] have revealed that the total energy of an apical O(4) atom, as a function of the position along the crystallographic axis **c** has two minima. Note that pyro- and piezoelectricity have been detected in single crystals of $YBa_2Cu_3O_{7-\delta}$, which suggests that there is macroscopic polarization along the **c** axis (see, e.g., Ref. 42). The occurrence of macroscopic polarization is usually attributed to the anharmonic motion of O(4) ions.[42,43]

The nontrivial dynamics of the strongly correlated apical O(4) atoms in a double-well potential must also directly manifest itself in the nature of the interaction between the vibrational states of the atoms and the electron subsystem of the crystal, which in addition to the participation of apical atoms in charge transfer from the basal plane to the CuO_2 plane may be one of the reasons for the formation of high-T_c superconductivity.[44–46] More than that, as we will show shortly, by allowing for the interaction between the bistable oxygen sublattice and the vibrational states of the matrix lattice we can explain a number of experimentally established phenomena: the temperature hysteresis of the specific heat and thermal conductivity and of the speed and absorption of ultrasound in yttrium and bismuth cuprates.

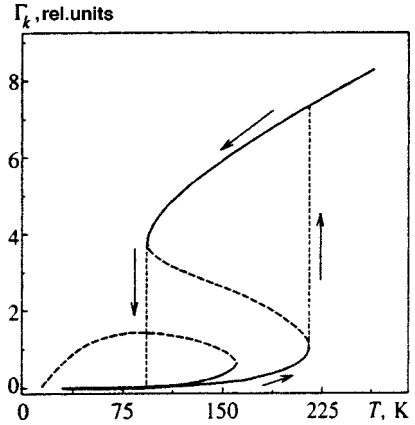

FIG. 5. Temperature dependence of the decay of the acoustic modes of the matrix in a crystal with a bistable sublattice; $u_1=0.03$ eV, $q_1=0.073$ Å, $q_2=0.14$ Å, and $T_0=173$ K.

5. TEMPERATURE HYSTERESIS OF THE SPEED AND DECAY COEFFICIENT OF ULTRASOUND IN HIGH-T_c OXIDE CUPRATES. COMPARISON OF THEORY AND EXPERIMENT

The temperature hysteresis of the speed of ultrasound was observed by the methods of ultrasound spectroscopy[1–14] (see also the review by Lubenets *et al.*[15]) soon after the discovery of high-T_c superconductivity in a number of oxide cuprates, including the compounds $YBa_2Cu_3O_{7-\delta}$ and $Bi_2Sr_2Ca_1Cu_2O_8$. This phenomenon can be observed not only polycrystals but also in single crystals, not only superconducting but also in nonsuperconducting high-T_c compounds. The temperature interval of the hysteresis changes from sample to sample and depends on the oxygen nonstoichiometry and the way in which the sample is prepared. The values 55 and 215 K were fixed as the most reliable limits of hysteresis at the lower and higher ends of the temperature interval, although the upper limit was found to often move up to 270 K. Interestingly, Kim *et al.*[8] (see also Ref. 4) also observed distinct temperature hysteresis of the absorption of

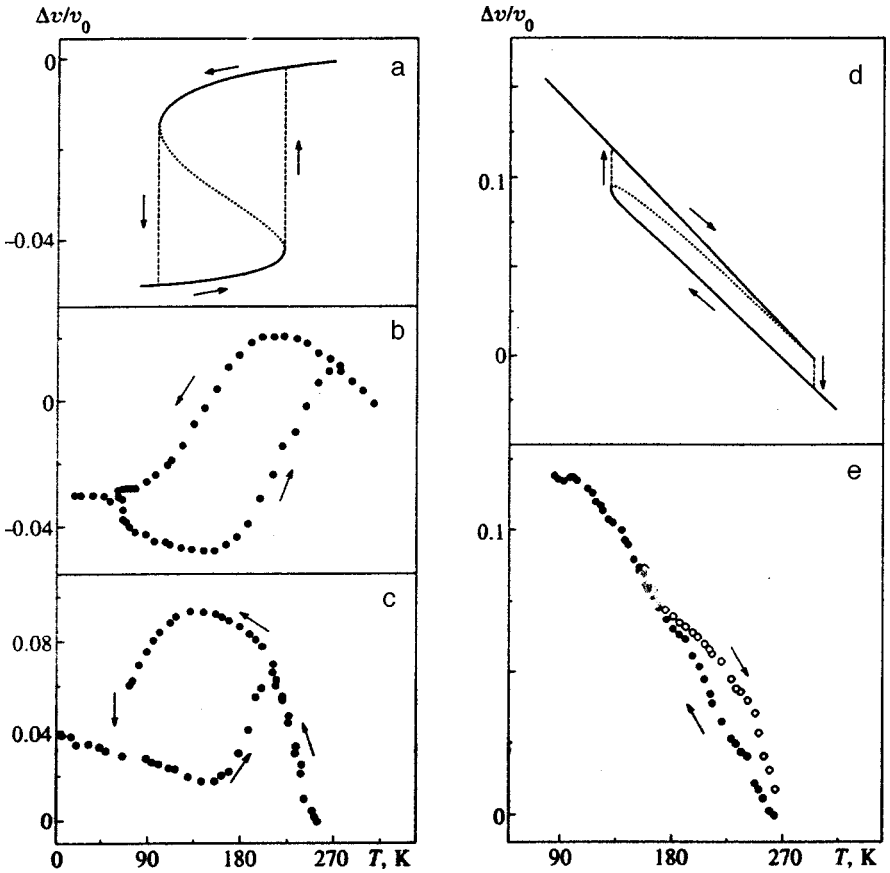

FIG. 6. Temperature dependence of the relative variation of the speed of an ultrasonic wave: (a) and (d) represent the results of calculations for a crystal with a bistable sublattice (the dotted curves represent an unstable solution), with (a) $u_1=0.03$ eV, $q_1=0.073$ Å, $q_2=0.14$ Å, and $\lambda_{kk}/\mu_k\omega_k^2=8.3$ Å^{-2}, and (d) $u_1=0.04$ eV, $q_1=0.073$ Å, $q_2=0.14$ Å, and $\lambda_k=3.74\times10^{-21}$ eV·Å^{-3}; $\Delta v/v_0=\varepsilon_k(T)/\varepsilon_k(300\text{ K})-1+(A-BT)$, with $A=0.19$ and $B=7.05\times10^{-4}$ the constants (determined from experiments[6]) of the linear dependence approximating the contribution of the main lattice. The experimental data for $YBa_2Cu_3O_7$ in the direction of the crystallographic axis **c** at 12 MHz (Ref. 2) are depicted in Fig. 6(b), and those at the frequency 1.25×10^5 s^{-1} in Fig. 6(c). Finally, the experimental data for $Bi_2Sr_2Ca_1Cu_2O_8$ (see Ref. 6) are depicted in Fig. 6(e).

ultrasound in a single-crystal $YBa_2Cu_3O_{7-\delta}$ sample, and the regions of hysteresis of the speed and the absorption of ultrasound were found to coincide.

Various mechanisms for explaining these phenomena have been proposed. Among these are the redistribution of oxygen,[11,13] the motion of twinning boundaries,[47] and the presence of a ferroelectric[10] or martensitic[9] phase transition. However, no satisfactory and consistent interpretation of the temperature hysteresis of the speed and absorption of ultrasound was proposed in these papers.

We believe that a qualitative explanation of these phenomena can be found if we assume that yttrium and bismuth cuprates have a bistable oxygen sublattice that modulates the phonon spectrum of the matrix lattice. Indeed, as we showed within the scope of the general theory in Sec. 3, in this case the renormalized frequencies of the matrix and their imaginary parts acquire a hysteresis temperature dependence, so that the elastic constants of the crystal will vary in the same manner. To compare the theoretical curves with the experimental data we only need to know the empirical values of the parameters of the matrix and the sublattice and the nature of the interaction between the two.

5.1. Hysteresis of the speed of ultrasound in $YBa_2Cu_3O_{7-\delta}$

The temperature-dependent renormalized frequency of the long-wavelength phonons, ε_k, directly determines the speed $v(T)$ of an ultrasonic wave:

$$v(T)=\text{const}\cdot\varepsilon_k(T), \tag{43}$$

where k is the wave vector of the mode at whose frequency the ultrasonic measurements are carried out. As shown by further investigations, for an yttrium cuprate the quartet interaction $H_{\text{int}}^{(4)}$ in the Hamiltonian H_{int} [see Eq. (11)] is the most probable one. Here, in the first-order perturbation in $H_{\text{int}}^{(4)}$ we have [see Eqs. (33) and (34)]

$$\varepsilon_k(T)\approx\omega_k[1+\lambda(\sigma+\langle q\rangle^2)], \tag{44}$$

where σ and $\langle q\rangle^2$ are calculated self-consistently by Eqs. (4)–(6) and we have allowed for the fact [see Eq. (40)] that $\lambda_{kk}=\lambda\mu_k\omega_k^2$. Figure 6(a) depicts the temperature dependence of the speed of ultrasound, $v(T)$, at the frequency 1.25×10^5 s^{-1} calculated theoretically from the formulas (43), (44), and (40) with allowance for (4)–(6). As applied to $YBa_2Cu_3O_{7-\delta}$, the model parameters were specified in the following way: the ''bare'' frequency of an O(4) ion in the global minimum, $(\alpha/m)^{1/2}\approx600$ cm^{-1}, was determined from the spectra of Raman scattering of light;[38] the position of the second, local, minimum, $q_2\approx0.14$ Å, was found from the measured radial distribution function;[48] $q_1\approx q_2/2$; the height of the potential barrier, $u_1\approx0.03$ eV, was chosen such that the hysteresis would land into the 100–200 K temperature range; and the coupling constant $\lambda\approx8.3$ Å^{-2} was chosen such that the theoretical values of the maximum difference in the speeds of ultrasound on the bistable branches of the hysteresis curve would agree with the experimental data. In Figs. 6(b) and 6(c) we depict, for the sake of comparison with the calculated curves, the experimental data[2,3] on the temperature dependence of the speed of longitudinal ultra-

sound propagating in the ceramic and single-crystal samples of $YBa_2Cu_3O_{7-\delta}$. The theoretical curves reflect fairly well the experimentally observed behavior of the speed of ultrasound: in the hysteresis region, high speeds are realized in the cooling mode and low speeds in the heating (''thawing'') mode. Thus, at reasonable values of the model parameters, not only do the size and temperature interval of the hysteresis loop agree with the experimental data but so does the sense of tracing of the hysteresis loop in the cooling-heating cycle.

5.2. Hysteresis of the speed of ultrasound in $Bi_2Sr_2Ca_1Cu_2O_8$

The experimentally observed pattern of the temperature hysteresis of the speed of ultrasound in bismuth cuprates, in particular, in $Bi_2Sr_2Ca_1Cu_2O_8$ differs from that for $YBa_2Cu_3O_{7-\delta}$: the higher values of the speed of ultrasound in the hysteresis region are realized in heating, while the lower values are realized in cooling.[6,7] Such behavior of the elastic properties in bismuth cuprates can be explained if we assume that third-order anharmonicity dominates in the interaction between the matrix lattice and the bistable oxygen sublattice, so that we can assume that $\lambda_{kk}=0$. In this case, according to (34) and (36), the renormalized frequency is given by the formula

$$\varepsilon_k(T)\approx\omega_k-\frac{\lambda_k^2}{2m\Omega^2\mu_k\omega_k}(\sigma+4\langle q\rangle^2), \tag{45}$$

where we have allowed for the fact that $\Omega\gg\omega_k$ holds in the experiment and that we have $\tilde{\omega}_k=\omega_k$ at $\lambda_{kk}\approx 0$.

Figure 6(d) depicts the temperature dependence of the speed of ultrasound at 7.5 MHz calculated by (43) and (45) with allowance for the self-consistent equations (4)–(6) for both heating and cooling. When making numerical estimates, we assumed the parameters of the bistable potential to be the same as those for yttrium compounds. We did, however, adjust the height of the potential barrier, which, like the cubic coupling constant λ_k, was chosen so that the calculated values of the temperature integral and size of hysteresis would agree best with the experimentally observed values. For the sake of comparison, in Fig. 6(e) we depict the corresponding experimental dependence[6] for ultrasonic longitudinal waves propagating in a $Bi_2Sr_2Ca_1Cu_2O_8$ single crystal. The theoretical curve represents fairly well the features of this dependence: the cooling curve lies below the heating curve, while the coincidences of the size and interval of the hysteresis in which the hysteresis loop is observed are realized at reasonable values of the parameters of the bistable and matrix lattices.

Thus, we conclude that the interaction between the metastable states of the strongly correlated oxygen sublattice (the apical O(4) atoms) and the matrix lattice in the high-T_c compounds $YBa_2Cu_3O_{7-\delta}$ and $Bi_2Sr_2Ca_1Cu_2O_8$ results in renormalization of the elastic constants of the matrix lattice and, in the final analysis, an experimentally observable temperature hysteresis of the speed of ultrasound in the 60–270 K temperature range. The inversion of the hysteresis branches when yttrium cuprates are replaced by bismuth cuprates is a consequence of the change in the interaction between the bistable sublattice of O(4) atoms and the matrix.

5.3. Hysteresis of the absorption coefficient of ultrasound in $YBa_2Cu_3O_{7-\delta}$

As noted earlier, Kim *et al.*[8] clearly detected a temperature hysteresis in the absorption of a longitudinal ultrasound wave with a frequency of 5 MHz propagating in a single-crystal sample of $YBa_2Cu_3O_{7-\delta}$. A small hysteresis of the damping constant was also observed by Pal'-Val' *et al.*[4] in a ceramic $YBa_2Cu_3O_{7-\delta}$ at $1.25\times10^5\ s^{-1}$. The absorption of ultrasound was greater when the sample was heated than when the sample was cooled. The hysteresis regions for absorption and for the speed of the wave were found to coincide, but the hysteresis loops were traced in opposite directions. The explanation of these facts follows directly from our previous discussion. Indeed, the absorption coefficient α_k of an ultrasound wave with a wave vector k is given by the formula

$$\alpha_k=\frac{\gamma_k}{v(T)}, \tag{46}$$

where γ_k is the attenuation of the wave, which includes as a component the decay coefficient of the hysteresis type, Γ_k, reflecting the presence of the quartet interaction $H_{\rm int}^{(4)}$ (see Eqs. (39), (41), and (42) and Fig. 5) and the contributions from other scattering mechanisms: due to the nonlinearity of the matrix lattice proper, the effect of defects, the boundaries of the sample, etc.; $v(T)$ is the speed of ultrasound given by formulas (43) and (44). But, as noted earlier, the decay constant Γ_k in the ultrasonic frequency range is extremely small, i.e., other scattering mechanisms are effective. Hence the anomalous temperature behavior of the absorption coefficient can be related only to the hysteresis dependence of the speed $v(T)$ of ultrasound. Figure 7(a) shows the results of a theoretical calculation of the absorption coefficient α_k by Eqs. (46), (43), and (44) (the attenuation γ_k is assumed temperature-independent), and Fig. 7(c) depicts the experimental data. We see that there is not only qualitative agreement between the experimental data and the theoretical estimates (the extent and the sense of tracing of the hysteresis loop) but also a correspondence in the relative discrepancy between the absorption on the heating and cooling curves.

Note that temperature hysteresis of the speed and absorption of ultrasound of a similar type was observed by Borisov *et al.*[17] in $LiKSO_4$ crystals.

6. HYSTERESIS BEHAVIOR OF THE THERMAL PROPERTIES OF HIGH-T_c OXIDE CUPRATES IN THE NORMAL STATE

6.1. Hysteresis behavior of specific heats

In the process of doing precision measurements, Vargas *et al.*[18,19] detected a temperature hysteresis of the specific heat at constant pressure, $C_P(T)$, for the high-T_c cuprates $YBa_2Cu_3O_{7-\delta}$ ($\sigma=0-1$) in the 190–230 K temperature range. The heating curve was found to have a sharp peak at 220 K, while the cooling curve was found to have a fairly broad ($\sim$10 K) maximum at 205 K [see Fig. 8(b)]. Kumar

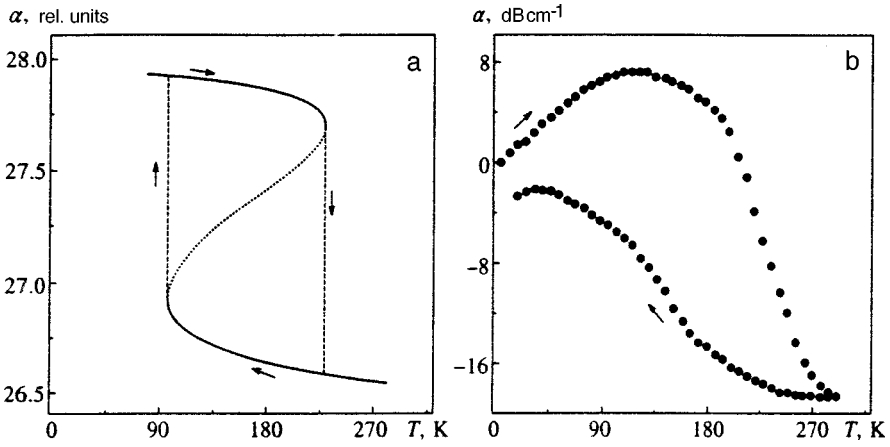

FIG. 7. Temperature dependence of ultrasound absorption: (a), the results of calculations of $\alpha \propto 1/\nu$ for a crystal with a bistable sublattice with $u_1=0.03$ eV, $q_1=0.073$ Å, and $q_2=0.14$ Å; (b), the experimentally measured absorption coefficient for the longitudinal C_{33} mode at 5 MHz in the $YBa_2Cu_3O_7$ crystal.[8]

et al.[20] obtained similar results. Vargas *et al.*[18,19] and Kumar *et al.*[20] suggest that such anomalies are due to the lattice instability; in particular, they relate the narrow peak in the heating curve to the disordering of oxygen atoms in the Cu(1)-O(1)chains, assuming that the interaction of these degrees of freedom and the lattice modes leads to a structural phase transition. We believe that these experimental facts can be explained within the scope of the idea that compounds of the form Y–Ba–Cu–O contain a strongly correlated bistable oxygen sublattice. Experiments have shown that the hysteresis interval may change by several tens of kelvins depending on oxygen content (i.e., on the way in which the sample, chiefly ceramic, is prepared). Later we will return to the problem of finding the hysteresis interval for the specific heat C_p measured in the experiments.[18,19] Here we determine the contributions introduced by the bistable sublattice to the general value of the lattice specific heats at constant pressure, $C_p^{\rm an}$, and at constant volume (constant mean displacement $\langle q\rangle$), $C_{\langle q\rangle}^{\rm an}$. Using formulas (3) and (7), we can find the expressions for the specific heats at constant displacement and at constant pressure:

$$C_{\langle q\rangle}^{\rm an}=k_B\left(\frac{\partial\langle H_{\rm anh}\rangle}{\partial\Theta}\right)_{\langle q\rangle}=k_B\left(m\Omega^2+\frac{3}{2}\gamma\sigma\right)\left(\frac{\partial\sigma}{\partial\Theta}\right)_{\langle q\rangle}, \tag{47}$$

$$p=-\left(\frac{\partial F}{\partial\langle q\rangle}\right)_{\Theta}=-\alpha\langle q\rangle+\beta\langle q\rangle^2 -\gamma\langle q\rangle^3+\sigma(\beta-3\gamma\langle q\rangle). \tag{48}$$

These expressions make it possible to calculate the specific heat at constant pressure:

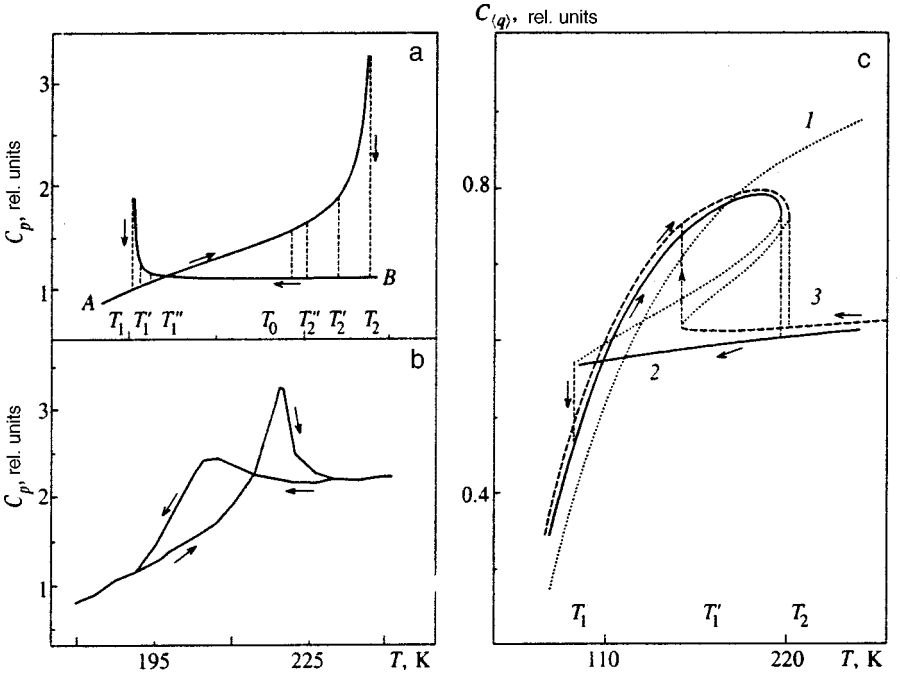

FIG. 8. Temperature dependence of the specific heat at constant pressure [(a) and (b)] and constant volume (c) for a crystal with a bistable sublattice: for a harmonic crystal (curve *1*), and for a crystal with a bistable lattice (curves *2* and *3*). Curves *2* and *3* were calculated for different values of the bistable potential. The dotted curves represent unstable solutions; $u_1=0.03$ eV, $q_1=0.073$ Å, and $q_2=0.14$ Å. Figure 8(b) depicts the experimental results for $YBa_2Cu_3O_7$ (see Ref. 19).

$$C_p^{\rm an}=C_{\langle q\rangle}^{\rm an}-k_B\frac{\Theta(\partial p/\partial\Theta)_{\langle q\rangle}^2}{(\partial p/\partial\langle q\rangle)_\Theta}=k_B\left(m\Omega^2+\frac{3}{2}\gamma\sigma\right)\times\left(\frac{\partial\sigma}{\partial\Theta}\right)_{\langle q\rangle}-k_B\frac{\Theta[(\beta-3\gamma\langle q\rangle)(\partial\sigma/\partial\Theta)_{\langle q\rangle}]^2}{(\beta-3\gamma\langle q\rangle)(\partial\sigma/\partial\langle q\rangle)_\Theta-m\Omega^2},\tag{49}$$

$$\left(\frac{\partial\sigma}{\partial\Theta}\right)_{\langle q\rangle}=\frac{2m\Omega^2}{\Theta}\varphi(\Omega,\sigma,\Theta),$$

$$\left(\frac{\partial\sigma}{\partial\langle q\rangle}\right)_\Theta=2(\beta-3\gamma\langle q\rangle)\varphi(\Omega,\sigma,\Theta).\tag{50}$$

$$\varphi(\Omega,\sigma,\Theta)=\frac{4m^2\Omega^2\sigma^2+4\Theta m\sigma-\hbar^2}{8\Theta^2m^2\Omega^2+3\gamma(4m^2\Omega^2\sigma^2+4\Theta m\sigma-\hbar^2)}.$$

The temperature dependence of the specific heat at constant volume (at constant mean displacement) $C_{\langle q\rangle}^{\rm an}$, constructed from (47) with allowance for (50), exhibits temperature hysteresis [Fig. 8(c)]. The hysteresis curve is transformed according to the shape of the anharmonic potential U in which the atoms O(4) move: it consists of one loop if the metastable minimum lies fairly high above the global minimum and of two loops if the metastable minimum moves downward so that the potential becomes more symmetric [see the part of the caption referring to Fig. 8(c)].

Figure 8(a) depicts the temperature dependence of the specific heat at constant pressure, $C_p^{\rm an}$, constructed from the above formulas. When the sample is cooled and the point T_1 is reached from the right, $C_p^{\rm an}$ becomes infinite and then suddenly drops to the finite value A; when the sample is heated and the point T_2 is reached from the left, $C_p^{\rm an}$ again becomes infinite and then suddenly drops to the value B. The size of the hysteresis interval (T_1,T_2) depends on the values and ratios of the parameters α, β, and γ of the oxygen sublattice (in this specific case they were selected equal to the values in Sec. 5). In the state of thermodynamic equilibrium the system has no memory and the hysteresis disappears, i.e., the function $C_p^{\rm an}(T)$ becomes single-valued; it has, however, a singular point at the temperature T_0 at which the values of the free energies for the cooling and heating curves coincide. The interval (T_0,T_1) in heating and the interval (T_0,T_2) in cooling determine the temperature range in which the system (sublattice) passes through a sequence of alternating unstable (metastable) states.

In real compounds to which our model can be applied, it is impossible to reach the theoretical boundary points of ''supercooling'' (T_1) and ''overheating'' (T_2). The longer the system is left to itself in the region of metastable states, the higher the probability that, thanks to fluctuation processes, it will go over to the other, stable, branch of the hysteresis curve before it reaches the boundary point T_1 or T_2 and hence the narrower the hysteresis region: $T_1\to T_1'>T_1$ and $T_2\to T_2'<T_2$. If the lifetime of a given metastable state of the oxygen sublattice at a certain temperature T_1'' (the cooling curve) or T_2'' (the heating curve) exceeds the time the system is kept in the given state, i.e., the rate of scanning of the temperature in the experiment is such that fluctuation processes are unable to initiate the transition of the sublattice from the metastable state to a stable state, the interval (T_1'',T_2'') exactly determines the real interval of hysteresis behavior of the specific heat and other properties of high-T_c compounds with a bistable sublattice in conditions of the given experiment. This is the reason why in their experiments Vargas *et al.*[18,19] observed a hysteresis interval for the function $C_p^{\rm an}(T)$ that was narrower [Fig. 8(b)] than the theoretical interval [Fig. 8(a)]. Unfortunately, we know of no experimental data on the C vs. T curves for temperature scanning rates so different that the ''shrinking'' of the hysteresis loop can be followed as the temperature scanning rates change from high to low, i.e., as the thermodynamic parameters become more quasistatic.

6.2. Temperature hysteresis of the thermal conductivity in high-T_c cuprates

In the process of doing precision measurements, Jezowski *et al.*,[21–23] Terzijska,[24] and Cohn[25] found a temperature hysteresis of the thermal conductivity of the high-T_c compounds $YBa_2Cu_3O_{7-\delta}$ (1:2:3) and $RBa_2Cu_4O_8$ (1:2:4; R=Dy, Gd, and Eu) in the 70–230 K temperature range. The maximum relative discrepancy between the values of the thermal conductivity on the upper and lower branches of the hysteresis curve amounts to more than 5%. What is remarkable is that the shape of the hysteresis curve, which is single-loop for all 1:2:4 superconductors, for 1:2:3 compounds depends on the index of oxygen nonstoichiometry. At $\delta=0$ a single loop is observed in experiments, while for oxygen-depleted ($\delta=1$), nonconducting, compounds the hysteresis curve consists of two loops with a definite sense of tracing of the contour when the sample is first cooled and then heated. The effect of the hysteresis behavior of the thermal conductivity is unusual and interesting not only in itself but also because the reason why it appears is related to the mechanism of high-T_c superconductivity.

In the high-T_c compounds 1:2:3 and 1:2:4 studied in the above experiments, heat is transferred primarily by long-wavelength acoustic phonons[49] (see also our attempt in Ref. 50 to relate the anomalies in the thermal conductivity of the 1:2:3 compound to optical excitations of the sublattice of apical oxygen atoms O(4)). Below we will show that the hysteresis of the thermal conductivity can be directly related to the scattering of these acoustic modes by vibrational excitations of the bistable oxygen sublattice of the O(4) ions.

6.2.1. Remarks about the formula for thermal conductivity

The Kubo formula for the kinetic thermal conductivity K can be approximately expressed in terms of the square of the one-particle Green's function G_k for acoustic phonons transferring heat:

$$K=\frac{k_B}{3\pi V\Theta}\sum_k\omega_k^2v_k^2\times\int_{-\infty}^{\infty}d\omega\,\frac{\exp\{\omega/\Theta\}}{(\exp\{\omega/\Theta\}-1)^2}[\operatorname{Im}G_k(\omega+i\varepsilon)]^2,\tag{51}$$

where $v_k=\nabla_k\omega_k$ is the phonon group velocity, and V is the volume occupied by the system. We write the Green's function as $G_k(\omega)=[\omega-\omega_k-\Delta_k(\omega)+i\gamma_k(\omega)]^{-1}$. The fre-

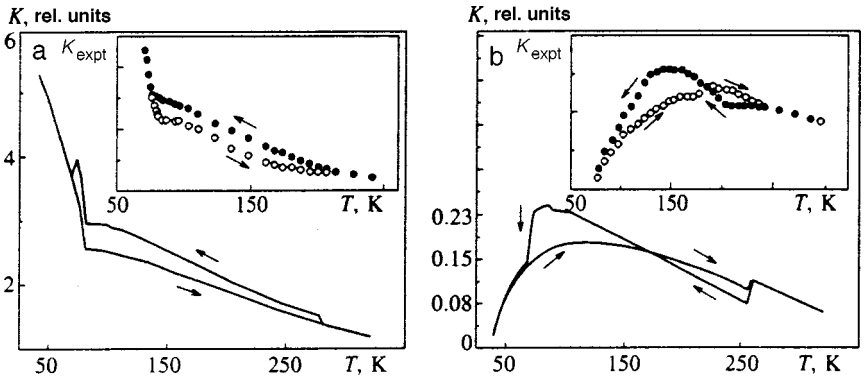

FIG. 9. Theoretical curves representing the temperature dependence of specific heat in crystals with a bistable sublattice: (a), cubic interaction between the bistable sublattice and the matrix is dominant; (b), competition between cubic and quartet interactions; $u_1=0.03$ eV, $q_1=0.073$ Å, $q_2=0.14$ Å, $T_0=173$ K, $\lambda_{kk}/\mu_k\omega_k^2=8.3$ Å^{-2}, and $\lambda_k=3.74\times10^{-21}$ eV·Å^{-3}. The insets schematically depict the experimental results[21–23] for $RBa_2Cu_4O_8$ and $YBa_2Cu_3O_7$ (a) and $RBa_2Cu_3O_6$ (b).

quency shift $\Delta_k(\omega)$ of the kth acoustic mode is due to the interaction between this mode and the other modes of the matrix, defects, electrons, etc., and the bistable sublattice. The same can be said of the rate of scattering of an acoustic phonon, γ_k, related to the lifetime of this phonon by the formula $\tau_k(\omega)=[2\gamma_k(\omega)]^{-1}$. We write the Green's function approximately as $G_k(\omega+i\varepsilon)\approx(\omega-\omega_k-\Delta_k+i\gamma_k)^{-1}$, replacing ω in the expressions for the frequency shift and decay by ω_k: $\Delta_k=\Delta_k(\omega_k)$ and $\gamma_k=\gamma_k(\omega_k)$.

Then for $\gamma_k/\omega_k\ll1$ we can write the expression (51) in the form[51]

$$K=\frac{k_B}{3\Theta^2V}\sum_k \omega_k^2 v_k^2\frac{\exp\{\varepsilon_k/\Theta\}}{(\exp\{\varepsilon_k/\Theta\}-1)^2}\frac{1}{2\gamma_k}, \tag{52}$$

where $\varepsilon_k=\omega_k+\Delta_k$. This formula differs from the standard expression[52] in explicitly allowing for the effect of the frequency shift Δ_k. Below we assume that the frequency shift $\Delta_k=\Delta_k(T)$ is caused solely by the interaction between the acoustic phonons and the vibrations of the bistable oxygen sublattice, with the frequency renormalization due to the nonlinearity of the matrix lattice proper, the interaction with defects and charge carriers, etc., included in the definition of ω_k. Such separation of contributions is natural since they differ qualitatively: as established earlier, the contribution of the bistable sublattice to the renormalization of the frequencies of the matrix lattice is of a hysteresis nature, i.e., its temperature behavior when the sample is cooled differs from its behavior when the sample undergoes heating (''thawing'').

The formula that is commonly used[53] to calculate the thermal conductivity of the high-T_c 1:2:3 and 1:2:4 compounds must be modified; into the exponents we must introduce the frequency shift $\Delta_k(T)$ caused by the heat-transferring phonons scattered by the bistable oxygen sublattice, which means that in the reciprocal phonon lifetime, τ_k^{-1}, we must separate this additional relaxation channel, which leads to decay, earlier denoted by Γ_k. As a result we have a formula for numerical calculations:

$$K=\text{const}\cdot T^3\int_0^{\omega_{\max}/k_BT}dx\,x^4\exp\left\{x+\frac{\Delta(x,T)}{k_BT}\right\}$$
$$\times\left(\exp\left\{x+\frac{\Delta(x,T)}{k_BT}\right\}-1\right)^{-2}\tau(x,T), \tag{53}$$

where

$$\tau^{-1}=\tau_b^{-1}+\tau_d^{-1}+\tau_e^{-1}+\tau_{mp}^{-1}+\tau_{bs}^{-1}=A+BT^4x^4$$
$$+CTxg(x,T/T_c)+Ex^2T^3+D\Gamma(x,T). \tag{54}$$

The terms in the expression for the relaxation time describe the scattering by the boundaries, point defects, electrons, matrix phonons, and the bistable sublattice, and $g(x,T/T_c)$ is the ratio of the relaxation times in the normal and superconducting states;[53] the functions $\Delta(x,T)$ and $\Gamma(x,t)$ are specified by Eqs. (37), (33), (36), (39), and (41), with $x\equiv\omega_k/k_BT$. When necessary, we can allow for the contribution to second order in $H_{\text{int}}^{(4)}$ in the frequency shift by using Eqs. (34) and (38). The effect of $\Gamma(x,t)$ is not appreciable against the background of other mechanisms, but it is obvious that it is the frequency shift $\Delta(x,t)$ of the acoustic modes, which enters into the exponential factors in (53), that basically determines the temperature dependence of $K(T)$.

6.2.2. Numerical estimates of the thermal conductivity

Numerical calculations of $K(T)$ by formula (53) with allowance for the expressions (37), (33), (36), (39), (41), and (54) and for the self-consistent equations (4)–(6) show that when the cubic interaction between the matrix and the bistable sublattice dominates, the hysteresis curve for the thermal conductivity in the 70–230 K temperature range consists of a single loop [Fig. 9(a)]. It is this hysteresis loop that is observed in experiments involving stoichiometric ($\delta=0$) samples of 1:2:3 and 1:2:4 compounds [see the inset in Fig. 9(a)]. What is interesting is that when the quartet interaction dominates ($\lambda_{kk}>0$), the temperature dependence of $K(T)$ basically retains its shape, although the hysteresis cycle is traced in the opposite direction, with the sense of tracing of the cycle agreeing with the one observed in experiments if $\lambda_{kk}<0$. As a result of the competition of comparable contributions of the cubic and quartet interactions, which probably occurs for nonconducting ($7-\delta=6$) samples of 1:2:3 compounds, the hysteresis part of the K vs. T curves becomes two-loop [Fig. 9(b)] and the temperature regions of the two-loop and the one-loop hysteresis coincide, as they do in experiments.

Thus, the interaction between the metastable states of the sublattice of apical O(4) atoms and the matrix gives rise to a temperature hysteresis of the renormalized frequencies of the heat-transferring acoustic modes, which in the final analysis is the reason for the hysteresis behavior of the thermal conductivity in the high-T_c 1:2:3 and 1:2:4 compounds; here the

region of bistable thermal conductivity coincides with the region of bistability of the oxygen sublattice. The form of the hysteresis curve, one-loop or two-loop, and the sense of tracing of the hysteresis loops for $K(T)$ depend on the ratio of the contributions, to the renormalization, of the cubic and quartet terms in the interaction between the matrix and the bistable sublattice.

7. CONCLUSION

We have examined the situation in which in a crystal lattice with a multiatomic basis the atoms of a certain species perform optical vibrations in an asymmetric double-well potential generated by the field of the matrix lattice. If the motion of such atoms is strongly correlated, i.e., is of a cooperative nature, this suppresses fluctuation above-barrier transitions of separate atoms from one energy minimum to other, in view of which (and because of broken symmetry of the potential) metastable states may become realizable for the ensemble of atoms considered, thus producing a bistable sublattice. The critical temperature of the transition of such a sublattice from vibrations inside the global minimum to above-barrier vibrations under heating does not coincide with the temperature of the inverse transition from above-barrier vibrations to intrawell vibrations in the inverse process of cooling. Here the temperature hysteresis is characteristic both for the dynamic parameters of the sublattice and for the statistical-thermodynamic parameters of the sublattice, and the size of the hysteresis interval and its position on the temperature axis are determined primarily by the energy difference of the global and local minima and the height of the potential barrier. Due to the nonlinear interaction between the metastable states of the sublattice and the vibrational states of the matrix lattice, a hysteresis temperature dependence becomes a characteristic feature of the renormalized frequencies of the lattice modes and the decay coefficients for these modes, which in the final analysis gives rise to hysteresis in the elastic and thermal properties of crystals. The ideas developed in this paper result in an interpretation (which agrees fairly well with the experimental data) of the hysteresis temperature behavior of the speed and absorption of ultrasound and of the specific heat and thermal conductivity in superconducting yttrium and bismuth cuprates.

The analysis has shown that in bismuth cuprates the cubic interaction between the degrees of freedom of the bistable sublattice and the matrix lattice is realized, while in yttrium cuprates the interaction is of the quartet type. In accordance with this it occurs that the higher values of the speed of ultrasound are observed in bismuth cuprates under heating, while the lower values are observed under cooling; the situation is the opposite in yttrium cuprates.

Our estimates made it possible to establish that the hysteresis of the absorption of ultrasound is due entirely to the hysteresis dependence of the speed of ultrasound. Hence, knowing the nature of the behavior of the latter, we can predict when the absorption is higher: under heating or under cooling.

While the bistability of the specific heat of the crystal is related to the anharmonic contribution of the bistable sublattice to the total specific heat and the shape of the hysteresis, one-loop or two-loop, depends on the relative positions of the global and local minima of the potential, the reason for the hysteresis of the thermal conductivity is the hysteresis temperature renormalization of the heat-transferring acoustic modes and the shape of the bistable thermal-conductivity curve (one-loop or two-loop), and the sense in which the loops are traced depends on the ratio of the contributions, to the renormalization, of the cubic and quartet terms in the interaction between the matrix and the bistable sublattice.

In all the effects examined in this paper, the region of temperature hysteresis coincides with that of the hysteresis of the bistable sublattice, since the elastic and thermal properties of the crystal depend on the main parameters of the sublattice: its statistical-mean displacement, the displacement variance, and the effective vibration frequency.

ACKNOWLEDGMENTS

This research was supported by the Belorussian Fund for Basic Research (Grant Φ96-342).

*)E-mail: ifttpanb@ifttt.basnet.minsk.by

[1] S. Ewert, S Guo, P. Lemmens *et al.*, Solid State Commun. **64**, 1153 (1987).
[2] M.-F. Xu, A. Schenstrom, Y. Hong *et al.*, IEEE Trans. Magn. **25**, 2414 (1989).
[3] V. D. Natsik, P. P. Pal'-Val', J. Engert, and H.-J. Kaufmann, Fiz. Nizk. Temp. **15**, 836 (1989) [Sov. J. Low Temp. **15**, 463 (1989)].
[4] L. N. Pal'-Val', V. I. Dotsenko, P. P. Pal'-Val', V. D. Natsik, I. F. Kislyak, and A. A. Shevchenok, Sverkhprovodimost': Fiz., Khim., Tekhnol. **3**, 1244 (1990) [Supercond., Phys. Chem. Technol. **3(6)**, S92 (1990)].
[5] Z. Zhao, S. Adenwalla, A. Moreau, and J. B. Ketterson, J. Less-Common Met. **149**, 451 (1989).
[6] Y.-N. Wang, J. Wu, H.-M. Shen, J.-S. Zhu, X.-H. Chen, Y.-F. Yan, and Z.-X. Zhao, Phys. Rev. B **41**, 8981 (1990).
[7] X.-D. Xiang, M. Chung, J. W. Brill *et al.*, Solid State Commun. **69**, 833 (1989).
[8] T. J. Kim, J. Kowalewski, W. Assmus, and W. Grill, Z. Phys. B **78**, 207 (1990).
[9] X. Chen, Y. Wang, and H. Shen, Phys. Status Solidi A **113**, K85 (1989).
[10] V. Müller, C. Hucho, K. deGroot *et al.*, Solid State Commun. **72**, 997 (1989).
[11] L. G. Mamsurova, K. S. Pigal'skiĭ, V. P. Sakun, A. I. Shushin, and L. G. Shcherbakova, Zh. Éksp. Teor. Fiz. **98**, 978 (1990) [Sov. Phys. JETP **71**, 544 (1990)].
[12] E. Biagi, E. Borchi, R. Garre *et al.*, Phys. Status Solidi A **138**, 249 (1993).
[13] T. Fukami, R. Kondo, T. Kobayashi *et al.*, Physica C **241**, 336 (1995).
[14] P. P. Pal-Val, L. N. Pal-Val, V. D. Natsik, I. F. Kislyak, and V. I. Dotsenko, Fiz. Nizk. Temp. **22**, 1452 (1996) [Low Temp. Phys. **22**, 1103 (1996)].
[15] S. V. Lubenets, V. D. Natsik, and L. S. Fomenko, Fiz. Nizk. Temp. **21**, 475 (1995) [Low Temp. Phys. **21**, 367 (1995)].
[16] L. Godfrey and J. Philip, Phys. Rev. B **54**, 15708 (1996).
[17] B. F. Borisov, E. V. Charnaya, and M. Ya. Vinogradova, Phys. Status Solidi B **199**, 51 (1997).
[18] R. A. Vargas, M. Chacon, J. C. Trochez, and I. Palaciox, Phys. Lett. A **139**, 81 (1989).
[19] R. A. Vargas, P. Prietto, M. Chacon, and J. C. Trochez, Phys. Lett. A **152**, 105 (1991).
[20] S. Kumar, S. P. Pai, R. Pinto, and D. Kumar, Physica C **215**, 286 (1993).
[21] A. Jezowski, J. Klamut, R. Horyn, and K. Rogacki, Supercond. Sci. Technol. **1**, 296 (1989).
[22] A. Jezowski, Solid State Commun. **71**, 419 (1989).
[23] A. Jezowski, J. Klamut, and B. Dabrowski, Phys. Rev. B **52**, R7030 (1995).

[24] B. M. Terzijska, Cryogenics **32**, 60 (1992).
[25] J. L. Cohn, J. Supercond. **8**, 457 (1995).
[26] V. Kh. Kozlovskiĭ, Zh. Éksp. Teor. Fiz. **30**, 766 (1956) [Sov. Phys. JETP **3**, 601 (1956)].
[27] Yu. N. Gornostyrev, M. I. Katsnel'son, and A. V. Trefilov, JETP Lett. **56**, 529 (1992).
[28] N. M. Plakida, in *Statistical Physics and Quantum Field Theory* [in Russian], N. N. Bogolyubov (Ed.), Nauka, Moscow (1973).
[29] D. N. Zubarev, *Nonequilibrium Statistical Thermodynamics*, Consultants Bureau, New York (1974).
[30] J. D. Jorgensen, B. W. Veal, A. P. Paulikas, L. J. Nowicki, G. W. Crabtree, H. Claus, and W. K. Kwok, Phys. Rev. B **41**, 1863 (1990); **42**, 995 (1990).
[31] R. J. Cava, A. W. Hewat, E. A. Hewat *et al.*, Physica C **165**, 419 (1990).
[32] V. E. Gusakov, Fiz. Nizk. Temp. **21**, 805 (1995) [Low Temp. Phys. **21**, 621 (1995)]; Ya. S. Bobovich, Usp. Fiz. Nauk **167**, 973 (1997) [Phys. Usp. **40**, 925 (1997)].
[33] D. Conradson, I. Raistrick, and A. R. Bishop, Science **248**, 1394 (1990).
[34] De Leon J. Mustre, S. D. Conradson, I. Batistić, and A. R. Bishop, Phys. Rev. Lett. **65**, 1675 (1990).
[35] E. A. Stern, M. M. Qian, Y. Y. Yacoby *et al.*, Physica C **209**, 331 (1993).
[36] J. Remmel, O. Meyer, J. Geerk, J. Reiner, and G. Linker, A. Erb and G. Müller-Vogt, Phys. Rev. B **48**, 16168 (1993).
[37] R. P. Sharma, L. E. Rehn, P. M. Baldo, and J. Z. Lin, Phys. Rev. Lett. **62**, 2869 (1989).
[38] D. Mihailovic, K. F. McCarty, and D. S. Ginley, Phys. Rev. B **47**, 8910 (1993).
[39] G. Ruani, C. Taliani, M. Muccini *et al.*, Physica C **226**, 101 (1994).
[40] L. V. Gasparov, V. D. Kulakovskii, V. B. Timofeev, and E. Sherma, J. Supercond. **8**, 27 (1995).
[41] H. A. Mook, B. C. Chakoumakos, M. Mostoller, A. T. Boothroyd, and D. McK. Paul, Phys. Rev. Lett. **69**, 2272 (1992).
[42] D. Mihailovic, I. Poberaj, and A. Mertelj, Phys. Rev. B **48**, 16634 (1993).
[43] T. Galbaatar, N. M. Plakida, and S. L. Drechsler, Preprint JINR E17-94-299, Dubna (1994).
[44] R. Aoki, H. Murakami, and T. Kita, Physica C **235–240**, 1891 (1994).
[45] A. P. Saĭko, V. E. Gusakov, and V. S. Kuz'min, JETP Lett. **56**, 411 (1992).
[46] A. P. Saĭko and V. E. Gusakov, Zh. Éksp. Teor. Fiz. **108**, 757 (1995) [JETP **81**, 413 (1995)].
[47] J. F. Smith and D. Wohlleben, Z. Phys. B **72**, 323 (1988).
[48] J. Ranninger, Z. Phys. B **84**, 167 (1991).
[49] B. M. Anderson and B. Sundqvist, Phys. Rev. B **48**, 3575 (1993).
[50] A. P. Saĭko, V. E. Gusakov, and V. S. Kuz'min, JETP Lett. **57**, 116 (1993).
[51] B. Deo and S. N. Behera, Phys. Rev. **141**, 738 (1966).
[52] P. Carruthers, Rev. Mod. Phys. **33**, 92 (1961).
[53] B. M. Anderson, B. Sundqvist, J. Niska, and B. Loberg, Phys. Rev. B **49**, 4189 (1994).

Translated by Eugene Yankovsky